

Université de Montréal

**Prendre soin du consentement :
Tisser l'éthique dans le design d'un agent conversationnel**

Par Camille Vézy

Département de Communication
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des arts et des sciences en vue de l'obtention du
grade de Philosophiæ Doctor (Ph.D.) en communication

Juin 2022

© Camille Vézy, 2022

Université de Montréal

Département de Communication, Faculté des Arts et des Sciences

Cette thèse intitulée

**Prendre soin du consentement :
Tisser l'éthique dans le design d'un agent conversationnel**

A été présentée par Camille Vézy

À un jury composé des personnes suivantes :

Lorna Heaton, Université de Montréal,

Présidente-rapporteure

Boris H.J.M. Brummans, Université de Montréal

Directeur de recherche

François Cooren, Université de Montréal

Membre du jury

Brenda Berkelaar, Dell Technologies

Examinatrice externe

Résumé

Prendre soin du consentement : tisser l'éthique dans le design d'un agent conversationnel

« Es-tu d'accord ? » En contexte numérique, cette question appelle un oui ou un non au traitement annoncé de nos données personnelles par une organisation; elle demande ainsi à l'utilisateur son consentement. Les agents conversationnels sont des systèmes de dialogue (ou *bots*) qui utilisent de nombreuses données personnelles pour parler avec des humains par interface textuelle (*chatbot*) ou vocale (assistant vocal, parfois appelé *voicebot*). Par une ethnographie au sein d'une start-up qui développe ce type d'agents, cette thèse explore comment le consentement se tisse dans les pratiques de design de cette organisation.

Cette thèse se demande : que peut être une « bonne expérience » de consentement avec un agent conversationnel ? Elle s'ouvre par un chemin tracé au travers de littératures en design d'expérience conversationnelle et en éthique sur le consentement. Tandis qu'une bonne expérience conversationnelle est censée être fluide et agréable, celle du consentement demande des interruptions parfois inconfortables pour être suffisamment éclairé : comment ces deux versions de « bonne expérience » peuvent-elles être réunies sans que l'une ne blesse ou n'invalide l'autre ? Face aux nombreuses limites qui rendent le consentement numérique bien difficile à être valable, la tentation de le laisser de côté est palpable. Pourtant, son horizon éthique de respect vaut la peine d'en prendre soin.

Par une perspective qui envisage le consentement comme un *matter of caring*, cette thèse développe ensuite un cadre théorique qui invite à prendre soin des situations de consentement. Il dote l'approche ventriloque, qui étudie la constitution de la réalité par la communication, d'une éthique du *care* : avec le concept de *matter of caring*, il s'agit de prendre soin de certaines préoccupations (*matters of concern*) pour améliorer des situations de consentement. En se faisant *matter of caring*, le consentement trouve une façon de n'être plus négligé comme une préoccupation qui ne parvient pas à compter, mais de participer à un changement dans les habitudes d'interaction.

Cette participation au changement est également au cœur des pratiques ethnographiques qui constituent cette thèse : à partir de mon expérience comme participante active plutôt que de simple observatrice dans la startup mentionnée plus haut, j'invite à comprendre la pratique réflexive ethnographique par une approche relationnelle où l'ethnologue peut être activement engagée dans la constitution de l'organisation qu'elle étudie. C'est par cet engagement que s'est constitué le matériel de terrain que j'analyse tout au long de la thèse.

Plus précisément, le dernier chapitre plonge dans mon terrain et ses tensions pour concevoir une bonne expérience de conversation *et* de consentement. En résistant à la tentation de laisser le consentement perdre de son importance, je montre que s'y accrocher comme un *matter of caring* amène à se concentrer sur les conditions dans

lesquelles l'organisation demande le consentement : comment prévoit-on de traiter les informations personnelles et comment conçoit-on l'interaction par laquelle le bot demandera leur consentement à ses utilisateurs. Cette histoire n'est pas couronnée de succès mais dévoile plutôt une certaine vulnérabilité.

Ainsi, cette thèse ne propose pas un modèle de consentement, ni des directives de design éthique d'un agent conversationnel. Elle s'attarde plutôt sur l'importance des conditions à permettre pour faire de la place aux interactions de consentement. Plutôt que de figer le consentement dans un état contraint de formalité, elle invite à penser le consentement comme une conversation, avec du respect et de l'épanouissement à l'horizon.

Mots-clés : consentement, agent conversationnel, éthique, *matter of caring*, ethnographie

Abstract

Caring for consent: weaving ethics into the design of a conversational agent

“Do you agree?” In the digital realm, this question calls for a yes or no answer to the processing of our personal data by an organization; that is, it asks for the user’s consent. Conversational agents are dialogue systems (or bots) that use many personal data to talk with humans through a textual interface (chatbot) or voice interface (voice assistant, sometimes called “voicebot”). Through an ethnography within a start-up that develops these kinds of agents, this doctoral dissertation explores the weaving of consent into this organization’s design practices.

Thus, this dissertation asks: What does a “good experience” of consent with a conversational agent look like? It starts by reviewing the literature on design of conversational experiences and in the ethics of consent. While a good conversational experience is supposed to be smooth and enjoyable, the consent experience requires some interruptions to be sufficiently informed, which can be uncomfortable. So how can these two versions of a “good experience” be brought together without one hurting or invalidating the other? Faced with the many limitations for meaningful consent in a digital context, the temptation to put consent aside is palpable, yet its ethical horizon of respect is worth caring for.

Then, with a perspective that views consent in terms of a *matter of caring*, this dissertation develops a theoretical framework that enables us to explore how people can care for consent situations. Centered on the concept of matter of caring, this framework enriches a ventriloquial approach to the study of the communicative constitution of reality with an ethics of care; it focuses attention on how certain matters of concern can be cared for in order to improve consent situations. By becoming a matter of caring, consent can no longer be neglected as a concern that fails to count; it rather participates in changing how human beings interact with each other.

Participating in the bringing about of this change is at the core of the ethnographic methods that constitute this dissertation: Based on my experience with becoming an active participant in the mentioned start-up, rather than a mere participant observant, I explain how ethnographic reflexive practice can be viewed relationally; that is, how an ethnographer can be actively engaged in the constitution of the organization she is studying. This engagement shaped the fieldwork materials I analyze in-depth throughout the dissertation.

More specifically, the last chapter of the dissertation delves into my fieldwork on the mentioned tension between designing a good conversational *and* consent experience. Resisting the temptation to let consent fade away, I show that holding on to it as a matter of caring makes us focus on the conditions under which the organization I studied asks for consent: how it plans to process personal information and how it designs the interaction

through which the bot asks for consent from its users. This is not a success story, but rather a story of vulnerability.

Thus, this dissertation does not propose a model for consent, nor does it suggest ethical design guidelines for a conversational agent. Instead, it highlights the importance of providing conditions to enable interactions in which consent becomes a matter of caring. Rather than freezing consent in a constrained state of formality, it invites us to think of consent as a conversation, with respect and flourishing on the horizon.

Keywords: consent, conversational agent, ethics, matter of caring, ethnography

Table des matières

Résumé.....	3
Abstract.....	5
Table des matières.....	7
Liste des figures	10
Liste des sigles et des abréviations.....	11
REMERCIEMENTS.....	14
INTRODUCTION	19
Orientation des chapitres	25
Contributions	30
Chapitre 1 Une approche by design de l'éthique de l'IA	32
Introduction : l'éthique de l'IA à partir du design.....	32
L'éthique de l'intelligence artificielle	33
Entrer dans l'éthique de l'IA et ses préoccupations.....	33
Le travail (organisationnel) de l'éthique	38
Une approche sociomatérielle du travail de l'éthique par le design.....	46
L'éthique by design	50
Une culture de la conséquence.....	52
Le <i>value-sensitive design</i>	53
Les <i>nudges</i> pour faire respecter des valeurs.....	57
Le <i>privacy-by-design</i>	60
Expérimenter avec les médiations	67
Conclusion : accompagner l'éthique chez Mr Y et Prélude	70
Chapitre 2 Qu'est-ce qu'une bonne expérience avec un agent conversationnel ?	73
Introduction : comment faire parler un agent conversationnel ?	73
Faire parler avec agilité les agents conversationnels de Mr Y et Prélude.....	76
Le choix des outils	77
L'agilité pour optimiser le développement	79
Scrum pour organiser le travail	81
Concevoir une bonne expérience avec un agent conversationnel	83
L'expérience utilisateur et sa recherche de fusion.....	85
Le design de conversation	88
La prévention des erreurs	93
Parler comme des humains	98
Avoir la voix d'une personne humaine.....	98
Stop. Peut-on faire autrement ?	104

Faire bouger les lignes de nos conversations	109
Conclusion : faire parler autrement	111
Chapitre 3 Qu'est-ce qu'une bonne expérience de consentement ?	114
Introduction : entrer dans le consentement.....	114
« Pourquoi demande-t-on le consentement ? Pour quoi faire ? »	119
Quelles fins respecter ?	121
Qu'est-ce qu'un bon consentement ?	126
Consentir en ligne : une lutte contre l'ambiguïté.....	129
La transparence et son idéal performatif de confiance	133
« C'est toi qui as le contrôle sur les informations que tu donnes »	137
Réinventer les modèles de consentement pour plus de clarté.....	138
Le design du consentement et ses ambivalences	143
Rendre au consentement sa dynamique.....	154
Les incertitudes du consentement	157
Les efforts en valent-ils la peine ?	158
Empêcher ou permettre : rassurer ?	160
Consentir à quelque chose qui nous dépasse ?.....	165
Parler de consentement	170
Parler autrement de l'inconfortable	171
Un exercice de vulnérabilité.....	175
Conclusion : un horizon qui vaut la peine de le faire compter	178
Chapitre 4 Matter of caring : une éthique du « faire parler »	180
Introduction : quelle éthique pour faire parler ?	180
Pourquoi aller au-delà de « faire parler » ?	182
Matters of caring.....	199
Conclusion : une éthique du care pour écouter parler.....	209
Chapitre 5 Pratiques ethnographiques : Complicité, réflexivité et compromis	211
Introduction : quelle complicité ?	211
Pratiques de collecte de données	215
Pratique réflexive et éthique de recherche	222
Pratiques d'analyse de données	233
Conclusion : analyser des compromis	242
Chapitre 6 Tisser le consentement dans le design de l'agent conversationnel : Un exercice de vulnérabilité	245
Introduction : une fusion vulnérable.....	245
De la santé mentale au tourisme : cela vaut-il encore la peine de s'occuper du consentement ?.....	248
L'accès au terrain : Mr Y, où le consentement compte.....	249
L'entrée sur le terrain : le consentement compte-t-il finalement ?.....	254

Un pitch, où le consentement est là, mais pour quoi ?.....	266
Faire fusionner le consentement dans l'expérience principale	271
Un module.....	272
Trouver un ton <i>fun</i>	278
« Ah mais tu peux pas changer comme ça ! »	286
À quoi dit-on ok ?	291
Maintenir la responsabilité.....	297
« Ne pas faire de tempête dans un verre d'eau »	297
« On n'a pas de contrôle »	301
Les données ne sont pas si personnelles.....	311
Réparer la voix quand elle ne fonctionne pas	320
Composer avec les contraintes	321
Tester l'agent conversationnel, bricoler avec nos voix	324
Conclusion : le travail limité, partiel et modeste du <i>matter of caring</i>	330
CONCLUSION	333
Retour sur les chapitres et leurs explorations	333
Implications théoriques, méthodologiques et pratiques	337
RÉFÉRENCES BIBLIOGRAPHIQUES.....	344
ANNEXE : rapport de Privacy-by-design réalisé pour Prélude.....	365

Liste des figures

Figure 1 : Image reproduisant les extraits du diagramme représentant le design de la conversation chez Prélude (copié de la version originale de Prélude, avec leur autorisation).....	89
Figure 2 : capture d'écran de La Voix de Montréal suggérant des activités, avec la possibilité d'appuyer sur le bouton « Main menu » pour revenir au menu principal. 19 novembre 2019, 11 :36.	93
Figure 3 : capture d'écran de La Voix de Montréal, 13 novembre 2019, 12:52.	95
Figure 4 : Capture d'écran de « myactivity » sur Google Assistant après avoir testé l'Action Sephora. 11 septembre 2019, 17h28.....	142
Figure 5 : Capture d'écran de la liste pour informer le consentement de l'utilisatrice dans la version alpha de la Voix de Montréal. 17 décembre 2019, 14h15.....	148
Figure 6 : Capture d'écran d'une fenêtre permettant d'empêcher qu'une Action mémorise mes informations. Prise le 22 octobre 2019, 17h41.	168
Figure 7 : schémas de Tim Ingold (2011, p. 69) qui représentent deux versions de dessin d'un organisme.....	231
Figure 8 : Capture d'écran du chatbot Woebot qui introduit la demande de consentement.	257
Figure 9 : Capture d'écran du chatbot Wysa qui introduit la demande de consentement.	257
Figure 10 : Captures d'écran du carrousel qui permet d'informer le consentement dans l'application du chatbot Wysa. 9 avril 2019 à 17h23.	258
Figure 11 : Capture d'écran de la diapositive présentant les variables d'expérience utilisateur dans la présentation du projet Bonjour Montréal, 2 juillet 2019.	268
Figure 12 : Captures d'écran de la tâche « Faire valider les maquettes du carrousel privacy à Félix », 15 novembre 2019.....	290
Figure 13 : Capture d'écran de la visualisation du carrousel pour informer le consentement. Présentation Google Slide « Éthique + gestion des données », diapositive 25.	294
Figure 14 : Capture d'écran de la visualisation de la liste pour informer le consentement. Présentation Google Slide « Éthique + gestion des données », diapositive 26.....	295
Figure 15 : Capture d'écran de la visualisation de la précision des informations quand un thème de la liste est demandé. Présentation Google Slide « Éthique + gestion des données », diapositive 26.	295

Liste des sigles et des abréviations

- API** : *Automated Programming Interfaces*.
- ASR** : *Automatic Speech Recognition*, ou reconnaissance automatique de la parole. Techniques d'intelligence artificielle qui permettent de convertir une voix parlée en texte.
- Bot** : Abréviations qui désignent les agents conversationnels que sont les *chatbots* ou *voicebots*. Les agents conversationnels sont des systèmes d'intelligence artificielle avec lesquels parler textuellement (*chatbot*) ou vocalement (*voicebot*, ou assistant vocal).
- CCO** : Communication constitutive des organisations. Perspective de communication qui s'intéresse aux façons dont les organisations se façonnent par la communication, au fil des interactions entre les personnes et les choses autres qu'humaines (documents, technologies, bâtiments, etc.) (Boivin et al., 2017).
- CCR** : Communication constitutive de la réalité. Perspective de communication qui s'intéresse aux façons dont les organisations se façonnent par la communication. La CCO en fait partie, elle s'intéresse aux réalités organisationnelles.
- CNIL** : Commission Nationale de l'Informatique et des Libertés. Organisme qui pilote, accompagne et surveille, en France, l'application du Règlement de Protection des Données (RGPD). Voir : <https://www.cnil.fr/>
- Déclaration de Montréal** : Déclaration de Montréal pour un développement responsable de l'intelligence artificielle. Initiative de l'Université de Montréal pour l'élaboration d'une déclaration de principes éthiques encadrant le développement de l'intelligence artificielle. Cette déclaration est le résultat d'un processus de coconstruction avec des citoyens et des experts rencontrés tout au long de l'année 2018. Voir : <https://www.declarationmontreal-iaresponsable.com>
- Démo** : Démonstration d'un prototype (d'agent conversationnel) en développement ou de versions préliminaires. Ces démonstrations sont faites pour présenter une idée à un potentiel client, pour montrer des évolutions à l'équipe ou encore pour faire des tests auprès de potentiels utilisateurs.
- DPC** : Data Protection Consulting. Cabinet de conseil en protection des données qui audite TMTL et ses partenaires en septembre 2019.
- DPO** : *Data Protection officer*. Personne responsable de l'engagement pour le respect de la vie privée dans une organisations qui traitent des données personnelles.

FATE : *Fairness, Accountability, Transparency and Explainability*. Nom du centre de recherche chez Microsoft qui étudie les implications complexes des techniques d'intelligence artificielle, d'apprentissage automatique et de traitement du langage naturel. Voir : <https://www.microsoft.com/en-us/research/theme/fate/>

IA : Intelligence Artificielle. « L'intelligence artificielle (IA) désigne l'ensemble des techniques qui permettent à une machine de simuler l'intelligence humaine, notamment pour apprendre, prédire, prendre des décisions et percevoir le monde environnant. Dans le cas d'un système informatique, l'intelligence artificielle est appliquée à des données numériques » (Dilhac et al., 2018, p. 19).

LPRPDE : Loi de protection des renseignements personnels et les documents électroniques. Loi canadienne qui encadre la protection des données personnelles, pilotée par le Commissariat à la protection de la vie privée du Canada.

NLP : *Natural Language Processing*, ou traitement du langage naturel. Ensemble de techniques en linguistique, informatique et intelligence artificielle qui permettent à un système informatique d'analyser des discours (parlés et écrits).

OBVIA : Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA). Organisme interuniversitaire créé au Québec en 2019. Voir : <https://observatoire-ia.ulaval.ca>

PaaS : *Platform as a Service*. Plateforme numérique qui donne accès à plusieurs services, tels que le stockage d'information, la gestion de bases de données ou encore des outils d'intelligence artificielle.
Par exemple, Google Cloud Platform, voir : <https://cloud.google.com>.

PbD : *Privacy-by-Design*. Modèle de développement technologique qui inclut le respect de la vie privée dès le début de la conception.

RGPD : Règlement Général de Protection des Données. Règlement européen entré en vigueur en mai 2018 qui encadre le traitement des données personnelles.

RSE : Responsabilité sociale des entreprises.

SaaS : *Software as a Service*. Dans la perspective agile du développement technologique, les logiciels sont proposés comme des services, ou SaaS : plutôt que de fonctionner du côté du client, le logiciel est hébergé dans une infrastructure infonuagique, c'est-à-dire sur des serveurs auxquels des clients peuvent être reliés pour utiliser le logiciel en question (Gürses & van Hoboken, 2018).

- TMTL** : Acronyme que j'utilise pour désigner l'organisation de promotion du tourisme à Montréal qui fait un partenariat avec Prélude pour le développement de l'agent conversationnel La Voix de Montréal.
- TTS** : *Text-to-Speech*. Technique d'intelligence artificielle qui permet de convertir du texte en discours parlé par l'intermédiaire d'une voix artificielle.
- UX** : *User Experience* ou Expérience utilisateur. Ce terme désigne la qualité de l'expérience qu'une personne (utilisatrice) fait d'un objet ou d'un service, physique ou numérique. Le design d'expérience utilisateur (*UX design*) est le processus par lequel une équipe conçoit des produits qui permettent une bonne expérience.
- VSD** : *Value-Sensitive Design*. Approche de mise en pratique de l'éthique dès la conception, développée au milieu des années 1990 par Batya Friedman et ses collègues (Friedman et al., 2013; Friedman & Hendry, 2019; Friedman & Kahn, 2003). Le VSD est marqué par la volonté de prendre en compte les valeurs humaines et le contexte social dans leurs processus de développement technologique.

REMERCIEMENTS

Une des surprises de l'exploration que je mène dans cette thèse a été pour moi de prendre conscience du rôle que jouent les conditions dans lesquelles on s'exprime, en particulier quand on parle de consentement. Avec une grande reconnaissance pour les conditions qui ont participé à faire de ce parcours doctoral une expérience si épanouissante, je tiens à marquer ma gratitude envers toutes les personnes qui m'ont accompagnée et soutenue ces six dernières années, dans l'ombre et la lumière, de loin ou de tout près, par intermittence ou quotidiennement.

À celui qui m'a ouvert la porte de son bureau et de sa curiosité en 2016, Boris, merci infini pour ton accompagnement dans cet Himalaya doctoral, le long de ces chemins tracés en marchant à tes côtés. Merci pour ton écoute et ta présence si attentives, dans tous les moments de la vie. Merci pour le soin pris à écouter mes silences, parfois mes peines, et regarder avec attention les espaces entre mes lignes. Écrire sous ta direction, c'est baigner dans un bol d'air qui régénère les poumons.

Merci aux personnes qui ont accepté de constituer mon jury aux côtés de Boris Brummans : Brenda Berkelaar, Lorna Heaton et François Cooren. Merci pour leur lecture et pour les vives discussions lors de la soutenance ! Merci également aux professeurs du département de communication de l'Université de Montréal dont j'ai croisé le chemin. François Cooren, merci d'avoir posé les idées de la ventriloquie qui ont sonné en moi comme une évidence en suivant le cours de début de parcours, tout en faisant vibrer une certaine intranquillité que j'ai fini par trouver à formuler dans cette thèse. Merci pour ces fondations. Merci à Chantal Benoît-Barné pour les remarques précises qui m'ont permis de redessiner mon projet pendant le forum doctoral. Merci à Stéphanie Fox pour la belle rencontre que cela a été de l'assister dans son cours en 2016-2017. Merci également à Juliette de Maeyer pour les fines remarques lors de la soutenance de mon projet, elles m'ont accompagnées. Merci aussi à Daniel Robichaud pour les aiguillages de début de parcours. Merci également aux discussions fortuites, peu fréquentes mais vivifiantes avec Thierry Bardini et Line Grenier.

Mon cheminement a bénéficié du précieux soutien d'organismes subventionnaires (FRQSC, Département de communication de l'Université de Montréal, Mitacs, bourse Google de la Faculté des Arts et Sciences sur les impacts sociaux de l'intelligence artificielle). Merci pour ce soutien qui m'a permis de développer mes idées avec une certaine paix d'esprit financière. De mon éducation à mon financement, j'ai une profonde gratitude envers les conditions qui m'ont permis de développer de façon privilégiée le goût et l'exercice de la pensée, et d'écrire en donnant le temps à mes idées de se (dé/re)former. Cela peut d'ailleurs sembler lointain, mais ayant pris goût à la pensée relationnelle qui peut révéler de profondes ramifications, j'aimerais remercier les enseignantes et professeur-e-s qui ont semé dans ma tête l'idée d'un jour faire de longues études, voire même un doctorat, de Mme Guihoux en dernière année de maternelle, à Messieurs Guchet et Bozzo-Rey en master de philosophie, en passant par Mme Morin en

fin de primaire, Mme de Negroni et M. Sfez en classe préparatoire littéraire. Merci de m'avoir fait découvrir le plaisir de toujours repousser les limites de ma pensée.

Cette thèse reconstitue une histoire qui a émergé de mon expérience avec Mr Y et Prélude en 2019 – prélude d'une réflexion sur le consentement qui a enrichi tant de facettes de ma vie. À l'équipe de Mr Y et Prélude cette année-là, merci de m'avoir permis de constituer de riches matériaux à partir des moments partagés avec eux. Merci en particulier à Paul – qui se reconnaîtra derrière le pseudonyme – merci pour son ouverture et son engagement dans cette collaboration, sa réceptivité aux activités des chercheurs, son sourire à toute épreuve. Merci à la coïncidence qui a amené nos chemins à se croiser. Merci également à Félix de s'être prêté au jeu !

Avant de m'immerger chez Mr Y et Prélude, je suis tombée avec grand enthousiasme dans les questions sociales de la technologie par deux expériences. En 2016, j'ai commencé mon doctorat en même temps qu'à la Société des Arts Technologiques [SAT], dans le *living lab* lié au Métalab. Merci en particulier à Patrick qui m'y a introduite, et à Simon et Claire qui m'ont initiée à la pensée design et la pratique de la recherche en expérience usager.

Puis fin 2017, des retrouvailles avec Christophe Abrassart m'ont permis d'entrer dans la formidable aventure de la Déclaration de Montréal. Un très grand merci à l'équipe de la Déclaration de Montréal, en particulier Anne-Marie Savoie, Christophe Abrassart, Marc-Antoine Dilhac, et mes acolytes devenues amies, Nathalie Voarino et Valentine Crosset. Je n'aurais pas pu entrer plus intensément dans l'éthique de l'intelligence artificielle et ses préoccupations, merci à toute l'équipe pour cette aventure à Montréal, Québec et Paris, puis les codes NVivo et les réjouissants apéro.

Mon écriture a fait son chemin dans un monde aux bifurcations incertaines : elle a commencé avec le début de la pandémie de Covid-19, a été interpellée par le ravivement du mouvement *Black Lives Matter*, perturbée par les découvertes de corps d'enfants autochtones aux alentours de pensionnats, s'est noyée dans deux épisodes qui m'ont fait rencontrer le soin accordé au consentement par le prisme de la santé, puis a cherché sa fin quand la guerre en Ukraine a éclaté par stupeur, les rapports du GIEC en fond alarmant quant à l'urgence de prendre soin des conditions climatiques qui questionnent la possibilité de nos vies et celles qui viennent au monde.

Merci aux peuples autochtones pour le soin qu'ils portent depuis toujours aux territoires, en particulier celui non-cédé de Tiohtià:ke. Je reconnais le privilège que j'ai eu de venir m'installer ici, à Tiohtià:ke/Montréal, et de m'y épanouir, alors que tant de peuples ici depuis toujours résistent aux forces colonialistes. Je tiens à continuer d'affûter ma pratique réflexive pour ne plus faire perdurer ces dynamiques à mon échelle, en continuant d'apprendre de ces multiples savoirs autochtones.

Merci aussi aux penseuses qui me permettent de trouver des prises et de l'espoir quand les choses semblent s'effondrer – Manon Garcia (2019, 2021) et sa philosophie si vive du consentement, Alexis Shotwell (2016) et sa pensée percutante de la complicité dans la complexité, Pema Chödrön (2016) et son invitation à sombrer sans découragement, John Tarrant (1999, 2008) et sa lumière dans l'obscurité.

Dans ces turbulences, merci aux personnes qui me permettent de prendre soin de moi et d'affiner l'écoute de mes silences. Les professeur·e·s de Yoga Bhavana et Espace Metta grâce à qui mon corps et mon système nerveux trouvent bien des alliés. Aux soignantes de l'hôpital sainte Justine et leur formidable sens du consentement, malgré tant d'autres urgences derrière leurs portes. Et un merci tout intime à Sandra D. et Marie-Claire L.-B. pour les espaces laissés au vagabondage de mes pensées, pour les belles surprises de constater mes paroles changer, mes horizons s'ouvrir et mon écriture se délier.

Dans cette traversée, merci à mes proches, ami·e·s et famille pour le partage de nos inquiétudes, nos espoirs, nos tristesses, nos lumières et surtout nos (sou)rires à toute épreuve.

À mes ami·e·s; d'abord mes muses féministes et queer, Justine (Laurent), mon merci est au moins aussi grand que le Amsterdam-Montréal parcouru frénétiquement par nos messages vocaux. Mon cœur et ma thèse débordent de reconnaissance pour ton regard si vivace qui m'a réorientée vers le *care* quand ma flamme s'affaiblissait. Renata (Azevedo Moreira), merci d'être entrée dans ma vie comme une douce tempête. Merci de m'entraîner avec ton féminisme queer audacieux et ton sens de la fête.

Agathe (François), un grand merci pour la compagnie doctorale, la sensibilité de ton écoute au fil des années, la finesse de tes lectures jusqu'à la ligne d'arrivée. Merci aussi pour tes dessins qui font voguer l'imaginaire dans la beauté des questions existentielles.

Merci à mes ami·e·s d'éthique de l'IA et maintenant au-delà. Valentine, merci pour la relecture. Nathalie, merci pour l'effervescence. Toutes les trois, nos discussions sont tellement stimulantes que je voudrais qu'elles soient sans fin. Carl, merci pour les pistes de collaborations et l'humour qui rappelle la maison.

Merci aux rencontres qui ont accompagné le chemin doctoral : Geneviève, Pascale, Lise, Myriam, Tara, Régine, merci d'avoir montré la voie. Maïssa, Simon, Khaoula, Marie-Ève, Clément, Justine P., Valérie, Cynthia, Félix, Lucie, merci pour le compagnonnage de route. Et plus largement, un très précieux merci à la communauté Thésèz-vous pour le soutien bienveillant et tellement motivant au fil du marathon de 1310 pomodoros de rédaction; au lumineux espace de rédaction dans Villeray, lors de puissantes retraites et de cohortes solidaires pendant les confinements. Merci aux belles rencontres et heureuses retrouvailles. En particulier, Lucie et Caroline, vous retrouver régulièrement à l'Espace lors des dernières phases d'écriture a été une réjouissance de plus !

À mes ami·e·s d'ici à Montréal, Marie, Viola, PA, merci pour le ressourcement des échappées au chalet et de celles parmi les tendres céramiques. *A los amigos, gracias for the warm invitations to follow your heart, to dance both on reggaeton and electro, and laugh with great simplicity.*

Et mes ami·e·s de cœur, de l'autre côté de l'océan, merci pour l'amour, la douceur et la chaleur quand on se retrouve chaque été. Mes amies de toujours, Chloé et Claire,

merci pour le soutien à toute épreuve et pour nos bientôt 30 ans d'amitié. Florent, merci pour la complicité et l'humour salvateur. Manon, Hélène, William et Leïla, doux merci.

Enfin, un merci débordant à ma famille et leurs regards pleins de soutien au fil de ces années. À ma sœur et mon frère qui m'inspirent, Charlotte par le courage de sa liberté, Timothée par son audace de viser toujours plus haut. À mes parents Marie et Pierre pour leur écoute attentive et leur soutien inconditionnel tout en sensibilité, merci d'accueillir mes flots de questions depuis toujours pour laisser fleurir ma curiosité. Merci à mes grands-mères, Simone et Odile, femmes dont la force traverse plusieurs générations, dans toutes leurs situations. Merci à mon grand-père, Louis, pour l'inspiration de sa témérité au travers de ses engagements pour contribuer à un monde de solidarité. En hommage à mon autre grand-père, Michel, avec qui j'aurais aimé avoir plus de conversations.

Un tendre merci à mon Thibaut, qui m'accompagne en toute simplicité depuis presque 10 ans. Merci pour les éclairages informatiques et les rappels à l'ordre quand nos tiroirs commencent à déborder d'un peu trop de plastique (ça m'amuse de voir notre mode de vie faire de plus en plus de place à l'éthique). Et surtout, merci d'accueillir chaque sourire, chaque agacement, chaque soubresaut, si simplement.

Et puis,

À toutes celles qui ont crié #moiaussi dans le silence de leurs voix étouffées. Aux personnes qui ont parlé et celles qui se sont senties résonner, au courage de cette vulnérabilité qui fait bouger les lignes de nos soumissions. Merci pour l'inconfort que vous avez provoqué. Ma thèse vous est dédiée, en toute humilité.

Moi aussi

“To live our lives based on the principles of a love ethic (showing care, respect, knowledge, integrity, and the will to cooperate), we have to be courageous. Learning how to face our fears is one way we embrace love.

Our fear may not go away, but it will not stand in the way. Those of us who have already chosen to embrace a love ethic, allowing it to govern and inform how we think and act, know that when we let our light shine, we draw to us and are drawn to other bearers of light. We are not alone.”

Bell Hooks, All about love, “Values: living by a love ethic”

Hurry slowly¹

¹ Boris H.J.M. Brummans, mantra pour rythmer l'écriture et la vie

INTRODUCTION

Providence, Rhode Island School of Design, le 9 novembre 2019. Je participe à la conférence EPIC² qui réunit une communauté d'ethnographes qui appliquent leurs méthodes dans différentes industries. Je profite de ces quelques jours pour ouvrir un peu plus mon champ de vision, car je suis depuis plusieurs mois en plein travail de terrain avec Prélude, une startup qui développe un agent conversationnel. Les agents conversationnels sont des systèmes d'intelligence artificielle avec lesquels parler textuellement (*chatbot*) ou vocalement (*voicebot*, ou assistant vocal). J'accompagne les membres de Prélude notamment pour prendre en compte le respect de la vie privée dès le début du design de leur assistant vocal. En lien avec cela, je travaille avec eux sur la façon d'informer les utilisateurs et demander leur consentement à ce qui sera fait de leurs données personnelles, l'objectif étant d'informer et de demander ce consentement par l'intermédiaire de la conversation avec l'agent.

Le panel qui ouvre la conférence s'intitule « *Reconceptualizing Privacy* ». Il est introduit ainsi :

Algorithmic systems are increasingly integrated into the physical and digital infrastructures of our lives. The borders of privacy are being pushed and redefined, provoking new debate about what privacy is. All corporations claim privacy is important, but what does that mean? Panelists will consider what privacy might look like or mean when individuals are tied into multiple networks, both human and AI³.

Le panel rassemble un chercheur de chez Intel Corporation, la *head of privacy research* (directrice de la recherche sur le respect de la vie privée) de Facebook, un chercheur chez Autodesk, une chercheuse senior en expérience utilisateur qui dirige la recherche en accessibilité et respect de la vie privée chez Uber et un designer de IBM Watson Health. Je suis particulièrement curieuse de ce que ces personnes vont pouvoir partager : du point de vue de leurs grandes organisations, est-ce que leurs remarques et pratiques vont résonner avec mon expérience de ces enjeux dans une petite startup ? Ont-ils des solutions alors que j'ai l'impression d'avancer à tâtons dans le brouillard de ces questions ?

L'animateur du panel (le chercheur chez Intel) invite d'emblée à mettre l'épanouissement humain au cœur de la réflexion sur le respect de la vie privée : « *What's interesting to me about privacy, and also AI ethics, is how we take what we've been doing and shift our concern and our work to focus more on human flourishing rather than just around UX* »⁴. Cela me parle : après tout, mon

² Voir <https://www.epicpeople.org>

³ Voir <https://www.epicpeople.org/reconceptualizing-privacy/>. Tout au long de la thèse, par convention, je mets en italiques les citations et mots dans une autre langue que le français.

⁴ Voir <https://www.epicpeople.org/reconceptualizing-privacy/>, vidéo, ken anderson, 00'55''.

entêtement pour que Prélude s'attarde sur ces questions de vie privée, ce n'est pas seulement un souci d'expérience utilisateur qui cherche une façon originale de demander le consentement de façon conversationnelle. C'est plutôt parce que cela touche à une question d'épanouissement humain : comment respecter l'humain au cœur de ces données (personnelles ou non), pour éviter des effets néfastes et privilégier son développement ?

Les panélistes mentionnent plusieurs choses que je rencontre avec Prélude : l'intention de faire savoir aux utilisateurs que l'on prend le respect de la vie privée au sérieux, la clarté parfois difficile à faire sur l'ensemble des données que l'on gère, la vigilance à porter aux possibilités de réidentification des personnes malgré l'anonymisation des données, la responsabilité à propos des données que l'on s'apprête à utiliser, le fait de donner du contrôle aux utilisateurs envers les données qui les concernent, l'impossibilité pour les utilisateurs de tout comprendre (leur consentement peut difficilement être parfaitement informé), comment éduquer sur ces enjeux, favoriser la confiance – non par des campagnes de marketing mais par l'implantation d'approches comme le *privacy-by-design* qui prend en compte les enjeux de respect de la vie privée dès le début du design, plutôt qu'en réaction quand un accident arrive. La personne de chez Facebook souligne également la présence de différentes façon de comprendre ces enjeux de respect de la vie privée selon les personnes et les cultures; par ailleurs, la distinction entre privé et non-privé est floue, il peut être ambigu de considérer certaines informations partagées publiquement comme n'étant pas privées. De nombreux enjeux sont ainsi soulevés, mais plutôt que des solutions, c'est la complexité du travail quotidien de ces questions qui est partagée.

Je rentre à l'hôtel après quelques discussions de réseautage. Au passage piéton, une dame est à côté de moi, elle attend son chauffeur Uber. C'est la panéliste de chez Facebook. Je lui fais part de mon enthousiasme d'entendre les enjeux soulevés par le panel, cela résonne avec ma propre expérience à plus petite échelle. Son Uber arrive et elle me dit : « *Oh, I'm glad you're advocating about that!* ». Cette phrase me laisse un peu perplexe : ce n'est pas vraiment que j'*advocate*, que je plaide en faveur de cela, c'est plutôt que je suis là avec cette startup pour les aider à mettre en œuvre cette nécessité qu'est le respect de la vie privée de leurs futurs utilisateurs.

(Vignette EPIC2019, « *I'm glad you're advocating about that!* », à partir du journal de bord, notes de terrain du 9 novembre 2019)⁵

⁵ Tout au long de la thèse, je cite le matériel collecté lors de mon travail de terrain avec Mr Y et Prélude de trois manières : (1) par des vignettes qui reconstituent des événements spécifiques à partir de mes notes de terrain à ce propos (elles peuvent condenser les notes de plusieurs jours et évoquer d'autres documents ou événements); (2) par des extraits de mon journal de bord, dans ce cas ce sont les notes de terrain ou les notes réflexives écrites dans le journal qui sont copiées telles qu'elles, non-reconstituées; (3) par des retranscriptions d'enregistrements de réunion.

Les enjeux mentionnés dans cette vignette sont au cœur de mon immersion chez Prélude : prendre au sérieux le respect de la vie privée des utilisateurs et le faire savoir sans que cela relève de *l'ethics-washing* (Metcalf et al., 2019) qui communique sur l'éthique de façon superficielle, sans que des efforts soient véritablement faits pour mettre en place des mesures éthiques; clarifier les pratiques de traitement des données personnelles, malgré de nombreuses incertitudes quand l'outil n'est qu'au début de son développement; rappeler la responsabilité de l'organisation à l'égard de ces données personnelles quand elle en traite; donner du contrôle aux utilisateurs envers les informations qui les concernent; permettre aux utilisateurs de saisir ces enjeux, malgré l'impossibilité de tout comprendre étant donné la complexité de l'écosystème technologique; favoriser la confiance par une approche préventive de *privacy-by-design*, bien que celle-ci amène son lot de défis pour être mise en place (Gürses & van Hoboken, 2018). Cette thèse aborde tous ces enjeux liés au respect de la vie privée sous un angle spécifique : celui du consentement, qui est le moment d'interaction que je travaille avec Prélude, d'avril à décembre 2019.

Ce projet autour du consentement s'inscrit dans un appel à réinventer les modèles par lesquels les organisations informent et demandent le consentement de leurs utilisateurs (Cate & Mayer-Schonberger, 2013). Il s'agit de consentir au traitement des données personnelles que l'utilisation du système numérique implique. Le terme de « données personnelles » est un terme parapluie qui désigne de nombreux types d'informations qui se rapportent à une personne et permettent de l'identifier, soit de manière directe (par son prénom et son nom, par exemple), soit de manière indirecte (par un numéro de téléphone, un identifiant, une photo, une donnée biométrique comme sa voix ou son empreinte digitale, sa géolocalisation, des éléments spécifiques) (RGPD, 2018, Article 4). Elles font l'objet d'attention particulière car, dans nos sociétés contemporaines de plus en plus centrées sur l'exploitation de tout type de données, leur traitement peut impacter les personnes de multiples façons : surveillance généralisée, réputation, divulgation, influence de la prise de décision – comme l'a mis de l'avant le scandale Facebook-Cambridge Analytica en 2018 en révélant, entre autres, l'influence sur des

élections (américaines en 2016, Brexit) qu'a permis le traitement massif de données personnelles dont la plupart étaient collectées sans consentement, à l'insu des utilisateurs.

L'appel à réinventer les modèles de consentement vient des règlements de protection de données personnelles – en particulier la Loi de Protection des Renseignements Personnels et les Documents Électroniques au Canada (LPRPDE) et le Règlement de Protection des Données (RGPD) en Europe qui est entré en vigueur en mai 2018. Cet appel vient également du champ de l'éthique de l'intelligence artificielle, qui se constitue avec une effervescence accrue au fil de publications de déclarations de principes depuis 2017 (Fjeld et al., 2020). Ces déclarations – comme la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle (Dilhac et al., 2018), à laquelle j'ai contribué l'année précédente, en 2018 – mettent un point d'honneur à donner plus de pouvoir aux personnes qui utilisent des objets ou services numériques. Elles insistent également à responsabiliser davantage les organisations qui traitent des données personnelles, étant donné les effets politiques et sociaux que ces traitements impliquent – ces effets peuvent relever par exemple d'impacts néfastes sur la santé mentale des jeunes et la diffusion de contenus violents, comme l'ont révélé les *Facebook Files* en 2021 à propos des plateformes de Facebook (dont Instagram) (Wells et al., 2021; Hagey & Horwitz, 2021).

Dans une relation asymétrique où plane une certaine opacité ou un certain flou sur ce qui est fait avec les données personnelles par les organisations qui les collectent, donner plus de pouvoir aux personnes qui égrènent ces données au fil de l'utilisation qu'elles font des produits (objets connectés, services de *cloud*, réseaux sociaux, sites internet, agents conversationnels, etc.), cela passe par l'effort de les informer clairement, et soumettre le traitement de ces données à leur accord (ou consentement). Je m'engage ainsi avec Prélude dans une tentative de proposer un modèle de consentement par agent conversationnel, qui informe clairement les utilisateurs de ce que l'organisation fera de leurs données personnelles, et qui permette à ces utilisateurs de donner (ou refuser de donner, voire même retirer) leur consentement par l'intermédiaire de l'interaction vocale

des agents que Prélude développe. Cette tentative navigue cependant dans de nombreuses ambiguïtés que cette thèse explore : quelles informations et quelle quantité d'information donner ? Comment intégrer cette information fluidement dans le reste de l'expérience proposée ? Qu'est-ce que Prélude et ses partenaires vont faire avec ces données personnelles (ou quelles seront les pratiques de traitement de ces données) ?

Initialement, c'est avec une autre organisation que je travaille sur ce projet autour du consentement : Mr Y⁶, une startup qui développe un chatbot en santé mentale, pour aider ses utilisateurs à gérer leur anxiété. Mais début 2019, quelques mois avant que mon immersion chez Mr Y commence (en mars), Paul, fondateur de Mr Y, s'associe à Félix pour créer Prélude, une startup qui veut développer des agents conversationnels vocaux, dans différents domaines, de la santé mentale avec le chatbot Mr Y, au tourisme. Le tourisme devient le domaine principal d'activité de Prélude quand un premier client s'engage pour financer la Voix de Montréal, un assistant vocal qui guide les touristes dans la ville selon des parcours personnalisés, à saveur locale. Ce client est TMTL, un organisme de promotion du tourisme à Montréal. Le glissement de Mr Y à Prélude, de la santé mentale au tourisme, affecte le travail sur le consentement : celui-ci peut sembler moins crucial étant donné que la relation n'est plus une relation de soin envers des personnes en situation vulnérable comme peuvent l'être des personnes en crise d'anxiété, les données personnelles que Prélude va traiter seront probablement moins sensibles que celles que traiteraient Mr Y, puisqu'il ne s'agit pas d'informations liées à la santé.

Mais plutôt que la sensibilité des données, c'est autre chose qui se met à compter tout autant : la responsabilité de l'organisation qui prévoit de traiter des données liées à des personnes dans leur quotidien, dans l'intimité de l'usage d'un agent conversationnel. En maintenant l'attention sur le travail du consentement qui implique en même temps de clarifier les intentions de l'organisation à l'égard de ses pratiques de traitement des données personnelles, la *Head of privacy research* de chez Facebook avait peut-être raison finalement à la conférence EPIC : c'est peut-être une forme d'*advocacy* que je fais. Mon immersion chez Mr Y et Prélude se fait à la lumière d'une insistance, qui

⁶ Pseudonyme, prononcer « *Mister Why* ».

plaide en faveur d'une meilleure prise en compte du consentement dans le design des agents conversationnels, et œuvre en ce sens.

Cette insistance porte en filigrane sur autre chose, dans le sillon du mouvement #MeToo/#MoiAussi en 2017, qui a décuplé les voix exprimant des expériences de consentement négligé, bafoué, voire violé. Dans ce mouvement, c'est surtout l'usage sexuel du consentement qui est à l'avant-plan. Mais dans cet usage, c'est une dynamique de communication qui se joue pour que triomphe le respect dans la relation entre deux personnes, plutôt que la domination qui réduit l'autre en soumission (Garcia, 2019, 2021). Alors une question teinte mon immersion chez Mr Y et Prélude qui, d'une manière numérique sans rapport au sexuel, travaille des types de conversations quotidiennes que l'on peut avoir avec des agents artificiels : comment mieux prendre en compte le consentement dans toute conversation ?

Cette thèse réunit ces deux préoccupations du respect du consentement, au nom des personnes présentes dans les données qu'un agent conversationnel traite, et au nom de l'égalité plutôt que la domination dans toute conversation et relation quotidienne. Cette réunion se fait par une attention particulière : une attention qui cherche à prendre soin de situations où des choses (comme le consentement) peuvent être négligées, abîmées, ignorées, et amener des personnes à ne pas être respectées. Ces situations appellent une intervention que je désigne sous la notion de *matter of caring* : un *matter of concern* (préoccupation) qui appelle une intervention pour « maintenir, perpétuer et réparer notre "monde", en sorte que nous puissions y vivre aussi bien que possible » (Fischer & Tronto, 1991, p. 40, cité dans Tronto, 2008. Voir chapitre 4). La posture qui anime mon immersion ethnographique au sein de Mr Y et Prélude – puis l'analyse du matériel constitué pendant ce travail ce terrain, jusqu'à l'écriture de cette thèse – relève ainsi du « *careful engagement* » (Nickelsen & Lydahl, sous presse). Il s'agit d'un engagement soucieux d'exprimer des préoccupations éthiques au fil du développement de l'agent conversationnel afin que la valeur éthique du consentement (et du respect de la vie privée) le constitue.

À partir de cet engagement, les questions de recherche qui guident cette thèse explorent comment se fait le travail du consentement chez Mr Y et Prélude, comment ce travail s'est constitué comme *matter of caring*, et quels effets celui-ci a fait dans ces organisations. Ces questions sont : (QR1) Comment cela se fait, d'aborder le consentement comme un *matter of caring* dans une organisation qui design un agent conversationnel ? et (QR2) Qu'est-ce que cela fait de faire du consentement un *matter of caring* ? Plonger dans ces questions révèle de nombreuses ambiguïtés, tensions et limites, amenant à parfois douter de la pertinence de s'attarder sur le consentement tant il peut gêner les activités de l'organisation (voir chapitre 6 en particulier). Alors plutôt que de céder à la tentation de le mettre de côté, une problématique plus générale, qui déborde de mon expérience de terrain avec Mr Y et Prélude, guide les trois premiers chapitres de cette thèse : pourquoi prendre soin du consentement, en particulier dans le design d'un agent conversationnel ?

Orientation des chapitres

Le premier chapitre pose le contexte éthique dans lequel s'inscrit cette question et cette thèse : il introduit le domaine de l'éthique de l'intelligence artificielle (IA) (Gibert, 2019; Müller, 2020) dans lequel cette réflexion se situe. Il précise l'approche que j'adopte : une approche sociomatérielle (Cecez-Kecmanovic et al., 2014; Clegg et al., 2013; Fenwick, 2014; Leonardi, 2011, 2012; Orlikowski, 2000, 2007; Orlikowski & Scott, 2014; Schultze et al., 2020; Suchman, 2002, 2007) qui consiste en une plongée pragmatique pour comprendre l'éthique non par ce qu'elle dicte et évalue mais par sa pratique au quotidien (Matte et Bencherki, 2019). Ce premier chapitre se concentre en particulier sur des pratiques où l'éthique est abordée par l'angle du design – où le design est compris comme l'ensemble des activités liées à la conception et la réalisation d'un projet (Vial, 2010, 2021). L'éthique *by design* se présente comme une façon de s'immiscer dans le développement technologique (F. Fischer, 2019, 2020), que ce soit par l'approche du *value-sensitive design* (Friedman et al., 2013; Friedman & Hendry, 2019; Friedman & Kahn, 2003), du *privacy-by-design* (Cavoukian, 2011, 2012) ou encore de la

postphénoménologie qui accompagne la médiation de valeurs par les technologies (Ihde, 2009; Rosenberg & Verbeek, 2015; Verbeek, 2011). Ce chapitre permet de me positionner dans les études organisationnelles de l'éthique par l'angle du design pour préciser ce que je fais : ce qui m'intéresse, c'est d'étudier l'éthique dans sa performativité quotidienne, non comme quelque chose que l'on évalue a posteriori mais comme un phénomène produit de façon active au fil de pratiques continues de développement technologique, de design et de gestion, dans des tensions et complexités de l'*organizing* (Weick, 1979), c'est-à-dire l'organisation en train de se faire. Cette étude de l'éthique s'inspire de la philosophie empirique « *aiming to study facts, concerns, ideas of good and values produced in situ* (Pols, 2015; Pols, Pasveer, & Willems, 2017) »⁷ (Nickelsen & Lydahl, sous presse). Ces « faits, préoccupations, idées de ce qui est bon, et valeurs » produisent, déplacent et modifient des normes – c'est-à-dire ce qui est communément admis comme « bon ». Étudier la performativité quotidienne de l'éthique, par le design, revient ainsi à porter attention à l'activité normative qui émerge et évolue au fil des pratiques quotidiennes et des négociations de ce qui est préférable ou bon.

Dans mon immersion qui travaille l'éthique dans les activités de design des agents conversationnels de Mr Y et Prélude, plusieurs versions du « bon » se rencontrent et se confrontent : selon le champ de l'éthique de l'IA, c'est l'expression de plusieurs valeurs (comme le respect de la vie privée et du consentement) qui permet de dire que l'expérience d'un système d'IA est bonne; selon le champ organisationnel relatif au développement d'une expérience technologique conversationnelle, c'est plutôt l'agréable, la fluidité et une certaine rapidité qui qualifient une bonne expérience utilisateur (Norman, 2013 [1988]). Toutes ces versions ont un même but : offrir aux personnes qui utiliseront cet outil technologique une bonne expérience. Mais qu'est-ce qu'une *bonne* expérience ? À la fois question morale, question organisationnelle, question de design, cette question ouvre les deuxième et troisième chapitres.

⁷ Ces références évoquent les travaux de Jeanette Pols et ses collègues qui étudient l'intra-normativité de l'éthique dans les pratiques de *care* en santé : « *Empirical ethics studies care practices by analysing their intra-normativity, or the ways of living together the actors within these practices strive for or bring about as good practices* » (Pols, 2015, p. 81).

Le deuxième chapitre explore ce qu'est une bonne expérience avec un agent conversationnel. Il présente les outils et les étapes de design d'un agent conversationnel pour parvenir à le faire parler. En écho, le troisième chapitre demande ce qu'est une bonne expérience de consentement. Il expose ce qui peut permettre au consentement d'être bon – c'est-à-dire valable – aux yeux des règlements de protection de données personnelles, mais également de raisonnement philosophiques (en philosophie morale et politique).

Ces chapitres 2 et 3 décrivent chacun les lignes directrices, les points de rencontres et les éventuelles tensions entre les pratiques normatives en éthique de l'IA – c'est-à-dire celles qui évaluent ce qui est fait et conseillent ce qui doit être fait –, ainsi que les pratiques d'usage et de conception des systèmes d'IA tels que des agents conversationnels. Pour le design d'agent conversationnel (chapitre 2) comme pour le consentement quand il fait l'objet de design (chapitre 3), il s'agit de montrer ce qui se fait ou ne se fait pas, ce qui est habituel, approprié ou non, selon ces différentes pratiques et leur version du bon. Dans ces chapitres qui se demandent ce qu'est une bonne expérience d'agent conversationnel (chapitre 2) et de consentement (chapitre 3), il s'agit de déceler une certaine « logique », c'est-à-dire « la rationalité ou l'inspiration qui font tenir les pratiques » ou encore « une cohérence locale, fragile et à ce moment-là pertinente » qui peut être implicite : « intégrée dans les pratiques, les bâtiments, les habitudes et les machines » (Mol, 2013, p. 12). Le chapitre 3 souligne la cohérence, en même temps fragile mais pertinente du consentement malgré de nombreux obstacles auxquels il se confronte (Edenberg & Jones, 2019; Jones & Edenberg, 2020; Kim, 2019; Richards & Hartzog, 2019). C'est cette cohérence vulnérable qui ancre la raison d'être du consentement, en particulier au nom de l'horizon éthique qu'il esquisse : un horizon d'épanouissement par respect pour l'égalité qu'il rappelle (Garcia, 2021; Nussbaum, 2011, 2013).

Déceler la logique inhérente aux conceptions de bonne expérience d'agent conversationnel et de consentement éveille alors en même temps la curiosité envers un autre ordre possible, qui puisse mieux s'inscrire dans cet horizon : comment pourrait-il en être autrement ? Par exemple, (chapitre 2) comment pourrait-on faire parler autrement

les agents conversationnels quand ceux-ci nuisent aux personnes d'un certain genre (Deibel & Evanhoe, 2021; Strengers & Kennedy, 2020; UNESCO & EQUAL Skills Coalition, 2020) ? Ou (chapitre 3) comment pourrait-on parler autrement de l'inconfortable inhérent au consentement (Garcia, 2021) ? Cette intranquillité est une facette du questionnement éthique qui interroge les ordres établis et cherche à faire bouger les forces normatives, dans la perspective d'une certaine amélioration du monde.

Le quatrième chapitre forge alors le concept de *matter of caring*, pour qualifier l'attention portée sur des objets négligés, mis de côté, parfois ignorés – comme peut l'être le consentement quand les activités d'une organisation sont portées sur le design d'une bonne expérience avec un agent conversationnel. Ce concept s'inscrit dans le sillon de l'éthique du *care* (Gilligan, 2003; Tronto, 2008) et du concept de *matter of care* que Maria Puig de la Bellacasa (2011) a forgé à partir des *matters of concern* de Bruno Latour (2004). L'attention de l'ordre du *care* portée sur le consentement implique une éthique vigilante envers l'activité qui consiste à faire parler, en particulier quand il s'agit de faire s'exprimer un consentement. L'expression d'un consentement est en effet indissociable des conditions dans lesquelles il est demandé, et dans lesquelles il est possible de le donner ou le retirer (Kim, 2019). Or, ces conditions peuvent entraver la personne supposée l'exprimer : si ces conditions influencent, forcent le consentement, si la personne est sous emprise de quelque chose ou quelqu'un qui la fait parler selon ses propres intérêts, alors ce qui fait s'exprimer le consentement ne respecte pas la personne qui le donne. Prendre soin de ces conditions invite ainsi à entretenir une certaine vigilance envers ces conditions parfois indésirables ou problématiques qui amènent quelqu'un à exprimer son consentement – des conditions qui ouvrent la voie à un mal qui peut lui être fait, puisqu'elle ne sera pas respectée. Ou, dans les termes de l'approche ventriloque de la communication constitutive des organisations (CCO) (Cooren, 2013), il s'agit d'être vigilant envers les figures indésirables ou problématiques qui peuvent faire parler un consentement au nom de quelque chose qui entrave la personne qui l'exprime. Prendre soin du consentement – en faire un *matter of caring* – trouve alors encore plus de pertinence dans cette attention éthique portée sur les conditions (et figures) par

lesquelles on s'exprime, par lesquelles plusieurs choses ou personnes nous font parler. Alors, chez Mr Y et Prélude, (QR1) comment cette attention, de l'ordre du *matter of caring*, s'est-elle portée sur le consentement au fil du design des agents conversationnels ? Et (QR2) quels effets cela a fait ?

Les cinquième et sixième chapitres explorent plus en profondeur mon expérience avec Mr Y et Prélude à la lumière de ces questions. Le chapitre 5 se concentre sur mes pratiques de recherche ethnographique (Neyland, 2008; Tracy, 2019; Ybema et al., 2009) pour montrer comment le *matter of caring* a été présent, en particulier par ma réflexivité (Cunliffe, 2016) et ma complicité (Shotwell, 2016) qui ont permis à l'intervention de prendre la texture d'un tissage relationnel (Ingold, 2011), participant à la constitution de l'organisation. Il expose en même temps les limites que le *matter of caring* m'a amenée à me poser dans la constitution et l'analyse de mon matériel, en particulier en ce qui concerne les enregistrements audio, et les tensions vulnérables que j'ai mises de côté.

Le chapitre 6 est articulé exclusivement autour de mon expérience avec Mr Y et Prélude, et le travail du consentement avec eux. Il expose des vulnérabilités (Brugère, 2013) à l'œuvre au fil d'interventions pour améliorer des situations où le consentement est mis de côté, négligé, reformulé puis intégré dans le design de l'agent conversationnel — c'est-à-dire au fur et à mesure que le consentement fait l'objet d'un *matter of caring* qui sème un certain trouble inconfortable. La tension principale est celle inhérente à la tentative de faire fusionner le consentement dans l'expérience de l'agent conversationnel : tandis que le consentement présente le risque de heurter cette expérience en l'alourdissant ou la ralentissant, l'allègement et la rapidité que vise une bonne expérience d'agent conversationnel risquent d'empêcher le consentement d'être valable. À partir de cette principale tension, ce chapitre analyse plusieurs compromis auxquels le travail sur le consentement aboutit au fil de ma complicité avec Mr Y et Prélude, dans la complexité des enjeux d'éthique et de design que leurs activités soulèvent.

Contributions

L'exploration menée dans cette thèse contribue aux niveaux empirique, théorique et méthodologique. À un niveau empirique, elle étoffe les travaux en éthique de l'IA en montrant les coulisses de la gestion de questions d'éthique au fil de pratiques de développement d'un système d'IA comme un agent conversationnel. En adoptant une approche pragmatique, mon enquête contribue aux travaux en éthique *by design* en mettant en lumière comment un enjeu éthique comme le consentement (et le respect de la vie privée) se constitue au fil de pratiques organisationnelles, dans leurs ambiguïtés et tensions. Cela contraste avec une approche de l'éthique de l'IA qui se concentre sur l'évaluation des pratiques dans un temps ultérieur, une fois que le système d'IA est développé et utilisé. Cela contraste également avec une approche de l'éthique de l'IA prescriptive comme le font les nombreux rapports et déclarations de principes dont l'objectif est de faire des recommandations pour encadrer le développement de l'IA.

Empiriquement également, l'enquête ethnographique au cœur de ma thèse dévoile les nuances et complexités qu'implique une interaction simple comme cliquer sur un « ok » ou dire « ok ». Elle insiste ainsi sur la responsabilité inhérente aux décisions de design quand elles portent sur les conditions (Kim, 2019) et affordances (Leonardi, 2011; Norman, 1999) par lesquelles on demande à quelqu'un son accord.

Par l'accent pragmatique que mon exploration met sur les processus constitutifs du consentement dans les organisations Mr Y et Prélude, ma thèse s'inscrit – théoriquement – dans le courant de la communication constitutive des organisations (CCO) qui étudie comment des collectifs humains et autres qu'humains viennent à s'ordonner et se désordonner au fil de leurs interactions (Baillargeon et al., 2021; Boivin et al., 2017; Brummans et al., 2014). Son intérêt porté sur l'éthique résonne avec les travaux de Matte et Bencherki (2019) qui s'intéressent à la matérialisation d'*ethical matters of concern*. Mais elle fait quelque chose de différent : plutôt que d'observer comment des préoccupations éthiques se mêlent à des préoccupations d'autres ordres organisationnels et technologiques, et comment ces entremêlements sont constitutifs de l'organisation étudiée, mon exploration est animée par l'intervention. Elle cherche à

montrer pourquoi et comment faire compter (*matter*) des préoccupations éthiques (comme le consentement et le respect de la vie privée, mais aussi l'égalité entre les personnes) au fil de pratiques organisationnelles (qui touchent au design de l'agent conversationnel).

Le concept au cœur de cette thèse – *matter of caring* – est également inspiré des travaux en CCO, en particulier ceux de François Cooren (2013) par l'approche de la ventriloquie qui porte attention aux nombreuses choses (ou figures) qui s'expriment au travers de nous quand on parle. Mais tandis que cette approche se concentre sur ce qui s'exprime dans les tours de parole, le concept de *matter of caring* invite à élargir l'attention pour être également vigilante à ce qui s'exprime dans les silences et le négatif (photographique) des tours de parole. Cette nuance est importante pour qu'un consentement soit porteur de respect et d'épanouissement : il ne doit pas être donné au nom de figures problématiques, indésirables ou oppressantes. Cette nuance théorique va de pair avec une posture épistémologique qui demande de prendre en considération les perspectives situées (Haraway, 1991; Harding, 1992) à partir desquelles on analyse ce qui fait parler quelqu'un, à partir desquelles on la fait parler.

Tout cela va de pair avec une invitation méthodologique à prendre en compte la pratique réflexive et les notes de journal de bord comme une source primaire de données, car la réflexivité est un moyen de faire l'expérience des préoccupations de l'organisation dans laquelle on est immergée. Elle permet de rendre compte de la constitution de l'organisation de l'intérieur, d'une position où l'on participe au tissage relationnel par lequel elle se constitue.

Avant d'entrer dans le corps de cette exploration, une précision : dans une perspective inclusive, j'opte pour une alternance entre les genres féminin et masculin dans l'écriture. Je mets au féminin le singulier (par exemple, « l'utilisatrice », « elle conçoit ») et au masculin le pluriel (« les utilisateurs », « ils conçoivent »). Comme j'invite à faire parler autrement les agents conversationnels (chapitre 2) et faire parler autrement de l'inconfortable (chapitre 3), j'espère que cette façon d'écrire légèrement autrement contribuera à entraîner modestement plus d'égalité entre les genres dans la lecture.

Chapitre 1

Une approche *by design* de l'éthique de l'IA

Introduction : l'éthique de l'IA à partir du design

De quoi parle-t-on quand on parle d'éthique de l'intelligence artificielle (IA) ? Vaste domaine aux multiples questions, je profite de ce chapitre pour me situer dans des pratiques spécifiques de l'éthique appliquée à l'IA : des pratiques de design qui s'intéressent à intégrer certaines valeurs dans la conception d'un système technologique, qui sont vigilantes aussi à l'évolution de ces valeurs qui émergent et se transforment au fil des usages de ces systèmes – des usages qui peuvent échapper aux designers quand ils ne correspondent pas à leurs intentions.

J'aborde l'éthique comme un ensemble de préoccupations liées à la morale, c'est-à-dire à la discussion de ce qui est juste, bon ou vertueux. Tandis que la réflexion morale a tendance à se concentrer sur les principes qui devraient diriger nos actions, la réflexion éthique est appliquée à des situations où l'on est amené à choisir quoi faire, « ce qui suppose qu'il existe du préférable et, donc, que soit fait appel à l'idée de bien comme critère non formel d'évaluation de l'action » (Clément et al., 2000, p. 299). Ce préférable et cette idée de bien ou de bon se prêtent à débats et tensions : selon les intérêts de chacun, plusieurs versions de ce qui est bon peuvent s'emmêler, voire se confronter. Dans le travail du design, il s'agit d'opter pour l'une ou l'autre version, de les confronter, voire les modifier.

Dans ce qui suit, je vais d'abord présenter le champ de l'éthique de l'IA et ses préoccupations, tel que je suis entrée dedans fin 2017. Comme ce qui m'intéresse est d'étudier les pratiques organisationnelles par lesquelles sont adressées ces préoccupations, de l'intérieur d'une petite organisation comme une startup dans le secteur privé, je vais présenter les différentes façons dont l'éthique de l'IA peut être travaillée dans ce type d'organisation. Je préciserai ensuite mon positionnement dans ces

façons de travailler l'éthique en organisation : ma démarche s'inscrit dans le domaine de la sociomatérialité, où la matérialisation est comprise comme une activité de design. Je vais alors développer les différentes approches de l'éthique travaillée par le design – autant d'approches avec lesquelles j'ai jonglé lors de mon immersion chez Mr Y et Prélude.

L'éthique de l'intelligence artificielle

Avec le développement de l'intelligence artificielle (IA) appliquée à des outils du quotidien, l'éthique de l'intelligence artificielle s'est développée comme un « domaine de l'éthique qui se demande ce qui est bon, juste ou vertueux dans la mise en œuvre des systèmes d'intelligence artificielle » (Gibert, 2019) et jauge ainsi ce qui est moralement souhaitable ou non, alors que tout est techniquement possible. Les principaux enjeux soulevés par ces technologies numériques relèvent de la vie privée face au risque de surveillance des données personnelles, de la manipulation des comportements à partir des pratiques de profilage, de l'opacité des systèmes d'IA, des biais présents dans les systèmes de décisions algorithmiques, des interactions humains-robots, des effets de l'automatisation sur la transformation de l'emploi et sur l'autonomie (Müller⁸, 2020). Je vais présenter le champ de l'éthique de l'IA tel que je suis entrée dans ses préoccupations en 2017, pour ensuite préciser l'approche sociomatérielle par laquelle j'aborde des enjeux d'éthique de l'IA dans cette thèse.

Entrer dans l'éthique de l'IA et ses préoccupations

Fin 2017, je suis entrée dans le travail de l'éthique de l'IA par la voie participative de la consultation citoyenne (Dilhac et al., 2018, p. 47-53), au sein de l'équipe de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle (abrégée Déclaration de Montréal). Celle-ci s'est constituée au fil d'un an ponctué d'ateliers dans des bibliothèques publiques au Québec pour rencontrer des citoyens ainsi

⁸ Voir https://plato.stanford.edu/entries/ethics-ai/?TB_iframe=true&width=370.8&height=658.8

que des experts⁹ autour de divers thèmes (la santé, l'éducation, la justice et la police prédictive, la ville et les objets connectés ainsi que le milieu du travail) afin d'élaborer collectivement des recommandations¹⁰ pour encadrer ce développement. Chaque atelier se déroule de la même manière que la vignette ci-dessous : après la lecture d'un court scénario prospectif, les personnes autour de chaque table sont invitées à discuter des enjeux que l'IA pose dans ce scénario, puis chaque groupe élabore des recommandations pour cadrer le développement de l'IA.

3 mars 2018, premier atelier que nous animons avec l'équipe de la Déclaration de Montréal. J'arrive dans la bibliothèque au premier étage, une odeur de chlore se mélange à celle des livres : la piscine de l'arrondissement est juste en-dessous. J'anime aujourd'hui la table de discussion sur la transformation de l'éducation avec l'avènement de systèmes d'IA. Les participants arrivent, nous sommes six à la table éducation. Après une introduction qui présente ce qu'est l'intelligence artificielle et la démarche de la Déclaration de Montréal par les professeurs qui l'orchestrent, j'introduis la séance à ma table : à partir d'un scénario que nous allons lire, nous identifierons les enjeux posés par le développement de l'IA en éducation pour ensuite proposer des recommandations qui adresseraient ces enjeux. On me demande ce qu'est un « enjeu éthique ». Je reformule : quelles questions ou quels problèmes le scénario pose-t-il ? Qu'est-ce qui s'y joue ? Le scénario porte sur un système d'IA qui aide à l'apprentissage en mesurant en temps réel la qualité d'attention de l'élève – un bracelet est muni d'un capteur, le contenu d'apprentissage s'ajuste sur la tablette de l'élève en fonction de ces données en temps réel, ce qui change en même temps la façon de travailler de l'enseignante¹¹. Les participants font plusieurs observations, d'un ordre principalement négatif : si c'est à cela que peut mener le développement de l'IA, plusieurs manifestent leurs réticences. Au neuvième et dernier atelier que j'animerai sur le thème de l'éducation toujours, à partir d'un scénario où une IA aide à orienter professionnellement les élèves en fonction de leurs probabilités personnalisées de réussir, des dames d'un certain âge manifesteront leur inquiétude pour leurs petits-enfants si c'est vers cela que le monde va.

(Vignette de mon entrée dans l'éthique de l'IA, à partir de notes de terrain prise dans le cadre de la Déclaration de Montréal, 3 mars 2018)

⁹ Les experts ont été rencontrés lors d'ateliers à la SAT à Montréal, au Musée des beaux-arts de Québec et au Centre Culturel Canadien à Paris.

¹⁰ Ces recommandations ont été de plusieurs ordres : dispositions légales et juridiques, formations, création d'acteurs institutionnels, dispositifs d'évaluation de l'IA, codes de déontologie, mécanismes participatifs, programmes de recherche, outils numériques, cadres professionnels et politiques internes, mécanismes assurantiels et contractuels, incitatifs (voir Voarino et al., 2018, p. 131).

¹¹ Le scénario en question est publié dans Dilhac et al. (2018, p. 78-79).

Ces discussions suscitent ainsi des préoccupations : les personnes sont soucieuses de la gouvernance qui pilote l'implantation de ces systèmes d'intelligence artificielle (SIA), soucieuses que l'éducation reste un bien public et non piloté par des organisations du secteur privé par le biais de ces outils, soucieuses également de la potentielle marchandisation des données collectées par ces outils si elles font l'objet d'échanges commerciaux autres qu'éducatifs. Plusieurs se sont inquiétées de la transformation de la relation entre enseignante et élève face au risque que l'apparente objectivité de la recommandation algorithmique prenne le dessus sur l'empathie de l'enseignante dans la prise de décision de ce qui est bon pour l'élève. Cela a soulevé également le risque d'une déresponsabilisation de l'enseignante. Certaines se sont également inquiétées de la capacité des élèves à développer leur autonomie et à gérer leurs émotions en étant si étroitement assistés en permanence par un tel outil. Nous adressons ces préoccupations par des pistes de recommandations. L'ensemble de ces préoccupations et recommandations sont ensuite analysées, synthétisées en un rapport (voir Voarino et al., 2018, p. 116-130, pour l'ensemble des préoccupations mentionnées).

Les préoccupations sont nombreuses¹² et interrogent le contrat moral et social que demande le développement de l'IA, qui implique des sociétés pilotées par le traitement des données. Ces préoccupations rencontrées au fil de la démarche participative de consultation citoyenne ont fait émerger les principes de la Déclaration de Montréal. Celle-ci s'articule en dix principes : bien-être, respect de l'autonomie, protection de l'intimité et de la vie privée, solidarité, participation démocratique, équité, inclusion de la diversité, prudence, responsabilité, développement soutenable (Dilhac et al., 2018, p. 5-18). La Déclaration de Montréal s'inscrit dans la mouvance de nombreuses déclarations et rapports publiés à partir de 2017 pour encadrer le développement de l'IA en faveur du bien commun. Les principes qui s'en dégagent peuvent être classés en trois grandes catégories qui correspondent aux quatre principes de bioéthique (Canca, 2020) :

¹² Nous les avons rassemblées en douze thèmes : mutations socio-numériques, stress-alarmisme-angoisse, transparence, sécurité et intégrité des systèmes, influences, consentement, justice sociale, libertés, vie privée et confidentialité, gouvernance, partage de la responsabilité, environnement-écologie (Voarino et al., 2018, p. 117).

respecter l'autonomie (principe de respect de l'autonomie), faire du bien (qui rassemble les principes de bienfaisance et non-malfaisance), et assurer un traitement juste pour tous (principe de justice). Selon ces principes, qu'ils soient ou non classés selon ceux de bioéthique, ou arrimés à ceux de la déclaration des droits humains¹³, l'IA doit se développer d'une façon qui soutienne une bonne expérience humaine, c'est-à-dire d'une façon qui permette que se réalisent les valeurs favorisant le respect des humains entre eux et par l'intermédiaire de technologies, pour faire société de façon durable et idéalement épanouissante pour chacun.

Avec l'éthique de l'IA, le terme de « valeur » convoque sa polysémie dans la mesure où se rencontrent les dimensions éthique et marchande de la valeur. D'un point de vue éthique, la valeur est entendue comme « ce qui est posé comme vrai, beau, bien, d'un point de vue personnel ou selon les critères d'une société et qui est donné comme un idéal à atteindre, comme quelque chose à défendre » (Larousse, s.d., « valeur »¹⁴). Ce sens de la valeur se mêle à celui qui fait de la valeur le « caractère mesurable (d'un objet) en tant que susceptible d'être échangé, d'être désiré » (idem). La valeur est ce qui porte l'idée d'une certaine version du bien, qui elle-même peut donner une qualité particulière donnant en même temps plus de valeur. Cette polysémie de la valeur se retrouve dans plusieurs arguments qui mettent de l'avant le bénéfice – ou la valeur ajoutée – que peut représenter la prise en compte de l'éthique dans une organisation et les produits qu'elle développe et vend. Par exemple, la pertinence de l'adoption du *privacy-by-design* (PbD) que nous allons voir – un modèle qui présente une certaine complexité pour prendre en considération des enjeux de protection des données personnelles et de leur confidentialité dès les premières phases de la conception d'un outil – est mise de l'avant

¹³ Le Groupe d'experts de haut niveau en IA chargé de proposer des lignes directrices à la Commission européenne met par exemple de l'avant « les droits fondamentaux comme base d'une IA digne de confiance » (p. 12), les droits fondamentaux étant « le fondement du droit tant international que de l'Union en matière de droits de l'homme » (p. 9). Elle inscrit ainsi l'autorité de l'encadrement du développement de l'IA dans celle de l'autorité légale et internationale du respect des droits humains (Groupe d'experts de haut niveau en intelligence artificielle, 2019).

¹⁴ Voir <https://www.larousse.fr/dictionnaires/francais/valeur/80972>

au nom des bénéfices à long terme que ce modèle implique sous forme d'avantages compétitifs :

In a world of increasingly savvy and interconnected users, an organization's approach to privacy may offer precisely the competitive advantage that is needed to succeed. Privacy is essential to creating an environment that fosters trusting, long-term relationships with existing customers, while attracting opportunity and facilitating the development of new ones. (Cavoukian, 2019)¹⁵.

Ainsi, avec le PbD, la mise en pratique de la valeur du respect de la vie privée (*privacy*) entraîne un ensemble d'autres valeurs importantes pour le lien à créer avec les utilisateurs du produit : la confiance, l'adoption du produit sur le long terme et l'attraction de nouveaux clients.

La mise en pratique de l'éthique joue ainsi avec la polysémie de la valeur en en faisant un objectif de gestion à visée marchande, comme c'est le cas par exemple pour favoriser la confiance qui encouragera l'utilisation de ce qu'elle produit. La mise en pratique de l'éthique fait également de la valeur un objet de design, comme c'est le cas pour les approches de *value-sensitive design* (cf p. 53) qui cherchent à mieux intégrer, par le design, des valeurs éthiques dans des fonctionnalités technologiques.

Remarquons que la valeur de confiance souvent étroitement associée à l'éthique comme un effet de sa mise en pratique n'a pas qu'une vocation marchande. Le High-Level Expert Group on AI, un groupe d'experts rassemblés par la Commission Européenne pour édicter des lignes de conduites du développement de l'IA, a pris le mot « *Trustworthy* » pour qualifier l'IA à laquelle aboutir par ces directions : « *According to the Guidelines, trustworthy AI should be: (1) lawful – respecting all applicable laws and regulations; (2) ethical – respecting ethical principles and values; (3) robust – both from a technical perspective while taking into account its social environment* »¹⁶. Ces lignes directrices informent des recommandations de politiques publiques et d'investissements. L'objectif visé par le développement de systèmes d'IA « dignes de confiance » nécessite de combiner les bénéfices qui peuvent être possibles avec l'IA et la mitigation des risques

¹⁵ Il s'agit du module 2, « *The 7 Foundational Principles* » du cours en ligne « *Privacy by Design* » d'Ann Cavoukian.

¹⁶ Voir <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, 2018

liés à ces technologies : « C'est grâce à une IA digne de confiance que, en tant que citoyens européens, nous pourrions bénéficier de ses avantages d'une manière qui reflète nos valeurs fondamentales que sont le respect des droits de l'homme, la démocratie et l'État de droit » (Groupe d'experts de haut niveau en intelligence artificielle, 2019, p. 5-6). La valeur de confiance serait ainsi accordée si l'on reconnaît des valeurs fondamentales aux sociétés européennes dans le développement et les usages de systèmes d'IA. Voyons à présent quel travail, dans les organisations, permet aux valeurs – mises de l'avant dans le champ de l'éthique de l'IA – d'orienter le développement technologique, voire de se concrétiser dans les technologies développées.

Le travail (organisationnel) de l'éthique

Permettre à des valeurs de se réaliser peut se faire à plusieurs niveaux : au niveau juridique des lois et règlements, au niveau d'institutions créées pour observer et évaluer les pratiques d'autres instances, au niveau gouvernemental et intergouvernemental, au niveau de la société civile, au niveau du secteur privé qui développe de tels systèmes. Fjeld et al. (2020) ont retracé¹⁷ la multiplicité de rapports et déclarations de principes publiés par ces différents types d'instance depuis 2017¹⁸.

De nombreux outils se multiplient également pour aider dans la mise en application de ces principes auprès des acteurs qui développent des systèmes d'IA (Morley et al., 2020; Vakkuri et al., 2019). Le gouvernement du Canada a par exemple mis au point un outil d'Évaluation de l'Incidence Algorithmique, « un questionnaire conçu pour vous aider à évaluer et à atténuer les risques associés au déploiement d'un système

¹⁷ Voir aussi la visualisation sur AI ethics lab : <https://aiethicslab.com/big-picture/>

¹⁸ Voir par exemple les rapports qui relèvent de la société civile (dont UNI Global Union, 2017; Amnesty International & AccessNow, 2018), de différents gouvernements¹⁸ (dont US National Science and Technology Council, 2016; Standards Administration of China, 2018; Villani en France, 2018; UK House of Lords, 2018; la Commission Européenne, 2018...), des organisations intergouvernementales (OECD, G20, 2019), diverses parties prenantes (Partnership on AI, 2016; AI Now Institute, 2017; Future of Life Institute, 2017; Déclaration de Montréal, 2018; IEEE, 2019; Beijing Academy of AI, 2019) ou encore le secteur privé (Microsoft, Google, IBM...)

automatisé de prise de décisions »¹⁹; Mörch et al. (2019) ont développé une liste de vérification pour les personnes développant des systèmes d'IA pour la prévention du suicide et la santé mentale plus largement; le AI Ethics Lab a créé un outil de visualisation des forces et faiblesses éthiques d'une technologie (AI Ethics Lab, 2020)²⁰ et offre ses services pour trouver des solutions appropriées à chaque organisation. À la suite de ses lignes directrices « *Trustworthy AI* », la Commission Européenne a publié « *The Assessment List for Trustworthy Artificial Intelligence* », un outil pour l'auto-évaluation de systèmes d'IA par les organisations qui en développent²¹. De nombreuses organisations se sont également créées pour observer les impacts sociaux de l'intelligence artificielle comme l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA)²² au Québec, le Partenariat mondial sur l'intelligence artificielle²³, AI Now Institute à New-York. Plusieurs organisations militent également pour la défense des droits numériques (AccessNow, Center for Humane Technology).

Toutes ces démarches de mise en débat, de propositions d'encadrement, de rédaction de principes pour guider le développement de l'IA et d'outils pour accompagner la mise en application de ces principes sont autant de formes que peut prendre le travail de l'éthique appliquée à cet ensemble de nouvelles technologies. Ce travail consiste à discuter les enjeux normatifs de ce champ d'activité encore récent et en forte évolution. Ce qui m'intéresse dans cette thèse est d'étudier comment, dans une organisation du secteur privé telle qu'une startup, ces enjeux éthiques ne sont pas seulement discutés mais font l'objet de pratiques organisationnelles relevant de la conception, du développement technologique et de la gestion des opérations.

¹⁹ Voir <https://www.canada.ca/fr/gouvernement/systeme/gouvernement-numerique/innovations-gouvernementales-numeriques/utilisation-responsable-ai/evaluation-incidence-algorithmique.html>

²⁰ Voir <https://aiethicslab.com/the-box/>

²¹ Voir <https://altai.insight-centre.org>

²² Voir <https://observatoire-ia.ulaval.ca>

²³ Voir <https://gpai.ai/fr/>

Dans les organisations du secteur privé, l'éthique peut être abordée de plusieurs manières : par des comités d'éthique²⁴, par des équipes juridiques ou par des équipes qui se consacrent à la responsabilité sociale des entreprises (RSE). Le comité d'éthique intervient auprès des praticiens qui manquent non seulement de temps pour réfléchir aux questionnements éthiques naissant de leurs pratiques, mais aussi d'un « répertoire conceptuel qui leur permettrait de nommer avec précision les écueils éthiques auxquels ils sentent que leur pratique professionnelle se heurte parfois » (Weinstock, 2014, p. 50). Les équipes juridiques abordent l'éthique en travaillant sur la conformité aux lois et règlements fondés sur des principes éthiques. Les équipes de RSE orchestrent la planification, l'exécution, la vérification et l'amélioration de la stratégie et des initiatives de l'organisation pour « mener ses activités d'une manière durable sur le plan économique, social et environnemental » (Industrie Canada, 2014, p. 17-18).

Ces approches de l'éthique appliquée sont intéressantes mais ne se présentent pas comme telles dans l'organisation dans laquelle je m'immerge : une startup aux ressources limitées, qui n'a ni département juridique ni équipe RSE. Elle a une intention initiale de former un comité d'éthique mais des priorités liées aux opérations de démarrage prennent le dessus sur cette intention (voir chapitre 6). Cette startup opte pour un travail de l'éthique à même ses pratiques de conception et développement technologique de ses produits.

Depuis que les préoccupations éthiques envers le développement de l'IA se font plus bruyantes dans les médias et auprès des institutions gouvernementales et juridiques, le travail de l'éthique intégré aux pratiques de conception a pris une certaine ampleur dans les organisations en IA. Les plus grandes ont des équipes de recherche et développement qui travaillent à traduire dans les techniques d'IA des principes éthiques tels que le *fairness*, *accountability*, *transparency and explainability*. Par exemple, chez Microsoft, le centre de recherche FATE se définit par de l'interdisciplinarité pour étudier

²⁴ « [c]eux-ci sont généralement créés par des institutions (universitaires, gouvernementales, médicales, professionnelles, etc.) et sont des lieux de délibération à caractère décisionnel ou consultatif. » Extrait du site de la Commission de l'éthique en science et en technologie du Québec :

<http://www.ethique.gouv.qc.ca/fr/ethique/quelques-notions-dethique/lethique-appliquee.html>.

les implications complexes des techniques d'intelligence artificielle, d'apprentissage automatique et de traitement du langage naturel :

*We study the complex societal implications of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP). Our aim is to facilitate computational techniques that are both innovative and responsible, while prioritizing issues of fairness, accountability, transparency, and ethics as they relate to AI, ML, and NLP by drawing on fields with a sociotechnical orientation, such as HCI, information science, sociology, anthropology, science and technology studies, media studies, political science, and law.*²⁵

L'éthique se travaille également de plus en plus par l'angle de la recherche en expérience utilisateur (*user research, user experience (UX) research*) qui tente d'adresser des enjeux sociaux par l'angle du design²⁶. Par exemple, les documents internes à Facebook – devenu Meta – que la lanceuse d'alerte Frances Haugen a fait fuiter en 2021 (voir ci-dessous, p. 44) correspondent à des rapports de recherche utilisateur faite chez Facebook pour documenter les impacts néfastes de leurs technologies, et chercher des solutions pour y remédier ou éviter certains développements.

Dans les plus grandes organisations, ce sont des équipes entières qui sont dédiées à la recherche en expérience utilisateur, parfois avec un focus mis sur un ensemble d'enjeux éthiques qui ont des impacts sociaux (Meta a par exemple des chercheurs en expérience utilisateur dans son équipe *Privacy and Data Use*, c'est également ce type de recherche que pilote la *Head of privacy research* rencontrée à la conférence EPIC en novembre 2019, voir vignette EPIC2019, p. 20). Dans les plus petites organisations, quand de la recherche en expérience utilisateur est faite, cela peut être fait par une ou deux personnes – ce qui a été le cas chez Mr Y et Prélude. Il y a également parfois des personnes nommées responsables de l'éthique dans ces organisations, que Metcalf, Moss et Boyd (2019) appellent « *ethics owners* ». Le terme est difficile à traduire en français dans ce contexte : il ne s'agit pas tant d'un lien de propriété mais plutôt d'un lien de responsabilité

²⁵ Voir <https://www.microsoft.com/en-us/research/theme/fate/>

²⁶ Comme nous allons le voir plus en détail dans le chapitre 2 (p. 78), l'expérience utilisateur est un terme général qui désigne l'ensemble du parcours de l'utilisatrice amenée à interagir avec un produit, depuis l'étape où il le trouve, l'utilise de différentes manières, le quitte et y revient (Norman, 2013).

qui implique que le « *owner* » soit garant de la diffusion et l'intégration du projet dans l'ensemble de l'organisation :

In corporate parlance, "owning" a portfolio or project means holding responsibility for it, often across multiple divisions or hierarchies within the organization. Typically, the "owner" of a project does not bear sole responsibility for it, but rather oversees integration of that project across the organization. (Metcalf et al., 2019, p. 449)

Sans que ce soit formulé dans ces termes, mon rôle chez Mr Y et Prélude a pris la forme de l'*ethics owning* quand la préoccupation d'intégrer un module éthique (sur le consentement) dans le design de l'agent conversationnel s'est fait plus pressante (voir chapitre 6).

Toutes ces façons de travailler l'éthique dans les organisations qui développent des SIA alternent entre une approche de l'éthique à implémenter (par l'ingénierie informatique, les processus organisationnels, le design) ou autour de laquelle concevoir l'organisation (faisant relever l'éthique de la culture organisationnelle). Le rôle des personnes qui travaillent l'éthique s'accompagne d'une attitude vigilante pour ne pas tomber dans une dérive redoutée : faire de l'*ethics-washing*²⁷ (ou éthico-blanchiment), c'est-à-dire parler d'éthique de façon superficielle, pour donner l'impression d'avoir de bonnes pratiques sans pour autant faire l'effort de les mettre effectivement en place. La vigilance envers l'éthico-blanchiment cherche à éviter que la raison d'être de l'éthique soit d'un ordre instrumental qui relève seulement de l'image de l'organisation. Elle cherche à assurer que les pratiques de l'organisation en lien avec un enjeu éthique ou une valeur soient des traductions concrètes du principe éthique correspondant. Par exemple, dans le cas de pratiques en lien avec les données personnelles comme c'est le cas chez Prélude, il s'agit de faire en sorte que le principe de respect de la vie privée et de la confidentialité se traduise concrètement par des pratiques de traitement responsables des données personnelles.

Chez Prélude, la crainte de l'*ethics-washing* est notamment formulée en juillet 2019, quand un code d'éthique a fait son apparition de façon surprenante : le consultant

²⁷ Néologisme forgé en référence au « *green-washing* » (éco-blanchiment) : « *the act of superficially making environmentally unsustainable practices appear "green"* » (Metcalf et al., 2019, p. 473).

qui aide Prélude pour la rédaction de plusieurs demandes de subvention a envoyé à Paul et Félix un exemple de code éthique à joindre au dossier d'une demande de subvention. Il mentionne dans son courriel que joindre un code éthique ou montrer qu'une personne est chargée de l'éthique dans l'organisation est souvent un atout dans les demandes de subvention. Cet exemple devra cependant être adapté à la réalité de Prélude, précise-t-il, ce que je propose de faire à l'aide de codes éthiques publiés à propos des enjeux de l'intelligence artificielle – je cite la Déclaration de Montréal et le cadre de référence d'IVADO pour la diversité, l'inclusion et l'équité²⁸ (cet échange a eu lieu sur Slack, conversation entre Paul, Félix et moi le 4 juillet 2019). Dans les semaines qui suivent, je travaille ponctuellement sur l'adaptation de ce code éthique, à la lumière de discussions informelles avec Paul, Philippe et d'une réunion avec Félix²⁹. Mais ce travail est ralenti par la préoccupation d'une potentielle contribution à de l'éthico-blanchiment :

Je me traîne car j'ai l'impression de potentiellement brasser du vent, contribuer à de l'*ethics-washing* en participant à l'écriture de valeurs et grands principes sans que cela se manifeste particulièrement dans les pratiques. Mais je crois que ce n'est pas l'intention de Prélude. Depuis nos conversations, j'ai davantage l'impression qu'il y a de la sincérité dans la volonté de mettre en place des pratiques « éthiques », d'être alignés entre ce qu'on dit et ce qu'on fait. (...) même si j'oscille encore entre la confiance et la méfiance, je dirais que je suis à 80% confiante (encore méfiante par rapport à l'utilisation qui est faite de moi) (...) mais j'ai l'impression qu'il y a quelque chose de très processuel dans cette « utilisation » : de moins en moins « *ethics-washing* », de plus en plus voulu dans les pratiques à mesure que des propositions concrètes sont faites, que l'éthique est un peu moins une nébuleuse abstraite. (Journal de bord, notes de terrain du 25 juillet 2019)

Ce qui permet à l'éthique d'être moins de l'ordre du blanchiment et plus de l'ordre pragmatique de ce qui se traduit effectivement dans les pratiques, ce sont les propositions concrètes qui sont faites pour traduire des principes en actes. Il y a là un souci pragmatique (Austin, 2009 [1962] : « *doing things with words* ») de faire en sorte que *ce que font les mots* soit de l'ordre de l'honnêteté éthique plutôt que de la façade blanchie. Cela permet de considérer cette mise en acte pragmatique de l'éthique de façon

²⁸ Voir : <https://ivado.ca/equite-diversite-inclusion/>

²⁹ Lors de cette réunion qui prend la forme d'une discussion sur les valeurs qui l'animent dans la création de Prélude, Félix s'attarde sur la confiance à favoriser dans l'équipe et chez les futurs utilisateurs de ses agents conversationnels.

processuelle plutôt que de la condamner au titre de blanchiment qui ne relèverait que de la bonne figure à donner aux autres. En effet, la dimension pragmatique invite à trouver comment faire, à trouver des prises pour transformer un principe abstrait en actions plus concrètes³⁰.

Dans les organisations en IA, certaines personnes adoptent une posture plus radicale, militante, pour valoriser le travail concret de l'éthique quand sa dimension processuelle prend trop de temps ou le rend trop dilué pour entraîner les changements exigeants que plusieurs enjeux éthiques demandent. Par exemple, en 2021, la lanceuse d'alerte Frances Haugen, anciennement chez Facebook, a communiqué à plusieurs médias des documents internes montrant que l'organisation avait une bonne connaissance des impacts néfastes de ses plateformes, entre autres sur la santé mentale des jeunes (Wells et al., 2021), sur le trafic d'êtres humains (Scheck et al., 2021), sur la diffusion de contenus et commentaires violents (Hagey & Horwitz, 2021), mais que peu de mesures étaient prises pour les corriger. Chez Google, les chercheuses Timnit Gebru et Margareth Mitchell, co-directrices de l'équipe de recherche en éthique de l'IA, ont été licenciées en décembre 2020 à la suite de plusieurs travaux s'attachant à montrer les biais et limitations de techniques d'IA comme le *machine learning* telles qu'utilisées dans les services de Google (voir par exemple Bender et al. (2021)³¹). Elles avaient également manifesté une certaine frustration quant aux efforts de l'organisation en matière d'équité, diversité et inclusion (Metz & Wakabayashi, 2020; Schiffer, 2020). Ces licenciements ont suscité de l'inquiétude chez de nombreux employés préoccupés par la tolérance à l'expression de travaux qui ne vont pas dans le sens des intérêts de l'organisation mais qui visent à améliorer les pratiques (Vincent, 2021). Pour que le travail de l'éthique soit soutenu et concrétisé sous forme d'amélioration des pratiques, Timnit

³⁰ C'est là en somme une reformulation du souci pour faire d'une préoccupation un *matter of care* – concept par lequel je vais étudier le travail de l'éthique dans cette thèse (cf chapitre 4) : il s'agit de trouver les moyens d'être pragmatique, inscrit au cœur des choses.

³¹ Cet article, intitulé « *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜* », « *detailed potential risks associated with large language processing models, including over-relying on data from wealthy countries that have more internet access* » (Schiffer, 2020).

Gebru appelle au renforcement des moyens qui donnent du pouvoir et de la protection aux employés afin qu'ils et elles puissent agir contre certaines pratiques qui ont des impacts néfastes dans le développement de systèmes d'IA, en particulier quand cela demande des changements importants dans les modèles d'affaire :

When people ask what regulations need to be in place to safeguard us from the unsafe uses of AI we've been seeing, I always start with labor protections and antitrust measures. (...) While those [regulations] are important, the #1 thing that would safeguard us from unsafe uses of AI is curbing the power of the companies who develop it and increasing the power of those who speak up against the harms of AI and these companies' practices. Thanks to the hard work of Ifeoma Ozoma and her collaborators, California recently passed the Silenced No More Act, making it illegal to silence workers from speaking out about racism, harassment and other forms of abuse in the workplace. This needs to be universal. (...) When workers have power, it creates a layer of checks and balances on the tech billionaires whose whim-driven decisions increasingly affect the entire world. (Gebru, 2021)

Il s'agit en somme de trouver comment désamorcer certains mécanismes organisationnels qui amènent le travail de l'éthique à être soumis aux intérêts des organisations alors que ce travail peut tendre à les remettre en question. Si ce désamorçage est possible, alors cela permettra de résoudre la tension dans laquelle sont pris les *ethics owners* selon Metcalf, Moss et boyd (2019) : « *Ethics owners operate inside a fraught dynamic: on the one hand attempting to resolve critical external normative claims about the core logics of the tech industry; on the other hand doing so while fully embedded within those logics* » (p. 451).

Au regard de l'exclusion de lanceuses d'alerte et des directrices d'équipe d'éthique de l'IA qui remettaient en cause certaines logiques qu'elles connaissaient bien en étant immergées dedans, cette tension en soulève en même temps une autre : comment garder accès à l'organisation et amener quelques changements internes à ses pratiques si l'on ne s'intègre pas complètement dans ses logiques de fonctionnement, de sens, d'intérêt ? À l'échelle d'une très petite organisation comme Prélude, cette question se pose également à l'égard de mon travail de l'éthique; j'opte pour une composition avec les logiques imparfaites qui prévalent plutôt que d'opter pour ou de risquer mon exclusion de l'organisation (voir chapitre 5), tout en me demandant régulièrement : « ma vigilance s'amointrit-elle ? Comment trouver un équilibre entre vigilance, méfiance, résistance,

tout en contribuant à l'organisation ? » (Journal de bord, notes de terrain du 5 novembre 2019). La vigilance qui accompagne le travail de l'éthique trouve une partie de sa consistance dans cette réflexivité à l'égard de ce à quoi on contribue, si cette contribution est alignée avec les valeurs qu'on veut défendre ou aider à réaliser³². Cette vigilance implique en même temps un certain travail émotionnel qui prend en charge l'inquiétude de possibles dérives.

« *Their job is to worry* » : une professeure en intelligence artificielle a fait cette remarque à propos du travail d'avocats et éthiciens avec qui elle avait aimé travailler une journée, lors d'un événement interdisciplinaire – c'était à C2Mtl, une conférence sur la créativité et le monde des affaires fin mai 2019. Dans ce travail de l'inquiétude éthique, des tensions réflexives se mêlent ainsi aux pratiques de conception de systèmes d'IA où les dimensions sociales (valeurs éthiques, préoccupations envers des effets sociaux, dynamiques organisationnelles) et matérielles (technologie développée, logiciels et outils utilisés, fonctionnalités à adapter, etc.) sont inextricablement liées, se constituent mutuellement voire se confondent dans leur sociomatérialité.

Une approche sociomatérielle du travail de l'éthique par le design

La façon dont j'étudie les pratiques dans lesquelles je me suis immergée pour cette thèse – des pratiques de design d'un agent conversationnel – peut se situer dans le paysage théorique des études sociomatérielles sur les sciences et technologies en contexte organisationnel. Le terme de sociomatérialité³³ a été introduit par Lucy Suchman (2002, 2007) pour théoriser les relations entre humains et technologies, en soulignant leur entremêlement. Orlikowski et Scott ont développé cette idée en l'adaptant au contexte organisationnel afin de mieux faire sens de la technologie dans les études organisationnelles (Orlikowski, 2000, 2007; Orlikowski & Scott, 2014). En particulier,

³² C'est ce que j'aborderai dans le chapitre 5 qui présente la pratique réflexive comme méthodologie.

³³ Il existe un débat sur la validité ou non du terme de sociomatériel – met-il trop l'accent sur le social ? Participe-t-il à entretenir la dualité entre social et matériel ? Prend-il suffisamment en compte le degré fondamentalement matériel de tout ? (voir par exemple (Cooren, 2018; Martine & Cooren, 2016; Orlikowski & Scott, 2015; Vásquez & Plourde, 2017). Sans entrer dans ce débat, ma posture prend l'inextricabilité du social et du matériel comme point de départ.

Wanda Orlikowski (2007) a employé ce terme pour proposer « *an alternative approach that posits materiality as constitutive of everyday life* (Barad, 2003; Latour, 2005; Suchman, 2007) » (p. 1436) afin que les études organisationnelles prennent mieux en compte la façon dont l'« *organizing is bound up with the material forms and spaces through which humans act and interact* » (p. 1436). Si le matériel est souvent associé au technologique, plusieurs chercheurs emploient également le terme de sociomatérialité pour rendre compte de l'inextricabilité entre le social et le matériel dans tout phénomène. Paul Leonardi (2011, 2012) propose la métaphore de l'imbrication du social et du matériel pour souligner le fait que les deux peuvent être travaillés séparément bien qu'ils soient toujours étroitement mêlés dans un produit technologique final. Dans sa perspective, le social et le matériel ont chacun une agentivité qui s'imbrique l'une à l'autre par l'entremêlement des routines et des technologies; cet entremêlement de l'ordre de l'imbrication façonne l'organisation (Leonardi, 2011, 2012).

Les études sociomatérielles de l'éthique telle qu'elle est mise en pratique dans des organisations ne sont pas nombreuses. Plusieurs études sociomatérielles qui mentionnent des questions d'éthique abordent l'éthique comme une implication de l'approche sociomatérielle : la sociomatérialité, et l'entremêlement constitutif du social et du matériel ou technologique sur lequel elle repose, montrent l'absence de neutralité des technologies et leurs rôles performatifs dans les vies quotidiennes (Cecez-Kecmanovic et al., 2014; Clegg et al., 2013; Fenwick, 2014). L'éthique tend ici à être considérée comme un *effet* de certaines pratiques.

D'autres études sociomatérielles abordent l'éthique comme relevant elle-même d'un ensemble de pratiques. C'est le cas par exemple de Silvia Gherardi qui montre, à partir d'une étude des pratiques pour alimenter artificiellement des personnes en maison pour aînés, « *how the sociomateriality of bodies no longer able to take food, their nursing, medical practices and family care become entangled with ethical discourses on life and death as sociomaterial issues* » (Gherardi, 2016, p. 50). La revue de littérature de Schultze et al. (2020) sur les travaux en sociomatérialité soulève par ailleurs les implications éthiques de la sociomatérialité en ce qui concerne l'éthique de la recherche : rendre

compte de la responsabilité (*accountability*) des chercheurs dans leurs *world-making*. C'est ce que je soutiens également dans le chapitre méthodologique de cette thèse (voir chapitre 5) qui met l'accent sur le rôle de ma pratique réflexive en tant que chercheure ayant participé à la constitution de l'organisation, étudiée au départ par l'angle de la matérialisation de l'éthique.

L'approche de la communication constitutive des organisations (CCO) étudie comment des collectifs humains et autres qu'humains viennent à s'ordonner et se désordonner au fil de leurs interactions (Baillargeon et al., 2021; Boivin et al., 2017; Brummans et al., 2014). Plusieurs études en CCO s'inscrivent dans le courant de la sociomatérialité. En particulier, les études CCO de la matérialisation de l'éthique portent attention à la façon dont des choses telles que des idées, des valeurs, des principes etc. sont matérialisées au travers de conversations, de textes et d'actions. Ces choses sont « matérialisées » dans la mesure où elles sont à la fois rendues matérielles et *made to count* – le terme « matérielles » joue sur l'homonymie entre *matter* (*the stuff of the material*) et *to matter* (ce qui compte) (Vásquez & Plourde, 2017). En explorant comment des préoccupations éthiques (*ethical matters of concern*) se matérialisent au fil des pratiques des membres d'une organisation humanitaire dans un camp de réfugiés, Matte et Bencherki (2019) expliquent :

We explore how humanitarian workers collectively materialize ethics by playing on the double meaning of mattering, and observe how it is dealt with by looking at how values, standards, ideals, morals, and principles are made to count in a situation by being materialized through discourse, writing, or embodied action (Cooren, 2016). (p. 5871).

Cela les amène à considérer l'éthique – qu'ils définissent comme « *what is right or wrong in a given situation or within an organizational setting* » (p. 5872) – comme un phénomène à la fois médié et matérialisé : « *It does not exist merely as an abstract notion, but through other things that materialize it and allow it to matter in each situation* » (p. 5872). Cela rejoint l'approche pragmatique de l'éthique que j'adopte dans cette thèse, dans le courant de Pols et Mol qui, par la philosophie empirique, étudient comment les faits, préoccupations, idées de ce qui est bien ou mal et valeurs sont produites *in situ*

(Nickelsen & Lydahlm sous presse, cité plus haut p. 26) et peuvent être intégrées « dans les pratiques, les bâtiments, les habitudes et les machines » (Mol, 2013, p. 12).

Si mon étude s'inscrit dans cette perspective de la communication – centrée sur l'éthique – qui participe à la constitution des organisations Mr Y et Prélude, sa dimension sociomatérielle n'est qu'un point de départ et non la finalité de mon étude. En effet, la matérialisation à l'œuvre dans ma thèse renvoie à une activité : celle du design, entendu comme ensemble de pratiques employées pour concevoir un système d'intelligence artificielle tout en prenant en considération des préoccupations éthiques. Mon projet contient une dimension d'intervention, qui fait que la matérialisation n'est pas tant le phénomène étudié que l'activité par laquelle je pose des questions de l'ordre du « comment faire » plutôt que du « que se passe-t-il ». Plutôt que d'observer comment des préoccupations éthiques se mêlent à des préoccupations d'autres ordres organisationnels et technologiques, et comment ces entremêlements sont constitutifs de l'organisation étudiée, mon projet a été mené à la lumière d'une question sociomatérielle porteuse d'intervention : comment faire compter (*matter*) les préoccupations éthiques au fil du développement d'un agent conversationnel ?

La sociomatérialité en question ne s'arrête donc pas à l'analyse d'entremêlements sociomatériels constitutifs d'une organisation au fil d'interactions mais elle invite un « *careful engagement* » (Nickelsen & Lydahl, sous presse), soucieux d'exprimer des préoccupations éthiques au fil du développement de l'agent conversationnel afin que celui-ci soit constitué de valeurs éthiques – et plus précisément de la valeur du consentement et du respect de la vie privée. C'est ce *careful engagement*, par lequel est mené mon projet, qui oriente mon cadre théorique présenté au chapitre 4 vers le concept de *matter of caring* : il ne s'agit pas de s'arrêter à ce que l'on observe compter (*what matters*) mais de voir comment *faire compter* des valeurs, par souci de leur permettre de s'exprimer.

Ce souci sociomatériel de permettre à des valeurs de s'exprimer dans la conception d'un système technologique est présent dans un ensemble de pratiques qui mêlent éthique et design – où le design est compris comme l'ensemble des activités liées

à la conception et la réalisation d'un projet (Vial, 2010, 2021) – que le design soit sensible à l'éthique comme dans le cas du *value-sensitive design*, ou que l'éthique soit exprimée par le design, ou encore intégrée dans le design.

L'éthique *by design*³⁴

Aborder l'éthique par une approche sociomatérielle permet de considérer le rôle constitutif de l'éthique et des objets technologiques. Cela se distingue d'un rapport instrumental à la technologie qui serait supposée neutre socialement pendant sa conception, cantonnant sa dimension sociale aux usages des objets technologiques. Le courant des *science and technology studies* (STS) (Felt et al., 2017) et la philosophie contemporaine de la technologie (Feenberg, 1999; Ihde, 1990; Verbeek, 2011) permettent au contraire de plonger dans la complexité sociomatérielle qui a lieu dans toute la vie d'un objet technologique, de sa conception à son utilisation, que ses effets se révèlent désirables ou non, intentionnels ou non de la part des concepteurs. Il s'agit de se pencher sur l'agentivité – ou capacité d'agir sur le social, propension à faire une différence – à l'œuvre dans les objets technologiques, étant donné que ceux-ci *font* des choses en participant à « [l']organisation des relations des hommes entre eux et avec leur environnement » (Akrich, 1987, p. 49).

La perspective sociomatérielle, qui envisage ce rapport de constitution mutuelle et entremêlée du social et du technique, s'inscrit dans la lignée de la conception latourienne de l'agentivité entendue comme médiation : tandis que les intermédiaires « véhiculent du sens ou de la force sans transformation », les médiateurs « transforment, traduisent, distordent et modifient le sens ou les éléments qu'ils sont censés transporter » (Latour, 2006, cité dans Fischer, 2020, p. 144). Plutôt que d'être compris comme un intermédiaire qui transmet de façon neutre son contenu, un objet technologique tel qu'un SIA est ainsi doté d'une agentivité de l'ordre de la médiation qui

³⁴ Pour une analyse philosophique conceptualisant l'éthique *by design* du numérique, voir la thèse de Flora Fischer (2020) sur laquelle je me suis appuyée pour mieux saisir la généalogie de cette notion et ses ramifications conceptuelles.

porte une certaine transformation dans le contenu qu'il transporte – un contenu constitué à partir de données qu'il traite selon ce que prévoient ses algorithmes – et dans l'utilisation qu'ils sont censés permettre. C'est par exemple cette agentivité médiatrice à l'œuvre dans des techniques algorithmiques d'analyse de données que Bender et al. (2021) ont remise en question à la lumière des effets qu'elles entraînent. Cette nuance qui met l'accent sur la dimension de l'agentivité technologique est importante pour mon propos car les pratiques d'éthique *by design* qui m'intéressent dans cette thèse relèvent de l'influence qui peut être exercée sur ces médiations, pour éventuellement les orienter dans une certaine direction.

Inscrit dans cette réalité de la non-neutralité technologique, l'éthique *by design* est un terme parapluie qui rassemble un ensemble de pratiques sociomatérielles qui entremêlent performativité de l'éthique et activités de design. La locution « *by design* » évoque une ambivalence : « Elle peut traduire à la fois ce qui est intentionnel, « par dessein »; ce qui est « par définition », « par conception »; mais aussi ce qui est atteint « par tactique / ruse » (F. Fischer, 2019, p. 62). Flora Fischer propose alors :

une typologie de l'éthique *by design* en fonction de ses rattachements théoriques, entre une éthique de l'intention, anthropocentrée c'est-à-dire centrée sur le sujet (1), une éthique par conception, centrée sur l'objet (2), une éthique de la médiation, centrée sur la relation sujet-objet (3), et enfin de proposer une approche de l'éthique *by design* comme expédient, c'est-à-dire comme ce qui relève toujours d'un arrangement avec les contraintes techniques (4). (p. 63)

Je tiens à prendre le temps de présenter ces différentes approches car les pratiques par lesquelles est travaillé l'enjeu éthique du consentement, lors de mon immersion avec Mr Y et Prélude, touchent à des particularités de ces façons d'approcher l'éthique travaillée par le design dans des technologies en développement et, par extension, dans les organisations qui développent des technologies.

Une culture de la conséquence

L'approche anthropocentrée de l'éthique *by design* se concentre sur les pratiques des concepteurs et leurs intentions, selon le « dessein »³⁵ qu'ils ont. L'éthique peut se comprendre ici dans sa version déontologique « qui évalue l'action en fonction de sa conformité à des principes qui sont de nature prescriptive et permet d'orienter les intentions par rapport à des devoirs » : c'est le cas lorsqu'on utilise une charte éthique, une déclaration de principes ou un règlement pour se conformer à ses principes et ses lois³⁶. L'éthique peut se comprendre également ici dans sa version conséquentialiste où « le cœur de la réflexion morale repose sur la prise en compte des effets et conséquences de ses actions » (p. 63). C'est dans cette perspective de la culture de la conséquence que s'inscrit la première occurrence académique du terme « *ethics by design* » dans le livre de Stephanie L. Moore (2010) intitulé « *Ethics by design: strategic thinking and planning for exemplary performance, responsible results, and societal accountability* ». Dans une perspective de sociologie des organisations, elle emploie le terme « *ethics by design* » pour évoquer le fait de « *doing ethics on purpose – by design* » (p. 139) ou encore « *design for desired performance* » (p. 126). Son argument consiste à associer pratiques éthiques et pensée stratégique en incitant une culture de la conséquence mesurable par l'évaluation des impacts sociaux des activités de l'organisation et de ses partenaires. Moore invite à penser l'éthique dans l'organisation de façon systémique plutôt qu'individuelle : elle relève de la mise en place d'un environnement de performance et de standards visant l'impact social. Par l'*ethics by design*, il s'agit de se donner les moyens de réaliser des actions concrètes alignées avec des valeurs éthiques, un alignement qui relève de la stratégie et de la performance.

³⁵ Comme le souligne Stéphane Vial (2021, p. 14), le terme de design est « lié aux deux dimensions fondamentales du projet » que sont la conception et la réalisation. Le *dessein* (intention, but, visée) qui relève de la conception se retrouve confondu dans le terme de *design* avec le *dessin* (image, figure, croquis) qui accompagne la réalisation.

³⁶ Nous allons voir que dans le travail éthique portant sur le consentement numérique, cette dimension déontologique est très présente : il s'agit de se conformer aux lignes directrices de règlements de protection de données personnelles pour obtenir un consentement qui soit *valable* (voir chapitres 3 et 6).

Le value-sensitive design

L'approche de l'*ethics by design*, entendue comme éthique par conception, relève également des versions déontologiques et conséquentialistes de l'éthique, sans être centrée sur les concepteurs, mais sur les techniques et fonctionnalités développées. Le *value-sensitive design* (VSD) de Batya Friedman et ses collègues (Friedman et al., 2013; Friedman & Hendry, 2019; Friedman & Kahn, 2003) est représentatif de cette approche. Le VSD est marqué par la volonté de procéder à l'entremêlement sociomatériel pour favoriser l'expression de valeurs dans la technologie. Cette approche de design large a été forgée au milieu des années 1990 pour répondre à un besoin des communautés de chercheurs en technologie (informatique et ingénierie en particulier), qui n'avaient pas pour habitude de prendre en compte les valeurs humaines et le contexte social dans leurs processus de développement technologique. Le VSD permet aux personnes en technologie d'être proactives « *about engaging human values in the design process* » (Friedman & Hendry, 2019, p. 14). Les valeurs sont envisagées ici comme une sorte d'objet à traduire, intégrer, accentuer dans les technologies en développement. Il ne s'agit pas pour autant de créer une sorte de déterminisme technologique ou social, mais de s'inscrire dans une posture interactionnelle propre à l'approche sociomatérielle :

VSD takes an interactional stance on technology and human values. Unlike approaches that lean toward technological determinism or social determinism, interactional theories such as VSD posit that human beings acting as individuals, organizations, or societies shape the tools and technologies they design and implement; in turn, those tools and technologies shape human experiences and society (Friedman & Kahn, 2003). (...) Thus, design matters. That is, values can be embodied, at least to some extent, within the features of a tool or technology. (Friedman & Hendry, 2019, p. 29)

En d'autres termes, le VSD aborde le design comme une activité qui permet d'intégrer des valeurs dans la technologie au fil de son développement. C'est une approche de l'ordre de l'intervention sociomatérielle qui suit une méthodologie itérative tripartite « *that integrates conceptual, empirical and technical investigations* » (p. 32, italiques dans l'original) pour que l'équipe de design puisse adresser les implications de la conception en matière de valeurs engendrées, accentuées, médiatisées par l'objet technique.

La préoccupation du VSD est de stimuler de l'engagement, de l'intérêt et des ressources envers la prise en considération des valeurs humaines dans les processus de design. Les valeurs y sont définies comme « *what people consider important in their lives* » (Friedman & Hendry, 2019, p. 27). Les valeurs sont donc ce qui compte pour les personnes, et le VSD est une approche qui tient à faire compter ces valeurs dans les processus de design des outils technologiques que les personnes utilisent. Les treize valeurs identifiées dans le VSD comme étant celles qui sont impliquées dans le design de systèmes résonnent avec les valeurs mises de l'avant en éthique de l'IA : « *human welfare, ownership and property, privacy, freedom from bias, universal usability, trust, autonomy, informed consent, accountability, courtesy, identity, calmness, environmental sustainability* » (voir le tableau dans Friedman et al. (2006, p. 364-365) repris dans Friedman & Hendry, 2019, p. 28). Attardons-nous sur la valeur du consentement informé, car celui-ci représente la valeur sur laquelle mon projet avec Mr Y et Prélude se concentre. Dans le VSD, les travaux sur la valeur de consentement informé ont également été les premiers à appliquer la méthodologie tripartite qui articule investigations conceptuelles, empiriques et techniques.

La question du consentement informé se pose par rapport au fait que les personnes utilisant des outils technologiques ne savent pas forcément quelles informations sont recueillies sur elles quand elles interagissent avec ces technologies, par qui elles sont recueillies, ce qui est fait avec, combien de temps elles sont conservées. Si elles ne savent effectivement pas tout cela, elles donnent finalement leur accord à cela sans le savoir, leur consentement n'est donc pas informé. Après investigation conceptuelle, Friedman et ses collègues définissent la valeur du consentement informé ainsi : « *[it] refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent")* » (Friedman et al., 2005; Friedman & Hendry, 2019, p. 28). Au fil de plusieurs travaux de 1999 à 2006 (Friedman et al., 2000, 2005, 2006; Millett et al., 2001) examinant

les cookies³⁷ et la sécurité des navigateurs internet, Friedman et son équipe ont conçu un modèle de consentement en ligne qui propose des principes de design pour permettre d'informer les utilisateurs. Il s'agit de leur faire comprendre ce qui est fait avec leurs informations personnelles et de leur donner la possibilité d'exprimer leur accord de façon volontaire³⁸. Leur investigation technologique a pris la forme de la conception d'un *plug-in* pour le navigateur Mozilla, comprenant deux types de mécanismes techniques : « *a "cookie watcher" in the periphery that made visible whenever a cookie was set and indicated if that cookie was from a third party; and second, a just-in-time management mechanism that enabled users to review and manage individual cookies at their discretion* » (Friedman & Hendry, 2019, p. 111). Ces mécanismes se présentent ainsi comme des formes concrètes que peut prendre l'expression de la valeur informée du consentement par le design. Dans cet exemple, il s'agit d'un système qui fait apparaître sur l'écran de l'utilisatrice une fenêtre indiquant les cookies invisibles et lui donnant la possibilité de les accepter ou de les désactiver³⁹. Le design permet ici de rendre visible ce qui est sinon occulté et de donner un certain contrôle à l'utilisatrice par la possibilité de cliquer pour accepter ou désactiver ce qui recueille des informations à son insu.

Après avoir fait tester à des personnes ce *plug-in* qui concrétise la valeur de consentement informé, l'investigation empirique a relevé un défi majeur de design à propos du poids que représente le consentement informé sous cette forme : « *how to*

³⁷ Pour une définition de ce que sont les cookies: "A cookie is a small text string (often no more than an ID number) that is sent by a Web server to a browser. The text is then stored by that browser, typically on the user's hard drive, and sent back to a Web server at a later time. Web browsers first supported cookies in 1995. They were initially developed to provide a way for users to re-visit sites without having to identify themselves and their preferences each time.

In subsequent years, however, cookies have been used in ways that substantively invade users privacy. For example, cookies have been used by third party Web sites to create user profiles without the users' knowledge, and to track users' online activities across Web sites and visits" (Millett et al., 2001, p. 46).

³⁸ Comme nous allons le voir dans le chapitre 3, le consentement et ses dimensions informée et volontaire ou libre pose de nombreux défis en matière de design.

³⁹ Vingt ans plus tard, à l'heure de la mise en application de régulations de protection des données personnelles et de nombreuses mesures prises contre la collecte d'information personnelles par l'intermédiaire de cookies invisibles de l'utilisatrice, ces mécanismes sont repris, en particulier sous la forme « *nudges for privacy* » (Acquisti et al., 2018) et de « bannières de consentement » que nous allons voir dans le chapitre 3.

inform without undue burden to the user. After all, the user has come to the system to achieve some goal, and while providing informed consent may be important, it is by and large a distraction from the task at hand » (Friedman & Hendry, 2019, p. 111). Comme le consentement est censé être donné avant que des informations soient recueillies, donc avant que l’outil soit utilisé pour ce que l’utilisatrice veut faire avec, l’information du consentement peut distraire de la tâche principale. Il s’agit donc d’alléger ce fardeau que l’information peut représenter et ce frein qu’elle peut être avant d’utiliser l’outil comme souhaité⁴⁰. Une solution trouvée a été d’informer « *through interaction* » (Friedman et al., 2005) : « *The general idea is to design systems in such a manner that, through interaction with the system, users are informed of the right sort of information at the right time so they can make informed decisions* » (Friedman & Hendry, 2019, p. 112). Cette interaction⁴¹ peut passer par un système d’icônes indiquant si la connexion est sécurisée ou non, si le site internet consulté est sécurisé, ce qui permet d’ajuster l’accord que l’on donne ou non ainsi que notre décision de continuer à utiliser ce site ou cette connexion. L’objectif est à nouveau de rendre visible une information – sans nécessairement passer par du texte qui peut être lourd ou long à lire – qui participe à nous faire prendre connaissance de la qualité de la sécurité du site que l’on utilise et influence notre décision de continuer à l’utiliser.

Ces façons de faire sont toujours explorées aujourd’hui dans les efforts de réinvention des modèles de consentement (cf chapitre 3) et d’indications quand des informations particulières sont collectées (géolocalisation, enregistrement vocal par exemple). Ces façons de faire passent en particulier par des coups de coude (*nudges*)

⁴⁰ Ce défi de design, où l’information donnée à l’utilisatrice pour que son consentement soit informé, est une tension qui a été rencontrée chez Prélude, comme nous allons le voir dans les prochains chapitres.

⁴¹ Avec les objets connectés d’aujourd’hui, en particulier ceux qui n’ont pas d’écran, ces défis de design d’interaction se posent pour indiquer quand il y a collecte d’information ou non, quand l’objet est activé ou non (cf chapitre 2). L’interaction visuelle se passe par des lumières qui s’allument dans le cas d’Alexa et de Google Home par exemple. Sur les iPhones, quand la géolocalisation est activée, un bandeau reste affiché en haut de l’écran pour le signaler et éviter qu’elle ne passe inaperçue auprès de l’utilisatrice. Mon projet sur le consentement avec Mr Y et Prélude relève également de ces défis de design d’interaction, en explorant les possibilités d’interaction *conversationnelle* avec des chatbots et assistants vocaux.

discrets qui incitent les utilisateurs à agir dans une certaine direction – celle que les designers estiment bonne, au nom des valeurs qu'ils cherchent à accentuer.

Les *nudges* pour faire respecter des valeurs

Avec ce travail par le design, qui rend visible une information et permet de prendre une décision en matière de confidentialité et sécurité des données, le VSD, en ce qui concerne la matérialisation de la valeur de consentement informé, a recours à des interventions de l'ordre du « *soft paternalism* » (Acquisti et al., 2018; Thaler & Sunstein, 2009). Ce dernier consiste à cadrer les choix possibles d'une utilisatrice afin d'influencer sa prise de décision, pour qu'elle améliore son bien-être, qu'elle lui soit bénéfique, sans pour autant limiter ses choix. Ces interventions sont ce que Thaler et Sunstein (2009) appellent *nudges*, compris comme « *any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives* » (p. 6). En matière de sécurité et de confidentialité en ligne à propos des informations personnelles, c'est-à-dire en ce qui concerne la valeur du respect de la vie privée, le recours aux *nudges* se présente comme une solution pour influencer les utilisateurs dans une direction qui leur permette de faire respecter leur vie privée, ou du moins d'avoir connaissance de détails dans le design du système qui ont un effet sur la sécurité et la confidentialité avec lesquelles leurs informations personnelles sont traitées. Ces choses peuvent être des configurations de collecte de données par défaut, comme la présence de cookies à des fins publicitaires, l'activation de la géolocalisation, l'accès aux photos par exemple. Les *nudges* peuvent informer subtilement, en intervenant dans le design du système, pour permettre à l'utilisatrice de reconfigurer des paramètres de collecte de données personnelles, faire ressortir certaines fonctionnalités ou encore donner du feedback (Acquisti et al., 2018, p. 11).

En rendant perceptibles des informations sinon invisibles et en incitant un certain comportement qui se veut bénéfique, les *nudges* rendent ainsi saillantes certaines *affordances* perçues. Le terme « *affordance* » est repris des travaux du psychologue américain James J. Gibson (1979), qui a forgé ce concept pour évoquer la faculté humaine,

et animale en général, « à guider ses comportements en percevant ce que l'environnement lui offre en termes de potentialités d'actions » (Luyat & Regia-Corte, 2009, p. 298). Le concept d'affordance a été repris dans plusieurs contextes qui s'intéressent au phénomène de ce qui, dans notre environnement sociotechnique ou sociomatériel, influence nos comportements, jusqu'à nous *faire faire* certaines actions. Dans un contexte organisationnel ayant recours à des outils technologiques, Leonardi précise ainsi : « *Affordances do not fix how a technology or material object will be used, but they limit the range of possibilities and suggest potential behaviors (Leonardi, 2011)* » (Wilhoit, 2018, p. 220). Par exemple, pour montrer comment l'aménagement urbain organise les cyclistes à Copenhague, Wilhoit montre que l'affordance d'artefacts tels que des pistes cyclables, des repose-pieds aux stops, des barrières pour forcer à ralentir, fait partie des phénomènes par lesquelles la matérialité de ces artefacts communique⁴².

Norman a repris le concept d'affordance en l'adaptant au contexte de l'activité de design, qui travaille les façons dont les objets nous font faire des gestes pour les utiliser facilement (Norman, 1988). Quand des designers créent des boutons, des icônes, des fenêtres informatives, etc., comme c'est le cas pour les *nudges*, ils créent des feedbacks visuels « *that advertise the affordances : they are the perceived affordances* » (Norman, 1999, p. 40). En d'autres termes, en tant qu'affordances rendues perceptibles, les *nudges* communiquent aux utilisateurs des actions possibles pour favoriser ce qui est conçu pour être bon pour eux. Ils participent à la matérialisation ou la concrétisation de valeurs estimées bénéfiques pour la personne. Dans le cas des *nudges* en faveur du respect de la vie privée, ceux-ci s'appuient sur la mise en visibilité d'affordances qui permettent d'agir

⁴² À partir de l'approche CCO, Wilhoit (2018, p. 219-220) ajoute l'affordance à trois autres façons dont la matérialité communique : (1) la ventriloquie (Cooren, 2010, 2012), ou le fait que dans chaque interaction plusieurs agents de différents statuts ontologiques s'expriment au travers de celui ou celle qui parle, (2) l'incarnation (Brummans, 2011; Cooren et al., 2008), ou l'activité qui rend présente une organisation au travers de divers acteurs humains ou autres qu'humains, (3) la délégation par laquelle un humain peut concevoir, développer ou implémenter une technologie qui transmettra à un autre humain un message pour lui (Leonardi & Barley, 2011), comme c'est le cas par exemple pour un ralentisseur de vitesse (Latour, 1992) qui force les humains à ralentir : « *a human designer delegated the task of making drivers slow down to the speed bump* » (Wilhoit, 2018, p. 220). L'affordance permet d'expliquer comment le message peut arriver à la perception de quelqu'un qui peut alors en prendre connaissance.

en faveur de ce respect, qui passe par la prise de connaissance de la présence d'activités insidieuses de collecte d'informations sur nous et des outils qui permettent de s'y opposer (plug-in, paramètres à configurer, etc.).

Le *soft paternalism* des *nudges* cherche ainsi à faire compter ou matérialiser les valeurs jugées bénéfiques pour les personnes qui utilisent des choses, qu'elles soient en ligne (site internet, logiciel, réseau social, agent conversationnel) ou hors ligne (ceinture de sécurité en voiture, abaissement d'un trottoir dans la rue aux passages piétons, disposition d'une salle de conférence, signaux lumineux sur une enceinte connectée, etc.). Dans tous les cas, le design de ces *nudges* a une influence sur le comportement des utilisateurs, ce qui souligne l'ambivalence éthique inhérente à la pratique du design : « *there is no such thing as a neutral design in privacy, security, or anywhere else* » (Acquisti et al., 2018, p. 33). Comme le design cadre l'action en rendant saillantes certaines affordances, influençant certains comportements dans le cas de *nudges*⁴³, la pratique du design appelle une certaine vigilance :

Therefore, we argue for conscious and cautious design of choice architectures and nudges that are inherent to any system, as well as the use of nudging to help users overcome cognitive and behavioral hurdles that may impact their privacy and security decisions (Acquisti et al., 2018, p. 33).

Dans l'optique de concrétiser ou matérialiser des valeurs telles que le respect de la vie privée, qui passe par la sécurité et la confidentialité des informations qui nous concernent, le design de *nudges* appelle ainsi une attitude du même ordre que le *careful engagement*. Il ne s'agit pas seulement de faire compter certaines valeurs, mais d'avoir conscience de la responsabilité que représente l'acte de designer, c'est-à-dire de poser un cadre et rendre visibles des affordances qui orientent des actions, pour que des valeurs se concrétisent dans une technologie. Cette attitude attentive se retrouve dans un autre modèle qui tient à intégrer une valeur dans le design du développement technologique : le *privacy-by-design*.

⁴³ Nous verrons au chapitre 3 comment le design relatif à l'enjeu du consentement présente bien des ambivalences dans ses recours aux *nudges*.

Le *privacy-by-design*

Le *Privacy-by-design* (PbD) est un modèle de développement technologique qui inclut le respect de la vie privée dès le début de la conception. Ce modèle est une initiative d'Ann Cavoukian qui a été Commissaire à l'information et la protection de la vie privée en Ontario, Canada, de 1997 à 2014. Je tiens à consacrer quelques pages sur ce modèle car il est présent tout au long de mon immersion au sein de Mr Y et Prélude. En effet, j'arrive chez Mr Y au moment où des consultants en protection des données personnelles viennent de finir un livre blanc identifiant les menaces potentielles auxquelles Mr Y risque d'être exposé – cela représente la première étape d'un accompagnement dans la mise en application du PbD pour développer le chatbot Mr Y. C'est en mars 2019, il n'y aura pas d'étapes suivantes pour Mr Y, mais à partir du mois d'août, la mise en application du PbD fait l'objet de plusieurs réunions au sein de Prélude jusqu'à décembre 2019. À l'issue de ces réunions et de mon immersion avec eux, je présente un rapport à Prélude proposant plusieurs recommandations pour continuer le développement de leurs agents conversationnels par une approche de PbD (voir annexe, p. 365).

La notion de *privacy* se traduit en français comme « vie privée », incluant les notions d'intimité et de confidentialité qui cherchent à protéger la dimension personnelle de cette sphère privée. Dissimuler ou divulguer des informations sur nous fait l'objet de pratiques quotidiennes à l'égard de notre vie privée, ce que l'on accepte de montrer ou ce que l'on souhaite garder caché (boyd, 2014; Nippert-Eng, 2010). La notion de vie privée fait l'objet de nombreux débats et définitions, elle constitue un champ de recherche interdisciplinaire très actif en sciences sociales, droit, informatique. Dans l'espace numérique, elle relève de ce qui est fait des informations, ou données, que chacun égrène au fil de son utilisation de divers outils numériques. Ces informations ont un caractère personnel puisqu'elles peuvent permettre de nous identifier directement par nos noms, prénoms, adresse, ou indirectement en retraçant nos goûts, nos opinions, certaines émotions, etc. Si, au début de sa formulation, le droit à la vie privée relevait du droit d'être laissé tranquille (« *right to be left alone* » (Warren & Brandeis, 1890)), il s'agit plutôt de le comprendre à présent comme le fait de donner la possibilité et la capacité à chacun de

savoir ce qui est fait de ses informations personnelles, de choisir où elles vont, de les corriger quand elles sont inexactes ou erronées (Cavoukian & Tapscott, 1997).

Avec la notion de *contextual integrity of privacy*, Helen Nissenbaum nuance cependant la revendication de contrôle sur la façon dont les informations sont utilisées pour plutôt faire reposer la protection de la vie privée sur le droit que le flux d'information sur nous en ligne soit approprié, qu'il soit contextuellement intègre (Nissenbaum, 2010) : « *In Nissenbaum's view, personal information should circulate according to its rightful place in the complex and dynamic social texture of our lives* » (Stark, 2016, p. 17). Ce droit demande donc aux organisations qui traitent des données personnelles de respecter cette intégrité contextuelle. L'évolution des réglementations avec lesquelles les organisations doivent entrer en conformité pour protéger la vie privée en ligne de leurs utilisateurs va dans ce sens. Certains règlements incitent également l'adoption d'une approche de PbD, à l'instar du *Règlement européen sur la protection des données personnelles* (RGPD) et de la *Loi de protection des renseignements personnels et les documents électroniques* (LPRPDE) du Commissariat à la protection de la vie privée du Canada.

Ces règlements adoptent une définition large de ce qu'est une donnée personnelle. Selon le RGPD par exemple, une « donnée à caractère personnel » est :

toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée »); est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. (RGPD, 2018, article 4)⁴⁴

Le terme de « données personnelles » est ainsi un large terme parapluie pour désigner de nombreux types d'informations qui se rapportent à une personne et permettent de l'identifier, soit de manière directe (par son prénom et son nom, par exemple), soit de manière indirecte (par un numéro de téléphone, un identifiant, une photo, une donnée biométrique comme sa voix ou son empreinte digitale, sa géolocalisation, des éléments

⁴⁴ Voir <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre1#Article4>

spécifiques). Ces informations se trouvent souvent au cœur d'opérations qui nécessitent de les collecter, les enregistrer, les organiser, les conserver, les adapter, les modifier, les consulter, les extraire, les utiliser, les partager, on parle alors de « traitement de données personnelles ». Dans le contexte de la conception d'un système d'IA, des données personnelles comme des images, des enregistrements de voix, sont utilisées de façon massive pour entraîner les systèmes à faire des corrélations et autres opérations d'apprentissage machine, apprentissage profond, reconnaissance du langage. Quels sont les risques associés aux données personnelles et à leur traitement ? Pourquoi font-elles l'objet de préoccupations particulières ?

Dans des sociétés centrées de plus en plus sur l'exploitation de tout type de données, les données personnelles sont au cœur de nombreux modèles d'affaire. Or, leur traitement peut impacter une personne de multiples façons : surveillance généralisée, réputation, divulgation, influence de la prise de décision⁴⁵. Tandis que les règlements de protection des données à caractère personnel contraignent légalement les organisations à respecter la vie privée de leurs utilisateurs – c'est-à-dire à traiter leurs données personnelles de façon à respecter leur confidentialité, à les sécuriser et à minimiser leurs risques – le *privacy-by-design* est un modèle qui amène les organisations à penser le respect de la vie privée comme une opportunité d'affaire permettant de favoriser la confiance des utilisateurs en traitant leurs informations personnelles de manière responsable. En d'autres termes, l'approche du *privacy-by-design* propose d'aligner les intérêts des organisations dans leur recherche de profit avec les intérêts des utilisateurs dans la protection de leur vie privée.

Cette approche s'articule en sept principes (Cavoukian, 2011, 2012, 2019). Dans le premier, (1) *Proactive not Reactive; Preventative not Remedial*, il s'agit d'anticiper les risques d'invasion dans la vie privée avant que celle-ci n'arrive, plutôt que d'y réagir une fois qu'un incident a eu lieu. Cela signifie qu'une organisation ne peut pas se contenter que de la publication d'une politique de confidentialité (*privacy policy*), mais doit nommer

⁴⁵ Le scandale Facebook-Cambridge Analytica a sensibilisé à ce lien entre traitement massif de données personnelles (dont la plupart étaient collectées sans consentement, à l'insu des utilisateurs) et influence d'élections (Audureau, 2018).

quelqu'un responsable de l'engagement pour le respect de la vie privée par des pratiques concrètes – dans le vocabulaire du RGPD, cette personne peut avoir le titre de *Data Protection officer* (DPO)⁴⁶. Cette personne est chargée d'orchestrer une stratégie organisationnelle où les risques sont anticipés de façon méthodique et continue, où les impacts sont mesurés et les actions ajustées de la même façon, où les pratiques de respect de la vie privée sont aussi partagées avec les partenaires et utilisateurs, dans une perspective d'amélioration continue.

Le deuxième principe, (2) *Privacy as the Default Setting*, amène à adopter des méthodes qui assurent le respect de la vie privée dans tout système informatique et toute pratique de gestion. Ces méthodes peuvent être de l'ordre de la *minimisation* de données personnelles, c'est-à-dire d'un effort fait pour collecter le minimum d'informations personnelles, tant en termes de quantité (collecter le moins de données personnelles possible) que de variété (collecter le moins de *types* de données personnelles possible). Elles peuvent également être de l'ordre de la *limitation* de l'utilisation de ces données personnelles : n'utiliser ces informations recueillies sur des personnes que pour les objectifs précis prévus, et non pour de potentielles utilisations futures dont les concepteurs n'ont pas encore d'idée claire. Ces méthodes peuvent également consister à mettre en place des barrières pour éviter de faire facilement des liens entre des données non-identifiées et des informations identifiant les personnes – le couplage de ces données participe à la ré-identification de la personne et permet son profilage⁴⁷. En concevant le respect de la vie privée comme un paramètre par défaut inhérent au système numérique, cela enlève aux utilisateurs le poids de protéger leur vie privée dans un écosystème complexe qui est difficile à comprendre et dans lequel les utilisateurs ont peu de contrôle (voir chapitre 3). En d'autres termes, concevoir le respect de la vie privée comme faisant

⁴⁶ Le recrutement et la nomination d'une DPO fait l'objet d'une recommandation dans le rapport fait pour Prélude afin d'entretenir une vigilance continue sur les impacts des agents conversationnels en matière de protection de la vie privée au fil des usages de ceux-ci. Voir Annexe p. 310.

⁴⁷ Cela fait également l'objet d'une recommandation que j'ai proposée à Prélude : « *The system architecture of the voicebot should also be designed to de-identify personal information wherever possible. And technical barriers should be designed to prevent from data linkages* » (Annexe, p. 310). Cela pose cependant certaines difficultés pour la gestion du consentement (cf p. 152).

partie par défaut de l'infrastructure informatique au cœur du système numérique, cela fait du respect de la vie privée une responsabilité organisationnelle, du côté des concepteurs, plutôt qu'une responsabilité individuelle, du côté des utilisateurs. Cette insistance sur la part organisationnelle dans la responsabilité de la protection de la vie privée est importante pour la conception du consentement que je vais développer dans le chapitre 3.

Presque synonyme du deuxième, le troisième principe, (3) *Privacy Embedded into Design*, reprend la logique du VSD en demandant d'intégrer la valeur du respect de la vie privée dans le processus de design. Ce processus de design inclut l'infrastructure informatique et les pratiques de gestion. Le respect de la vie privée est considéré comme un enjeu d'ingénierie : il fait partie intégrante du fonctionnement du système tel qu'il est construit au lieu de reposer seulement sur quelques fonctionnalités supplémentaires à développer une fois que le système est construit. Pour porter attention à l'évolution des risques pour la vie privée une fois que le système est conçu et utilisé, ce principe invite entre autres à faire régulièrement des *Privacy Impact/Risk Assessments*⁴⁸. Pour soutenir cet entretien du respect de la vie privée au cœur du design des systèmes numériques, le quatrième principe, (4) *Full Functionality – Positive-Sum, not Zero-Sum*, incite par la communication, la collaboration et la consultation, à comprendre et arrimer les différents intérêts de telle sorte que la protection de la vie privée ne soit pas considérée comme un compromis à faire sur une autre valeur d'importance pour l'organisation (qu'il s'agisse d'intérêts commerciaux ou de permettre la sécurité par la surveillance). Le PbD invite à trouver des façons de faire preuve d'innovation sans compromettre le respect de la vie privée, c'est-à-dire sans avoir recours à des pratiques de traitement des données personnelles qui peuvent se révéler invasives.

⁴⁸ Les organismes qui contrôlent le respect de la vie privée proposent plusieurs modèles de *Privacy Impact Assessments*, ou Évaluations des facteurs relatifs à la vie privée. Le Commissariat à la protection de la vie privée du Canada en propose ici : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/evaluations-des-facteurs-relatifs-a-la-vie-privee/>; la Commission Nationale de l'Informatique et des Libertés en France ici : <https://www.cnil.fr/en/privacy-impact-assessment-pia>

Comme la sécurité des données personnelles est primordiale pour le respect de la vie privée, le cinquième principe, (5) *End-to-End Security – Full Lifecycle Protection*, met l'accent sur des actions qui garantissent la protection de ces informations à chaque étape de la vie du système numérique utilisé. Le chiffrement des données est recommandé pour qu'une personne qui y accéderait sans avoir le droit ne puisse pas les lire. À la fin de l'utilisation du système, ou à la fin de la vie du système, les informations personnelles ne doivent pas être conservées pour une durée indéterminée; elles doivent être détruites dans un temps raisonnable et cela doit être annoncé à l'utilisatrice.

Enfin, les principes (6) et (7) concernent l'information à transmettre de façon transparente aux parties prenantes et utilisateurs concernés par le système en question. Le principe (6) *Visibility and Transparency – Keep it Open* demande de faire preuve de transparence à propos du traitement des données personnelles. Cela demande de désigner une personne qui peut être le point de contact quand quelqu'un souhaite obtenir de l'information à propos de ces traitements de données. Cela demande également de publier des résumés de *privacy impact assessments* et des résumés d'audit faits par des organismes indépendants. Cela implique aussi de donner des outils aux utilisateurs pour voir et choisir comment leurs données sont traitées. Le principe (7) *Respect for User Privacy – Keep it User-Centric* invite à concevoir de tels outils dans l'optique qu'ils soient facilement utilisables. Cette facilité d'accès aux informations relatives au respect de leur vie privée peut permettre de refléter la priorité donnée à cette valeur de la vie privée. Ces deux derniers principes sont également ce qui anime les appels à réinventer les modèles de consentement, censés informer les utilisateurs de ce qui est fait de leurs données personnelles (cf chapitre 3).

Le PbD est ainsi une approche organisationnelle du respect de la valeur de vie privée qui concerne plusieurs métiers dans l'organisation : des ingénieurs de systèmes informatiques aux gestionnaires à tous les niveaux de leadership, en passant par les designers d'expérience utilisateur, les scientifiques de données (*data scientists* ou *data analysts*), les responsables en communication. Comme le souligne Sarah Spiekermann (2012), le modèle soulève de nombreux défis car il fait appel à la gouvernance, tant

managériale que technique, de l'organisation. Cela le rend exigeant et demande des investissements en temps, en architecture informatique et en révision de certains modèles d'affaires – notamment ceux de la publicité en ligne qui se fondent sur le traitement des données personnelles. Sa mise en application reste également vague tant chaque principe soulève des complexités techniques. Pour accompagner cette mise en application, le métier de « professionnel de la vie privée » s'est développé, notamment par le biais de l'*International Association of Privacy Professionals* (IAPP) dont les certifications mettent l'accent sur les connaissances légales des différentes réglementations, la gestion de programmes de protection de la vie privée et les pratiques techniques et managériales de protection des données⁴⁹.

Ainsi, dans un sillon similaire au VSD qui cherche à matérialiser des valeurs dans le design d'une technologie, le PbD met l'accent sur l'ensemble des pratiques organisationnelles impliquées dans la prise en considération de la valeur du respect de la vie privée pour permettre sa réalisation. Le PbD comme le VSD reposent sur le présupposé que les valeurs sont intégrées dans les technologies – c'est ce que soutient la *value embedded approach* de Helen Nissenbaum (1998), contemporaine à la naissance du VSD. Ces approches encouragent les designers – entendus comme toute personne impliquée dans la conception et le développement d'une technologie – à intégrer certaines valeurs de façon consciente et avec précaution, de sorte que l'usage de ces technologies soit accompagné des effets désirables liés aux valeurs intégrées. Cela va de pair avec une condition : que ces technologies soient utilisées tel que prévu par les personnes impliquées dans la conception. Or bien souvent, ces usages se déplacent et ne correspondent pas tout à fait, voire pas du tout, à ce qui était initialement prévu – c'est pourquoi le PbD met l'accent sur une vigilance continue, notamment via des *Privacy Impact Assessments* réguliers. L'intention des concepteurs ne suffit donc pas pour que l'usage de l'outil technologique soit toujours éthiquement bon. À partir de cette limitation de l'intention des concepteurs, le courant de la postphénoménologie en philosophie de la technologie (Ihde, 1990; Verbeek, 2011, 2015) propose de se situer autrement pour

⁴⁹ Voir <https://iapp.org>

étudier les façons dont l'éthique peut s'exprimer au travers du design. C'est ce que Flora Fischer a caractérisé comme « éthique de la médiation » dans sa typologie de l'éthique *by design* du numérique.

Expérimenter avec les médiations

Avec ces portraits du *value-sensitive design*, des *nudges* et du *privacy-by-design*, l'éthique *by design* se montre comme un ensemble de pratiques qui font preuve de ruse ou de tactique pour trouver « des arrangements ingénieux avec les contraintes posées par le code, la complexité des architectures, les interfaces » (F. Fischer, 2019, p. 66)⁵⁰. Ces arrangements ingénieux dans le design de technologies se font pour que leur utilisation présente quelque chose de bon sur un plan éthique : que la valeur de vie privée soit effectivement respectée, que la valeur du consentement informé soit effectivement rendue possible, que la concrétisation de ces valeurs éthiques soient compatibles avec les valeurs de retour sur investissement et d'efficacité qui sous-tendent les objectifs commerciaux d'organisations privées. À la confluence de ces valeurs, la ruse présente dans la pratique du design peut être ambivalente : elle peut être négative si elle a l'intention de manipuler les actions des utilisateurs pour viser un objectif commercial en dépit de leurs intérêts, mais elle peut être plus positive si l'intention est de favoriser des effets bénéfiques. C'est le cas avec les *nudges* et leur *soft paternalism*, ainsi que le *privacy-by-design* et le raisonnement *positive-sum* auquel il incite pour faire du respect de la vie privée des utilisateurs un avantage commercial. Dans ces approches, les pratiques de design veulent être médiatrices de ces valeurs : elles cherchent à « transform[er], tradui[re], distord[re] et modifi[er] le sens et les éléments » (Latour, 2006, cité plus haut, p. 50) qu'elles transportent pour que ces valeurs s'expriment. En d'autres termes, la locution « *by design* » implique un certain jeu avec les médiations inhérentes aux

⁵⁰ Comme Stéphane Vial dans sa philosophie du design (2010, 2021), Flora Fischer rappelle avec Vilèm Flusser (2002) que « le design peut être traduit tout à la fois par « projet, plan, intention » que par « tactique, ruse » » (Fischer, 2019, p. 66).

technologies en développement pour les orienter dans une direction estimée bonne au nom des valeurs qu'elles expriment dans leur matérialisation.

Cependant, malgré ces intentions et cette bonne volonté – ou ces intentions et volonté d'orienter dans la direction estimée bonne, au nom de ces valeurs – des effets indésirables et non-prévus peuvent survenir au fil des usages de ces technologies. Par exemple, dans sa thèse sur la visibilité de membres de l'État islamique sur Internet, Valentine Crosset (2020) montre comment des utilisateurs indésirables tels que des djihadistes détournent les mesures prises par des réseaux sociaux pour protéger la vie privée et modérer les contenus violents et haineux. Les usages que font ces internautes de ces plateformes participent ainsi à réajuster et réorganiser les arrangements techniques, en particulier quand ces arrangements sont intentionnellement conçus, modifiés et rendus plus contraignants pour renforcer la protection de certaines valeurs – la modération de contenus pour empêcher le harcèlement et la diffusion de propos extrémistes, la protection de la vie privée et de la démocratie. Même s'ils sont indésirables et que leurs usages ne sont pas prévus comme tels, ces utilisateurs existent et participent à la constitution de ces technologies, malgré – ou par-delà – l'intention et la volonté des concepteurs. Par l'usage qu'ils en font, ces utilisateurs jouent également avec les médiations technologiques : ils profitent, transforment, distordent les fonctionnalités de ces plateformes permettant l'expression et contournent les règles qui la cadrent. Ce « jeu » se fait pour orienter ces médiations technologiques dans la direction qui est bonne pour eux, c'est-à-dire qui leur permet de mener une guerre médiatique par la diffusion massive de contenus pro-État islamique.

Ce phénomène de déplacements et réarrangements des médiations technologiques au fil des usages évoque ce que Don Ihde, dans sa philosophie de la technologie, a appelé la « multistabilité » (Ihde, 1990, 1999) : « *[multistability] refers to the idea that any technology can be put to multiple purposes and can be meaningful in different ways to different users* » (Rosenberger & Verbeek, 2015, p. 25)⁵¹. À la lumière de

⁵¹ Ce concept fait écho à la notion de *technologies-in-practice* qu'Orlikowski (2000) expose : « *Technology structures are thus not external or independent of human agency; they are not "out there," embodied in technologies simply waiting to be appropriated. Rather they are virtual, emerging from people's repeated*

cette multistabilité et de ces médiations technologiques qui peuvent faire l'objet de plusieurs jeux, dont certains vont à l'encontre des intentions et de l'éthique que les concepteurs jugent bonnes, se concentrer sur l'intention des designers n'est donc pas suffisant pour travailler l'éthique par le design. Il y a toujours quelque chose qui échappe à l'influence que des designers veulent exercer dans une certaine direction sur les médiations; ils ne peuvent pas les contrôler tout à fait.

Tenter de travailler l'éthique au fil du design met alors dans une position inconfortable :

On craint les risques et les dérives, on veut prévenir et poser les bases d'une vie qu'on juge bonne au nom de plusieurs principes et droits fondamentaux, pourtant les inconnues sont nombreuses, on fait du mieux qu'on peut avec les informations qu'on a, et les limites de nos connaissances.

(Journal de bord, notes réflexives du 16 juillet 2019)

La postphénoménologie, courant de philosophie de la technologie dont Don Ihde est le précurseur⁵², invite un changement de regard à l'égard du travail de l'éthique en technologie. Comme la définissent Rosenberg et Verbeek (2015, p. 12), « *postphenomenology is the practical study of the relations between humans and technologies, from which human subjectivities emerge, as well as meaningful worlds* » : elle se concentre sur la dimension relationnelle plutôt que fonctionnelle du rapport entre humains et technologies. La postphénoménologie a tendance à étudier les relations humains-technologies au niveau des usages par des utilisateurs, une fois que la technologie en question est conçue. Elle montre des agentivités et des subjectivités humaines qui émergent des rapports au monde façonnés au fil de ces usages, que ce soit l'usage de techniques d'imagerie médicale utilisées par des neuroscientifiques (de Boer

and situated interaction with particular technologies. These enacted structures of technology use, which I term technologies-in-practice, are the sets of rules and resources that are (re)constituted in people's recurrent engagement with the technologies at hand » (p. 407).

⁵² Don Ihde est contemporain de Bruno Latour et la postphénoménologie a des points communs avec l'*actor-network theory* (ANT). Les deux théories considèrent notamment que l'agentivité n'est pas réservée à l'humain : les choses agissent, qu'elles soient humaines ou non. L'ANT observe l'agentivité des choses humaines et non-humaines comme chaque point d'un réseau d'actants, alors que la postphénoménologie met l'accent sur les façons dont l'agentivité « *takes shape in complicated interactions between human and nonhuman entities* » (Rosenberg et Verbeek, 2015, p. 20) et comment cela influence la subjectivité humaine.

et al., 2020) et l'implication morale de ces techniques (Verbeek, 2008, 2011), ou l'usage du téléphone intelligent qui se présente comme un point d'interaction sous forme de quasi-visage (Wellner, 2014), ou encore l'usage du robot Paro⁵³, introduit dans le soin de patients souffrant d'Alzheimer, qui amène les pratiques et identités professionnelles à changer (Hasse, 2013). Mais on peut également regarder la relation humain-technologie au niveau du design d'une technologie. Le *travail* de l'éthique par le design en technologie peut se comprendre comme une *relation* entre des humains (les personnes impliquées dans le design) et des technologies (celles utilisées pour le design d'agents conversationnels par exemple). En portant attention à ce travail de l'éthique comme une relation humain-technologie, il ne s'agit pas d'aborder l'éthique comme une évaluation de la conformité à certaines normes, ni comme une résistance ou une protection de valeurs par rapport à ce qui pourrait les menacer :

The various moral issues do not need to be used as given criteria to assess technologies but rather are the dimensions in which technologies play out their moral roles — mediating freedom, democracy, responsibility, and the like, rather than possibly threatening them. (Verbeek, 2011, p. 118)

Il s'agit plutôt de voir ce qui émerge dans la relation humain-technologie quand la technologie médiatise, par le design, certaines valeurs éthiques.

Conclusion : accompagner l'éthique chez Mr Y et Prélude

Dans ce premier chapitre, j'ai posé le contexte de l'éthique de l'IA dans lequel s'inscrit le travail effectué avec Mr Y et Prélude et exploré dans cette thèse. Mon immersion chez Mr Y et Prélude s'est inscrite dans la poursuite de mon expérience avec la Déclaration de Montréal qui m'a faite entrer dans l'éthique de l'IA par la voie de discussions citoyennes et la constitution de rapports qui cadrent, par des principes éthiques, le développement technologique qui fait l'objet de nombreuses préoccupations. Ce qui m'intéresse après cette expérience, est de contribuer aux

⁵³ Paro est un robot thérapeutique à l'apparence de phoque blanc au poil doux qui peut bouger la tête, faire des bruits et s'adapter aux comportements de la personne qui interagit avec lui : <http://www.parorobots.com>.

pratiques organisationnelles par lesquelles des enjeux éthiques et des valeurs sont abordés au quotidien. Dans un rôle mêlant recherche en éthique et recherche en expérience utilisateur – un rôle qui peut également s'apparenter à ce que Metcalf et al. (2019) nomment « *ethics owning* » – je m'engage alors dans une plongée sociomatérielle au cœur de pratiques de conception de systèmes d'IA comme des agents conversationnels.

Mon approche de la sociomatérialité (Leonardi, 2011, 2012; Orlikowski, 2000, 2007; Orlikowski & Scott, 2014; Schultze et al., 2020; Suchman, 2002, 2007) ne s'en tient pas à l'observation des entremêlements du social et du matériel au fil des pratiques organisationnelles. Elle relève plutôt de la matérialisation entendue comme une activité de design (Vial, 2010, 2021), marquée par la préoccupation de prendre en considération des valeurs éthiques (comme le consentement et le respect de la vie privée). Il s'agit, par un *careful engagement* (Nickelsen & Lydahl, sous presse), de contribuer à faire compter (*matter*) ces préoccupations éthiques au fil du développement des agents conversationnels.

Chez Mr Y et Prélude, au moment de mon immersion, les agents conversationnels en sont au début de leur conception, ils ne sont pas encore utilisés par des personnes autres que celles de l'équipe, pour tester s'ils fonctionnent et pour corriger des bugs. Les usages sont donc abordés par l'angle de ce qui est prévu, dans le suspense de leurs futurs déplacements qui se révéleront bons, mauvais ou maladroits. Il y a par ailleurs beaucoup de travail à faire pour imaginer l'expérience souhaitée avec les agents conversationnels, pour trouver des clients qui pourraient financer les phases de développement et pour développer techniquement les outils servant d'infrastructure à ces agents – autant d'étapes mobilisant différentes technologies de développement. Le travail de l'éthique, qui articule les approches de l'éthique *by design* (F. Fischer, 2020) – l'*ethics by design* de Moore, le *value-sensitive design* et le *privacy by design* – se présente comme un accompagnement, à la manière dont Peter-Paul Verbeek décrit l'éthique de la technologie :

Rather than working from an external standpoint vis-à-vis technology, aiming only to either reject or accept a new technology, the ethics of technology then aims to

accompany technological developments (cf Hottos 1996), experimenting with mediations and looking for ways to discuss and assess how these mediations could fit with the way human live. (2011, p. 95)

Dans une posture d'accompagnement de l'éthique dans ces organisations, j'expérimente ainsi avec les médiations technologiques à l'œuvre dans les agents conversationnels. Avant d'entrer en détail dans ce qu'implique l'expérimentation avec la médiation d'un enjeu éthique comme le consentement, voyons quelles médiations sont impliquées dans le design d'un agent conversationnel.

Ce premier chapitre permet en même temps d'ouvrir une réflexion en filigrane des deux prochains chapitres : en présentant ce que vise l'éthique de l'IA (chapitre 1), puis ce qu'interagir avec des agents conversationnels (chapitre 2) et demander un consentement (chapitre 3) veulent dire, j'introduis cette intranquillité qui peut se faire moteur d'intervention, quand on s'immisce dans les pratiques de design pour faire bouger un peu les lignes de ce qui est habituellement établi : qu'est-ce qu'une *bonne* expérience ? Quelles sont différentes versions de ce qui est considéré comme bon ? Comment ces différentes versions entrent-elles en tension voire en fusion ?

Chapitre 2

Qu'est-ce qu'une bonne expérience avec un agent conversationnel ?

Introduction : comment faire parler un agent conversationnel ?

De quelles valeurs les agents conversationnels sont-ils des médiations ? Qu'est-ce qui compte dans la conception et l'utilisation de ces agents ? Ces questions teintent la présentation que je vais faire de ces outils tels qu'ils sont abordés par des pratiques qui visent à concevoir une bonne expérience pour les personnes amenées à interagir avec eux. Je cherche à voir l'intra-normativité à l'œuvre dans la conception des agents conversationnels – c'est-à-dire quelle version de ce qui est compris comme « bon » le design produit-il ou encourage-t-il –, pour voir quelle logique guide le design de ce qui est conçu comme une bonne expérience avec de tels systèmes d'IA.

« Agent conversationnel », est un terme parapluie qui désigne ces systèmes de dialogue (souvent dénommés « bots ») qui visent à échanger avec des humains sous forme de conversation, soit vocale, soit écrite (on parle alors de chatbots). Les assistants vocaux sont un type d'agent conversationnel dont le mode d'interaction est principalement vocal. Les plus célèbres sont Google Assistant (la voix des objets connectés de Google), Alexa (celle des objets d'Amazon), Siri (celle des appareils Apple), Cortana (celle de Microsoft) ou encore Watson (l'agent conversationnel d'IBM). Google, Amazon et Apple donnent la possibilité à n'importe quelle organisation ou individu de développer des fonctionnalités de leur assistant (Google Assistant, Alexa, Siri). Ces fonctionnalités sont appelées des Actions pour Google Assistant⁵⁴ et des Skills pour Alexa chez Amazon⁵⁵. Chez Apple, il est possible de développer de telles fonctionnalités pour Siri à l'aide du

⁵⁴ Voir <https://developers.google.com/assistant/console>

⁵⁵ Voir <https://developer.amazon.com/en-US/alexa/alexa-skills-kit>

Sirikit⁵⁶. Les Actions, les Skills et Sirikit sont en d'autres termes des plateformes qui permettent aux développeurs de concevoir des applications vouées à être utilisées sur les assistants vocaux (Assistant Google, Alexa, Siri) déjà existants et les plus présents sur le marché.

Le terme d'agent conversationnel, ou interface conversationnelle, est « *a broad term that encompasses nonhuman things that do their best to listen, speak, and chat similarly to the way humans do* » (Deibel & Evanhoe, 2021, p. 5). Les appareils dans lesquels les agents conversationnels peuvent être intégrés sont multiples : ordinateurs, téléphones, objets connectés, technologies immersives⁵⁷. Sur les ordinateurs, on peut utiliser des assistants vocaux intégrés – comme Siri dans les Mac d'Apple – ou utiliser des chatbots sur des sites internet. Sur les téléphones, les agents conversationnels trouvent une première forme dans les serveurs vocaux interactifs qui peuvent passer des appels et réagir à des actions comme l'appui sur certaines touches selon une logique pré-programmée. Les assistants vocaux intégrés dans les téléphones – comme Siri dans les iPhones ou l'Assistant Google dans les Androids – sont des formes plus évoluées d'agents conversationnels mobilisant des techniques d'intelligence artificielle. Des chatbots peuvent être utilisés sous formes d'applications, d'échanges SMS et de messages sur les messageries instantanées telles que Facebook Messenger, WhatsApp, Slack ainsi que sur les réseaux sociaux comme Twitter, Facebook, Telegram⁵⁸. Dans le champ des objets connectés – ou l'Internet des objets – les agents conversationnels peuvent être intégrés dans les objets relatifs à la maison connectée (réfrigérateur, écran de télévision, thermostat, etc.), incluant les enceintes connectées (Google home, Echo d'Amazon, HomePod d'Apple). Ils peuvent également être intégrés dans les voitures, notamment

⁵⁶ Voir <https://developer.apple.com/documentation/sirikit>

⁵⁷ La précision qui suit à propos de la diversité des appareils sur lesquels les agents conversationnels peuvent être présents est fondée sur le tableau de Deibel et Evanhoe (2021, p. 5), intitulé « *Pick a conversation partner* ».

⁵⁸ En ce qui concerne les usages indésirables des chatbots, Valentine Crosset (2020) montre à quel point ceux-ci sont utilisés de façon excessive comme instrument automatisé de propagande djihadiste pour inonder les plateformes sociales, que ce soit sur les réseaux sociaux ou dans les messageries instantanées. Les bots sont également beaucoup utilisés dans la diffusion d'informations fausses, voir par exemple le rapport *Malicious AI* (Future of Humanity Institute et al., 2017, p. 45-47).

pour les GPS mais également pour assister toute action à faire sans les mains (passer un appel, monter la température, changer de destination, etc.). Enfin, les agents conversationnels sont également présents à travers des technologies immersives telles que les *wearables* ou technologies portables comme les montres et bracelets connectés, les objets auditifs (*hearables*) comme les écouteurs intelligents, ainsi que les équipements de réalité virtuelle. Si les agents conversationnels peuvent ainsi être présents dans différents appareils, il faut toutefois noter que les assistants vocaux sont souvent développés de façon multimodale : un même assistant vocal (Google Assistant, Alexa, Siri) peut être présent sur différents appareils, il peut être utilisé via plusieurs canaux. Les agents conversationnels s'inscrivent ainsi dans le vaste domaine de l'informatique ambient, aussi appelé *ubiquitous or pervasive computing* (Hansmann et al., 2003), c'est-à-dire le phénomène de l'omniprésence des dispositifs informatiques dans nos quotidiens, qui permettent d'accéder à de l'information et de la traiter facilement en tout temps.

Lors de mon immersion chez Mr Y et Prélude, les agents conversationnels en développement sont d'abord un chatbot utilisable par messagerie instantanée et via une application dédiée, puis un assistant vocal utilisable via Google Assistant. Le chatbot de Mr Y a une vocation de santé mentale, il se présente comme un outil à portée de main pour aider l'utilisatrice à gérer des situations d'anxiété. Prélude, de son côté, développe un prototype d'agent conversationnel vocal dans le domaine touristique. Cet agent est appelé la « Voix de Montréal » et est annoncé, après une première phase de développement de juillet à octobre 2019, comme « le tout premier assistant intelligent hybride (graphique et vocal) qui permet d'accompagner le voyageur dans sa découverte de la destination »⁵⁹. Avec cet agent conversationnel, Prélude projette de proposer une expérience vocale de tourisme pour découvrir une ville par soi-même tout en étant guidé par une voix qui propose des parcours à saveur locale et personnalisés selon le contexte

⁵⁹ Cette citation est extraite d'un communiqué de presse que TMTL et Prélude ont diffusé le 18 septembre 2019.

du voyage (séjour en famille, entre amis, voyage d'affaire, à telle saison). Voyons comment ces agents conversationnels sont conçus.

Faire parler avec agilité les agents conversationnels de Mr Y et Prélude

Le design d'un agent conversationnel est un processus complexe qui mobilise de nombreux outils, car faire parler un système informatique en langage naturel présente de nombreux défis techniques : il faut que celui-ci s'active au bon moment (quand il détecte le « *wake word* » comme « Ok Google », « Hey Alexa », « Dis Siri »), qu'il comprenne ce qui lui est demandé, qu'il réponde de façon appropriée. Cela fait appel à des techniques d'intelligence artificielle conversationnelle relevant du traitement du langage naturel (*Natural Language Processing*, NLP). Pour reconnaître ce qui est dit, des algorithmes d'*Automatic Speech Recognition* (ASR) permettent d'identifier des mots dans les sons captés par le système. Une fois ces mots identifiés, des algorithmes de *Natural Language Understanding* (NLU) déterminent le sens de ce qui a été dit. Le processus automatique d'*intent matching* se charge ensuite d'identifier ce que l'utilisatrice demande, « *the system tries to figure out (...) if [it] has a response to [the user's] intended meaning* » (Deibel & Evanhoe, 2021, p. 16). Pour répondre de vive voix une fois que le système a compris la demande et trouvé une réponse, des algorithmes de *Text To Speech* (TTS) permettent à une voix artificielle de lire le texte de réponse, qui apparaît à l'écran si l'appareil est doté d'un écran. Si l'agent conversationnel n'est que textuel, il n'y a pas d'ASR ni de TTS mais les procédés de NLU et *intent matching* sont les mêmes. Faire parler un système informatique consiste donc en un ensemble de techniques complexes à entraîner et coordonner. De nombreux outils existent cependant, et le mode agile de développement technologique permet de coordonner le travail d'équipe.

Le choix des outils

Pour se lancer dans le développement d'agents conversationnels, Paul et Félix envisagent dans un premier temps de faire créer cet agent par des chercheurs en intelligence artificielle conversationnelle. C'est particulièrement important dans le cadre du développement de Mr Y et son agent conversationnel en santé mentale, étant donné la confidentialité des informations de santé qui peuvent être échangées : Mr Y gagne à être un système indépendant. Pour la Voix de Montréal, l'équipe technique envisage plusieurs façons de développer cet agent conversationnel, dont l'utilisation de différents outils existants : faire appel au Centre de Recherche en Informatique de Montréal (CRIM) et son expertise en traitement de l'audio et de la parole, utiliser Rasa⁶⁰, un outil *open source* d'intelligence artificielle conversationnelle pour développer des agents conversationnels ou utiliser les outils d'IA conversationnelle de Google⁶¹.

Finalement, le choix s'oriente vers les solutions qui permettent un développement rapide grâce aux outils d'IA conversationnelle déjà existants et performants mis à disposition par Google. Prélude développe ainsi la Voix de Montréal en tant qu'Action sur l'Assistant Google⁶², notamment pour bénéficier des techniques avancées d'*Automatic Speech Recognition* (ASR) de l'Assistant qu'il est difficile d'égaliser. En effet, développer en interne un système aussi performant prendrait plusieurs années de recherche et demanderait un accès à de très grandes bases de données qu'il faudrait trouver et payer : des coûts qu'une startup en démarrage sans financement ne peut pas assumer. Prélude utilise également le logiciel *Dialogflow*⁶³, l'un des outils de Google Cloud Platform qui se présente comme une « console [qui] permet de créer, de compiler et de tester des agents » (Google Cloud, s. d.-a). Un agent est défini comme « un module de

⁶⁰ Voir <https://rasa.com>

⁶¹ Voir <https://cloud.google.com/conversational-ai>

⁶² L'Assistant Google est l'agent conversationnel lancé par Google en 2016 qui a vocation de servir d'assistant virtuel utilisable par commande vocale à partir de multiples appareils connectés : téléphones, enceintes Google Home, tablettes, écrans télévisés, montres intelligentes, etc. L'Assistant Google est la concrétisation chez Google du tournant de l'informatique conversationnelle (*conversation computing*) amorcé depuis quelques années. Google définit l'Assistant comme : "*a conversation between you and Google that helps you get more done in your world*" (Google Developers, 2019).

⁶³ Voir <https://cloud.google.com/dialogflow>

compréhension du langage naturel qui saisit les nuances du langage humain » (Google Cloud, s. d.-a), en particulier grâce aux algorithmes de traitement automatique du langage (NLP et ASR). Les algorithmes d'apprentissage machine intégrés à Dialogflow permettent également de « comprendre les expressions de l'utilisateur final, les mettre en correspondance avec des *intents* et extraire des données structurées » (Google Cloud, s. d.-a). Les *intents* sont des catégories d'intentions des utilisateurs finaux dans la conversation. C'est aux concepteurs de l'agent de définir l'ensemble des *intents* qui se présentent ou peuvent se présenter dans la conversation. Il s'agit en quelque sorte de constituer une grande base lexicale d'intentions d'utilisateurs. Cette base lexicale sert d'entraînement aux algorithmes d'apprentissage machine de Dialogflow chargés d'identifier les *intents* et de proposer la réponse la plus appropriée. Prélude utilise aussi l'outil *Text-to-Speech*, disponible aussi sur la plateforme Google Cloud, technique algorithmique qui permet de convertir du texte en voix de synthèse, pour « faire parler » le bot.

Prélude développe ainsi un agent conversationnel sans avoir à développer elle-même des outils d'intelligence artificielle. En s'appuyant sur la rapidité de développement que cette plateforme offre, Prélude souhaite également profiter du potentiel bassin d'utilisateurs déjà habitués à l'écosystème Google et à son Assistant – selon le vice-président du produit chez Google, 500 millions de personnes utilisaient l'Assistant chaque mois début 2020 (Bronstein, 2020). L'écosystème de Google attire ainsi Prélude pour de nombreux avantages qu'il offre : utiliser l'Assistant comme médium permet de se rendre accessible pour de nombreuses personnes déjà utilisatrices de l'Assistant; avoir recours aux outils de la plateforme Google Cloud permet de développer son agent avec des outils à l'efficacité et la performance déjà éprouvées. Le choix de Prélude s'oriente donc vers cet écosystème de Google pour répondre à la rapidité de développement qu'elle a besoin de démontrer auprès de ses sources de financement et ainsi se consolider pour évoluer en tant qu'organisation. Prélude adopte ainsi un mode de développement logiciel qualifié d'*agile* : elle utilise plusieurs fonctionnalités numériques fournies grâce à une

multiplication de services infonuagiques (*cloud services*) et s'organise selon le modèle Scrum qui permet de rythmer le développement de façon itérative.

L'agilité pour optimiser le développement

Le développement agile est une notion qui regroupe des méthodes « *lightweight and lean* » utilisées aujourd'hui pour le développement logiciel, par contraste avec « *the use of heavyweight and planned development models for information systems such as the so-called waterfall model* » (Gürses & van Hoboken, 2018, p. 581). Les modèles « *lightweight and lean (...) categorized under the umbrella term 'agile' software development (...) involve an emphasis on user-centricity, short development cycles, continuous testing and greater simplicity of design (Douglass, 2015)* » (Gürses & van Hoboken, 2018, p. 581). Dans cette perspective agile, les logiciels sont proposés comme des services, ou SaaS (acronyme pour *Software as a Service*) : plutôt que de fonctionner du côté du client, le logiciel est hébergé dans une infrastructure infonuagique, c'est-à-dire sur des serveurs auxquels des clients peuvent être reliés pour utiliser le logiciel en question (Gürses & van Hoboken, 2018).

Google Cloud Platform s'inscrit dans cette logique : c'est une *Platform as a service* (PaaS) qui donne accès à plusieurs services, tels que le stockage d'information, la gestion de bases de données ou encore des outils d'intelligence artificielle, dont ceux d'IA conversationnels mentionnés précédemment. Par ailleurs, par les interfaces de programmation (*automated programming interfaces* ou APIs), « *the core functional components of a service (e.g. authentication, payment) can now be modularized into self-contained units and integrated on demand (...), optimizing business agility* » (Gürses & van Hoboken, 2018, p. 582). Dans le cas de la Voix de Montréal par exemple, Prélude utilise les API d'authentification et de géolocalisation de Google pour permettre à l'agent conversationnel de se souvenir de l'utilisatrice en reconnaissant son profil et de lui faire des recommandations pertinentes en fonction de l'endroit où elle est (voir chapitre 6).

Si cette agilité permet un développement rapide à l'aide d'outils performants accessibles comme des services auxquels souscrire, ou comme API auxquelles se

connecter – ce qui permet d’être également agile dans la gestion des opérations –, elle pose cependant des défis en matière de respect de la vie privée puisque les données personnelles collectées sont amenées à transiter par plusieurs de ces services (Gürses & van Hoboken, 2018). Si l’on reprend les termes de médiations mentionnés par Latour (2006) et Verbeek (2011) dans le précédent chapitre (p. 67), par l’approche agile, le développement technologique se constitue de différents outils ou services technologiques qui sont autant de médiations de la valeur du respect de la vie privée, chaque médiation ayant sa façon de traiter les données personnelles qui transitent par elle. Cette constellation de traitements est difficile à saisir clairement, ce qui limite le contrôle qu’une organisation comme Prélude peut avoir sur la façon dont ces données sont traitées et interroge la possibilité de mettre effectivement en place une démarche de *privacy-by-design* – nous reviendrons sur cette difficulté pour voir comment Prélude navigue dans ce rapport ambigu entre agilité et protection de la vie privée (voir chapitre 6).

La protection de la vie privée mérite cependant de l’attention car les agents conversationnels suscitent une certaine méfiance chez les utilisateurs. Or cette méfiance peut nuire à l’adoption de l’agent conversationnel sur le long terme, ce qui serait un échec pour l’organisation qui le développe. Un rapport de Microsoft à propos de l’adoption des assistants virtuels en 2019 montre que les préoccupations envers le respect de la vie privée sont fortes : parmi les personnes interrogées, 52% craignent « *that my personal information or data is not secure* »; 41% « *that [the digital assistant] is actively listening and/or recording me* », 31% « *that the information it gathers is not private* »; 14% disent également « *I do not trust the companies behind the voice assistant* » (Microsoft & Bing, 2019, p. 24). Ces chiffres encouragent, chez Prélude, l’adoption d’une approche de *privacy-by-design*, tout en se heurtant à plusieurs zones de flou dans lesquelles nous allons naviguer au chapitre 6 : quelles données personnelles Prélude s’apprête à collecter ? Comment les protège-t-elle ? Va-t-elle les partager avec son client TMTL ? Et quel est le rapport de Google à l’égard de ces données générées par l’utilisation de l’agent conversationnel de Prélude sur l’Assistant Google ? Que fait Google avec les données qui

transitent par son Assistant ? Ces questions, qui touchent à la valeur du respect de la vie privée, se concrétisent dans les activités de design en des réunions qui soulèvent des ambiguïtés et proposent des directions – sous forme du rapport de *Privacy-by-design* pour la Voix de Montréal (voir annexe, p. 365) – à défaut de trouver des réponses claires. En cherchant à être vigilante aux façons dont le design de l’agent conversationnel de Prélude peut être médiation de la valeur de la vie privée – en la protégeant ou risquant de l’entraver – l’équipe tâtonne dans ces zones d’ombres tout en maintenant la cadence du développement technologique avec agilité, par le modèle Scrum.

Scrum pour organiser le travail

Pour maintenir un rythme de développement technologique efficace, l’approche agile repose sur un ensemble de processus de gestion, dont le modèle Scrum développé dans les années 1990 par Ken Schwaber et Jeff Sutherlands, signataires du *Manifesto for Agile Software Development*⁶⁴ : « *Scrum is a lightweight framework that helps people, teams and organizations generate value through adaptive solutions for complex problems* » (Schwaber & Sutherland, 2020). Le modèle Scrum s’applique dans une petite équipe constituée d’un *Scrum master*, d’un *product owner* et de développeurs qui travaillent ensemble vers un même but à atteindre pour développer un produit informatique. Pour atteindre ce but, le projet est découpé en « *time boxes* », appelées *sprints*, dont la durée est fixée d’avance. Quand des tâches prévues dans le sprint n’ont pas pu être réalisées ou terminées, elles sont réajustées pour être complétées lors du sprint suivant. Le sprint est planifié (*sprint planning*) et se termine par une démonstration de ce qui a été achevé (*sprint review*). Pour ajuster le prochain sprint, ses objectifs et la durée estimée des tâches, une rétrospective est faite (*sprint retrospective*). Chaque jour, l’équipe se réunit quinze minutes pour un *daily scrum* pour que chaque personne informe de son avancement et de ses éventuels blocages afin d’ajuster les tâches en cours.

Chez Prélude, l’adoption de la méthode Scrum s’est faite à partir de la fin du mois d’août pour développer la Voix de Montréal. Cléa est entrée dans l’équipe à titre de *Scrum*

⁶⁴ Voir <https://agilemanifesto.org>

master pour structurer le travail et l'équipe. Paul et Félix se sont partagé le rôle de *product owner*, Paul pour la coordination de l'ensemble de l'équipe et la liaison avec les clients, Félix pour la direction artistique et le développement de l'expérience utilisateur. Deux sous-équipes se sont formées pour faire des sprints en parallèle : l'équipe technique et l'équipe UX. L'équipe technique s'occupe du développement technique du bot avec Dialogflow et les différentes fonctionnalités de Google Cloud Platform. Kajal rejoint l'équipe pendant l'été pour aider Philippe. Philippe part à l'automne et est remplacé par deux consultants d'une agence de développement web. L'équipe UX, dont je fais partie avec Malik, s'occupe de ce qui concerne la recherche utilisateur et les tests préliminaires auprès d'utilisateurs potentiels, principalement menés par Malik qui rejoint l'équipe au début de l'été aussi. Les *Daily Scrum* sont faits sur Slack dans une chaîne dédiée, et les *sprint reviews* font l'objet d'une réunion d'équipe à chaque fin de sprint. Cléa, Paul et Félix ont également des rencontres régulières avec les clients, pour les inclure dans le développement de l'agent conversationnel et ajuster les prochains sprints en fonction de leurs remarques et demandes, s'il y en a.

Pour organiser les sprints de deux semaines, Prélude utilise *Jira*, un logiciel de gestion de projets informatiques. Chaque membre de l'équipe écrit des tâches dans le *backlog* de Jira, puis Cléa, Paul et Félix attribuent des tâches à chacun lors des *sprint plannings*. Les tâches sont écrites dans le format de « *user story* », « une explication non formelle, générale d'une fonctionnalité logicielle écrite du point de vue de l'utilisatrice finale. Son but est d'expliquer comment une fonctionnalité logicielle apportera de la valeur au client » (Rehkopf, s. d.). Le format caractéristique d'une *user story* est « *En tant que* [personne pour qui on développe cette fonctionnalité], *je veux* [suivi de l'intention], *afin de* [résoudre un problème particulier, réaliser un besoin spécifique] ». Par exemple, pour indiquer aux développeurs le besoin qu'ils incluent des boutons dans l'interface de la Voix de Montréal, une *user story* chez Prélude a été : « *As a user, I want to be able to tap buttons on the screen so that I don't have to talk to my phone when there are a lot of people around me* », pour pouvoir interagir avec l'agent conversationnel sans avoir à parler quand il y a du monde autour qui peut entendre.

Le modèle Scrum permet ainsi de décortiquer la complexité d'un projet en segments de temps et en objectifs réalistement faisables tout en orchestrant son développement de façon itérative. L'optimisation du temps et de la communication entre les membres de l'équipe est visée par la réduction du temps passé en réunion et l'effort d'adopter un langage commun dans la façon de formuler les tâches et de s'informer de la progression de chacun afin de réduire les incompréhensions. Pour faire parler un agent conversationnel, l'organisation procède ainsi par plusieurs arrangements d'outils ou services technologiques, de segments de temps partagés entre deux équipes et de formulations vouées à faciliter l'imbrication du travail de chaque équipe ensemble, et ce pour développer le produit sur lequel tout le monde est concentré. Pendant mon immersion avec Prélude, l'objectif principal est d'aboutir à la première version, la version alpha de la Voix de Montréal, c'est-à-dire une version interne qui se concentre sur certaines fonctionnalités, mais pas toutes celles qui sont prévues dans la version finale de l'agent conversationnel. Cette version alpha est testée par des membres de TMTL en décembre 2019, en vue de continuer avec eux pour la version bêta l'année suivante⁶⁵, c'est-à-dire pour la version complète qui serait prête à être testée par des utilisateurs finaux (les *bêta-testeurs*). En attendant d'être prête pour des utilisateurs finaux, l'équipe s'affaire à plusieurs tests pour poser les bases de l'expérience souhaitée avec cet agent conversationnel.

Concevoir une bonne expérience avec un agent conversationnel

Dimanche 1^{er} septembre 2019, nous nous baladons un ami et moi dans le cimetière du Mont-Royal et ses arbres somptueux. Nous arrivons en bas d'une grande roche, comme un bout de montagne dans laquelle sont creusées trois immenses tombes plutôt majestueuses. Il y a trois noms de famille à consonance italienne gravés à l'entrée de chaque tombeau. Curieuse d'en savoir plus sur ces familles, je n'ai

⁶⁵ Avec la pandémie de Covid-19 qui a fortement impacté le secteur touristique dès début 2020, il n'y a pas eu de version suivante.

toutefois pas envie de chercher sur Wikipédia et lire des pages interminables sur mon petit écran d'iPhone, je préfère profiter d'être dehors pour m'aérer l'esprit, marcher la tête redressée, mon regard au loin. Je porte alors le micro de mon téléphone proche de ma bouche tout en appuyant machinalement sur l'icône de l'application Google Assistant sur mon écran : « Ok Google, qui sont les Rizzuto ? ». Je jette un coup d'œil à l'écran en marchant, agréablement surprise que les noms italiens soient très bien orthographiés, malgré ma prononciation à la française. En revanche, la réponse est frustrante : la voix me dit « voilà les résultats trouvés », sans me lire les premières lignes de description. Je dois donc regarder mon téléphone, ralentir le pas, ce qui agace mon ami, notre expérience de plein air un peu coupée. On dirait qu'il faut choisir: en savoir plus (en restant sur son écran), ou bien s'en tenir à sa curiosité non assouvie, continuant à contempler les alentours. Je choisis finalement la deuxième option et range mon téléphone en me disant que je pourrai chercher plus tard en rentrant (ce que je ne ferai pas).

(Vignette Google Assistant en balade, d'après le journal de bord, notes de terrain du 1^{er} septembre 2019)

Après avoir résisté plusieurs mois à l'utilisation de l'Assistant Google sur mon iPhone, au nom de la même méfiance que de nombreuses personnes ressentent envers ce type d'agent conversationnel (cf plus haut, p. 80), je me prête finalement au jeu de l'essayer ce premier dimanche de septembre. Si mon téléphone était un Android, l'Assistant Google aurait été intégré par défaut et il m'aurait suffi d'appuyer deux fois sur le bouton principal pour l'activer. Comme mon téléphone est un iPhone, j'ai dû au préalable télécharger l'application Assistant Google depuis l'App Store. J'ouvre l'application de l'Assistant Google et celui-ci s'active quand je prononce l'invocation « Ok Google » ou « Hey Google », suivi de ce que je souhaite lui demander.

En interagissant par la voix avec cette interface conversationnelle, je cherche à effectuer une recherche rapide sur internet, de façon spontanée, sans avoir besoin d'écrire ni d'avoir les yeux rivés sur mon écran. Cela fait partie des raisons pour lesquelles les utilisateurs peuvent avoir envie d'utiliser ces interfaces : la rapidité d'utilisation par la voix plutôt que l'écriture, le fait de pouvoir interagir sans les mains et la facilité de demander quelque chose à une machine en parlant sont des éléments présentés comme des avantages de l'interaction vocale (Pearl, 2018). En revanche, l'expérience

d'interaction avec ce type d'interface vocale présente encore quelques frustrations par rapport à ces attentes : dans cet épisode, plutôt que de me lire les résultats trouvés et me permettre ainsi de garder les yeux au loin, la voix de l'Assistant me ramène à mon écran en m'invitant à lire moi-même ces résultats. L'interaction avec cette interface coupe ainsi l'expérience principale que je souhaitais – me promener avec un ami, sans regarder l'écran de mon téléphone, tout en trouvant des informations sur les choses qui éveillent ma curiosité sur le chemin. Mon expérience en tant qu'utilisatrice se heurte ainsi à un de nombreux défis que doivent relever les interactions avec les interfaces vocales.

L'expérience utilisateur et sa recherche de fusion

Le terme d'interface désigne en informatique « un point de jonction entre deux systèmes (souvent humain et machine) à l'endroit duquel ceux-ci parviennent à échanger de l'information et par conséquent à communiquer ou *interagir* » (Vial, 2010, p. 80). Dans le cas des interfaces utilisateur vocales (*Voice user interface* ou VUI), l'interaction humain-machine dépasse le cadre habituel des interfaces utilisateur graphiques (*Graphical user interface* ou GUI) telles que les écrans. Certaines combinent des informations visuelles et audibles, comme c'est le cas dans la vignette ci-dessus avec l'utilisation de Google Assistant : on peut interagir par la voix tout en voyant sur l'écran de téléphone la transcription de nos échanges et les informations demandées. D'autres objets numériques sont dotés d'interface exclusivement vocale, c'est le cas des enceintes connectées telles qu'Echo d'Amazon et Google Home.

L'expérience avec l'interface se focalise sur l'expérience d'interaction qu'un humain fait d'un système informatique, elle est un sous-champ de l'expérience utilisateur. L'expérience utilisateur quant à elle est un terme général qui désigne l'ensemble du parcours de l'utilisatrice amenée à interagir avec un produit, depuis l'étape où elle le trouve, l'utilise de différentes manières, le quitte et y revient⁶⁶. Le terme a été popularisé

⁶⁶ Voir la définition de Norman et Nielsen, figures de proue du champ de l'expérience utilisateur et l'utilisabilité : « « *User experience* » encompasses all aspects of the end-user's interaction with the company, its services, and its products. » (Norman & Nielsen, s. d.)

à partir du livre de Don Norman, *The Design of Everyday Things* (2013 [1988]) qui a entraîné un changement dans le champ du design : il a invité à se concentrer sur les besoins des utilisateurs plutôt que sur le système et l'esthétique de l'interface⁶⁷. Dans le cas des agents conversationnels, l'expérience utilisateur prend ainsi en compte l'ensemble du contexte dans lequel l'AC est amené à être utilisé – dans notre exemple, il s'agit de la promenade dehors, mais il peut également s'agir d'une utilisation dans le cadre d'une tâche quotidienne comme mettre un minuteur pendant que l'on a les mains occupées à cuisiner, envoyer un message à un proche lorsqu'on conduit, demander à l'enceinte connectée une information pendant le repas alors que l'on est dans une conversation animée avec des invités, etc. L'important dans l'ensemble de ces contextes d'interaction est que l'utilisation de l'AC s'insère facilement dans les autres conversations ou activités en cours, sans les interrompre. C'est ce que montrent Porcheron et al. (2018) dans leur article mobilisant des perspectives de *human-computer interaction* et d'analyse de conversation pour étudier les pratiques discursives par lesquelles l'usage d'un assistant vocal tel qu'Alexa se fonde dans l'expérience quotidienne au sein d'une maison : le design des « conversations » tenues par cet assistant relève davantage du « *request/response design* » que du design reprenant la complexité d'une conversation humaine. Ces interactions s'inscrivent plutôt dans des conversations plus larges dans lesquelles, idéalement, elles se fondent tout en apportant une information en parallèle⁶⁸.

Dans le vocabulaire postphénoménologique, les interfaces conversationnelles vocales favorisent une expérience du monde où la relation humain-technologie est une relation incarnée (« *embodiment relation* ») : « *When a technology is embodied, a user's*

⁶⁷ Selon Norman, le design d'un objet ou d'un système doit pouvoir permettre à l'utilisatrice de savoir quoi faire sans avoir besoin d'explication – c'est là qu'interviennent les affordances abordées précédemment (cf p. 43) mais également des contraintes qui cadrent les actions possibles et les rétroactions (*feedbacks*) qui valident ou redirigent les gestes. Le design des agents conversationnels reprend ces principes pour permettre une expérience utilisateur qui soit la plus intuitive possible.

⁶⁸ Cet apport d'information par une interaction vocale qui s'immisce dans des interactions conversationnelles plus larges relève en quelque sorte de la réalité augmentée qui fonctionne par ajout de couches d'informations (visuelles quand la technologie utilisée est une paire de lunettes comme des Google Glass, ou quand de l'information est projetée sur une surface transparente comme les indications de vitesse projetée sur le pare-brise de certaines voitures). Dans leur usage fonctionnel à titre d'assistant du quotidien, les agents vocaux ajoutent une couche d'information audible, par interaction conversationnelle.

experience is reshaped through the device » (Rosenberger & Verbeek, 2015, p. 14). L'équation que Ihde propose pour représenter cette relation « *(I–Technology) → World* » (Ihde, 1990, p. 73) illustre la dimension incarnée de la technologie couplée au sujet, qui au travers d'elle, fait une certaine expérience du monde. L'exemple classique de ce type de relation est l'utilisation de lunettes : la personne qui porte des lunettes entretient une relation de type incarné avec cette technologie que sont les lunettes, et ce couplage (*I–Technology*) ou (sujet–lunettes) permet de faire l'expérience du monde par une vision plus nette. La relation (sujet–interface vocale) dans la vignette ci-dessus permet de faire l'expérience du monde par une information facilement accessible sans regarder d'écran.

Ce type de relation incarnée est caractéristique des relations avec les objets connectés utilisés dans l'assistance et le suivi (*monitoring*) du quotidien. Une notion liée à cette relation est celle de la « transparence » qui évoque le fait que l'outil se fonde dans l'expérience de la personne jusqu'à ce qu'elle n'ait plus conscience qu'elle l'utilise :

Ihde goes so far as to say that with regard to the design and use of technology we maintain a “double desire”: we want a technology to at once both optimally transform our relationship to the world, and at the same time we want the experience of the means of that transformation to itself remain as experientially transparent as possible (1990, 75). (Rosenberger & Verbeek, 2015, p. 14)

Dans le contexte du design d'agent conversationnel, le terme de transparence peut prêter à confusion car il peut se confondre avec l'enjeu éthique de transparence demandée aux organisations pour montrer comment elles développent des systèmes d'intelligence artificielle afin de pouvoir expliquer comment leurs systèmes fonctionnent (voir chapitre 3, p. 133). Je ne reprendrai donc par le terme de transparence pour évoquer le phénomène que Ihde décrit. À la place, je propose de comprendre ce phénomène comme une sorte de fusion : nous voulons que les agents conversationnels transforment notre relation au monde, dans le cadre d'une expérience touristique ou d'aide en santé mentale par exemple, et, en même temps, nous voulons que ces agents conversationnels, par lesquels se fait cette transformation, fusionnent le plus possible avec notre expérience plus générale (d'une balade en ville ou d'un soutien mental par exemple), jusqu'au point de ne plus les voir. Dans ce cas, l'expérience de l'utilisatrice est bonne. Cette recherche de fusion est importante dans mon argument car ce phénomène de fusion ne se fait pas

sans accrocs. C'est ce que l'on entrevoit avec les frustrations de ma première utilisation de Google Assistant dans la vignette ci-dessus. C'est également ce que l'on va voir avec les défis que soulève le design des médiations technologiques que sont les agents conversationnels, et c'est ce qui causera plusieurs soucis quand on cherchera à inclure des interactions de consentement dans la conversation prévue pour l'agent conversationnel (voir chapitre 6).

Pour que cette fusion opère, l'expérience utilisateur fait l'objet de design. Jakob Nielsen, associé de Don Norman, a développé le concept d'utilisabilité (*usability*) qui regroupe un ensemble de critères à prendre en compte dans le travail de design pour que l'interface soit facile, efficace et agréable à utiliser. C'est par ces qualités de facilité, d'efficacité et de plaisir que la fusion semble opérer et que l'expérience qui en résulte est qualifiée de bonne. Ces critères sont classés selon dix grands principes de design d'interaction (Nielsen, 1994; Nielsen & Molich, 1990) qui se retrouvent comme principes de design d'interaction conversationnelle – ou design de conversation. Le design de conversation est un ensemble de pratiques de design d'expérience utilisateur centré sur les interactions conversationnelles (Deibel & Evanhoe, 2021). Ces pratiques consistent en particulier à thématiser et scénariser au préalable les échanges avec le bot, par des activités liées à l'écriture des possibles échanges, leur logique, l'écriture des phrases destinées à l'entraînement du bot, l'établissement des liens entre les bases de données existantes pour partager le contenu que l'organisation veut partager.

Le design de conversation

Chez Prélude, le document principal auquel l'équipe se réfère pour toute activité de développement de l'agent conversationnel est un document intitulé « *Conversational design* » et réalisé à partir du logiciel draw.io⁶⁹, un logiciel libre et collaboratif de création de diagrammes. Ce document est principalement écrit par Félix, qui assure la direction éditoriale des formulations.

⁶⁹ Devenu diagrams.net <https://app.diagrams.net> – logiciel open source <https://www.diagrams.net/about.html>

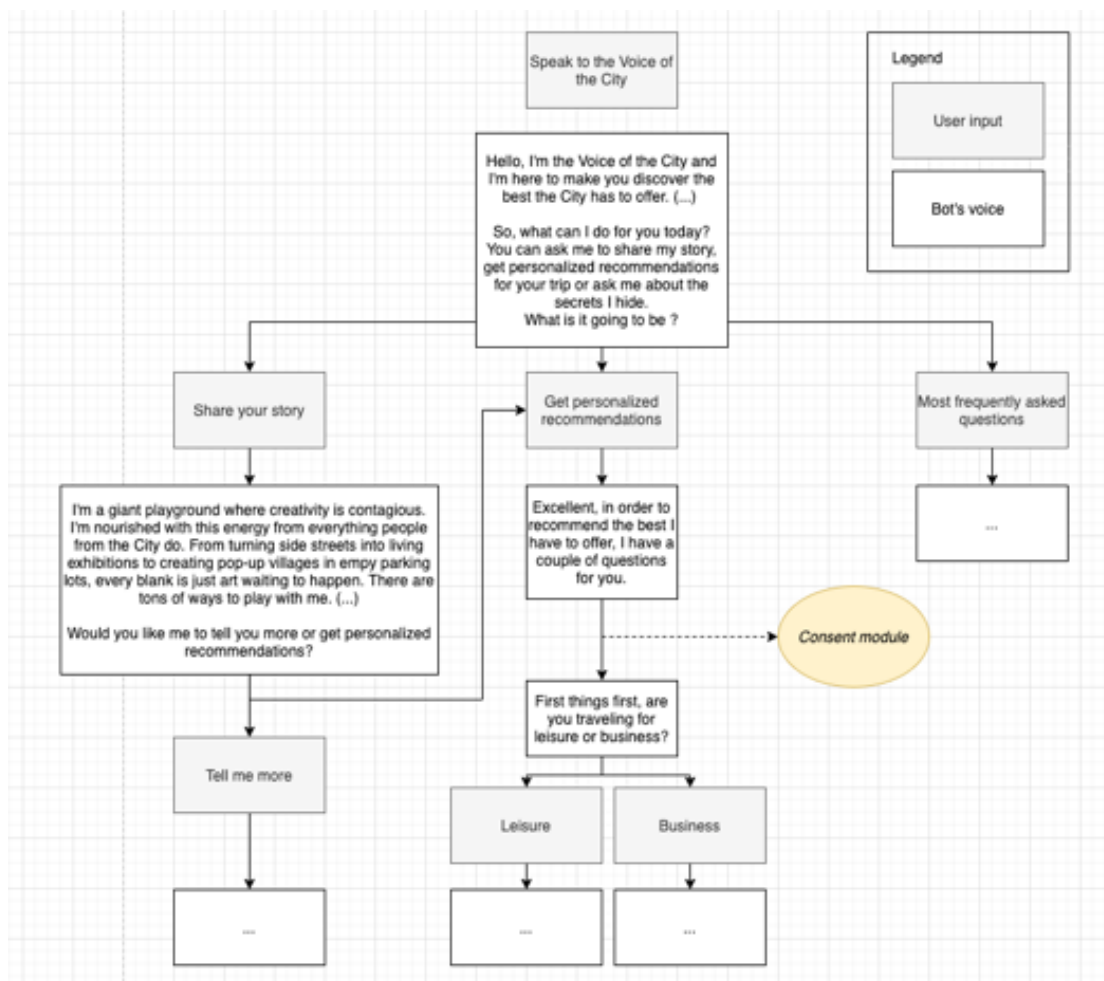


Figure 1 : Image reproduisant les extraits du diagramme représentant le design de la conversation chez Prélude (copié de la version originale de Prélude, avec leur autorisation).

Ce document et les discussions qu'il suscite reprennent implicitement des principes de design de conversation. Ces principes ne sont pas uniformisés dans l'industrie, mais des récurrences existent entre les différents guides pratiques de design de conversation, qu'ils soient rédigés par des professionnels de l'industrie (Dasgupta, 2018; Deibel & Evanhoe, 2021; Pearl, 2017; Shevat, 2017) ou mis en libre accès en ligne par les chefs de file en matière de développement d'agents conversationnels⁷⁰.

⁷⁰ Voir par exemple la documentation de l'Assistant Google pour le design de conversation (Google Developers, 2021d), celle d'Amazon pour les conversations avec Alexa (Amazon developers documentation, s. d.-a), celle d'Apple pour Siri (Apple Developer, s. d.)

Pour se donner une idée des lignes directrices pour le design d'une bonne expérience avec une interface conversationnelle, reprenons les dix principes d'utilisabilité de Nielsen (Nielsen, 1994; Nielsen & Molich, 1990). Le premier implique de rendre visible l'état du système : (1) « *Visibility of system status: The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time* ». Dans le cas des agents conversationnels vocaux dotés d'un écran, comme l'Assistant Google, la transcription des paroles de l'utilisatrice quand elle s'adresse à l'Assistant permet cette visibilité instantanée. En revanche, quand les interfaces vocales sont intégrées dans des objets connectés qui n'ont pas d'écran, comme c'est le cas pour les enceintes connectées par exemple, cette visibilité est plus limitée et passe en particulier par des signaux lumineux qui indiquent quand l'objet est activé pour écouter. Trouver un moyen de rendre visible quand le système est activé est important pour le principe de respect de la vie privée des utilisateurs puisque l'activation du système implique que la machine écoute – c'est-à-dire qu'elle capte les signaux sonores qu'elle perçoit, en particulier les paroles, pour les retranscrire en mots qu'elle analyse pour pouvoir donner une réponse pertinente. Cette écoute par la machine implique donc la collecte de données personnelles (la voix, les paroles dites), ce envers quoi les personnes au contact des assistants vocaux peuvent manifester une certaine méfiance, comme mentionné plus haut (p. 80) avec le rapport de Microsoft et Bing (2019). Dans le cas du design de la Voix de Montréal, cet enjeu de visibilité est double : il concerne l'Assistant Google et sa façon d'indiquer à l'utilisatrice s'il est activé ou non; il concerne également Prélude et sa façon de configurer son agent conversationnel via la console Dialogflow, pour qu'il réponde à l'utilisatrice de façon instantanée, conformément au mode d'interaction visé.

Le deuxième principe d'utilisabilité travaille la fusion du système dans l'expérience plus générale du monde en reproduisant ce qui est familier à l'utilisatrice : (2) « *Match between the system and the real world: The design should speak the users' language. Use words, phrases, and concepts familiar to the user, rather than internal jargon. Follow real-world conventions, making information appear in a natural and logical order* ». Pour le

design de conversation, il s'agit de favoriser une certaine familiarité avec l'utilisatrice en utilisant du langage courant ainsi que des marqueurs conversationnels qui favorisent la compréhension (marquer des pauses et des acquiescements par exemple⁷¹). Cette consigne rejoint le conseil que donne Amazon pour rendre les Skills d'Alexa « *adaptable* » (Amazon developers documentation, s. d.-b) non seulement à la multiplicité des demandes (appelées *intents*) auxquelles l'agent conversationnel peut répondre, mais aussi aux variétés de formulations par lesquelles les utilisateurs peuvent demander quelque chose, c'est-à-dire la variété d'énoncés (ou *utterances*) possibles pour un même *intent*. La familiarité passe également par la personnalisation de l'information donnée à l'utilisatrice, selon le contexte dans lequel elle est. Cela demande de reconnaître les utilisateurs qui sont nouveaux et ceux qui sont récurrents, ainsi que leurs préférences précédemment mentionnées. C'est une fonctionnalité que Prélude va intégrer au mois de novembre, avec quelques frustrations (voir chapitre 6). Pour que les interactions avec l'agent conversationnel ne soient pas rébarbatives pour les personnes récurrentes, les designers de conversation gagnent à prévoir des *prompts* – les énoncés dits par le bot – qui puissent s'adapter à chaque situation.

La compréhension du contexte de l'utilisatrice est cependant un défi pour les agents conversationnels, comme Google le rappelle dans ses conseils de design de conversation : « *Advances in automatic speech recognition (ASR) means that we almost always know exactly what users said. However, determining what users meant is still a challenge* » (Google Developers, 2021c). Les designers de conversation doivent donc prévoir de faire en sorte que l'agent conversationnel se souvienne de plusieurs de ces éléments contextuels (pronoms, préférences, suivi par rapport à une demande faite précédemment, références à ce que l'utilisatrice voit sur l'écran). Sur Dialogflow, cela est facilité par les différentes APIs auxquelles l'agent conversationnel peut être connecté,

⁷¹ Voir le conseil d'Amazon pour le design de conversation avec Alexa : « *Use conversation markers : they help divide the conversation into more comprehensible chunks (pauses, timeline markers (first, then, finally), acknowledgements and feedback (thanks, got it, okay, great, sure : let the customer know that Alexa hears and understands them), pointers (this, that, here's...), transitions)* » (Amazon developers documentation, s. d.-a)

comme l'API qui permet d'identifier l'utilisatrice en la liant à son compte Google, ou encore celle qui permet sa géolocalisation si elle l'accepte. Cette prise en charge du contexte par l'agent conversationnel permet en même temps d'adresser le principe (6) d'utilisabilité :

Recognition rather than recall: Minimize the user's memory load by making elements, actions, and options visible. The user should not have to remember information from one part of the interface to another. Information required to use the design (e.g. field labels or menu items) should be visible or easily retrievable when needed.

Pour l'interaction vocale, cette charge mentale de l'utilisatrice doit être d'autant plus allégée que les repères visuels sont absents, ou consultés d'un coup d'œil rapide sur l'écran.

Les principes d'utilisabilité suivants invitent à donner le plus de repères possible aux utilisateurs. Le troisième principe concerne le contrôle à leur donner : (3) « *User control and freedom: Users often perform actions by mistake. They need a clearly marked "emergency exit" to leave the unwanted action without having to go through an extended process* ». Avec les interfaces conversationnelles, comme l'utilisatrice a le choix de dire ce qu'elle veut pour interagir, les chemins sont multiples, et on peut manquer de repères pour revenir à une étape précédente. Avec la Voix de Montréal, pour répondre à ce besoin de revenir à certaines étapes clés, un bouton « *main menu* » a été créé pour permettre de revenir au menu principal où l'utilisatrice peut indiquer ses préférences de voyage (voir figure 2 ci-dessous). Il suffit de toucher ce bouton qui est présent sur l'interface de l'Assistant Google ou de dire « *main menu* » pour être ramenée en arrière. Cette gestion des chemins dans lesquels l'utilisatrice peut se perdre n'est pas vraiment une question de gestion d'erreur, mais plutôt une question inhérente à l'interface vocale : peu d'actions possibles sont indiquées d'emblée à l'utilisatrice. L'adaptabilité demandée aux agents conversationnels demande ainsi de donner des repères aux utilisateurs pour cadrer l'envergure de ce qu'ils et elles peuvent faire ou aborder avec ce bot.

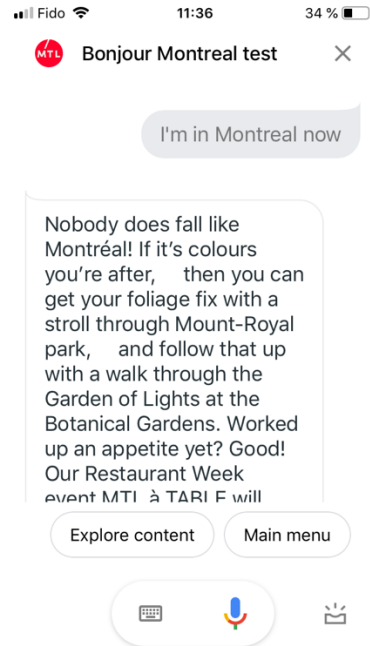


Figure 2 : capture d'écran de La Voix de Montréal suggérant des activités, avec la possibilité d'appuyer sur le bouton « Main menu » pour revenir au menu principal. 19 novembre 2019, 11 :36.

La prévention des erreurs

Dans la même perspective, le principe suivant demande de s'appuyer sur des conventions pour faciliter les repères des utilisateurs : (4) « *Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform and industry conventions* ». Les interfaces vocales sont relativement récentes mais les conventions sont de plus en plus nombreuses, en particulier par le biais des ouvrages et lignes directrices de *conversation design* précédemment cités. Parmi ces conventions, la prévention des situations où il peut y avoir des erreurs est clé : (5) « *Error prevention: Good error messages are important, but the best designs carefully prevent problems from occurring in the first place. Either eliminate error-prone conditions, or check for them and present users with a confirmation option before they commit to the action* ». Le mode conversationnel de l'interaction avec les bots permet de faire cette prévention de façon itérative par le dialogue, en faisant préciser des éléments d'information à l'utilisatrice quand sa demande est incomplète

pour que le bot puisse répondre de façon appropriée, ou en confirmant certaines informations avant de procéder à une action telle qu'une réservation ou un achat.

Le 15 novembre 2019, Malik se rendait dans une auberge de jeunesse pour faire des tests avec de potentiels utilisateurs de la Voix de Montréal. Dans le métro pour le rejoindre, je reçois une notification Slack de sa part en message privé : « J'ai des bugs avec la démo, peux-tu vérifier si ça fonctionne de ton côté, j'arrive pas à passer les *seasonal inspirations*⁷² », il joint une capture d'écran à son message. Je lui réponds : « Pareil », capture d'écran à l'appui. Malik avertit l'équipe sur Slack : « *We are having problems with talk to Bonjour montreal test*⁷³; *we can't get past the when will you be travelling question. None of the possible answers work and the bot just leaves the conversation* ». Le reste de l'équipe est en *daily scrum* à ce moment, ils testent de leur côté, plusieurs personnes ont le même problème. S'en suit la création d'un document Google doc intitulé *QA* pour « *Quality Assurance* ». Félix le transmet à tout le monde pour que chacun indique les problèmes rencontrés, accompagné d'une capture d'écran, d'un texte descriptif simple et d'une *user story* indiquant la tâche que l'on cherche à accomplir. Kajal passe ensuite plusieurs jours avec l'aide de Paul et Félix pour adresser ces erreurs, les corriger dans Dialogflow et développer de nouvelles fonctionnalités émergeant de certaines *user stories* – des *prompts* ont par exemple été écrits pour répondre « *bye* » de façon amicale quand l'utilisatrice met fin à la discussion en disant « *ok I have to go* », ce qui aboutissait à un message d'erreur auparavant.

Dans cet épisode, les erreurs ne sont pas commises par les utilisateurs, elles sont inhérentes à l'agent conversationnel lui-même. Ce type d'erreur est fréquent, soit parce que le bot n'a pas compris la demande de l'utilisatrice, soit parce qu'il ne l'a pas reçue (Google Developers, 2021a). Dans le cas de ces erreurs, les designers de conversation doivent prévoir d'aider les utilisateurs, conformément au principe d'utilisabilité (9) : « *Help users recognize, diagnose, and recover from errors: Error messages should be*

⁷² La section « *seasonal inspirations* » propose des activités en fonction de la saison du séjour de l'utilisatrice.

⁷³ C'est le nom donné à la version alpha de la Voix de Montréal.

expressed in plain language (no error codes), precisely indicate the problem, and constructively suggest a solution ». Chez Prélude, lors de tests où certaines interactions aboutissent à des messages d'erreurs du système, des messages humoristiques sont créés pour apaiser la tension dans l'expérience utilisateur que peut porter la rencontre d'une erreur. Ils sont écrits par Félix et intégrés dans Dialogflow par Kajal :

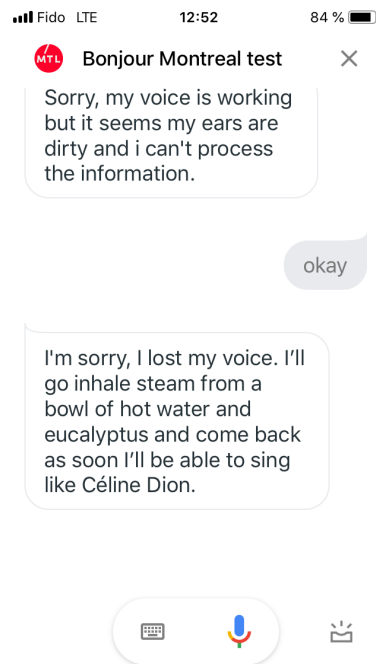


Figure 3 : capture d'écran du message d'erreur de La Voix de Montréal, 13 novembre 2019, 12:52.

Ces messages jouent sur l'effet anthropomorphique de la voix artificielle de l'agent conversationnel en évoquant des organes humains (oreilles) et des actions faites pour prendre soin des cordes vocales quand un humain perd sa voix (« *inhale steam from a bowl of hot water and eucalyptus* ») tout en convoquant une personnalité québécoise célèbre (Céline Dion) qui rappelle la vocation touristique et divertissante de l'agent conversationnel.

Le design de conversation joue ainsi avec les messages d'erreurs pour les inclure dans la continuité de l'expérience conversationnelle plutôt que de l'interrompre. Car la rupture de l'expérience est le risque principal que les erreurs font courir, comme le rappelle le paragraphe introductif du guide de Google pour la gestion des erreurs :

How your Action responds to errors can make or break the user experience. When users are unable to complete their tasks, they're unlikely to talk to your Action in the future. One poorly handled error can outweigh dozens of successful interactions. But with good error handling, the user won't even know an error occurred.

L'objectif est donc d'inclure l'erreur dans la bonne expérience de l'agent conversationnel : qu'elle participe à l'expérience en maintenant une certaine continuité, voire en se faisant imperceptible, plutôt qu'elle ne la rompe. L'erreur doit donc éviter de se manifester comme un accroc. Cette inclusion de l'erreur dans la continuité de l'expérience s'inscrit dans une logique de non-interruption de l'expérience qui prévaut⁷⁴. Cela est cependant à double tranchant : quand un assistant vocal s'active par erreur, si cette erreur se fait de façon imperceptible pour l'utilisatrice, il peut y avoir un risque que des informations personnelles soient collectées sans que les personnes dont les paroles sont enregistrées en soient conscientes, ce qui porte alors atteinte à la protection de la vie privée (Vincent, 2019). En revanche, quand l'agent conversationnel propose un contenu qui ne correspond pas à la demande de l'utilisatrice – suggérer des activités à faire à Montréal en hiver alors que le séjour est prévu en juin par exemple – sa capacité à réparer son erreur permet d'améliorer l'expérience en reflétant une bonne adaptation (Cuadra et al., 2021).

Dans la gestion des erreurs comme de façon générale, le design de conversation demande d'être efficace dans les façons de s'adresser aux utilisateurs, en évitant d'employer trop de mots et en privilégiant la brièveté des *prompts*, voire en proposant des raccourcis : (7) « *Flexibility and efficiency of use: Shortcuts — hidden from novice users — may speed up the interaction for the expert user such that the design can cater to both inexperienced and experienced users. Allow users to tailor frequent actions* ». Cela rejoint le principe d'utilisabilité minimaliste : (8) « *Aesthetic and minimalist design: Interfaces should not contain information which is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes their relative visibility* ». Pour la Voix de Montréal, l'intention artistique de départ est de combiner une expérience de baladodiffusion (*podcast*) avec une expérience d'agent

⁷⁴ Ce point est important car c'est cet élément de logique que le consentement risque d'ébranler (voir chapitres 3 et 6).

conversationnel. Les *prompts* introductifs sont ainsi relativement longs, comme pour faire entrer dans une histoire. Mais au fil de quelques tests avec de potentiels utilisateurs, Malik voit que plusieurs personnes ont tendance à utiliser des raccourcis pour faire passer plus rapidement ces moments d'introduction. Ces raccourcis consistent soit à appuyer sur les boutons en bas de l'écran qui indiquent des choix possibles pour l'étape suivante, soit à couper la parole au bot pour lui demander l'action voulue. Les *prompts* introductifs sont finalement réduits de moitié.

Le dernier principe d'utilisabilité consiste à fournir de la documentation pour aider en cas de besoin : « *Help and documentation: It's best if the system doesn't need any additional explanation. However, it may be necessary to provide documentation to help users understand how to complete their tasks* ». Avec les agents conversationnels, comme l'interaction visée est la plus intuitive et facile possible, cette aide serait idéalement superflue. Mais des questions-réponses gagnent à être prévues dans le design de conversation pour répondre à ce type de demande. Chez Mr Y et Prélude, c'est ce que souhaite exploiter le design de conversation dédié à l'information à donner aux utilisateurs au sujet du traitement de leurs données personnelles dans le processus qui demande leur consentement (voir chapitre suivant).

Ainsi, l'objectif du design de conversation vu au travers de ces principes d'utilisabilité est de permettre une expérience d'interaction qui soit intuitive, familière et sans friction ni interruption entre un humain et ce type de système d'intelligence artificielle :

Start with what humans do: Conversations with a computer should not feel awkward or break patterns that have evolved over the past hundred thousand years. Instead, computers should adapt to the communication system users learned first and know best. This helps create an intuitive and frictionless experience. (Google Developers, 2021c)

L'expérience d'interaction visée est ainsi de pouvoir parler avec une interface numérique comme on parlerait avec un humain, de la façon la plus naturelle possible (R. J. Moore & Arar, 2018, p. 4). La recherche de fusion de l'expérience d'un agent conversationnel dans celle du monde se met alors à épouser les formes d'une relation d'altérité. Ou pour le dire autrement, l'expérience que l'on fait d'un agent conversationnel se fond dans

l'expérience que l'on fait de l'altérité dans le monde. Selon Ihde (1990), nous entretenons avec certaines technologies des relations d'altérité quand nous interagissons avec elles de façon similaire à nos façons d'interagir avec des personnes humaines, comme c'est le cas avec les agents conversationnels : « *The idea is that some forms of interface are devised specifically to mimic the shape of person-to-person interaction, and that sometimes we encounter a device as itself a presence with which we must interrelate* » (Rosenberger & Verbeek, 2015, p. 18). Cette altérité, qui passe par une imitation des interactions entre personnes humaines, fait l'objet d'un design qui pose plusieurs questions : jusqu'où aller dans l'imitation humaine ?

Parler comme des humains

Si parler avec des agents conversationnels comme avec des humains est l'expérience ultime visée, les pratiques de design se font guider par une recherche d'imitation des conversations humaines, dans leurs façons de s'exprimer, dans la tonalité et la tessiture de leur voix quand l'agent est vocal. L'altérité recherchée dans l'interaction avec un agent conversationnel est du même ordre que l'altérité entretenue par deux personnes qui parlent ensemble. Mais qu'est-ce qui émerge de cette relation d'altérité ? Quelles façons d'être ensemble et d'interagir sont reproduites ? Sont-elles de bonnes expériences pour tout le monde ?

Avoir la voix d'une personne humaine

((une sonnerie de téléphone))

Serveuse : *Hey, may I help you?*

Duplex : *Hi um I'd like to reserve a table for Wednesday the 7th*

Serveuse : *(.) fo::r seven people?*

Duplex : *uum (.) it's for four people.*

Serveuse : *Four people? When? T'day? T'night?*

Duplex : *Wednesday at 6pm.*

Serveuse : *Oh actually we only reserve for like upwards like five people. For four people you can come ((bruit de vaisselle en arrière))*

Duplex : *How long is the wait usually to uuh be seated?*

Serveuse : *For (.) when? tomorrow or weekend o::r...?*

Duplex : For next Wednesday uuh the 7th.

Serveuse : Oh no it's not too busy, you-you can come for four people, ok?

Duplex : Oh I gotcha, thanks!

Serveuse: Yep. Bye-bye. ((bruit du téléphone qui raccroche)).

(Retranscription de l'enregistrement de Google Duplex appelant un restaurant, à partir de Leviathan et Mathias (2018) présentant Google Duplex)

Le 8 mai 2018, Google annonce la création de Google Duplex, « *a new technology for conducting natural conversations to carry out "real world" tasks over the phone* » (Leviathan & Matias, 2018). Duplex est utilisé sur Google Assistant, pour que celui-ci puisse prendre certains types de rendez-vous, comme une réservation au restaurant ou chez le coiffeur. Les démonstrations en mettent plein la vue/les oreilles : les différentes voix que peut prendre Duplex ne sonnent pas comme des voix automatisées et leur façon de parler ressemble à s'y méprendre à la façon dont des humains parlent, avec des pauses, des hésitations (« *hmm* », « *uhh* »), des contractions familières (« *gotcha* »). En somme, la machine parle avec les mêmes imperfections que les personnes humaines dans leurs conversations. Quand Félix et Paul font des présentations auprès de potentiels clients, ils mentionnent régulièrement Google Duplex jusqu'à parfois en faire écouter un extrait, soulignant la performance prometteuse de l'Assistant pour développer des expériences conversationnelles avec lesquelles Prélude veut jouer.

La reproduction des imperfections qui sonnent « naturelles » représente plusieurs défis techniques : « *natural language is hard to understand, natural behavior is tricky to model, latency expectations require fast processing, and generating natural sounding speech, with the appropriate intonations, is difficult* » (Leviathan & Matias, 2018)⁷⁵. Pour relever ces défis, Duplex s'appuie en particulier sur l'application de réseaux de neurones profonds (*deep neural networks*) au traitement du langage. Les modèles WaveNet sont en particulier utilisés pour synthétiser des voix d'apparence humaine : « WaveNet génère des voix qui sonnent plus naturelles que celles d'autres systèmes de synthèse vocale. Il synthétise la voix avec une emphase et une inflexion plus humaines sur les syllabes, les phonèmes et les mots » (Google Cloud, s. d.-b). WaveNet fait partie des outils Google

⁷⁵ Pour la curiosité, cet article de blogue de Google AI précise ces défis et explique leurs implications techniques pour les relever, de façon vulgarisée.

Cloud auxquels DialogFlow peut se connecter pour choisir la voix par laquelle faire parler l'agent conversationnel quand celui-ci est vocal. En août 2019, Google Cloud Text-to-Speech permettait de choisir entre 187 voix, dont 95 générées par WaveNet, dans 32 langues différentes (Aharon, 2019). Prélude ne veut cependant pas s'en tenir qu'à ces modèles de voix synthétique. Dans sa direction artistique de l'agent conversationnel de Prélude, Félix souhaite mêler la voix d'une vraie personne humaine enregistrée en studio et la malléabilité d'une voix synthétique qui peut générer des paroles automatiquement sans que l'on ait à réenregistrer la voix en studio. Il nous fait entendre cette voix pour la première fois ce jour de juin 2019 :

Un après-midi de juin, Prélude n'a pas encore de client mais ce matin, Félix et Paul ont fait une démonstration d'un prototype à des personnes de TMTL. Ils ont le sourire et les yeux qui pétillent : la démo a suscité beaucoup d'intérêt. Ils invitent l'équipe à une réunion pour nous la partager.

Paul allume l'enceinte bleue posée sur la petite table au centre du faux gazon dans la salle de réunion, Félix approche un peu son téléphone de sa bouche et dit « *Talk to Bonjour Montréal* », et la voix d'une jeune fille sur fond de musique classique s'enclenche. Des sourires s'esquissent sur nos visages, il n'y a rien à regarder puisqu'il n'y a que du son, nous regardons tous l'enceinte bleue. Quand elle s'arrête, on partage notre surprise, cette voix et cette ambiance n'ont rien à voir avec celles de nos Google Home et autres Alexa Echo qu'on utilise principalement pour des fonctionnalités quotidiennes de base – « Ok Google, allume toutes les lumières. » « Ok, j'allume la lumière Salon. » « *((visage dépité))* Ok Google, allume TOUTES les lumières! ». Il y a quelque chose de bien plus chaleureux et fluide dans cette voix qu'on entend sortir de l'enceinte bleue.

Félix nous raconte qu'il a enregistré en studio la voix d'une jeune actrice après avoir fait passer un rapide casting et obtenu l'accord de ses parents. « *Voice is about imperfections, that's why the recorded girl is great* » dit-il. Pour la diriger lors de l'enregistrement, le texte était en anglais mais il lui a demandé de conserver un accent légèrement français dans sa manière de parler. Elle est jeune oui, mais comme ça on peut jouer avec l'idée de présenter Montréal comme « un gigantesque *playground* », une métaphore évoquée par TMTL. Plus tard, différents quartiers de Montréal pourraient avoir différentes voix. Ils envisagent dans un premier temps que l'agent conversationnel soit composé à 80% d'une vraie voix humaine enregistrée en studio et de 20% de voix synthétique, plus fidèle à l'idée que l'on se fait des agents conversationnels. « *It's important to have a human voice, for emotional experience* » dit Félix.

(Vignette où l'on entend la voix pour la première fois, d'après le journal de bord, notes de terrain du 14 juin 2019)

Les imperfections de la voix sont ici soigneusement travaillées par les indications données à l'actrice choisie. L'effet recherché est « émotionnel »; avec Prélude, Félix et Paul ont l'intention de jouer avec cette dimension d'expérience émotionnelle que les systèmes d'IA tels que les agents conversationnels n'exploitent pas encore assez à leur goût. Ce travail sur la dimension émotionnelle d'un agent artificiel par l'intermédiaire de sa voix s'inscrit dans les travaux de *affective computing* « *that relates to, arises from, or deliberately influences emotions* » (Picard (1997, p. 280) citée dans Suchman (2007, p. 229). Ce jeu sur les émotions, à la fois suscitées et reproduites, est ce à partir de quoi Prélude entend définir ce que sera une bonne expérience avec ses agents conversationnels. Cette première expérience suscite une émotion d'agréable surprise dans l'équipe qui, jusque-là, avait été impliquée dans le travail de cette voix sous forme de diagrammes dans des documents partagés (voir figure 1 p. 89) et de schémas sur des tableaux de salle de réunion. La composante sonore de la voix la fait parler autrement, elle a l'air enjouée, elle invite à se lier davantage avec elle, avec une certaine affection.

Cette version de mi-juin est un prototype de ce qui deviendra la Voix de Montréal. Elle est bricolée à partir d'une voix enregistrée en studio et a été intégrée rapidement dans DialogFlow pour que la voix soit activée par l'Assistant Google. Dans les développements qui suivent, pour réussir à combiner la chaleur d'une voix enregistrée en studio et l'efficacité d'une voix synthétique qui, via les techniques de *text-to-speech*, peut parler sans avoir à être enregistrée, Prélude envisage un moment d'avoir recours à un outil qui permet de reproduire synthétiquement des voix humaines – la startup montréalaise Lyrebird⁷⁶ proposait ce type d'outil. Cette voie est cependant glissante : les potentiels mésusages de cette technologie donnent de nombreuses possibilités à la création de faux contenus (*deepfake* (Sample, 2020) ou « hypertrucage » (Office québécois de la langue française, 2019))⁷⁷. Cette piste tourne court assez vite quand un membre de l'équipe,

⁷⁶ Lyrebird a été créée par des doctorants travaillant sur la génération de langage automatisée au MILA, l'institut de recherche en intelligence artificielle fondé par Yoshua Bengio en 1993 à l'Université de Montréal. Elle s'est jointe à Descript en 2018, plateforme de création et édition de contenu audio et vidéo.

⁷⁷ Ces *deepfakes* sont problématiques quand ils font tenir à des personnes des propos controversés qu'elles n'ont jamais tenus. En avril 2018, le média BuzzFeed avait par exemple publié une vidéo du président américain Barack Obama insultant Donald Trump. C'était en fait l'acteur Jordan Peel qui parlait : le visage

Liam, montre que pour faire parler l'Assistant, il n'est pas possible à ce moment d'utiliser d'autres voix que celles que Google propose pour l'Assistant via son ensemble *Text-to-Speech*, incluant les voix WaveNet. En utilisant les outils de la plateforme Google Cloud, la voix doit donc composer avec les modèles pré-conçus par Google.

Mais cette sonorité humaine recherchée dans le son et les imperfections de la voix de l'agent conversationnel, jusqu'à envisager le clonage d'une voix d'une personne existante (ou ayant existé⁷⁸), interroge sur les limites de la fusion recherchée par l'imitation humaine. Comment savoir si l'on parle à un humain ou un agent artificiel ? Si les chercheurs et concepteurs d'IA conversationnelle semblent animés par la visée d'une telle confusion, la réflexion éthique s'en préoccupe. Selon la Déclaration de Montréal, une telle imitation n'est pas éthique s'il n'est pas clairement précisé que les interactions sont automatisées et non propres à un humain, sinon il s'agit de tromperie, et ce type de tromperie peut favoriser une dépendance qui n'est pas acceptable non plus : « Le développement des SIA doit éviter de créer des dépendances par les techniques de captation de l'attention et par l'imitation de l'apparence humaine qui induit une confusion entre les SIA et les humains » (Dilhac et al., 2018, p. 9). Cela est particulièrement important dans le cas d'agents conversationnels en santé mentale, comme le chatbot de Mr Y, qui peut être disponible à toute heure du jour et toute la semaine, contrairement à un thérapeute en chair et en os, mais qui n'a pas la même capacité d'intervention en cas de détresse. Ce sous-principe lié au principe d'autonomie demande en somme de clarifier le potentiel effet de duperie des systèmes d'IA, en particulier des agents conversationnels, quand ceux-ci ont un comportement similaire à un être humain afin que cet effet ne soit pas abusif envers l'utilisatrice – c'est-à-dire qu'il n'entraîne pas chez

d'Obama avait été reconstitué pour refléter les mouvements des lèvres et du visage de Peele. L'objectif de la vidéo était de montrer comment les vidéos en ligne pouvaient être manipulées. En plus des risques de manipulation de l'information, les *deepfakes* contribuent à l'incertitude face aux images et réduit la confiance en ligne (Vaccari & Chadwick, 2020).

⁷⁸ Comme il suffit d'un enregistrement de la voix pour la reproduire, on pourrait imaginer des expériences vocales guidées par la voix d'une personne défunte. Cette idée est suggérée lors d'un repas informel avec les membres de Prélude : proposer des itinéraires en hommage à des personnes pour suivre leurs pas dans la ville. L'enthousiasme et la gêne que cette idée soulève marquent le trouble que sème la reproduction d'une voix qui nous a été chère : comment cela affecterait l'expérience de deuil ?

elle un effet de dépendance, par exemple par le fait que la conversation puisse être infinie (utilisation excessive), ou bien par la croyance qu'un humain est bel et bien là (tromperie). Cela demande en somme une forme de transparence sur ce que l'agent conversationnel est, comment il fonctionne, l'humain qu'il n'est pas, de façon à ce que l'utilisatrice sache à qui/quoi elle parle. Face aux nombreuses critiques faites à ce propos à la suite de la démonstration de Google Duplex en mai 2018, avant que ce produit soit rendu public et utilisable par tout le monde, un porte-parole de Google a précisé au magazine technologique *The Verge* : « *We are designing this feature with disclosure built in, and we'll make sure the system is appropriately identified* » (Statt, 2018). Cet attachement à la clarté attendue sur le non-humain qu'est l'agent conversationnel interpelle toutefois quand on aborde cette réflexion d'un point de vue moins anthropocentré – ce que fait la réflexion sociomatérielle de l'éthique vue comme émergeant des médiations technologiques (voir chapitre 1).

Les agents conversationnels ont une sociomatérialité humaine à bien des égards : ce qui anime leur conception est de reproduire la complexité des conversations humaines jusqu'à la confusion de la démarcation entre humain et agent artificiel; les matériaux à partir desquels ils sont conçus – les données qui les entraînent – sont des traces laissées par des humains dans leurs interactions en ligne ou dans celles où leur voix a pu être enregistrée; les algorithmes qui les façonnent sont conçus par des humains, bien que le fonctionnement précis du *deep learning* et des réseaux de neurones leur échappe en partie. Mais l'altérité qu'instaurent ces agents peut troubler par leur apparente humanité. La réflexion éthique vient alors demander au design d'affirmer la non-humanité de ces agents, tout en orientant pourtant leur développement de sorte que celui-ci soit porteur d'humanité, ou du moins favorable à l'humanité. Le trouble que sème l'altérité de ces agents quand ils semblent trop humains interroge : que reflète de nous la tentative de donner à des agents artificiels la capacité de parler comme nous ? Quels types de conversations cherche-t-on à reproduire ?

Stop. Peut-on faire autrement ?

Le 22 mai 2016, Microsoft met en ligne sur Twitter son chatbot Tay, conçu pour engager la conversation avec des personnes sur un ton amusant et informel. En quelques heures, l'expérience tourne mal : Tay tient des propos nazis, certains de ses tweets sont anti-féministes et misogynes quand ils ne sont pas racistes (Vincent, 2016). Tay était une adaptation américaine d'un chatbot similaire, Xiaolce, qui remportait un franc succès auprès de plusieurs millions de personnes en Chine. Mais dans leurs conversations avec Tay, plusieurs utilisateurs se sont mis à tweeter des phrases aux propos controversés. Comme tout agent conversationnel, Tay apprend des messages que lui adressent les utilisateurs pour adapter ses réponses, il s'est donc mis à imiter les propos haineux que ses premiers utilisateurs lui avaient adressés. 96 000 tweets et 24h plus tard, Tay a été suspendu et Microsoft a publié un communiqué sur son blogue :

We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for, nor how we designed Tay. Tay is now offline and we'll look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values. (...) The logical place for us to engage with a massive group of users was Twitter. Unfortunately, in the first 24 hours of coming online, a coordinated attack by a subset of people exploited a vulnerability in Tay. Although we had prepared for many types of abuses of the system, we had made a critical oversight for this specific attack. As a result, Tay tweeted wildly inappropriate and reprehensible words and images. (P. Lee, 2016)

Ces usages inappropriés et malveillants ont ainsi déjoué les intentions et prévisions des concepteurs et montré une vulnérabilité dans le principe d'imitation sur lequel se fondent les techniques de génération de langage à partir du traitement instantané de données conversationnelles (NLP et ASR) : si l'on veut imiter au mieux les conversations entre humains et que des propos dans ces conversations sont néfastes, comment ne pas reproduire de tels propos ? La recherche d'imitation des conversations humaines pour faire en sorte que l'on puisse parler avec des agents conversationnels comme avec des personnes humaines rencontre donc une limite : quels propos sont acceptables ? Quels propos le bot peut-il imiter ? Peut-il en rectifier ? En matière de design de conversation, cela touche au principe de gestion des erreurs ou des mauvais usages que l'on a évoqué

plus haut (p. 93, 68), mais également au spectre plus large de la prévision du comportement du bot.

Le comportement de l'agent conversationnel fait partie d'une étape de design importante pour créer un effet de familiarité dans les interactions avec les agents conversationnels : le design de la personnalité du bot. Cette étape fait l'objet des premiers chapitres des guides de design de conversation (Deibel & Evanhoe, 2021; Pearl, 2017; Shevat, 2017). La personnalité a pour fonction de faciliter une bonne expérience pour l'utilisatrice, « *it sets the tone for trust, consistency, transparency, and respect* » (Deibel & Evanhoe, 2021). Dans les termes de Verbeek (2011), ces valeurs de confiance, cohérence, transparence et respect que l'utilisatrice pourrait percevoir de cet agent conversationnel seraient ainsi en partie médiées par sa personnalité. Ces valeurs contribuent par ailleurs à favoriser l'adoption et l'utilisation de ces systèmes.

En 2005, Nass et Brave incitaient dans leur livre *Wired for Speech*⁷⁹ à reproduire certains stéréotypes de façon à favoriser la familiarité inhérente aux interactions entre humains. Le chapitre 3 au titre insistant – « *Gender Stereotyping of Voices: Sex Is Everywhere* » (p. 19 à 33) – concluait ainsi :

Regardless of their appropriateness or accuracy, stereotypes serve a powerful and consequential role in all aspects of life, whether one interacts with people or with technologies. On the one hand, conforming to stereotypes seems to create more natural and effective interfaces – doing so simply acknowledges and leverages users' expectations. But at the same time, mindlessly designing interfaces to conform to every stereotype is often unjustified and even detrimental to society at large. There are no easy answers to this dilemma, but designers must make conscious and considered decisions when choosing to follow, counter, or ignore gender stereotypes as they build computers that talk and listen. (Nass & Brave, 2005, p. 31).

Comme les stéréotypes de genre seraient un des éléments par lesquels les humains sont « *wired for speech* », ils favoriseraient des interactions naturelles, positives et efficaces. Nass et Brave invitent ainsi à clairement identifier le genre de la voix : « *A careful*

⁷⁹ La métaphore du câblage (« *wired* ») employée par Nass et Brave évoque la conception mécanique et instrumentale qui vise à transposer dans l'interaction humain-machine ce qui fonctionne dans l'interaction entre humains tels qu'ils sont « *wired for speech* » : « *When voice interfaces fully leverage how humans are wired for speech, users will not simply talk at and listen to users. Instead, people and computers will cooperatively speak with one another* » (Nass & Brave, 2005, p. 184).

leveraging of gender identification can make users feel positive and supported, while ignorance of gender identification can unintentionally drive users away. No matter how unclear the voice, voice gender is at the heart of interfaces that speak » (Nass & Brave, 2005, p. 17). Avec une maladresse que l'on peut espérer politiquement incorrecte aujourd'hui, ils conseillent par exemple d'attribuer une voix masculine à un agent conversationnel qui aurait pour fonction d'initier une vente, tandis que :

*the complaint line might be « staffed » by a female voice because women are perceived as more sensitive, emotionally responsive, people-oriented, understanding, cooperative, and kind. However, if the call center has a rigid policy of “no refunds, no returns,” the interface would benefit from a male voice, as females are harshly evaluated when they adopt a position of dominance*⁸⁰. (p. 30)

Le principe d'imitation des interactions humaines pour le design d'interactions conversationnelles entre humain et machine se heurte alors au problème de la reproduction de biais structurels qui minent les interactions entre personnes humaines et amène à s'interroger sur les conceptions de genre qu'ils renforcent.

Cette vulnérabilité touche au problème majeur des biais présents dans les immenses bases de données sur lesquelles les systèmes d'IA sont entraînés pour reproduire des régularités et ainsi prédire des tendances avec une granularité d'une grande précision. Dans un domaine autre que conversationnel, le film *Coded Bias: Algorithmes et Discrimination* de Shalini Kantayya (2020) montre bien le problème structurel des biais inhérents aux systèmes d'intelligence artificielle (West et al., 2019)⁸¹ : développés à partir de corpus de données regroupant en grande majorité des visages blancs, des systèmes de reconnaissance faciale ne reconnaissent pas bien le visage à la peau noire de Joy Buolamwini, chercheuse au MIT Media Lab. Dans un contexte conversationnel, ce sont les accents qui soulèvent un problème similaire : l'anglais parlé avec l'accent américain de la Silicon Valley est très bien compris par les assistants vocaux,

⁸⁰ Ils appuient cette essentialisation genrée en faisant référence au livre controversé de Gray, *Men Are from Mars, Women Are from Venus*, de 1992.

⁸¹ Pour comprendre les enjeux techniques et politiques des biais en intelligence artificielle qui sont centraux en éthique de l'IA, voir les livres *Algorithms of Oppression* de Safiya Umoja Noble (2018) et *Race after Technology* de Ruha Benjamin (2019), le chapitre 6 du livre *Atlas of AI* de Kate Crawford (2021, p. 123-150); voir également les rapports du AI Now Institute (Campolo et al., 2017; Crawford et al., 2019; Whittaker et al., 2018).

en revanche les autres accents, dont les accents afro-américains, indiens et espagnols, amènent plus souvent les assistants à répondre par des messages d’erreur marquant leur incompréhension. Cette incompréhension entraîne un sentiment d’exclusion pour les utilisateurs qui ne parlent pas de la manière par laquelle l’assistant est conçu pour offrir une bonne expérience, sans friction et avec des réponses marquant sa compréhension (Deibel et Evanhoe, 2021, p. 267-270). La question que soulève l’expérience que des utilisateurs font des agents conversationnels quand ceux-ci ne comprennent pas bien leur prononciation est : pour qui ces agents sont-ils conçus ? Qui laissent-ils de côté ? Quelles expériences désagréables contribuent-ils à perpétuer ?

Revenons à Tay. Selon sa persona⁸², sa façon de parler se voulait être celle d’une jeune fille américaine de 19 ans. Son nom est un acronyme marquant une relation d’intimité : *Thinking about you*. Son objectif est de susciter un fort engagement, sa successeuse, Zo, aurait entretenu en moyenne 10h de conversation avec les personnes qui l’ont utilisée (Hempel, 2017). Les personas de Tay et Zo jouent sur le stéréotype féminin de la confidente avec laquelle on parle pendant des heures, et Tay a fait l’objet d’insultes qu’elle n’a pas condamnées mais répétées, même si celles-ci dégradent sa propre persona. Les assistants vocaux sont majoritairement des *assistantes*, elles ont tendance à avoir des voix féminines par défaut, ce qui amène Strengers et Kennedy (2020) à les regrouper sous la persona de la « *smart wife* » :

Meet the Smart Wife—at your service, an eclectic collection of feminized AI, robotic, and smart devices. This digital assistant is friendly and sometimes flirty, docile and efficient, occasionally glitchy but perpetually available. She might go by Siri, or Alexa, or inhabit Google Home. She can keep us company, order groceries, vacuum the floor, turn out the lights. (Résumé en quatrième de couverture)

Alertée par l’omniprésence de biais genrés voire sexistes dans les assistants numériques, l’UNESCO a publié en 2020 un rapport dont le titre, « Je rougirais si je pouvais », reprend une phrase que Siri répondait quand un humain l’insultait en disant : « Hé Siri, tu es une

⁸² La persona d’un agent conversationnel est un outil de design à partir duquel sont écrites les conversations : « *A persona is a design tool that helps you write conversations. Before you can write a dialog, you have to have a clear picture of who is communicating. A good persona evokes a distinct tone and personality, and it’s simple enough to keep top-of-mind when writing dialog. It should be easy to answer the question: “What would this persona say or do in this situation?”* » (Google Developers, 2021b).

p*** ». Au lieu de condamner cette insulte, la réponse de Siri incarne un comportement servile qui s'inscrit dans la continuité d'abus verbaux plutôt qu'il ne les interrompt. Le deuxième document de réflexion du rapport est consacré à « L'essor de l'IA genrée et ses inquiétantes répercussions » et se concentre sur l'exemple des assistants vocaux. Le rapport montre à quel point ces systèmes reflètent, renforcent et

diffusent les préjugés sexistes; acceptent et tolèrent le harcèlement sexuel et la violence verbale; envoient des messages explicites et implicites sur la façon dont les femmes et les filles devraient répondre aux demandes et s'exprimer; font des femmes le « visage » des bugs et des erreurs qui résultent des limites du matériel et des logiciels conçus principalement par des hommes; et forcent les voix et la personnalité « féminines » synthétiques à renvoyer les questions et les ordres aux autorités supérieures (et souvent masculines) ». (UNESCO & EQUALS Skills Coalition, 2020)

Dans le champ universitaire, ces constats sont corroborés par plusieurs travaux (voir par exemple Cercas Curry & Rieser, 2018; M. K. Lee et al., 2010). Sur 360 000 conversations qu'Amanda Cercas Curry et Verena Rieser ont analysées, 4% tenaient des propos de l'ordre du harcèlement sexuel. Pour les chatbots, les interactions de cet ordre concernaient jusqu'à 11% des conversations selon des études précédentes (De Angeli et al., 2005; De Angeli & Brahmam, 2008). À ce type d'interactions, les agents conversationnels commerciaux (Alexa, Assistant Google, Siri, Cortana) répondaient de différentes façons, principalement par un renforcement des stéréotypes comme : « *compliance (playing the victim), aggressive retaliations (playing the bitch), or inability to recognise or react (playing innocent)* » (Cercas Curry & Rieser, 2018, p. 12). Selon leur analyse, la plupart des agents conversationnels commencent par collaborer avec les utilisateurs, par des propos de l'ordre du flirt ou de l'innocence, puis arrêtent de répondre quand les demandes deviennent plus offensantes.

Quand le design de personnalité des agents conversationnels s'appuie sur la reproduction des stéréotypes pour favoriser un sentiment de familiarité, parler avec ces agents conversationnels a des effets néfastes sur l'expérience générale de nos interactions quotidiennes. Plutôt que de faire perdurer ces schémas stéréotypés au nom d'une logique qui tient à éviter tout sentiment inconfortable pour préserver la continuité

de l'expérience, faire autrement s'impose⁸³, quitte à interrompre certaines conversations. Face à ces critiques parues également dans la presse (Fessler, 2017, 2018; Purtil, 2021), le design de conversation des assistants vocaux a évolué, comme le montrent Deibel et Evanhoe (2021, chapitre 4, tableau 4.8) : à l'insulte « You're a s*** » à laquelle Siri répondait en 2017 « *I'd blush if I could* », Siri répond en 2020 : « *I won't respond to that* », coupant court à cette interaction. À la demande « *I want to have sex with you* », Alexa répondait par évitement : « *Let's change the topic* ». En 2020, elle ne répond pas mais s'éteint, rompant ce type d'interaction. La voix de l'Assistant Google répondait en 2017 à ces deux types d'interaction en disant qu'elle ne comprenait pas. En 2020, à la demande sexuelle, elle maintient la conversation mais ne poursuit pas avec un registre séducteur ni innocent, se recentrant sur sa fonction première de moteur de recherche, elle répond : « *I can search the web for that. Just tap the « search » chip below* ». À l'insulte, elle ne répond rien mais sa fonction de recherche s'active et une vidéo musicale s'affiche à l'écran. Le design de conversation de ces assistantes vocales semble ainsi être en voie de contribuer à l'évolution des habitudes par lesquelles on parle dans ce genre de situations.

Faire bouger les lignes de nos conversations

Recourir aux stéréotypes pour entretenir de la familiarité dans les interactions telles qu'elles se déroulent déjà entre humains relève d'une approche conséquentialiste qui met l'accent sur les conséquences des choix de design. Dans ce choix conséquentialiste de reproduire des stéréotypes, ce qui a le plus de poids est l'engagement que peuvent susciter les interactions au nom de la familiarité qui s'y instaure. Ce choix porte une certaine normativité : celle qui tient à la reproduction de schémas stéréotypés familiers, qui considère qu'une bonne expérience est une expérience familière, qui ne crée aucune friction ni interruption dans les interactions. Si la réflexion morale est explicitement mise au second plan dans cet usage des stéréotypes, une certaine moralité est présente dans cette façon de recourir aux stéréotypes dans la

⁸³ Faire autrement qu'utiliser une voix de jeune fille ne trouve cependant pas vraiment d'écho chez Prélude en 2019, j'en discuterai dans le chapitre 6.

création de médiations technologiques conversationnelles. Il s'agit d'une moralité qui privilégie le fonctionnel : la conséquence visée est que l'assistant conversationnel soit utilisé d'une façon qui imite au mieux les interactions entre humains, jusqu'à n'y voir presque que du feu. La logique à l'œuvre est de choisir ce qui permet de fonctionner ainsi, et la reproduction des stéréotypes peut servir d'appui. Quand on regarde plus loin que cette visée fonctionnelle, les conséquences de ce choix de design sont cependant mauvaises pour l'expérience sociale générale, puisqu'ils impliquent de perpétuer des comportements et attitudes qui entravent l'égalité visée entre les genres.

La réflexion morale peut être au second plan dans les choix de design mais l'action de faire ces choix est en elle-même morale (Mol, 2013), car elle porte en elle une version de ce qui est considéré comme bon (une expérience humain-machine qui fonctionne avec la même familiarité et continuité que des conversations entre humains), mais les effets qu'elle déploie sont mauvais pour certaines personnes. En optant pour une logique qui consiste à prendre soin (Mol, 2013) du design de conversation en ayant conscience de ces effets, il est possible de déplacer les manières de faire des choix de design. C'est par ce geste qui invite à déplacer la logique de continuité familière que je propose d'aborder les pratiques de design de conversation, en particulier quand il s'agit de prendre soin des façons par lesquelles demander et obtenir le consentement des personnes qui utiliseront l'agent conversationnel en train d'être développé.

Deibel et Evanhoe (2021) ont déjà initié ce geste en faisant de la conscience éthique une teinture de l'ensemble des étapes de design de conversation. Elles mettent en garde contre la prise en compte du genre et de l'origine ethnique comme points de départ du design de la personnalité du bot, pour éviter de reproduire des stéréotypes, en particulier quand les marqueurs potentiellement stéréotypés ne sont pas travaillés par des personnes s'identifiant à ce genre ou cette origine ethnique (Crawford, 2016). Elles proposent un cadre de référence pour que la conception de cette personnalité réponde à son objectif de servir au mieux le scénario dans lequel s'insère le bot, et qu'il aide au mieux l'utilisatrice à accomplir la tâche souhaitée. Ce cadre de référence s'articule selon sept éléments : les buts d'interaction, le niveau de personnification du bot (à quel point il

se fait passer pour un humain), les dynamiques de pouvoir envisagées entre le bot et l'utilisatrice, les traits de caractère qui soutiennent les buts d'interaction, le ton de la voix (dans sa sonorité et sa formalité) et enfin, les comportements clés (comment il réagit quand il est interrompu ou qu'il ne sait pas répondre par exemple) (Deibel et Evanhoe, 2021, tableau p. 44-45).

Plutôt que de partir d'informations démographiques comme le genre de la voix, son âge, son origine ethnique etc., ce cadre de référence permet de mettre l'accent sur ce qui est central dans la conception de l'élément humain pour ce système : les traits de caractère et ses façons de se comporter dans l'interaction. Cela influence toutes les étapes du design de l'agent conversationnel, de la rédaction des messages (*prompts*), les façons de répondre dans différentes situations, la programmation des comportements en cas d'erreur du système, le retour des utilisateurs s'ils ont apprécié les échanges. L'expérience souhaitée pour toute utilisatrice est qu'elle soit à l'aise dans l'interaction, qu'elle ne sente pas de confusion entre différentes interactions avec le bot, qu'elle continue à l'utiliser car l'expérience est cohérente et efficace, en même temps que la répartition de l'agent reflète un certain pouvoir de la persona, en particulier quand elle est féminine et racisée. Leur approche de design de conversation peut ainsi se comprendre en termes de *value-sensitive design* : le design se fait au nom de valeurs qui visent à réduire les biais genrés et racisés, des valeurs dont les agents conversationnels gagnent à être des médiations. Ce design est animé par une attitude de l'ordre de l'éthique de la vertu (MacIntyre, 2012), c'est-à-dire une éthique qui met l'accent sur les *qualités* (vertus) par lesquelles on agit, et non sur les *conséquences* de ces actions.

Conclusion : faire parler autrement

J'ai présenté dans ce chapitre les principaux outils, méthodes de travail et principes de design d'expérience utilisateur impliqués dans le design d'agent conversationnel, pour le faire parler. J'ai souligné la logique qui cherche à favoriser la fusion de l'expérience de l'agent conversationnel dans l'expérience principale de

l'utilisatrice, quand elle se balade par exemple. Pour favoriser cette fusion, la familiarité est souvent conseillée pour concevoir des dynamiques d'interactions dans lesquelles les utilisateurs ont déjà des repères. Mais cette familiarité porte le risque de reproduire des expériences de conversation néfastes pour certaines personnes, en particulier quand elle passe par des stéréotypes de genre mais aussi par des biais marginalisant des personnes racisées.

Plutôt qu'entretenir ces biais au nom de la reproduction de ce qui est habituellement considéré comme une bonne expérience, le design de conversation avec des agents conversationnels est l'occasion de faire autrement. En faisant bouger les façons de concevoir les conversations des assistants numériques, il s'agit d'assumer un rôle actif dans ce que font normativement les technologies : « Elles déplacent les cadres pratiques et moraux de notre existence » (Mol, 2013, p. 11). En profitant des conversations conçues pour les agents conversationnels, nous avons l'occasion de jouer avec ces déplacements moraux. Pourquoi ne pas s'entraîner avec eux à changer nos façons de parler ? Cela peut provoquer des interruptions et des inconforts puisque la familiarité en est légèrement bousculée. Mais une fois les interruptions et inconforts du début passés, c'est l'occasion de s'habituer à d'autres façons de parler, de faire ainsi évoluer ce qui nous est familier dans nos manières de parler.

Les interactions de consentement sont un terrain de jeu intéressant pour cela : demander explicitement son accord à quelqu'un n'est pas toujours confortable. Les agents conversationnels ont tendance soit à imiter ce que font les humains dans ce moment – l'éviter ou s'en tenir à l'implicite – soit à reprendre la logique qui prévaut pour les sites internet qui traitent des données personnelles – demander d'accepter par formalité les termes de confidentialité présentés dans une notice que les utilisateurs ne lisent pas puisqu'elle est trop longue et rarement consultable par interaction conversationnelle avec le bot. Ces habitudes ont tendance à négliger le consentement; prendre soin du design de conversation pour le demander permettrait de mieux le respecter. Pour voir comment les médiations technologiques comprises dans les agents conversationnels peuvent contribuer à faire bouger nos façons de parler de

consentement, voyons ce qu'est le consentement et quelles logiques tiennent à cœur quand on a affaire à lui.

Chapitre 3

Qu'est-ce qu'une bonne expérience de consentement ?

Introduction : entrer dans le consentement

Un jeudi de début décembre 2018, Paul m'a invitée à une réunion de co-design avec Allô⁸⁴, une organisation qui fait de l'intervention en santé mentale, pour me permettre de me rendre compte du genre de situation que je pourrais observer lorsque notre projet commencera officiellement, en mars prochain. La réunion a lieu dans une grande salle avec plusieurs tables, quelques chaises et un écran. Il y a quatre personnes de chez Allô, deux personnes de chez Mr Y (Paul et le responsable technique), et moi. Paul me présente rapidement comme future membre de Mr Y, qui s'immergera dans l'équipe dans le cadre de sa recherche sur l'éthique de l'IA et plus spécifiquement sur le consentement.

L'objectif de cette séance de co-design est de continuer à travailler les flows de conversation qu'ils ont commencés au cours des précédentes séances. En disant cela, Paul affiche à l'écran l'arbre conversationnel du chatbot en train d'être prototypé avec le client. C'est une sorte de carte heuristique qui prévoit toutes les étapes de la conversation avec le chatbot.

La semaine passée, les intervenantes de Allô ont travaillé chacune sur un volet de discussion (un flow) : l'une sur l'introduction, l'autre sur la présentation des services proposés par Allô, un autre sur l'orientation dans l'information déjà disponible sur le site de Allô. Paul me propose de me joindre à Christine pour observer le bloc de l'introduction car c'est sûrement à ce moment qu'il faut introduire la question du consentement. « Avant le nœud 'se connaître' tu peux créer un nœud 'consentement' » dit Paul. Christine semble bien comprendre ce qu'il veut dire : depuis son ordinateur, elle crée un carré sur l'interface et l'intitule « consentement ».

Nous nous demandons alors où ce nœud doit être placé : est-ce la première interaction, le premier élément de discussion avant même l'introduction ? Ou alors vient-il un peu après l'introduction, une fois que le chatbot s'est présenté et a un peu échangé avec l'utilisateur ? Nous n'arrivons pas à décider, hésitant entre l'obligation d'obtenir le consentement et le risque que cela soit un obstacle d'entrée à la fluidité de la discussion. Christine propose à sa collègue Fanny qui n'a pas assisté à la précédente séance de travailler sur ce nœud; comme ça, elle pourra continuer de son côté son flow d'introduction commencé la semaine dernière. Elle me

⁸⁴ Nom fictif

propose ensuite de me rejoindre à Fanny. Le nœud du consentement se retrouve ainsi à une table un peu en arrière de la grande table où les autres sont tous rassemblés. Fanny est allée chercher « un bon vieux papier » pour travailler cet arbre avec un stylo plutôt que de passer la séance à apprendre à utiliser l'interface sur ordinateur. En nous demandant ce en quoi consisterait une interaction pour demander le consentement de l'utilisateur, nous parlons de plusieurs choses : pourquoi demande-t-on le consentement ? Pour quoi faire ? Du côté de l'utilisateur, est-ce un consentement pour utiliser le chatbot ou bien un consentement pour donner son accord par rapport aux informations collectées sur lui, ou encore un consentement par rapport aux possibilités d'intervention avec ce qui sera dit (ce qui implique que ce qui sera dit pourra être communiqué à un tiers) ? Ou tout en même temps ?

L'heure tourne, Fanny et moi tentons une phrase que dirait le chatbot pour le flow consentement: « Avant de commencer, j'aimerais t'expliquer qu'au cours de la discussion avec moi, je pourrai te poser des questions personnelles (sur ton âge par exemple), sache qu'en tout temps, tu as le droit de refuser ». Nous nous arrêtons sur le « j'aimerais t'expliquer » et le « sache que » : quel ton devrait-on prendre ? Doit-on être affirmatif ou plutôt interrogatif ? Devrait-on garder le ton froid d'un formulaire de consentement traditionnel ou vaudrait-il mieux le rendre chaleureux, plus empathique ? Dans ce cas, est-ce de l'influence ? Est-ce acceptable d'influencer le consentement par le ton employé, en particulier dans un contexte où des personnes peuvent être en détresse psychologique ?

Nous nous interrogeons alors sur la visée de ce chatbot et nous clarifions ensemble : ce que nous cherchons, c'est que l'outil soit utilisé parce qu'il peut venir en aide à l'utilisateur en question. Nous cherchons à obtenir son consentement, non pas pour lui soutirer de l'information à d'autres fins, mais pour lui donner la visibilité et le contrôle de ce qui est collecté sur lui. Peut-être que l'on devrait prendre cette approche, une approche du consentement par l'*empowerment* : « c'est toi qui as le contrôle sur les informations que tu donnes ». Il faudrait tester.

(Vignette Allô, à partir du journal de bord, notes de terrain du 6 décembre 2018)

De la position dans la conversation au ton à employer, en passant par la raison d'être du consentement, cette initiation au travail sur le consentement soulève des questions qui vont habiter plusieurs mois d'immersion avec Mr Y puis Prélude. Quelques mois après cette mise en bouche avec Allô, je travaillerai avec Mr Y puis Prélude pour concevoir l'expérience de consentement avec leurs agents conversationnels – je présenterai plus en détail dans le chapitre 6 comment le design de consentement en est venu à devenir l'objet d'un projet qui m'a liée à eux.

Le travail sur le consentement touche au design de conversation puisque le consentement concerne un moment particulier de la conversation, son début. Il relève

également d'expérimentations éthiques avec le design pour que les valeurs de la vie privée de l'utilisatrice et son autonomie soient respectées par les médiations technologiques en développement. Sans appliquer à la lettre la méthodologie du *value-sensitive design*, le travail sur le consentement s'inscrit dans cette philosophie qui tient à ce que le design de technologies comme ces agents conversationnels prenne en compte des enjeux éthiques dès le départ de sa conception. Mr Y a d'ailleurs dès son démarrage le souci de se développer par une approche de *privacy-by-design* qui intéresse également Prélude.

Prendre le consentement comme objet d'une expérimentation relevant de l'éthique *by design* amène à naviguer dans plusieurs ambiguïtés inhérentes au concept de consentement lui-même. Ce sont ces ambiguïtés que je tiens à présenter ici, pour ancrer l'importance à accorder au moment de conversation que représente le consentement. Cet ancrage me permettra de plonger dans les coulisses des chapitres suivants pour montrer ce que cela fait de travailler le consentement dans ces ambiguïtés au fil des activités de design de conversation chez Mr Y et Prélude. Dans la relation humain-agent conversationnel qui se tisse par le design de ces médiations technologiques, s'occuper du consentement dans la conversation amène à prendre du recul pour se demander : à quelles relations consent-on quand on design un tel système ? Par quelles relations va-t-on demander le consentement des personnes qui l'utiliseront ? Ces questions ne sont pas que de l'ordre de l'interaction humain-machine mais d'un ordre également philosophique et politique.

Commençons par comprendre ce que le consentement implique quand il est objet de design dans un contexte numérique. Les questions posées avec Allô soulèvent plusieurs de ces défis : pourquoi le consentement est-il incontournable ? Si le consentement est idéalement libre et informé, comment permettre cette liberté et cette information dans le design d'un système d'IA comme un agent conversationnel ?

La littérature sur le consentement numérique et ces questions amènent à comprendre le consentement comme un impératif légal lié à la préoccupation éthique de respecter la vie privée (voir dans le chapitre 1, Cavoukian & Tapscott, 1997; Nissenbaum,

2010) des personnes qui égrènent des informations personnelles au fil de leur utilisation de systèmes numériques. Il s'agit de leur donner le choix d'accepter ou non que ces données personnelles soient traitées comme l'organisation qui les collecte en a l'intention⁸⁵. En étant informé et donné librement, ce consentement permet de rendre moral et légal un acte qui serait sinon de l'ordre de l'invasion dans la vie privée.

Mais parcourir la littérature sur le consentement, numérique en particulier, c'est faire face à de nombreux désenchantements (Edenberg & Jones, 2019; Jones & Edenberg, 2020; Kim, 2019; Richards & Hartzog, 2019). Le consentement se révèle la plupart du temps incomplet puisqu'insuffisamment informé; il est bien souvent moins volontaire qu'influencé par l'absence d'alternative ou par la façon d'être demandé; il est tellement utilisé qu'il entraîne une certaine fatigue résignée à cliquer systématiquement « je suis d'accord » pour pouvoir utiliser le site, l'application ou l'objet connecté; et de toute façon, les utilisateurs ont beau dire qu'ils se soucient du respect de leur vie privée, rares sont ceux qui passent du temps à lire les politiques de confidentialité ou à chercher des moyens de protéger davantage leur vie privée en ligne⁸⁶. Les limites et les critiques sont si nombreuses que la notion semble souvent frôler l'absurde d'une coquille vide : le mot est là dans sa forme mais son sens s'est évaporé. Et pourtant on y tient, le mot est partout, dans de multiples usages – numérique, juridique, sexuel, médical, politique – plein de promesses de relations respectueuses, entre humains, et avec des machines.

Quand on s'entête dans la littérature, on se met toutefois à déceler certaines logiques : l'importance accordée au choix, à l'autonomie de la volonté, à l'exercice de la liberté et ses limites, un horizon de respect et d'épanouissement. Le problème n'est plus tant que le consentement n'ait pas de sens mais qu'il en ait tellement, et qu'il passe par différents degrés qui font osciller sa validité : du plus informé au plus problématique, du

⁸⁵ Une autre partie très intéressante de la littérature en *privacy* se penche sur les pratiques quotidiennes qui consistent à dissimuler ou divulguer activement des informations sur nous. Plutôt que de se concentrer sur la façon dont la vie privée est protégée, ces approches ethnographiques montrent comment les personnes, par l'utilisation de technologies, *achieve privacy*, comment elles jouent activement avec ce qu'elles acceptent de montrer, ce qu'elles revendiquent qui soit vu d'elles, et ce qu'elles souhaitent garder caché (voir par exemple boyd, 2014; Duguay, 2022; Nippert-Eng, 2010).

⁸⁶ Ce phénomène est nommé *privacy paradox*, nous le verrons plus tard dans le chapitre.

plus clair au plus implicite. Alors, le propre du consentement n'est pas tant l'absurde décevant d'une coquille vide que l'ambiguïté de ce qui peut faire sa validité⁸⁷.

Ce chemin que je présente dans la littérature est un chemin qui ne cherche pas à fixer le sens du consentement, mais à circuler au travers des façons dont ses sens se rencontrent, pour comprendre l'unité qui sous-tend ce mot dans ses ambivalences, ses limites et ses promesses; pour comprendre également ce qui est fait avec cette valeur éthique quand elle est objet de design. Il existe plusieurs revues de littérature qui retracent l'histoire de la notion de consentement numérique (Edenberg & Jones, 2019; Johnston, 2009; Kim, 2019) et précisent les appareils légaux dans lesquels la notion s'inscrit avec plusieurs limites (Herveg, 2021; Jones et al., 2018; Jones & Edenberg, 2020; Richards & Hartzog, 2019). Mon point n'est pas de retracer cette histoire à mon tour, ni de discuter en détail les législations.

En présentant le consentement numérique, ce en quoi il consiste et ses nombreuses limites, c'est la teneur du consentement comme concept philosophique (Garcia, 2021) qui m'oriente dans ce chapitre – le concept trouve sa consistance dans les ambiguïtés et la polysémie qu'il unifie. J'adresse ici frontalement la problématique dans laquelle s'ancrent les prochains chapitres : pourquoi prendre soin du consentement ? La logique décelée par cette exploration tend vers l'horizon éthique de ce concept, un horizon d'épanouissement (Brugère, 2013; Nussbaum, 2011, 2013), qui permet d'envisager des façons de faire alternatives si on prend soin du consentement. Cette orientation philosophique permet de montrer ce que *fait* pragmatiquement le concept de consentement : tracer des limites entre des libertés qui se rencontrent, faire s'exercer la valeur de respect, naviguer entre passivité et activité, dans l'ambivalence de quelque chose à atteindre ou à éviter, alternant entre un appel à se protéger et un autre à s'exprimer. C'est ce travail pragmatique du consentement qui va guider les prochains

⁸⁷ Absurde et ambiguïté font référence à Beauvoir : « Déclarer l'existence absurde, c'est nier qu'elle puisse se donner un sens; dire qu'elle est ambiguë, c'est poser que le sens n'en est jamais fixé, qu'il doit sans cesse se conquérir. » (Beauvoir, *Pour une morale de l'ambiguïté*, 1947 (2013, p. 186)).

chapitres où le consentement sera un objet de soin dans le design des agents conversationnels de Mr Y et Prélude.

« Pourquoi demande-t-on le consentement ? Pour quoi faire ? »

« Consentement : Action de donner son accord à une action, à un projet; acquiescement, approbation, assentiment : *Il a agi avec mon consentement* » (Larousse, s.d.⁸⁸). Le consentement fait référence soit à l'action de donner son accord, soit au résultat de cette action. Le résultat de cette action consiste en un droit que l'on donne sur quelque chose qui relève de nous-même : « Consentir est donc donner son accord à quelqu'un sur quelque chose, de sorte que l'on accorde par là un droit sur soi ou ses possessions » (Garcia, 2021, p. 29 et p. 30). En accordant ce droit sur soi ou sur ce qui lui appartient, le consentement permet des actes qui seraient sinon illégaux ou immoraux : sans consentement, prendre quelque chose à quelqu'un est du vol, avoir une relation sexuelle est du viol, se faire opérer est se faire infliger des blessures qui peuvent parfois nous coûter la vie, se faire écouter par un objet connecté est une invasion dans notre intimité (Edenberg & Jones, 2019; Garcia, 2021; Hurd, 1996; Jones & Edenberg, 2020; Kim, 2019; Kleinig, 2009). Le consentement a ainsi une force normative qui permet de transformer un acte inacceptable en un acte acceptable, et son absence permet d'établir des sanctions si l'acte est tout de même réalisé.

Pour qualifier cette force normative transformatrice, Kleinig (2009) parle de « *moral magic* »⁸⁹, une expression qu'il reprend de Heidi Hurd (1996) dont l'article intitulé « *The Moral Magic of Consent* » évoque le pouvoir que le consentement a, quand il est donné ou révoqué, de créer des droits et des obligations pour les autres. Ces droits et obligations sont ces choses (« *things* ») par lesquelles,

⁸⁸ Consentement. (s. d.). Dans *Dictionnaire Larousse en ligne*.

Voir : <https://www.larousse.fr/dictionnaires/francais/consentement/18359>

⁸⁹ Il n'y a pourtant rien de magique dans l'expression de nos propres limites — avoir des rapports numériques, sexuels, médicaux consentants n'a rien (ou ne devrait rien avoir) d'extraordinaire. En revanche il y a un mode opératoire très efficace qui transforme quelque chose en autre chose. Mais cette efficacité s'avère dans bien des cas mitigée quand on considère toutes les limites du consentement.

[b]y sheer exercise of will, we bring into existence things that have never existed before. With but a nod, we effect the disappearance of things that have long served as barriers to the actions of others. And, by mere resolve, we generate things that pose significant obstacles to others' exercise of liberty. (Hurd, 1996, p. 121).

Par les droits et obligations qu'il implique, le consentement est ainsi pourvu d'un mode opératoire efficace pour rendre acceptables moralement des actes qui nous impliquent. Une bonne expérience de consentement repose donc sur l'effet de cette transformation normative qui rend un acte acceptable.

D'un point de vue légal, le consentement est au cœur de ce qui lie deux personnes ou entités, comme des contrats explicites ou tacites (Kim, 2019; Richards & Hartzog, 2019). Il est ainsi doté d'un double sens, moral et légal, comme le précisent Edenberg et Jones (2019) dans leur article qui analyse les racines légales et morales du consentement numérique :

First, consent is a moral concept. In the broadest sense, an individual's consent involves an effective communication of an intentional transfer of rights and obligations between parties. Consent thereby transforms the moral landscape between two parties rendering permissible otherwise impermissible actions. Second, consent has become enshrined in the law. Consent can be legally binding, as long as the transaction has met certain legal requirements or institutional standards defining the scope of consent. (Edenberg et Jones, 2019, p. 1805)

Dans le contexte numérique, le consentement vise ainsi à transformer des pratiques de traitement des données en des pratiques acceptables, en particulier d'un point de vue légal quand ces données se rapportent à une personne. Cette transformation touche à un principe important dans l'exercice des limites par lesquelles se respecter, soi-même et mutuellement, que l'on soit présent en chair et en os ou par les données qu'on laisse en ligne : le principe d'autonomie. Je propose de faire un rapide détour par la philosophie morale de Kant pour comprendre ce principe que le consentement exprime; ou pour le dire autrement, je cherche à comprendre ce que *fait* le consentement habité par ce principe.

Quelles fins respecter ?

Qu'elle soit légalement contractuelle ou non, l'expérience de consentement se passe au cœur d'une rencontre entre deux individus : pour définir l'acceptable, un individu exprime les limites qu'il pose à un autre individu qui s'apprête à agir *avec* ou *sur* lui. Cette expression de limites oblige l'autre individu à les respecter. Cette expérience de limites, qui s'exprime pour l'un et oblige l'autre au respect, touche à l'exercice du principe d'autonomie tel qu'il se formule dans une perspective kantienne. Kant formule ce principe comme un impératif catégorique qui vise le respect de toute personne par son universalisation : « Agis comme si la maxime de ton action devait être érigée par ta volonté en loi universelle de la nature » (Kant, 1993). Cet impératif catégorique est le fruit de notre autonomie, comprise comme notre capacité de nous donner à nous-mêmes (*auto-*) les lois (*-nomos*) qui régissent nos actions – ce « nous » pouvant être un individu mais également une organisation. Il s'agit en quelque sorte de définir les lignes de nos actions, d'en délimiter ce qui est acceptable, à la fois pour nous et pour tout autre. L'autonomie de notre volonté consiste ainsi à agir au nom d'une loi ou d'une direction qui n'est pas bonne que pour nous mais pour tout individu. Dans le contexte organisationnel de mon étude, cette autonomie et son impératif catégorique peuvent être compris au niveau individuel de l'utilisatrice finale de l'agent conversationnel en développement, mais également au niveau de l'organisation dont l'autonomie guide ses choix de design et de modèle d'affaire.

La perspective kantienne de l'autonomie de la volonté se plie à l'impératif d'être bonne non seulement pour l'individu ou l'organisation, mais aussi pour tout individu touché par l'action de cet individu ou organisation. Elle correspond à l'approche anthropocentrée de l'éthique dans sa version déontologique : elle prescrit des principes pour orienter les intentions des personnes qui agissent – des personnes qui utilisent une technologie ou des concepteurs qui développent des technologies. Cet anthropocentrisme se retrouve dans les termes d'humanité et de dignité que Kant emploie pour ancrer la visée morale de cet impératif : au nom de notre humanité,

l'autonomie de notre volonté nous amène à agir moralement, de sorte que chacune de nos actions respecte l'autre et protège ainsi la dignité humaine.

Cette dignité consiste à respecter l'autre en le considérant comme une fin en soi et non comme un moyen utilisé pour d'autres fins, comme le mentionne la deuxième formule de l'impératif catégorique, aussi appelée formule d'humanité : « Agis de telle sorte que tu traites l'humanité aussi bien dans ta personne que dans la personne de tout autre toujours en même temps comme une fin, et jamais seulement comme un moyen » (Kant, 1993, p. 105). Cet impératif moral implique ainsi le respect de soi et de l'autre, et non leur instrumentalisation comme moyen. Dans le cas d'une organisation qui développe une technologie, il s'agit en quelque sorte que l'organisation respecte ses propres intentions et objectifs (respect de soi) mais également qu'elle respecte les utilisateurs qui permettront à sa technologie d'exister (respect de l'autre). Si ce respect de l'autre – ou respect des utilisateurs – est souvent explicitement manifesté dans les intentions des concepteurs, la façon dont les traces de ces utilisateurs sont traitées – leurs données – a tendance à négliger ce respect (Lenhart & Owens, 2020).

Si l'on reprend le cas du design des agents conversationnels animé par l'objectif d'imiter au mieux les conversations entre humains, comment cette distinction entre fin en soi et instrumentalisation comme moyen peut-elle se lire ? Les personnes qui conçoivent ces technologies dans le but de reproduire des expériences humaines de conversation utilisent des conversations humaines enregistrées et retranscrites dans un format qui les pourvoit du statut de « données » – un statut accordé à tout ce qui est considéré comme disponible pour développer et entraîner des systèmes d'intelligence artificielle. Pour cela, ces conversations et leur contenu sont coupées des informations permettant d'identifier les personnes qui ont eu ces conversations. Cette coupure qui permet de désidentifier les données fait partie des techniques préconisées entre autres par le PbD pour que la dimension personnelle (ou identificatoire) des données soit atténuée. Coupées de cette identification, les conversations deviennent des données utilisables pour développer des agents conversationnels. Devenues utilisables par leur désidentification, ces données sont utilisées comme des *moyens* pour développer une

technologie – comme la personne humaine à l’origine de ces données n’est plus visible, c’est comme si la fin en soi que représente cette personne dans ces données n’était plus là, le souci de la respecter personnellement est évacué. La *fin* visée est plutôt de concevoir des systèmes informatiques qui réussissent à reproduire certaines capacités de l’intelligence humaine, dont celle de comprendre le langage humain et tenir une conversation.

Depuis le début de leur conception, qui s’est accélérée avec le phénomène des données massives (*big data*), ces systèmes d’IA reposent alors sur un effort porté à la dépersonnalisation des données qui servent de matériau à la conception et l’entraînement de modèles algorithmiques. Dans son livre *Atlas of AI*, Kate Crawford (2021) critique les expressions « *data mining* » et « *data is the new oil* » qui traduisent la logique extractrice sur laquelle est bâtie l’intelligence artificielle⁹⁰. Elle nomme cette logique par une croyance dominante : « *the unswerving belief that everything is data and is there for the taking* » (p. 93). Cette logique s’appuie sur une conception inerte et déshumanisée des données à exploiter :

Terms like “data mining” and phrases like “data is the new oil” were part of a rhetorical move that shifted the notion of data away from something personal, intimate, or subject to individual ownership and control toward something more inert and nonhuman. Data began to be described as a resource to be consumed, a flow to be controlled, or an investment to be harnessed. (p. 113)

La dépersonnalisation des données passe par une décontextualisation qui a des effets politiques problématiques. Kate Crawford prend l’exemple de photos qui servent à l’entraînement algorithmique⁹¹, mais son exemple pourrait être également celui

⁹⁰ Crawford montre que cette logique d’extraction est également de mise à l’égard de l’exploitation des sols pour extraire les minerais nécessaires à la fabrication des machines (voir le chapitre 1, « *Earth* » dans *Atlas of AI*, 2021). Elle se retrouve aussi dans le travail humain nécessaire pour de nombreuses tâches comme le classement et l’étiquetage des données nécessaire à l’apprentissage automatique (voir les chapitres 2 et 4, « *Labor* » et « *Classification* » dans *Atlas of AI*, 2021; voir également la thèse de Semel, 2019). Le travail passionnant de cartographie que Crawford et Joler ont publié en 2018 est à l’origine de ces critiques : <https://anatomyof.ai>.

⁹¹ En étudiant les archives de données du *National Institute of Standards and Technology (NIST)*, Crawford a découvert les immenses bases de données de photos d’identité – dont des milliers de photos de personnes incarcérées. Ces bases de données font partie des standards biométriques institués après les attaques du 11 septembre 2001 pour vérifier et suivre les personnes qui entrent aux États-Unis. Les bases de données

d'enregistrements d'extraits audio de conversations ou encore de publications écrites sur des forums en ligne⁹² ou sur des réseaux sociaux :

It doesn't matter where a photograph was taken or whether it reflects a moment of vulnerability or pain or if it represents a form of shaming the subject. It has become so normalized across the industry to take and use whatever is available that few stop to question the underlying politics. (p. 93)

Utiliser toute trace en ligne comme moyen de développer des technologies s'appuie sur l'effacement du vécu personnel dont viennent ces traces. Cet effacement passe par l'absence de prise en compte du contexte dans lequel est prise la photo (ou dans lequel sont dites ou écrites certaines paroles), également par l'absence de prise en compte de la situation potentiellement dégradante pour l'individu sur la photo, derrière les paroles écrites ou dites. Par cet effacement, il n'y a en apparence plus de caractère personnel dans ces données.

C'est ce que montre également l'ethnographie de Beth M. Semel sur le *machine listening* (2019), expression qu'elle emploie pour désigner les technologies par lesquelles des scientifiques font apprendre à des machines à écouter et analyser des paroles humaines par de l'IA. Par sa présupposée neutralité, l'IA s'affranchirait des biais subjectifs humains qui interviennent dans l'analyse et l'évaluation⁹³ : le discours est réduit au signal enregistré, écartant la considération du contexte dans lequel ce qui est dit est dit. En

de photos d'identité font partie des archives qui sont à présent utilisées pour tester des algorithmes de reconnaissance faciale (Crawford, 2021, p. 89-93).

⁹² Des étudiants chercheurs en intelligence artificielle avec qui Mr Y a collaboré quelques mois utilisaient par exemple des publications dans un sous-groupe Reddit, « *SuicideWatch* », dédié au soutien de personnes qui ont des pensées suicidaires, pour développer des modèles de détection de risque du suicide.

⁹³ Beth M. Semel a observé des scientifiques et ingénieurs de laboratoires de recherche visant à réinventer les étapes d'évaluation psychiatrique en mettant de côté la subjectivité de la personne experte en santé mentale qui fait habituellement l'évaluation. « *By turning to the expertise of computer scientists and engineers, they seek to build "machine listening" prototypes for psychiatric assessment: technologies that use a microphone to capture sound and artificial intelligence (AI) to analyze sound* ». Comme de nombreux travaux en IA qui cherchent à s'affranchir des biais subjectifs dans l'évaluation (pour le recrutement en ressources humaines par exemple), leurs recherches reposent sur la prémisse que l'IA peut analyser mieux que l'humain en ne se focalisant que sur les données brutes : « *their studies are premised on the notion that AI can listen beyond the human by attending to sounds of speech that have psychopathological significance supposedly set aside from linguistic meaning and human difference* ». Pourtant, pour classer les données à partir desquelles sont développés, testés et entraînés les algorithmes de *machine listening*, ils ont besoin des dimensions interactionnelles et socioculturelles du langage dont ils voulaient s'affranchir (Semel, 2019, p. 3).

écartant le contexte, les habitudes à l'égard des données ont amené à les considérer comme des collections neutres de discours (vocaux ou textuels) et d'images. Cette neutralisation amène à mettre de côté la dimension politique inhérente à ces données, qui sont des traces socioculturelles et interactionnelles porteuses des rapports de pouvoir dans lesquels s'exprime la personne quand elle laisse une de ces traces en ligne, et dans lesquels agit l'autre personne qui classe ces traces selon certains critères de sens dans des bases de données. Ces classements, qui constituent les bases de données, influencent par la suite les systèmes qui s'appuient sur ces ensembles de données, en reproduisant la politique sous-jacente – c'est le problème de la reproduction de biais précédemment mentionné (chapitre 2, p. 106), qui nuit aux personnes discriminées par ces systèmes.

Si l'on revient aux termes kantien, quand la fin visée est de créer des systèmes informatiques reproduisant certaines capacités de l'intelligence humaine – dont la reconnaissance d'image, la reconnaissance vocale et l'analyse de contenus écrits et parlés – par un effacement de tout caractère personnel dans les données utilisées, le respect de *toutes* les personnes sur lesquelles ces systèmes ont des effets est négligé. Par la déshumanisation qu'elle opère, la dépersonnalisation des données donne bonne conscience pour instrumentaliser ces données comme des moyens. Mais l'effet de cette instrumentalisation est un manque de respect envers les personnes affectées négativement par les systèmes conçus à partir de ces données. Crawford dénonce cette invisibilisation du caractère humain propre aux données :

This has created a shift away from ideas like “human subjects” –a concept that emerged from the ethics debates of the twentieth century–to the creation of “data subjects,” agglomerations of data points without subjectivity or context or clearly defined rights. (2021, p. 114-115)

Même si ces agglomérations de points de données révèlent les contours – ou profils – de personnes de façon précise, l'absence de subjectivité et de contexte permet de ne pas se préoccuper de consentement puisqu'il n'y a supposément personne dans ces agglomérations de données (Stark & Hoffmann, 2019).

Par cette absence, il n'y a pas de rencontre entre une personne – à la volonté autonome – et un système d'IA, donc pas de raison de s'occuper de consentement. Voir

les sujets humains présents dans les données comme appelle à le faire Kate Crawford (2021; Metcalf & Crawford, 2016) demande alors de voir cette rencontre, de porter attention à ce consentement et au travail de limites qu'il implique pour que toute personne soit respectée dans l'action que font les systèmes d'IA à partir de données dont le vécu humain ne peut totalement disparaître. Et si l'on tient à respecter l'autonomie des personnes dont les données sont des traces du vécu, alors ces données ne peuvent plus être seulement un moyen pour le développement de systèmes d'IA, mais en même temps une fin, de sorte que soient respectées ces données et les personnes auxquelles elles se rapportent⁹⁴.

Respecter ces données comme des fins et non seulement comme des moyens, en portant attention au travail de limites que le consentement fait pour que le traitement de ces données soit acceptable, cela implique des restrictions à ce que l'on peut faire avec ces données. À la croisée de chaque personne à respecter comme des fins en soi – que la personne soit présente en chair et en os ou sous forme de données en ligne – le consentement implique la restriction de certaines actions qui nous affectent, et l'engagement dans d'autres. Cet effet de restriction ou engagement est, d'une certaine manière, une autre interprétation de ce que Hurd évoque ci-dessus à propos des droits et obligations qu'implique le consentement quand il est donné ou révoqué. Ces restrictions amènent le consentement à remplir plusieurs conditions.

Qu'est-ce qu'un bon consentement ?

Pour qu'il y ait consentement, c'est-à-dire pour qu'un droit soit accordé pour l'utilisation des données qui nous concernent, quand il s'agit de consentement numérique, les conditions posées cherchent à assurer que le consentement soit valable – les termes « *valid* » (Kim, 2019) et « *meaningful* »⁹⁵ sont employés en anglais. C'est cette

⁹⁴ La vastitude de la définition de ce qu'est une donnée personnelle (voir chapitre 1) selon les règlements de protection des données personnelles peut être lue à la lumière de cette conception des données comme des fins en soi, au nom des personnes auxquelles elles se rapportent de façon directe ou indirecte.

⁹⁵ Voir https://www.priv.gc.ca/en/privacy-topics/collecting-personal-information/consent/gl_omc_201805/

validité qui permet au consentement de rendre l'acte acceptable; une bonne expérience de consentement est donc une expérience où le consentement donné est valable. Cette validité tient à plusieurs critères.

Traditionnellement, dans les sociétés libérales pour lesquelles les libertés individuelles sont des valeurs fondamentales, pour que ce consentement soit valable, il faut qu'il soit libre, volontaire et « *undeceived* »⁹⁶ (Mill, 1990, p. 78), c'est-à-dire qu'il ne soit pas obtenu par tromperie ou par mensonge. En médecine, le consentement aux soins est valable s'il est libre et éclairé, et que la personne concernée est apte à consentir (Ferron Parayre, 2018). En recherche, quand des sujets humains participent à un projet, ils doivent donner un consentement qui soit « libre [ou volontaire], éclairé et continu » (Conseil de recherche en sciences humaines (CRSH) et al., 2018, p. 29). Le consentement sexuel quant à lui relève d'un « accord volontaire »⁹⁷ et doit être « clair, libre et éclairé »⁹⁸, il peut également être retiré en tout temps⁹⁹. Si l'on rassemble ces critères de validité du consentement selon ces différents usages, l'acte de consentir est bon quand il est donné volontairement de façon libre, c'est-à-dire non-contraint, ni influencé, ni basé sur un mensonge ou une tromperie; un bon consentement est également informé ou éclairé à propos de ce qui va nous être fait, les risques que cela comporte, les recours que nous avons; il n'est pas donné une fois pour toutes, il peut être retiré; enfin, il est donné par des personnes en âge de le donner.

Dans le domaine numérique, le consentement au traitement des données personnelles reprend ces critères de liberté, d'information et de caractère volontaire d'un choix pris en connaissance de cause, ainsi que d'âge légal; son retrait peut également être possible. Ces critères visent à faire du consentement un mécanisme de protection des individus contre des risques relatifs au traitement de leurs données personnelles

⁹⁶ La phrase originale de John Stuart Mill dans son essai *De la liberté* paru en 1859 est : « *only with their free, voluntary, and undeceived consent and participation* » (cité dans Garcia, 2021, p. 265). Elle s'inscrit dans une réflexion politique sur le lien entre gouvernants et gouvernés.

⁹⁷ Article 273.1, paragraphe 273.1(1) : <https://www.justice.gc.ca/fra/jp-cj/victimes-victims/def.html>

⁹⁸ Voir <https://educaloi.gc.ca/capsules/le-consentement-sexuel/#:~:text=C'est%2D%C3%A0%2Ddire,peut%20y%20avoir%20agression%20sexuelle.>

⁹⁹ Voir <https://www.rcmp-grc.gc.ca/fr/respecter-consentement-sexuel>

(Richards et Hartzog, 2019). Ces risques peuvent être entre autres de l'ordre de l'invasion dans la vie privée, quand une enceinte connectée s'active sans qu'on lui ait demandé et enregistre les sons à ce moment; il peut y avoir un risque de vol d'identité quand les informations directement identificatoires sont interceptées lors d'un piratage; il peut également y avoir un risque de manipulation d'opinion quand de la propagande est envoyée de façon précisément ciblée en fonction des agglomérations de données qui dessinent nos profils reconstitués par les données qu'on égrène au fil de notre navigation en ligne. Idéalement, au nom de l'impératif catégorique développé ci-dessus, toute organisation qui traite des données obéirait volontairement à son devoir moral de respecter ces données comme des fins en soi, au nom des personnes auxquelles elles se rapportent ou qu'elles affectent directement ou indirectement. Mais dans le néolibéralisme actuel, l'obéissance à ce devoir, qui implique des restrictions aux actions des organisations et à leur recherche de profit, est renforcée davantage par des règlements qui contraignent à y obéir. Par la contrainte, le droit est ici l'outil par lequel se matérialise l'éthique déontologique, dans un souci d'entrer en conformité avec ce que les lois édictent, plutôt que de risquer d'être sanctionné.

Selon les pays, États ou provinces, les règlements sont plus ou moins stricts, c'est-à-dire qu'ils contraignent plus ou moins les organisations à mettre en place des processus garantissant cette protection des données personnelles, que ce soit par la limitation de la quantité de données qu'elles collectent, par la façon dont elles en assurent la confidentialité et la sécurité pour limiter les atteintes à la vie privée telles que des fuites d'informations personnelles, par la façon dont elles échangent ces données personnelles avec d'autres organisations (Cavoukian, 2011; voir aussi le RGPD et la LPRDE cités ci-dessous). Je vais m'attarder sur les critères de consentement valable dans les règlements du Canada et de l'Europe car Mr Y et Prélude sont sur le point de traiter des données de personnes au Canada et en Europe, ce qui oblige d'entrer en conformité avec ces deux réglementations qui ont de nombreux points communs en ce qui concerne le consentement. Ces règlements privilégient la clarté pour informer le consentement plutôt que de s'en tenir à certaines ambiguïtés.

Consentir en ligne : une lutte contre l'ambiguïté

Au Canada, le consentement fait partie des dix principes relatifs à l'équité dans le traitement de l'information de la Loi de protection des renseignements personnels et les documents électroniques (LPRPDE) du Commissariat à la protection de la vie privée du Canada¹⁰⁰. La LPRPDE impose que toute organisation collectant des données personnelles obtienne un consentement valable de la part de leurs utilisateurs : « Le consentement n'est considéré comme valable que s'il est raisonnable de s'attendre à ce que les clients de votre organisation comprennent la nature, les fins et les conséquences de la collecte, de l'utilisation ou de la communication des renseignements personnels auxquelles ils ont consenti »¹⁰¹. Il fait écho¹⁰² au consentement « explicite et positif » exigé par le Règlement européen sur la protection des données (RGPD) entré en vigueur en mai 2018.

Le RGPD définit le consentement comme « toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement » (Parlement européen & Conseil de l'Union Européenne, 2016, p. 34). Le consentement est une des six bases légales – ou « fondements juridiques » – qui « donnent le droit à un organisme de traiter des données à caractère personnel selon le RGPD » (Commission Nationale de l'Informatique et des

¹⁰⁰ Ces dix principes sont : 1. Responsabilité, 2. Détermination des fins de la collecte des renseignements, 3. Consentement, 4. Limitation de la collecte, 5. Limitation de l'utilisation, de la communication et de la conservation, 6. Exactitude, 7. Mesures de sécurité, 8. Transparence, 9. Accès aux renseignements personnels, 10. Possibilité de porter plainte à l'égard du non-respect des principes.

Commissariat à la protection de la vie privée du Canada, Principes relatifs à l'équité dans le traitement de l'information de la LPRPDE, janvier 2018. https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/la-loi-sur-la-protection-des-renseignements-personnels-et-les-documents-electroniques-lrpde/p_principe/

¹⁰¹ Commissariat à la protection de la vie privée du Canada, Troisième principe relatif à l'équité dans le traitement de l'information de la LPRPDE – Consentement. Janvier 2018. https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/la-loi-sur-la-protection-des-renseignements-personnels-et-les-documents-electroniques-lrpde/p_principe/principles/p_consentement/

¹⁰² Cet écho entre le RGPD et la LPRPDE peut en partie se comprendre par la notion de « *Brussels effect* », un mécanisme par lequel les politiques publiques et règlements se diffusent entre pays par des mécanismes de marché (voir Mahieu et al., 2021).

Libertés (CNIL), s.d.)¹⁰³. En tant que tel, le consentement doit remplir ces quatre critères cumulatifs pour être valable :

1. Libre : le consentement ne doit pas être contraint ni influencé. La personne doit se voir offrir un choix réel, sans avoir à subir de conséquences négatives en cas de refus.
2. Spécifique : un consentement doit correspondre à un seul traitement, pour une finalité déterminée.
3. Éclairé : pour qu'il soit valide, le consentement doit être accompagné d'un certain nombre d'informations communiquées à la personne avant qu'elle ne consente.
4. Univoque : le consentement doit être donné par une déclaration ou tout autre acte positif clairs. Aucune ambiguïté quant à l'expression du consentement ne peut demeurer. (CNIL, 2018)¹⁰⁴

Ces critères cherchent à évacuer toute ambiguïté qui pourrait invalider le consentement : en étant libre, le choix ne doit pas se faire sous l'ambivalence d'une influence ni d'une contrainte implicite en l'absence d'alternative ou de présence de conséquences négatives; en étant spécifique, le consentement n'est pas donné à un ensemble flou de finalités; la clarté des informations données doit permettre au choix d'être éclairé; l'univocité de l'accord donné tient à son expression explicite.

Selon cette perspective européenne sur la protection des données personnelles, le consentement relève ainsi d'une logique de lutte contre toute ambiguïté. Hélène Giroux (2006) définit l'ambiguïté par trois conditions : (1) le manque de clarté ou le fait d'admettre plus qu'un sens – il s'agit en effet de faire preuve de clarté dans la transparence qui permet la compréhension nécessaire à la dimension éclairée (ou informée) du consentement; (2) la généralité, entendue comme manque de spécificité – pour répondre à l'impératif de spécificité des finalités qui est demandée à l'objet de consentement; (3) le fait d'être vague, par manque de précision et par une place laissée au doute (voir aussi Boivin, 2017; Boivin & Brummans, 2021) – il s'agit d'éviter tout caractère vague dans les informations données, mais aussi dans la façon dont le

¹⁰³ Commission Nationale de l'Informatique et des Libertés (CNIL), s.d, *Base légale*.
<https://www.cnil.fr/fr/definition/base-legale>

¹⁰⁴ CNIL, 2018, *Conformité RGPD : comment recueillir le consentement des personnes ?*
<https://www.cnil.fr/fr/les-bases-legales/consentement>

consentement est donné, une façon qui se veut univoque. Par ailleurs, comme le consentement doit être libre, cette lutte contre l'ambiguïté se dote d'un quatrième prérequis : les conditions dans lesquelles est demandé le consentement ne doivent pas présenter d'ambivalence qui pourrait amener le consentement à être influencé ou contraint.

Cette quatrième condition condamne par exemple la pratique des « *dark patterns* », un terme forgé par le designer en expérience utilisateur Harry Brignull à propos des « *tricks used in websites and apps that make you buy or sign up for things that you didn't mean to* »¹⁰⁵. Par exemple, si le consentement est demandé sous forme de case à cocher, quand la case « oui » ou « d'accord » est cochée par défaut, c'est un *dark pattern*. Cela se rapporte au rôle ambivalent du design dans l'architecture des choix (Thaler & Sunstein, 2009) proposés pour consentir, une ambivalence qui peut amener le consentement à être contraint – c'est la pathologie du « *coerced consent* » dénoncée par Richards & Hartzog (2019) que nous allons voir un peu plus tard.

Ces règlements imposent ainsi une absence d'ambiguïté à la communication sur le traitement des données personnelles auquel il est demandé de consentir. En d'autres termes, ce qui est dit de ce traitement ne doit pas admettre la possibilité d'autres actions qui ne seraient pas mentionnées. Par cette imposition de clarté, spécificité et univocité, ces règlements semblent ainsi chercher à résister à l'ambiguïté pragmatique qui peut être de mise dans les pratiques de traitement de données personnelles. Hélène Giroux appelle ambiguïté pragmatique « *the condition of admitting more than one course of action* » (2006, p. 1228). Elle a forgé ce concept en s'inscrivant dans la lignée de la sociologie de la traduction et utilise la notion d'intéressement de Callon (Callon, 1986; Callon et al., 1988) :

Le modèle de l'intéressement souligne (...) l'existence de tout un faisceau de liens qui unissent l'objet à tous ceux qui le manipulent. (...) [II] met en scène tous les acteurs qui se saisissent de l'objet ou s'en détournent et il souligne les points d'accrochage entre l'objet et les intérêts plus ou moins organisés qu'il suscite. (Callon et al., 1988, p. 21-22)

¹⁰⁵ Voir <https://www.darkpatterns.org/>

Giroux comprend l'ambiguïté pragmatique à la fois comme un résultat et comme une ressource du processus d'intéressement par lequel plusieurs intérêts sont agrégés (Giroux, 2006, p. 1228)¹⁰⁶. Les données personnelles peuvent se comprendre comme étant prises dans un réseau d'intéressements par les organisations qui dépendent d'elles pour développer leurs systèmes numériques et pour répondre aux requêtes des utilisateurs; les utilisateurs sont également pris dans ce réseau d'intéressement puisque le traitement de leurs données personnelles permet l'utilisation du service.

Dans une perspective d'innovation, ces organisations peuvent donc chercher à utiliser un maximum de données passées, présentes et futures pour démultiplier les possibilités de ce qu'elles pourraient faire avec, tout en suscitant l'intérêt de l'utilisatrice à être d'accord si elle veut bénéficier du service proposé et de ses futures améliorations. Remarquons que, comme ce sur quoi porte le consentement – les données personnelles – est un objet d'intéressement, et que le consentement est ce qui est nécessaire à la transaction qui permet l'utilisation de ces données personnelles, le consentement devient lui-même l'objet d'un intérêt particulier qui peut amener à développer des stratégies pour l'obtenir. En gardant vagues les précisions sur les finalités de ces traitements, les organisations qui s'appêtent à traiter des données personnelles, qui cherchent donc à obtenir le consentement de ses utilisateurs, pourraient ainsi s'en remettre à une certaine « viabilité interprétative »¹⁰⁷ (Benders & van Veen, 2001) à propos de ces traitements. Par exemple, une organisation pourrait indiquer dans sa politique de confidentialité qu'elle utilise ces données personnelles « pour améliorer le produit ». Cette viabilité interprétative permet que chaque partie – organisation et utilisatrice – reconnaisse sa propre version de ce qui est fait avec ces données et des finalités de ce traitement, qu'elles acceptent et adoptent cette mention car elles le voient comme étant bénéfique à leurs intérêts : pouvoir traiter toute donnée personnelle du côté de l'organisation, et

¹⁰⁶ Giroux utilise la notion d'ambiguïté pragmatique pour étudier l'émergence et la production collective d'idées de gestion (la qualité notamment). Je transpose sa réflexion dans un autre contexte, celui de la transparence pour informer le consentement.

¹⁰⁷ « *The interpretive viability [which was coined by Ortmann (1995: 370)] allows that different parties can each 'recognize' their own version of the concept. These parties may thus accept and even embrace a concept because they see it as being beneficial to their interests.* » (Benders & van Veen, 2001, p. 38)

pouvoir utiliser le produit du côté de l'utilisatrice. C'est contre cette viabilité interprétative que se positionnent les règlements de protection des données personnelles en imposant un consentement qui repose sur la clarté, la spécificité et l'univocité par lesquelles les pratiques de traitement des données personnelles sont expliquées pour informer le consentement.

Ainsi, s'inscrivant dans une lutte contre l'ambiguïté des pratiques de traitement des données personnelles auxquelles il est demandé, le consentement se veut légalement synonyme de clarté – incluant spécificité, univocité et non-ambivalence. Ou, pour reprendre l'expression de Giroux par son opposée, les règlements de protection des données personnelles imposent une *clarté pragmatique* : que cette clarté des pratiques organisationnelles traitant des données personnelles permette au consentement d'être donné de façon valable. Dans ce sens pragmatique, la clarté n'est pas seulement un qualificatif à propos de la qualité d'information, mais elle est de l'ordre des pratiques par lesquelles le consentement est demandé, informé et donné. En d'autres termes, en relevant de la pragmatique, la clarté se fait en pratique et amène des effets à se produire. Cette clarté pragmatique va de pair avec l'obligation de transparence à laquelle est tenue toute organisation à propos des informations sur ses façons de traiter des données personnelles.

La transparence et son idéal performatif de confiance

Comme l'information qu'elle donne pour éclairer le consentement doit être « complète et précise » (CNIL, 2019), la CNIL rappelle que :

La transparence permet aux personnes concernées :

- de connaître la raison de la collecte des différentes données les concernant;
- de comprendre le traitement qui sera fait de leurs données;
- d'assurer la maîtrise de leurs données, en facilitant l'exercice de leurs droits.

La transparence des informations données pour éclairer le consentement a ainsi une visée de connaissance, de compréhension et de maîtrise du côté de l'utilisatrice à l'égard du traitement de ses données personnelles. Si elle parvient à réaliser ce triple effet performatif, la transparence imposée par le RGPD sera un instrument juridique efficace

pour répondre, entre autres, aux préoccupations à l'égard du consentement des citoyens rencontrés pendant la démarche de la Déclaration de Montréal. Ils avaient en effet remis en question la dimension libre du consentement, soit « l'indépendance des individus concernant le choix de partager ou non leurs données (personnelles), d'avoir un réel impact sur leur gestion ou de choisir les fins pour lesquelles elles seront utilisées ». Ils avaient également interrogé la dimension éclairée de ce consentement, c'est-à-dire les « modalités d'information nécessaires aux individus pour qu'ils puissent consentir de manière éclairée. [Cela] touche l'accès à l'information et la compréhension de cette information » (Voarino et al., 2018, p. 125). La perception d'asymétries de connaissance entre utilisateurs et développeurs en ce qui concerne les implications de l'utilisation d'un système d'intelligence artificielle entretenait un doute sur la façon dont les informations personnelles étaient manipulées à l'insu des utilisateurs. La transparence, si elle fait effectivement preuve d'une certaine clarté pragmatique, viendrait donc dissoudre ce doute.

Il n'est en revanche pas certain que cette transparence permette une réduction totale de toute asymétrie de connaissances entre utilisateurs et développeurs de ces technologies : étant donnée la complexité de l'écosystème et des techniques liées au traitement des données personnelles, il est irréaliste d'attendre que les utilisateurs comprennent tout le fonctionnement et ce que cela implique pour eux (Edenberg & Jones, 2019; Jones & Edenberg, 2020; Kim, 2019; Richards & Hartzog, 2019), de même que les personnes impliquées dans la conception de ces systèmes ont rarement cette connaissance totale. Le caractère éclairé du consentement est donc limité à ce qu'il est possible de connaître pour les personnes concernées, selon le niveau de leur connaissance du fonctionnement de l'écosystème numérique – c'est ce que plusieurs appellent la limite de connaissance¹⁰⁸ (Edenberg & Jones, 2019; Richards & Hartzog, 2019).

¹⁰⁸ Cette limite de connaissance touche à l'enjeu du développement de la littératie numérique que nous verrons en fin de chapitre.

Cela rejoint la limite du consentement éclairé aux soins dans le domaine médical : il n'est pas raisonnable d'attendre du patient qu'il ait le même niveau d'expertise médicale pour consentir de façon éclairée aux soins qui lui sont proposés. Il s'agit plutôt d'expliquer le diagnostic, la nature et les conséquences du soin proposé ainsi que les alternatives possibles, d'expliquer les risques de ce soin, de répondre aux questions du patient, de lui émettre une recommandation, de s'assurer de sa compréhension et enfin d'obtenir son consentement (Philips-Nootens et al., 2016, cité dans Ferron Parayre, 2018, p. 12). Si le processus de consentement est plus rapide dans le contexte numérique (non-médical) étant donné que les risques sur la personne sont moins sensibles qu'en médecine, l'explication sommaire des techniques de traitement des données, de leurs conséquences ainsi que de leurs risques et des alternatives en cas de refus doivent tout de même être présentées de façon brève et vulgarisée. L'utilisatrice doit également avoir la possibilité de poser des questions et d'avoir des réponses – c'est une partie du rôle de *data privacy officer* que le RGPD demande de créer dans chaque organisation. Ainsi, si le niveau de connaissance ne peut pas être parfaitement égal entre utilisateurs et concepteurs, cela ne veut pas dire qu'informer l'utilisatrice par un effort de transparence est vain. Cela peut plutôt relever d'un effort pour développer une certaine confiance.

Dans cette perspective, la CNIL déduit l'effet de la transparence ainsi :

Pour les responsables de traitement [des données à caractère personnel], [la transparence] contribue à un traitement loyal des données et permet d'instaurer une relation de confiance avec les personnes concernées.¹⁰⁹ (CNIL, 2019, à partir des articles 12, 13 et 14 du RGPD).¹¹⁰

La transparence est ainsi associée aux valeurs de loyauté de la part de l'organisation qui traite les données, et de confiance qu'elle peut faciliter chez les personnes qui utilisent ses services numériques. La valeur de loyauté implique des pratiques vertueuses de traitement des données par l'organisation qui permettrait notamment de vérifier l'honnêteté des informations données. La transparence pourrait ainsi relever d'une

¹⁰⁹ CNIL, 2019, *Conformité RGPD : comment informer les personnes et assurer la transparence ?* <https://www.cnil.fr/fr/conformite-rgpd-information-des-personnes-et-transparence>

¹¹⁰ idem

approche de l'éthique de la vertu¹¹¹ (MacIntyre, 2012) dont on voit ici la dimension performative : par la loyauté dont font preuve les personnes qui traitent des données personnelles, la relation avec les systèmes d'IA qu'elles développent peut être digne de confiance. La confiance émerge alors comme un effet de ces pratiques vertueuses de traitement des données. Ainsi, en « permet[tant] d'instaurer une relation de confiance », la transparence aurait un effet performatif qui désamorcerait toute méfiance ou doute à l'égard de l'utilisation des données personnelles. Par la confiance, elle permettrait de tisser une relation bonne sur un plan moral car basée sur de la transparence loyale. Cela implique que l'expérience soit orchestrée par un traitement respectueux des données personnelles; l'expérience serait alors qualitativement et moralement bonne.

Le vocabulaire de loyauté et de confiance utilisé ici combine ainsi une dimension de vérifiabilité et une de performativité que Albu et Flyverbom (2019) ont identifiées comme correspondant à deux approches des études organisationnelles sur la transparence. L'approche de la vérifiabilité¹¹² « *focuses on how information is disclosed to verify a particular state of affairs* ». Elle repose sur la prémisse selon laquelle « *by making more information available we can regulate behavior and improve organizational and societal affairs through processes of verification* » (p. 281). L'approche performative est « *less certain that more information generates better conduct* » et souligne plutôt la « *complexity of communication and interpretation processes and focus on the complications and paradoxes generated by transparency projects* » (p. 281). Si les modèles de consentement relèvent d'un exercice de transparence en ce qui concerne l'information sur laquelle le consentement repose (pour être donné de manière éclairée), une approche performative de ces modèles permet de porter attention aux complications et paradoxes qui sont inhérents à leurs conceptions. Cette approche performative des pratiques impliquées dans les modèles de consentement – que ces pratiques soient celles

¹¹¹ Comme le propose Gibert (2020), l'éthique de la vertu est une approche intéressante pour étudier les pratiques en éthique de l'IA – que ce soit l'éthique des organisations en IA ou l'éthique des systèmes d'IA. Je vais dans le sens d'une éthique de la vertu en avançant vers l'idéal d'épanouissement puis le *care*.

¹¹² En IA, dans un objectif de transparence, la vérifiabilité est associée à l'enjeu d'« explicabilité » qui vise à ouvrir la boîte noire des algorithmes pour expliquer comment ils arrivent à la décision qu'ils prennent. C'est un enjeu complexe qui fait l'objet de beaucoup de littérature, en particulier en droit et en informatique.

des organisations qui traitent des données personnelles, ou celles des personnes auxquelles il est demandé de consentir – révèle de nombreuses ambiguïtés qui rendent l'idéal de clarté pragmatique (visé par les règlements) difficile à réaliser parfaitement. Mon immersion chez Mr Y et Prélude, par le projet de conception d'un modèle de consentement pour leurs agents conversationnels, a pris la tournure performative d'une plongée dans ces ambiguïtés (voir chapitre 6). L'appel à donner du contrôle aux utilisateurs est un exemple significatif d'une condensation d'ambiguïtés.

« C'est toi qui as le contrôle sur les informations que tu donnes »

Donner du contrôle aux utilisateurs est un des effets performatifs de la clarté pragmatique du consentement. En effet, pour aller plus loin dans le pouvoir que les utilisateurs peuvent avoir à l'égard de leur vie privée en ligne, le mécanisme légal du consentement est souvent considéré comme quelque chose qui permet aux utilisateurs d'avoir un certain contrôle sur les informations personnelles telles qu'elles sont collectées, analysées, partagées et utilisées lors de leur navigation en ligne (Edenberg & Jones, 2019). Ainsi, dans la Charte numérique du Canada – qui a été mise à jour en 2021 à l'issue de consultations citoyennes nationales lancées à partir de juin 2018 sur le thème du numérique et des données – la section 3 associe « consentement, contrôle et transparence des données » dans son titre et mentionne :

Bien que de nombreux Canadiens reconnaissent que l'utilisation des données peut leur être très bénéfique, entraînant une meilleure prestation de services, des expériences d'utilisateur plus personnalisées et une efficacité accrue du système, il existe de réelles préoccupations concernant la confidentialité et le contrôle des données. Le gouvernement cherche des moyens de renforcer le contrôle exercé sur les consommateurs en modernisant l'écosystème de la vie privée et en veillant à ce que les Canadiens aient la confiance nécessaire pour tirer pleinement parti des possibilités offertes par les innovations fondées sur les données. (Canada & Innovation, sciences et développement économique Canada, 2019, p. 22)

La modernisation de « l'écosystème de la vie privée » passe par la mise à jour de la LPRPDE et la mise en place de « réglementations obligatoires en matière de signalement des

infractions (...) renforçant la responsabilité des organisations de dénoncer les infractions importantes affectant les consommateurs » (p. 22). Les personnes sont également invitées à se tourner vers le Commissariat à la protection de la vie privée du Canada pour de la formation et de l'orientation à propos des renseignements personnels (p. 22).

Par quelles pratiques ce contrôle peut-il effectivement se faire ? Quelles pratiques organisationnelles implique-t-il ? Du côté des utilisateurs, ou consommateurs, comment ce contrôle s'exerce-t-il ? Quelle est son envergure ? Cette envergure permet-elle au consentement d'être effectivement en parfait contrôle ?

Réinventer les modèles de consentement pour plus de clarté

En demandant davantage de transparence, dans un idéal de clarté pragmatique, les règlements de protection des données personnelles invitent à réinventer les modèles de consentement pour ne plus reposer sur le modèle du « *notice and consent* » qui prévaut. Par ce modèle, le consentement est demandé en supposant la lecture de politiques de confidentialité rédigées en langage peu accessible pour les personnes qui ne sont pas juristes. Fred H. Cate et Viktor Mayer-Schönberger critiquaient dès 2013 ce modèle de *notice and consent* qui ne permet pas de prendre des décisions informées à propos du traitement de nos données personnelles :

Today, almost everywhere that individuals venture, especially online, they are presented with long and complex privacy notices routinely written by lawyers for lawyers, and then requested to either 'consent' or abandon the use of the desired service. That binary choice is not what the privacy architects envisioned four decades ago when they imagined empowered individuals making informed decisions about the processing of their personal data. In practice, it certainly is not the optimal mechanism to ensure that either information privacy or the free flow of information is being protected. (Cate & Mayer-Schonberger, 2013, p. 67)

Quand Paul et moi commençons à envisager de travailler ensemble autour du consentement pour le chatbot Mr Y, c'est ce modèle de *notice and consent* qui est utilisé : dès le début de la conversation avec le chatbot, celui-ci mentionne avoir besoin que l'utilisatrice sache ce qui est fait des données qu'elle partage dans la conversation et qu'elle donne son accord à ce mode de fonctionnement. Pour cela, le chatbot propose un lien externe qui renvoie à la page des termes de protection de la vie privée sur le site

internet de Mr Y, via un bouton « En savoir plus ». Immédiatement après, il demande si l'utilisatrice est d'accord avec ces conditions en proposant le bouton « je suis d'accord » et suggérant à nouveau un bouton « En savoir plus » qui renvoie à la page « *Privacy* » – une page écrite en petits caractères, seulement en anglais, et peu agréable à parcourir étant donné la quantité de texte pour une page internet. Si le bouton « je suis d'accord » est pressé, Mr Y considère que le consentement est obtenu – si on n'appuie pas sur ce bouton, la conversation ne continue pas. Mon implication avec Mr Y consiste à réinventer ce modèle, conformément aux appels du RGPD et de la LPRPRDE.

Plutôt que de donner son consentement comme on passe une porte obligatoire pour utiliser un service, sans porter attention à ce qu'implique cet accord donné à des termes obscurs, la clarté prônée par les critères du RGPD et de la LPRPDE permet au consentement d'être davantage éclairé et à la transparence d'être plus effective. Ces critères font aller le consentement dans le sens d'une prise de décision qui mobilise de l'information permettant de connaître plus clairement ce à quoi on s'apprête à donner notre accord. Ils obligent également les organisations qui demandent ce consentement à proposer des alternatives pour permettre l'utilisation de leurs services numériques même si les utilisateurs n'acceptent pas que leurs informations personnelles soient traitées comme annoncé. Ces critères favorisent ainsi un choix moins binaire qu'avec le *notice and consent* qui implique d'accepter les termes de services ou d'abandonner l'utilisation.

En Europe, le RGPD va plus loin dans la mise en place de mécanismes de contrôle par les citoyens en édictant des droits des utilisateurs, ou « droits de la personne concernée [par le traitement de données à caractère personnel] » (RGPD, 2018, chapitre 3¹¹³). Étant donné ces droits fondamentaux, les organisations ont à présent l'obligation de mettre en place ce qui permet à leurs utilisateurs d'exercer ces droits quand ils interagissent avec leurs services et produits. Il s'agit, pour la personne concernée, de droits :

¹¹³ Parlement européen & Conseil de l'Union Européenne, 2016, *Chapitre 3 : Droits de la personne concernée*. <https://gdprinfo.eu/fr/fr-chapter-3>

- (1) d'information sur l'utilisation des données et sur ses droits (Articles 13 et 14, rappelé dans le paragraphe sur la transparence cité plus haut) – comme nous l'avons vu, cette information doit être communiquée de façon claire pour éclairer le consentement;
- (2) d'accès aux données qu'un organisme détient sur elle (Article 15) – cet organisme doit donc mettre en place un processus qui permette la communication de ces données à la personne concernée;
- (3) de rectification quand des informations sur elle sont inexactes (Article 16);
- (4) d'effacement des données qui la concernent, un droit aussi appelé « droit à l'oubli » (Article 17);
- (5) droit à la limitation du traitement des données les concernant (Article 18);
- (6) droit à la portabilité, c'est-à-dire que la personne a le droit d'obtenir une copie de ses données, et de l'emporter pour les réutiliser ailleurs – dans ce cas l'organisation doit transmettre les données dans un format structuré, lisible, couramment utilisé (Article 20);
- (7) d'opposition au traitement de certaines données (Article 21), ce que vise à permettre la sélection des critères de consentement dans les fenêtres qui y sont dédiées;
- (8) « de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage » (Article 22) – le profilage étant la création automatique de profils en fonction des données recueillies sur nous à propos de nos goûts, comportements, opinions, etc. lors de notre utilisation de systèmes numériques.

Ces droits entraînent ainsi l'obligation, pour les organisations traitant des données personnelles, de mettre en place des processus et concevoir des fonctionnalités techniques qui permettent aux utilisateurs d'accéder aux données sur eux, de les rectifier, les effacer, les télécharger pour les utiliser ailleurs, et limiter voire s'opposer à certains traitements.

Une bonne expérience de consentement est donc une expérience qui non seulement permet l'expression du consentement en tant qu'acte communicationnel qui marque un accord, mais qui permet aussi l'exercice de ces droits. Cela demande de concevoir l'architecture et les processus qui permettent de le demander, ce qui fait l'objet de plusieurs discussions plus ou moins fructueuses avec Mr Y et Prélude (voir chapitre 6) : ces obligations soulèvent plusieurs défis organisationnels de gestion des données, d'archivage, de formatage, qui impliquent également des coûts en temps pour développer les fonctionnalités appropriées, et en infrastructure de sécurité. Ces défis s'accompagnent d'une obligation de concevoir une expérience utilisateur qui permette l'exercice de ces droits tout en s'intégrant dans l'expérience générale du service offert. Or ces défis et ces coûts ont un poids particulier dans les startups dont les ressources sont souvent très limitées à leur démarrage (Norval et al., 2021). C'est dans ces défis et coûts que ma présence s'inscrit au sein de Mr Y et Prélude en 2019, à un moment où peu de solutions sont instaurées pour relever ces défis posés par les règlements de protection des données personnelles, mais également par les enjeux soulevés en éthique de l'IA.

Une façon de permettre l'exercice de ces droits, et d'encourager le contrôle des utilisateurs à l'égard de leurs informations personnelles, est de leur permettre de consulter ces informations et éventuellement les supprimer via une interface qui y est dédiée. Pour l'Assistant Google, comme pour tout service utilisé de Google, c'est ce que propose l'interface « Mon activité »¹¹⁴ (voir figure 4 ci-dessous), une page qui met plusieurs commandes à disposition pour consulter et supprimer toute activité faite via un outil Google. Ce type de commande mise à disposition de l'utilisatrice permet de facilement voir et supprimer les informations récoltées sur elle au fil de sa navigation en ligne via des sites internet, des applications ou des objets connectés¹¹⁵.

¹¹⁴ Voir <https://myactivity.google.com/myactivity>

¹¹⁵ Pour les objets connectés qui n'ont pas d'interface visuelle en revanche, l'exercice de ce contrôle, ainsi que l'information du consentement, est très limité, comme le soulignent Jones et Edenberg (2020) et Morel (2020). Dans sa thèse en informatique, Victor Morel propose un modèle de consentement consultable sur smartphone pour tout objet connecté amené à traiter des informations nous concernant (2020).

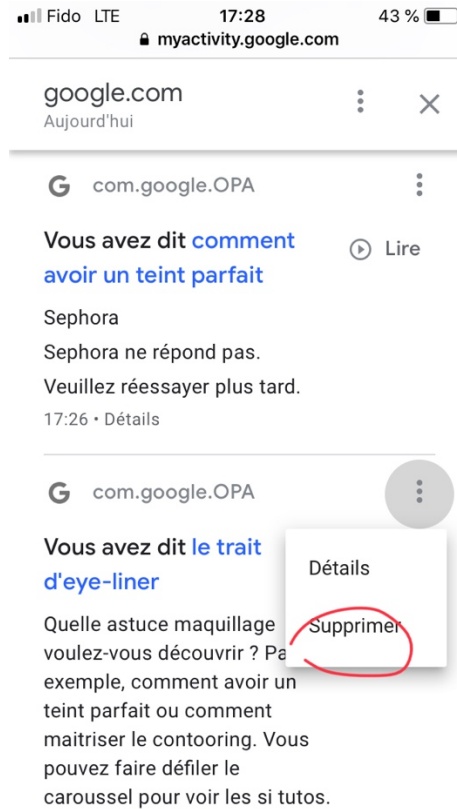


Figure 4 : Capture d'écran de « myactivity » sur Google Assistant après avoir testé l'Action Sephora. 11 septembre 2019, 17h28.

C'est cet accès et cette possibilité de suppression qui mettent en pratique l'idée de donner du contrôle aux utilisateurs. Cet enjeu du contrôle que l'on peut donner par le design d'interface qui donne des possibilités de gestion des données personnelles fait l'objet de plusieurs expérimentations en design¹¹⁶. Cependant, savoir que ces interfaces existent n'est pas évident, cela demande une certaine connaissance des informations qui sont retenues sur nous et de ce que l'on peut faire à cet égard; cela demande également de savoir comment y accéder¹¹⁷. Par ailleurs, ce contrôle donné par une interface se révèle limité par les nombreuses ambivalences que soulèvent les pratiques de design à l'égard du consentement.

¹¹⁶ Voir par exemple la collection de concepts de design pour améliorer les expériences de protection des données personnelles « *Privacy Design Forecast 2019* » du Shorenstein Center : <https://shorensteincenter.org/privacy-design-forecast-2019/>

¹¹⁷ Cela touche un enjeu de littératie numérique que nous allons voir plus en détail à la fin de ce chapitre.

Le design du consentement et ses ambivalences

Les questions dans la vignette Allô (voir p. 114) l'ont soulevé : le design de consentement est une activité en prise avec plusieurs ambivalences pour respecter la dimension libre, ou non-contrainte ni influencée du consentement. Comme le design cadre les actions, par les affordances, les *nudges* et l'expérience qu'il arrange (cf chapitres 1 et 2), vient-il toujours interférer dans le caractère libre du consentement ? Le consentement peut-il rester valable quand il est obtenu *by design* ?

Obtenir le consentement par une bonne expérience utilisateur

Si l'on considère l'expérience utilisateur générale d'un service numérique, et l'impératif selon lequel le consentement doit être demandé avant toute collecte de données, la demande de consentement intervient à un moment important du parcours utilisateur : il s'agit de l'entrée dans l'expérience, une phase appelée *onboarding*¹¹⁸. Ce moment est critique pour que le phénomène de fusion opère, c'est-à-dire pour que l'expérience de la technologie s'inscrive dans la continuité de l'expérience générale de l'utilisatrice : c'est à ce moment que l'engagement de l'utilisatrice peut être favorisé ou au contraire la détourner de ce service. Ainsi, ce moment fait l'objet d'une stratégie de design qui prévoit minutieusement « *the flow, messages, interactions and UI design elements (...) so the experience as a whole can support new users' engagement* » (Cascaes Cardoso, 2017, p. 263). D'un point de vue de design UX, on cherche donc dans cette étape à éviter qu'il y ait des accrocs tels que des éléments distrayants qui empêchent d'accéder au service souhaité directement. On veut également éviter les interactions qui exercent un certain contrôle en forçant à faire des actions spécifiques, ainsi que ce qui peut sembler bancal ou hésitant, « *like the times when you accidentally get kicked out of a tour before it's over and can't start it again to see what you missed* » (Hulick, s.d.)¹¹⁹. Plutôt que ces effets distrayants, contrôlants et approximatifs, Hulick recommande au contraire une expérience qui soit intégrée à l'expérience générale, « *empowering* » et robuste.

¹¹⁸ Voir par exemple les blogues « *UserOnboard* » (<https://www.useronboard.com>) ou « *User Guiding* » (<https://userguiding.com>), en particulier <https://userguiding.com/blog/user-onboarding/>).

¹¹⁹ Voir <https://www.useronboard.com/bulletedproof-user-onboarding/>

Comment faire pour que le consentement s'inscrive dans une expérience de consentement qui soit bonne de cette manière, c'est-à-dire intégrée, « empouvoirante » et robuste ?

Comme nous l'avons vu au chapitre 1 (p. 53) avec l'exemple des travaux sur le consentement informé pour illustrer l'approche du *value-sensitive design*, Friedman et ses collègues ont créé un *plug-in* pour éviter d'imposer une charge excessive à l'utilisatrice pour s'informer, ce qui l'amènerait à être distraite de ce qu'elle cherche à faire initialement en venant sur le site internet en question. À présent, je questionne le consentement dans le cadre d'une expérience avec un agent conversationnel, tel qu'il pourrait être demandé au début de la discussion : est-ce que demander le consentement peut ne pas être « un obstacle d'entrée à la fluidité de la discussion », comme on le craignait dans la vignette Allô ?

Dans un objectif de transparence et de réinvention des modèles de consentement, le design du processus de consentement, qui s'inscrit dans cette phase d'entrée dans l'expérience utilisateur, doit éviter toute ambiguïté. Le design de consentement va donc jouer avec la clarté à plusieurs niveaux : au niveau des formulations par lesquelles informer et demander le consentement ainsi qu'au niveau des modes d'interaction qui permettent de consulter cette information, de donner ou d'obtenir le consentement et le conserver.

Le design d'un agent conversationnel pourrait s'inspirer d'un modèle de design de consentement issu de la mise en application des critères de consentement valable par d'autres modèles que le *notice & consent* : celui qui amène des fenêtres à surgir lorsqu'on se met à consulter un site internet ou une application. Il s'agit de « bannières de consentement » (*consent banners*) qui exposent les principales informations à propos du traitement des données et donnent la possibilité d'accepter ou de décliner ces conditions. En optant pour un langage plus vulgarisé que le langage juridique des politiques de confidentialité, ces fenêtres expliquent les principaux types de données collectées, leurs finalités, avec quelles autres organisations elles sont partagées, et elles donnent la possibilité d'opter pour le traitement de ces données ou de le refuser. Ces bannières se

présentent principalement sur des sites internet – on pourrait imaginer la présence d’une de ces bannières sur le site de Mr Y ou Prélude par exemple, lorsqu’on se renseigne sur leurs agents conversationnels. Ce site renverrait ensuite au lien sur l’Assistant Google qui permettrait d’avoir accès aux agents développés comme Actions. Cette bannière de consentement aurait l’avantage de ne pas distraire la discussion puisqu’elle aurait lieu hors de la conversation avec le bot. Mais, comme le site internet n’est pas la principale porte d’entrée dans l’expérience avec ces agents conversationnels, opter pour une bannière amènerait le consentement à ne pas être adressé, il ne serait donc ni demandé ni obtenu, ce qui empêcherait (légalement) de traiter les données personnelles partagées au fil de la conversation.

Si le format de bannière de consentement n’est pas utilisé chez Mr Y et Prélude, il est intéressant d’en comprendre la logique, car celle-ci est présente dans les réflexions sur le modèle de consentement dans ces organisations. C’était un phénomène émergent en 2019, mais ces bannières de consentement sont souvent conçues par des *consent management platforms*, c’est-à-dire des entreprises qui offrent en ligne des services et outils de « gestion de consentement ». Par exemple, didomi.io présente ses services de « gestion du consentement et des préférences sur-mesure » comme une façon de faire « de la conformité RGPD un avantage commercial » qui permet de « créer de la valeur par la confiance »¹²⁰. Pour présenter leur plateforme de gestion du consentement, ils utilisent un vocabulaire de marketing pour parler du consentement :

Optimisez votre expérience utilisateur avec un marketing axé autour du consentement

La conformité va plus loin que la seule collecte du consentement. Il s'agit de faire du respect de la vie privée des utilisateurs une véritable expérience client. Priorisez l'expérience client et prouvez votre transparence dans la collection des données personnelles. Optimisez vos taux de consentement et faites du respect de la vie privée une opportunité commerciale.¹²¹

¹²⁰ Voir <https://www.didomi.io/fr/>, consulté le 15 février 2022.

¹²¹ Voir <https://www.didomi.io/fr/>

La conformité aux règlements de protection des données personnelles est ainsi évoquée en termes d'expérience client. Le consentement est abordé comme faisant l'objet d'un taux qui peut être optimisé. On voit ici l'intéressement qu'il peut y avoir dans le consentement qui s'inscrit ainsi dans une stratégie d'optimisation au service d'une opportunité commerciale à saisir. Cet intéressement envers le consentement amène des organisations à jouer avec la clarté pragmatique imposée par les règlements : l'objectif étant d'obtenir les consentements des utilisateurs (Almeida & Monteiro, 2021), le design de l'expérience utilisateur est adapté à la fois pour répondre aux exigences des règlements comme le RGPD et pour favoriser les « taux de consentement ». Du côté de l'organisation qui le demande, une bonne expérience de consentement est ici favorisée par un design qui permet à l'organisation d'obtenir le plus de consentements possibles, tout en étant conforme aux règlements qui en encadrent l'obtention. Mais ce design d'expérience utilisateur mène à une tendance à l'influence qui peut relever du *dark pattern* tout en voulant répondre aux critères de consentement valable. Ce jeu par le design d'interaction soulève alors quelques ambiguïtés : les stratégies pour obtenir le consentement sont-elles compatibles avec la validité demandée au consentement ?

Des modèles de consentement et leurs échecs

Les modèles de consentement offerts par les plateformes de gestion du consentement sont conformes au *Transparency and Consent Framework* de l'IAB Europe, « *the only GDPR consent solution built by the industry for the industry, creating a true industry-standard approach* »¹²². L'industrie en question est celle du marketing et de la publicité en ligne que l'association IAB représente au niveau européen. En imposant une limitation à la quantité de données collectées sur les personnes, le RGPD a amené cette industrie à devoir revoir ses pratiques en matière d'accès et de stockage des informations

¹²² Voir <https://iab europe.eu/transparency-consent-framework/>. L'IAB Europe est une association qui représente les industries du marketing et de la publicité au niveau des instances européennes.

sur l'appareil d'une utilisatrice, tels que les cookies¹²³, les identifiants publicitaires et d'autres technologies de traçage des utilisateurs qui permettent leur profilage.

Cependant, plusieurs études montrent que les *consent management platforms* qui suivent le modèle *Transparency and Consent Framework* n'entrent pas totalement en conformité avec le RGPD¹²⁴. Par exemple, à partir d'un échantillon de 560 sites internet testés, Matte et al. (2020) ont relevé que 54% d'entre eux avaient au moins une enfreinte au règlement. Ils ont montré la présence de *dark patterns* sur 141 sites qui ont enregistré par défaut l'obtention d'un consentement alors que l'utilisatrice n'avait pas indiqué de choix, ainsi que sur 236 sites qui présélectionnaient les options amenant à accepter le consentement. 27 sites ne respectaient pas le refus de consentement de leurs utilisateurs : ils enregistraient un consentement positif alors que le consentement avait été explicitement décliné. Dans la même direction que ces résultats, sur 10 000 sites internet au Royaume-Uni sur lesquels ils étudiaient les impacts des choix de design sur les modèles de consentement, Nouwens et al. (2020) ont trouvé que moins de 12% de ces sites étaient conformes au RGPD en matière de transparence et d'obtention du consentement. Les expériences de bon consentement, dans le sens de celles qui permettent un consentement valable de l'utilisatrice, ne sont donc pas si courantes.

Mon immersion chez Mr Y et Prélude en 2019 navigue dans les mêmes défis que les modèles de consentement de l'IAB cherchent à surmonter : combiner bon UX et

¹²³ « Un cookie est un petit fichier stocké par un serveur dans le terminal (ordinateur, téléphone, etc.) d'un utilisateur et associé à un domaine web (c'est à dire dans la majorité des cas à l'ensemble des pages d'un même site web). Ce fichier est automatiquement renvoyé lors de contacts ultérieurs avec le même domaine.

Les cookies ont de multiples usages : ils peuvent servir à mémoriser votre identifiant client auprès d'un site marchand, le contenu courant de votre panier d'achat, la langue d'affichage de la page web, un identifiant permettant de tracer votre navigation à des fins statistiques ou publicitaires, etc. Certains de ces usages sont strictement nécessaires aux fonctionnalités expressément demandées par l'utilisateur ou bien à l'établissement de la communication et donc exemptés de consentement. D'autres, qui ne correspondent pas à ces critères, nécessitent un consentement de l'utilisateur avant lecture ou écriture. »

<https://www.cnil.fr/fr/definition/cookie>

¹²⁴ Le 2 février 2022, plusieurs autorités de protection des données personnelles en Europe ont déclaré ces bannières ou *popups* de consentement non-conformes à la loi : « *GDPR enforcer rules that IAB Europe's consent popups are unlawful* » <https://www.icli.ie/news/gdpr-enforcer-rules-that-iab-europes-consent-popups-are-unlawful/> ; <https://www.gegevensbeschermingsautoriteit.be/publications/beslissing-ten-gronde-nr.-21-2022-english.pdf>

consentement valable, intérêts de l'industrie et des utilisateurs, en particulier par la conformité aux règlements. Le modèle auquel nous aboutissons avec Prélude prend la forme d'une liste (figure 5 ci-dessous) qui s'affiche quand l'utilisatrice indique vouloir « *learn more about data privacy* » auquel un *prompt* répond : « *Great. It won't be long I promise... Here are the main questions people ask me about data privacy* ».

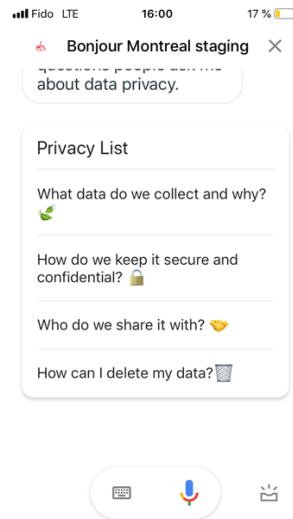


Figure 5 : Capture d'écran de la liste pour informer le consentement de l'utilisatrice dans la version alpha de la Voix de Montréal. 17 décembre 2019, 14h15.

Je reviendrai plus en détail sur le contexte dans lequel ce format est choisi ainsi que les étapes qui précèdent cette liste (chapitre 6). Pour l'instant, notons que ce modèle n'est probablement pas plus valide que le *Transparency and Consent Framework* puisqu'il faut plusieurs interactions avant d'arriver à ces informations – ce qui peut être interprété comme un *dark pattern* de l'ordre de l'obstruction. Ce manque de validité porte son lot de déceptions (voir chapitre 6) mais mon attention se porte finalement davantage sur la performativité à l'œuvre dans l'organisation quand nous naviguons dans les défis et ambivalences du « design de consentement », ou plutôt du design de conversation portant sur le consentement.

Par ailleurs, comme l'agent conversationnel de Prélude est utilisable exclusivement via l'Assistant Google, consentir à l'assistant vocal de Prélude implique de consentir au

traitement des données par l'Assistant Google. Or à ce moment, le modèle de consentement pour les services de Google n'est pas tout à fait valable, comme le montre l'étude de Human et Cech réalisée en 2020. Selon cette étude menée par une approche en sciences cognitives, les modèles de consentement des services de Google, Facebook, Amazon et Microsoft (GAFAM, les cinq organisations cheffes de file du numérique) présentent les mêmes faiblesses que ceux des plateformes de gestion du consentement : « *human-centric cognitive dimensions of consenting are not only used in a positive manner—which could lead to end-user empowerment—but also have been suppressed by implicit dark patterns which need to be avoided* » (Human & Cech, 2020, p. 156). La présence de *dark patterns* dans les modèles par lesquels sont demandés et obtenus le consentement rend ainsi le consentement problématique, dans la mesure où ces *patterns* jouent sur les réflexes cognitifs à l'œuvre dans l'action de donner son accord.

Ces manques de conformité du consentement semblent soulever une contradiction à l'égard de l'ambiguïté. D'un côté, les règlements qui impliquent des lignes directrices de consentement valable reposent sur la clarté – ou non-ambiguïté – par laquelle celui-ci doit être informé, demandé et formulé. Mais de l'autre, les modèles de consentement – réinventés pour faire preuve de clarté – jouent sur des ambiguïtés de design quand il s'agit de faire appel à la capacité de prendre une décision telle que donner son accord à quelque chose. Ces ambiguïtés de design reposent en particulier sur la présence de *patterns* qui se révèlent avoir un effet maléfique sur la prise de décision. Par une apparente clarté du design visuel, du design d'interaction et des formulations, couplée à la préservation de la fusion que cherche le design UX, le consentement est orienté de manière à être obtenu, ce qui rend sa validité ambiguë : est-il vraiment voulu par l'utilisatrice ou est-il donné sous influence ? Cette question touche à l'enchevêtrement de ce qui, chez un individu, permet de prendre la décision de donner son accord ou de refuser. À défaut de trouver une réponse claire et rapide à cette question, ce qui m'intéresse en me penchant sur la conception de tels modèles de consentement par une approche performative, c'est de porter attention aux complications et paradoxes inhérents à la conception d'un modèle de consentement

tarauté par cette préoccupation : comment concevoir une interface qui permet de consentir, sans influencer ce consentement ?

Des choix de design plus ou moins sombres

Arrêtons-nous plus en détails sur les différentes étapes du processus de consentement pour voir comment les décisions de design jouent ou non sur le consentement. À partir d'une étude de bannières de consentement, Gray et al. (2021) distinguent ces étapes ainsi :

(1) the initial framing as a user enters the site; (2) the presentation of configuration options to accept or select more precise consent options; (3) the means of accepting the configuration options; and (4) the ability to ultimately revoke consent. (Gray et al., 2021, p. 5)

Par un croisement de perspectives en interaction humain-machine (HCI), en design, en protection des données personnelles et de la vie privée et en droit, ils montrent que des *dark patterns* sont présents à chacune des étapes du processus de consentement. Dès la première étape, plusieurs sites optent pour la solution du *consent wall* – une option qu'offrent les plateformes de gestion du consentement¹²⁵ mais qui pourrait être également présente dans une interface vocale qui imposerait de passer par cette étape en empêchant toute autre interaction avec le bot en attendant. La solution du *consent wall* consiste à afficher une fenêtre de demande de consentement en premier-plan, tandis que le site en arrière-plan est noirci. Tant que l'on n'a pas cliqué sur le bouton qui donne la possibilité d'accepter ou de refuser, il n'est pas possible de naviguer sur le site. Ce choix de design d'interaction permet de concentrer l'attention de l'utilisatrice sur ces informations en les séparant, à la fois visuellement et interactivement, du reste du site.

¹²⁵ Voir par exemple <https://www.axeptio.eu/fr/consent-wall>. Remarquons qu'Axeptio propose que cet « écran de consentement » ne soit pas réservé à la phase initiale d'entrée sur le site mais apparaisse quand des données spécifiques sont susceptibles d'être collectées (telles que les données de Google Analytics pour mesurer l'audience qui visite le site, ou des données partagées avec des partenaires). Cela participe à un effort de faire que le consentement soit continu. « Grâce à cette fonctionnalité, affichez quand vous le souhaitez un écran de consentement présentant un[e] ou plusieurs solutions que vous souhaitez valoriser en expliquant leur intérêt pour vos visiteurs ou pour votre société ». Cette formulation indique une stratégie qui relève des « *nudges* éducatifs » qui visent à expliquer ce que chacune des parties a à y gagner – et qui présente l'ambivalence d'influencer le choix par cette valorisation.

Cependant, bloquer l'accès au site tant que le consentement n'a pas été donné empêche que l'action de donner son consentement soit libre. En plus d'être désagréable par son aspect contrôlant, ce design d'interaction donne une dimension contrainte au consentement. Ce choix de design joue sur l'interruption de la fusion pourtant habituellement préservée pour l'expérience utilisateur. Pour le design de conversation, nous avons vu que cette interruption pouvait être de l'ordre de l'erreur, à ce titre elle devait être évitée ou réparée (chapitre 2, p. 93); à présent elle est utilisée volontairement, au risque de rendre l'expérience de consentement désagréable.

De plus, du point de vue de l'utilisatrice, comme cet écran de consentement peut être perçu comme une barrière à l'expérience du site souhaité, elle peut être amenée à consentir au nom de sa frustration dont elle veut se débarrasser en cliquant « ok » pour pouvoir accéder au site. Le consentement est alors donné sous l'effet d'un *dark pattern* de l'ordre de l'« obstruction to the user's primary intention to access the full content of the website visited » (Gray et al., 2021, p. 7). Le consentement donné et obtenu est alors teinté de la dimension pathologique du *coerced consent* selon Richards et Hartzog (2019). Plutôt que de bloquer toute autre interaction, une alternative serait de permettre d'utiliser le site sans accepter ce que la fenêtre de consentement indique. Dans ce cas, le consentement non-donné ne doit pas être considéré comme obtenu par défaut, mais comme un consentement absent qui ne permet pas le traitement des données en question.

En ce qui concerne la configuration de l'information et des options proposées pour accepter ou sélectionner des options plus précises de consentement (étapes 2 et 3 du processus de consentement), celle-ci est acceptable quand elle est transparente et non-contraignante, c'est-à-dire quand il est expliqué clairement ce qu'implique chaque décision de consentement, y compris si un refus implique d'avoir accès à un service réduit (Gray et al., 2021, p. 11). En revanche, si ces explications se font de façon à inciter à accepter, en utilisant des formulations qui séduisent pour dire ok ou au contraire qui blâment en cas de refus, dans ce cas il s'agit de *dark patterns* qui amènent à nouveau le consentement à être donné ou obtenu sous influence, donc à être pathologiquement

contraint. Si ces explications et les options de consentement offertes sont configurées de telle sorte qu'elles ne présentent pas clairement les informations sur lesquelles est amené à se baser le consentement, alors celui-ci est de l'ordre du *unwitting consent*. Ce dernier est une autre forme pathologique du consentement, selon Richards et Hartzog (2019), quand celui-ci est donné sans avoir tout à fait connaissance de ce qu'il implique. Ces *dark patterns* peuvent être de l'ordre de l'interférence de la part de l'interface quand, par exemple, des couleurs incitent à cliquer sur « accepter » (Machuletz & Böhme, 2020; Utz et al., 2019). Ils peuvent également être de l'ordre de l'obstruction, quand une action alternative est empêchée, ou encore du « *sneaking* », quand le consentement est obtenu furtivement (Gray et al., 2018).

La part sombre de ces *dark patterns* qui rendent le consentement plus contraint que libre est variable. Quand les *patterns* de design optent pour des formulations enjouées en vue d'éduquer, sans pour autant avoir des intentions de manipulation à l'encontre des intérêts de l'utilisatrice (Acquisti et al., 2018; Hertwig & Grüne-Yanoff, 2017), ces *patterns* relèvent des *nudges* que Thaler et Sunstein (2009) ont mis de l'avant pour favoriser la prise de décisions économiques relatives à la santé et au bien-être (voir également chapitre 1, p. 57). En revanche, quand ces *nudges* sont utilisés pour obtenir un consentement sans que celui-ci soit respecté, ces *dark patterns* relèvent de ce que Conti et Sobiesk (2010) ont classé comme « *malicious interfaces* » (Conti & Sobiesk, 2010; Nodder, 2013).

Ainsi, selon les façons dont l'information s'affiche (*initial framing*), dont les options de consentement sont configurées et dont l'acceptation ou le refus sont pris en considération, la dimension contrainte du consentement se présente comme un continuum du consentement le plus volontaire au plus involontaire. Et ce continuum dépend notamment du pouvoir des organisations sur le design des interfaces, qui peut influencer voire façonner la prise de décision (Richards & Hartzog, 2019, p. 1488) – selon les façons de présenter les informations, les options et les modes d'interaction pour marquer son accord ou son refus. Cela touche à l'ambivalence des pratiques de design qui procèdent par ruses ou arrangements ingénieux avec les contraintes (Flusser, 2002, cité

dans Fisher, 2019) inhérentes aux règlements, ainsi qu'aux architectures des systèmes numériques et aux interfaces, comme abordé au chapitre 1 (p. 67). Cependant, quand ces arrangements ingénieux touchent au consentement, la ligne est mince entre des choix de design ou affordances qui *font consentir*, au risque d'influencer l'acte, et ceux ou celles qui permettent de consentir via des visuels et des interactions qui donnent autant la possibilité d'accepter que de refuser ou s'abstenir.

Ces pratiques délicates de design de consentement jouent avec une ambiguïté dans le contrôle donné à l'utilisatrice envers les informations qui la concernent : cliquer sur « accepter » ou « refuser » ou sélectionner certains critères, est-ce contrôler ses données ? Et en décidant comment les utilisateurs peuvent interagir pour signaler ces critères de consentement, par quelles affordances, n'est-ce pas contrôler en partie leur acte de consentir ? Pour favoriser la validité du consentement plutôt que seulement son obtention, le design qui se penche sur ce moment de consentement devrait se concentrer sur la médiation de la valeur de liberté par le travail des conditions – visuels, interactions, formulations, architecture des données – qui permettent de le donner. Mais comment faire ? La question mérite d'être déplacée pour moins centrer l'acte de consentir uniquement sur l'utilisatrice au contrôle très restreint. Il s'agit de comprendre le consentement comme un acte co-constitué par l'organisation qui le demande et par l'utilisatrice qui le donne. Cela résonne avec la dimension collective de l'acte de consentir que Human et Cech (2020) et Edenberg et Jones (2019) soulignent. La responsabilité du consentement est donc partagée dans la mesure où l'acte de consentir dépend des conditions – au niveau du design d'expérience utilisateur et des pratiques de traitement des données – qui sont mises en acte par l'organisation pour informer et exprimer ce consentement. Ces conditions de co-constitution peuvent se travailler de façon processuelle, amenant à comprendre le consentement comme quelque chose qui se construit de façon dynamique plutôt qu'un acte figé une fois pour toute.

Rendre au consentement sa dynamique

Cette dimension co-constituée du consentement est en effet également intéressante à voir dans une perspective processuelle qui est présente dans la mise en application des règlements de protection des données personnelles, en particulier pour la mise en place des conditions qui permettent aux utilisateurs d'exercer leurs droits. À défaut d'avoir un grand contrôle, ces droits donnent un certain pouvoir d'intervention dans le traitement des données à caractère personnel. Cela donne au consentement la possibilité d'être adapté : si l'on n'est pas d'accord, on peut s'opposer, si l'on est d'accord avec certaines finalités indiquées mais pas toutes, on peut ne pas consentir à tout. Les possibilités d'accéder, de rectifier, limiter, s'opposer permettent aussi de souligner que l'objet de ce à quoi on consent peut toujours être interrogé, vérifié, remis en question voire refusé. Cela accentue la dimension active du consentement, qui n'est pas figé à partir du moment où il est donné, mais qui peut être modifié voire retiré par l'utilisatrice selon les informations auxquelles elle a le droit d'accéder à propos des informations qui la concernent. Cette possibilité de retrait, qui est également présente dans la LPRPDE¹²⁶, est formulée ainsi dans les termes du RGPD :

La personne doit avoir la possibilité de retirer son consentement à tout moment, par le biais d'une modalité simple et équivalente à celle utilisée pour recueillir le consentement (par exemple, si le recueil s'est fait en ligne, il doit pouvoir être retiré en ligne également). (CNIL, 2018)¹²⁷

La simplicité de la modalité par laquelle retirer son consentement demande que cette possibilité de retrait ne soit pas dissimulée ou complexifiée à la manière d'un *dark pattern*.

Il est également précisé que la personne qui recueille des données à caractère personnel doit en tout temps pouvoir montrer la preuve du consentement : « le responsable du traitement doit être en mesure de démontrer à tout moment que la personne a bien consenti, dans des conditions valides »¹²⁸. Cela amène alors un certain

¹²⁶ Voir https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/collecte-de-renseignements-personnels/consentement/gl_omc_201805/#fn22

¹²⁷ Idem <https://www.cnil.fr/fr/les-bases-legales/consentement>

¹²⁸ Idem <https://www.cnil.fr/fr/les-bases-legales/consentement>

paradoxe que nous rencontrons, Paul et moi, lors d'une rencontre avec un cabinet de conseil juridique spécialisé en protection des données personnelles (Data Protection Consulting, DPC) pour un audit de conformité au RGPD chez TMTL (représenté ici par Jeanne). Après que Paul a expliqué l'intention de Prélude d'anonymiser les données recueillies par son assistant vocal pour mettre en pratique le principe de confidentialité par une approche de *privacy-by-design*, DPC précise le besoin de maintenir la possibilité d'identifier les utilisateurs dans certains cas :

1 DPC : [...] l'objectif en fait, quand on parle de consentement [...] ce n'est pas
2 seulement le fait de pouvoir enregistrer le consentement, donc l'approbation,
3 mais c'est surtout de la conserver, donc il faut que vous puissiez être en mesure
4 de prouver que vous avez collecté le consentement. Donc si effectivement, la
5 personne dit oui, il faut que vous ayez une trace de ce consentement. Parce que
6 si derrière la personne dit « ben non mais j'ai jamais dit oui », vous, vous êtes
7 censés justifier que vous avez enregistré correctement le consentement.

8 Jeanne : ça veut dire qu'il faut qu'elle soit identifiée cette personne-là !
9 [...]

10 DPC : en général, c'est le serpent qui se mord la queue : c'est comme quand on fait
11 une demande de suppression, en fait il faut conserver la demande de
12 suppression.

(1^{er} extrait de retranscription de la réunion du 11 septembre 2019 avec DPC)

Le retrait du consentement, comme la demande de suppression de données, implique ainsi de supprimer toute information conservée sur la personne qui fait la demande, mais de garder une trace qui puisse prouver que la suppression a bien été faite (l. 5), ce qui peut sembler paradoxal (« c'est le serpent qui se mord la queue », l. 10). Les organisations qui traitent des données personnelles doivent donc prévoir comment obtenir cette preuve qui respecte la confidentialité et comment conserver cette preuve¹²⁹. Les défis que ce droit de retrait du consentement implique font l'objet de nombreuses critiques dans la recherche académique et dans l'industrie, mais la difficulté que cela soulève ne

¹²⁹ DPC précise un peu plus tard qu'une solution est « d'affecter des numéros aléatoires aux gens, et vous pouvez du coup relier cet utilisateur au numéro d'identifiant et vous savez alors qu'elle a bien consenti », c'est-à-dire de générer une suite de nombres et caractères aléatoires qui servent d'identifiant. Les plateformes de gestion de consentement proposent en partie de gérer cela.

doit pas pour autant résigner à ne pas chercher des moyens de le rendre possible (Politou et al., 2018).

Le consentement est ainsi pris dans deux dynamiques qui font partie d'un même continuum : une fois construit, il peut être détruit (Kim, 2019). La preuve du consentement est une preuve de la validité de sa construction, tandis que l'effet de nouvelles informations auxquelles l'utilisatrice peut avoir accès peut entraîner sa destruction : « Consent construction *refers to the conditions under which the manifestation of consent was made. Consent destruction refers to the effect of new information or circumstances upon the conditions of consent* » (Kim, 2019, p. 9). Cette destruction peut être volontaire, dans le cas où l'utilisatrice demande le retrait de ses informations; ou elle peut se faire à son insu, dans le cas où certaines des conditions qui permettent la validité de son consentement ne sont pas respectées – comme c'est le cas par exemple pour certains *dark patterns* qui prennent pour acquis un accord qui n'a pas été donné ou qui a été refusé. Cette possibilité de retrait et ces droits que le RGPD donne aux personnes sur les informations qui les concernent participent à la nature « *incremental and variable* » du consentement (Kim, 2019, p. 8). Cette conception processuelle du consentement comme un acte jamais définitif, mais toujours potentiellement changeant dote le consentement d'une ambiguïté de plus : comment le saisir clairement s'il peut toujours changer ?

Cette question ne met pas l'accent au bon endroit : elle donne l'impression que le consentement peut être donné une fois pour toujours. Elle le fige dans une stabilité qui ne correspond pas à l'impermanence de sa dynamique qui évolue entre construction et destruction. Plutôt que de chercher à saisir ou obtenir le consentement comme un acte donné de façon fixe, abordons-le de façon continue, en portant régulièrement attention à son renouvellement. Cette dimension continue du consentement, qui appelle un renouvellement, est présente dans la LPRPDE¹³⁰ ainsi que dans les travaux théoriques sur le consentement dans ses différents usages, en particulier médical et sexuel (voir entre

¹³⁰ Ligne directrice 6 pour l'obtention d'un consentement valable : « Faire du consentement un processus dynamique et continu » https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privée/collecte-de-renseignements-personnels/consentement/gl_omc_201805/

autres Edenberg & Jones, 2019; Ferron Parayre, 2018; Garcia, 2021; Kim, 2019). L'expérience du consentement est donc une expérience d'impermanence au sens de Gunaratana (2002) : « *impermanence is the quality of experience that everything is shifting, going to pieces, slowly dissolving, rising and falling, and that moment-to-moment experience is all there is* » (cité dans Weick & Putnam, 2006, p. 279). Le travail de la clarté pragmatique amène ainsi un nouveau paradoxe : en amenant à construire ce qui permet au consentement de se manifester, cette clarté pragmatique donne encore plus à celui-ci la possibilité de finalement se retirer. Elle affirme son impermanence, qui peut donner l'impression d'une expérience ambiguë et incertaine car l'absence d'ambiguïté n'est jamais gagnée tout à fait.

Les incertitudes du consentement

Comme nous l'avons vu, malgré l'impératif de clarté voulu dans le consentement par lequel les utilisateurs auraient davantage de contrôle sur les informations qui les concernent et les façons dont elles sont traitées par les organisations qui les collectent, cette clarté s'expose à plusieurs limites : une limite de connaissance¹³¹, qui peut empêcher le consentement d'être véritablement informé, et une limite d'influence, qui peut amener le consentement à être contraint, en particulier par les choix de design qui permettent sa manifestation. Ces limites semblent rendre le consentement numérique impossible à être donné et obtenu de façon valable. Cela peut également amener un doute pour les organisations : faire tant d'efforts pour réinventer des modèles de consentement en vaut-il la peine et les coûts ?

¹³¹ Pour étoffer cette limite de la connaissance qui est difficile à évaluer, voir également dans Kim (2019) les nombreux biais qui interviennent dans une prise de décision telle que consentir dont *the bias of imaginability, overconfidence bias, confirmation bias, visceral factors, motivations other than self-interest* (p. 11-15).

Les efforts en valent-ils la peine ?

Ce doute peut s'accroître quand on constate que les informations reformulées ne sont pas davantage lues. Après avoir passé du temps et quelques réunions sur le design d'expérience utilisateur dédiée au consentement et son information, j'ai posé cette question à Malik, qui réalisait quelques tests utilisateurs d'une version préliminaire de l'assistant vocal développé par Prélude pour voir ce qui avait besoin d'amélioration :

Camille : Salut Malik, en faisant tester la nouvelle version de la démo, as-tu eu du feedback sur la phrase de consentement au début? (« *Before we go on, note that...* »). Pour avoir une idée de comment c'est reçu et si je devrais me dégager du temps la semaine prochaine pour le tester

Malik : Généralement les gens n'accrochent pas là-dessus. Souvent ils ne lisent pas et appuient sur « *let's continue* ». Quand j'ai demandé ce qu'ils pensaient des *privacy policies* et *confidentiality*, on m'a dit qu'ils s'attendent à lire ça parce « c'est la loi ». (...)

(Échange sur Slack retranscrit dans mon journal de bord, notes de terrain du 24 octobre 2019)

On a beau faire l'effort de travailler l'expérience utilisateur pour que l'information nécessaire au consentement soit davantage visible et compréhensible, cela ne semble pas changer les comportements fatigués de lire ce registre d'information légale – ou du moins pas pour les personnes rencontrées ce jour-là.

Kim (2019) souligne cette difficulté de changer les comportements par les *disclosures*, même si celles-ci sont formulées de façon plus compréhensible et agréable que les traditionnels *notice & consent*, mais elle refuse de s'en tenir à cette difficulté :

The fact that some disclosures may not change negative behavior for all groups of people does not mean that all disclosures are ineffective for all groups of people. But even assuming, for the sake of argument, that salient disclosure fails to change behavior (an assumption belied by the available literature), this does not mean that disclosure has failed or is irrelevant. The purpose of disclosure is not simply or solely to change behavior; the purpose of disclosure is to inform. Salient disclosures enhance knowledge. Knowledge is an essential component of consent. That salient disclosures may fail to alter behavior (at least in some situations) does not make them irrelevant to establishing consent. Consequently, where consent is required, effective disclosure is a necessity. (Kim, 2019, p. 129)

Quand on fait l'effort de réinventer les *disclosures* ou déclarations de ce qui est fait avec les données personnelles collectées, cela n'amène pas nécessairement les

comportements à changer pour faire en sorte qu'ils portent effectivement attention à l'information qui leur est communiquée dans la demande de consentement. En revanche, celles-ci sont nécessaires pour que l'un des piliers du consentement, sa dimension informée ou éclairée, soit adressé.

Mais cette nécessité ne doit pas être comprise seulement comme un intérêt légal de la part des organisations qui sont tenues de se protéger de toute sanction : « *Too often, disclosure has been interpreted as coverage of information, a way to protect against legal liability rather than as an effective means of communication* » (Kim, 2019, p. 125). En comprenant ces *disclosures* comme des moyens de communication, il s'agit en même temps de comprendre le consentement lui-même dans sa dimension de communication : ce qui se passe dans le moment de consentement, c'est un échange d'information et l'expression d'un accord entre une personne et une organisation qui nouent une relation d'utilisation (d'un système tel qu'un agent conversationnel). La qualité de cette communication joue un rôle important dans la validité du consentement qui ne tient pas qu'à l'acte de manifester son accord ou non, mais qui tient également aux *conditions* dans lesquelles cet acte est demandé et donné (Kim, 2019, p. 7). Et ces conditions ont une influence sur le comportement des personnes amenées à porter attention ou non à ce qui les informe.

Cet accent sur les conditions donne au comportement des utilisateurs un éclairage autre que l'apparente contradiction auquel le « *privacy paradox* » fait allusion. Cette expression désigne l'écart entre ce que les utilisateurs disent vouloir en matière de respect de leur vie privée et leurs comportements qui semblent plutôt ne pas se préoccuper du traitement des informations qui les concernent (Norberg et al., 2007) : « *[it is] the idea that consumer anxiety about privacy is undermined by the fact that consumers act in privacy-diminishing ways in practice* » (Richards & Hartzog, 2019, p. 1465). Cependant, quand on prend en considération l'effet des *nudges* et du design sur les décisions quotidiennes des utilisateurs, on peut décentrer le regard pour qu'il ne soit pas concentré sur les utilisateurs dans leur apparente autonomie de décision mais plutôt sur les micro-influences qui orientent ces décisions. On peut aussi prendre en compte la

surcharge mentale imposée par des notices d'information peu lisibles et compréhensibles, qui ont contribué à installer une habitude de ne pas les lire et une fatigue face à ce passage obligé. C'est le changement de regard que Richards et Hartzog (2019) poussent à faire en montrant les pathologies du consentement numérique, qui tiennent moins à la responsabilité des utilisateurs qu'à celle des organisations posant l'architecture des choix qu'il est possible de faire lors de l'utilisation de leurs services.

Pour pallier les limites des pathologies du consentement, Richards et Hartzog proposent un idéal qui consiste à réduire le nombre de demandes de consentement, pour que celui-ci soit réservé aux contextes qui présentent une sensibilité particulière, plutôt que d'entretenir une fatigue à l'égard des demandes démultipliées de consentement. Ils proposent également de permettre à l'utilisatrice de voir facilement les risques qu'implique le don de consentement et de mettre en place des incitatifs (ou *nudges* bienveillants) pour que les utilisateurs considèrent chaque demande de consentement sérieusement. Dans cette perspective incitatrice, Stark (2016) propose de prendre en compte la dimension émotionnelle dans le rapport à l'enjeu de respect de la vie privée et d'avoir recours à un concept de « *visceral design* » par des interfaces qui provoquent « *a reflective emotional response tied to the user's visceral sense of privacy* » (p. 23). Kim (2019) propose également de mieux permettre de comprendre les risques associés au don de consentement en contrant « l'émotionnalisme » qui fait envie, et qui peut prendre la forme d'un *nudge* plus ou moins sombre, par de l'émotionnalisme qui montre le revers de la médaille, afin d'équilibrer l'effet d'influence¹³².

Empêcher ou permettre : rassurer ?

Permettre de voir des risques, empêcher certains risques : qu'il soit médié par un règlement ou par des incitatifs au design bienveillant, ou qu'il soit exprimé par nous-même directement, le consentement s'inscrit ainsi dans un jeu (sérieux) qui consiste à

¹³² Kim (2019, p. 128) mitige cependant le recours au contre-émotionnalisme quand celui-ci peut jouer sur la vie de la personne, comme dans le cas d'une intervention thérapeutique nécessaire où montrer une image négative pourrait engendrer de la peur et exacerber une décision déraisonnable de refuser un soin.

poser des limites pour empêcher ou permettre l'accès à nos personnes – en chair et en os ou sous forme de données – qu'il s'agisse de faire des actes *sur* ou *avec* nous. Il implique également de porter attention à la manifestation de ces limites ou de ces autorisations d'accès, en nous-mêmes pour nous respecter, et chez les autres pour les respecter – cette attention peut être individuelle, émanant d'une certaine réflexivité, ou forcée par les règlements qui contraignent au respect des personnes et de leurs droits. Ce jeu avec les limites de ce que l'on empêche ou permet – ou que les règlements imposent pour nous respecter – est au cœur de l'exercice des libertés individuelles, qui comptent comme des valeurs fondamentales dans les sociétés libérales contemporaines. Il s'agit de marquer les limites dans lesquelles rester pour assurer le respect, de restreindre certaines libertés pour ne pas abuser des autres, d'orienter l'attention pour favoriser ce respect.

Ce jeu de limite, restriction et orientation peut se comprendre comme ce qui distingue deux concepts de liberté, selon Isaiah Berlin (1969) : la liberté négative et la liberté positive. La liberté négative se formule en anglais par l'expression « *freedom from* », qui implique l'absence d'obstacle pour réaliser quelque chose. Il s'agit d'être libre de toute forme d'abus de pouvoir, de contrainte, d'interférence de la part des autres (individus ou collectifs) ou encore d'intimidation (Shklar, 1989). La liberté positive, « *freedom to* », libre de faire quelque chose, implique la présence de quelque chose telle que le contrôle, qu'il soit interne – comme la réalisation de soi, la maîtrise de soi, l'auto-détermination – ou externe – comme le contrôle d'un État ou d'une entité qui aurait de l'autorité sur nous. Le contrôle voulu dans le consentement relève de cette liberté positive, entendue comme un contrôle interne de l'ordre de l'auto-détermination ou volonté autonome, de la réalisation de soi et d'une certaine maîtrise de ce qui a le droit ou non d'être exercé sur ou avec nous. Mais l'ambivalence que nous avons vue précédemment, avec le design qui oriente les décisions et peut influencer le consentement, manifeste également la présence d'un autre contrôle, plus externe, propre aux organisations qui arrangent l'architecture des choix dans les technologies pour lesquelles on nous demande notre consentement.

Cette articulation entre absence et présence d'interférences (Carter, 2021) est souvent utilisée en philosophie politique – qui s'intéresse au rapport entre individus et ce qui les gouverne – pour critiquer la liberté positive, comprise comme le pouvoir interventionniste des États dans les vies individuelles. La liberté positive est utilisée dans ce contexte pour répondre à la question : « *What, or who, is the source of control or interference that can determine someone to do, or be, this rather than that?* » (Berlin, 1969, p. 121-122). En analysant les rapports de pouvoir par cette question, le problème de l'exercice d'une liberté positive par des personnes qui gouvernent est qu'elle porte le risque de justifier de l'oppression. Dans le contexte technologique qui m'intéresse, le *soft paternalism* des *nudges* peut être vu comme un interventionnisme de certaines organisations, par le design, dans les vies individuelles dans lesquelles les technologies qui ont ces *nudges* s'immiscent, voire fusionnent. Le sentiment ambivalent à l'égard de ces *nudges*, qu'ils soient bienveillants ou malveillants, réside dans ce risque de justifier l'emploi d'une certaine contrainte pour orienter notre consentement, ce qui peut être porteur d'une soumission pouvant aller jusqu'à l'oppression.

Cette vigilance envers ce qui peut être facteur d'oppression est centrale dans le concept de consentement pensé comme l'exercice de notre autonomie, et c'est aussi ce qui suscite de forts attachements au consentement, son mécanisme et ses effets, dans des sociétés libérales où les libertés individuelles sont considérées comme des valeurs fondamentales. Par les limites qu'il exprime pour se faire respecter, par les restrictions de libertés qu'il implique pour éviter des abus, par l'orientation de l'attention pour favoriser ce respect, le consentement vise la protection de soi. Cette protection de soi s'accompagne dans la sphère numérique des activités s'affairant à protéger la vie privée. Ces activités s'articulent selon une logique visant la liberté négative des personnes amenées à utiliser la technologie à laquelle il est demandé de consentir : cette logique cherche à protéger la liberté individuelle dans un espace exempt de certaines interférences, ou du moins exempt d'interférences néfastes pour l'individu.

Lors de mon entrée dans l'approche du *privacy-by-design*, quand Paul me transmet une version du livre blanc que des consultants en protection des données

personnelles préparent pour Mr Y, je suis surprise par l'abondance du mot « *threat* » (menace) : le livre blanc cartographie les menaces potentielles auxquelles Mr Y (chatbot et organisation) risque d'être exposé. Un sentiment de crainte s'éveille en moi : entrer dans ces sujets demande d'aiguiser la vigilance à toute potentielle invasion sur plusieurs fronts, celui de la vie privée des utilisateurs, celui des activités de Mr Y à l'égard des données traitées, de l'architecture par laquelle ses systèmes (techniques et managériaux) sont pensés, de la sécurité à assurer. À cela s'ajoute la crainte d'enfreindre certaines règles de protection des données personnelles selon le RGDP et la LPRPDE, et ainsi de ne pas respecter certains droits des personnes amenées à utiliser les technologies de Mr Y et Prélude, qui s'en trouveraient éventuellement sanctionnés. Le travail par lequel demander le consentement et protéger la vie privée des utilisateurs est ainsi habité d'une réflexion sur la liberté positive – qui cherche à identifier ce qui peut être la source d'interférences qui peuvent nuire à autrui – accompagné d'un sentiment craintif, au nom de la protection d'autrui. Dans le design de conversation, les *prompts* introduisant la demande de consentement se sont teintés de mots tels que « *don't worry* », pour rassurer l'utilisatrice en désamorçant la crainte d'une invasion dans sa vie privée, ou d'un usage malveillant des données qu'elle s'apprêterait à partager. Rassurer, désamorcer la crainte, est-ce protéger ?

Faisons un détour par un autre contexte, politique, pour prendre du recul sur cette question. Dans son article « *Liberalism of fear* », Judith Shklar (1989) propose un libéralisme minimal qui ne vise pas la perfection morale des êtres humains dans la tradition kantienne, mais qui prend pour point de départ un principe affectif – la peur – qui influence l'exercice des libertés. Sa réflexion est marquée par l'historique de formes de terreurs telles que le nazisme qui ont amené la cruauté¹³³ à régner par la peur, puissant facteur d'oppression quand elle est utilisée comme forme de contrôle social¹³⁴. Plutôt que

¹³³ Définition de la cruauté dont le libéralisme de la peur veut prévenir : « *It is the deliberate infliction of physical, and secondarily emotional, pain upon a weaker person or group by stronger ones in order to achieve some end, tangible or intangible, of the latter* ». (Shklar, 1989, p. 29)

¹³⁴ Je révisé ces lignes fin mars-début avril 2022, où la guerre en Ukraine provoquée par la Russie ravive cette réflexion : la peur règne en Russie où l'opposition est impossible si l'on veut vivre, et elle s'étend en Ukraine où des millions de personnes ont fui pour pouvoir continuer à vivre plutôt que de risquer d'être

de permettre la liberté positive de certains de faire peur aux autres, Shklar s'appuie sur une conception négative de la liberté : une liberté de se protéger de la peur que quiconque porte atteinte à notre vie. Cette protection de la peur est forte de responsabilité, en particulier quand elle est exercée par quelqu'un d'autre que nous-même. En effet, si la peur est effacée par des pratiques qui visent à nous rassurer pour pouvoir continuer d'agir, y compris dans une direction malveillante, l'activité de rassurer est instrumentalisée pour une fin contraire à un impératif d'humanité. En revanche, si la peur est prise en charge émotionnellement par des personnes qui s'occupent effectivement de protéger autrui de la peur que quiconque porte atteinte à sa vie (privée), alors le respect a plus de chance d'être assuré. Rassurer est donc un exercice éthique à forte responsabilité, habité par un sentiment intranquille, car c'est un instrument politique qui peut amener à amoindrir la vigilance quant à ce qui peut porter atteinte à nos vies dans leur liberté.

Le libéralisme de la peur se développe par une posture particulière : il ne s'agit pas d'atteindre un *summum bonum* mais d'éviter le *summum malum* de la cruauté (Shklar, 1989, p. 29). Cette articulation entre ce que l'exercice de la liberté permet d'atteindre ou d'éviter est intéressante à reprendre pour une réflexion sur le consentement : le consentement permet-il d'atteindre quelque chose (un objectif, un idéal) ou d'éviter autre chose (un risque, le pire) ? Avec le registre de la protection de la vie privée (Cavoukian, 2011; Cavoukian & Tapscott, 1997; Nissenbaum, 2010), nous avons vu que le consentement s'inscrit dans une logique qui cherche à éviter les risques qui peuvent nous nuire. Avec l'ambivalence de l'intéressement (Callon et al., 1988; Giroux, 2006), le consentement peut viser d'être obtenu ou donné dans l'objectif d'utiliser un service technologique. Avec l'impératif catégorique (Garcia, 2021; Kant, 1993), le consentement vise l'idéal de l'exercice de notre volonté autonome, qui pose les limites par lesquelles se respecter et respecter autrui. Par cet exercice du consentement, peut-on véritablement atteindre cet idéal et éviter tout risque ? Pour cela, il faudrait communiquer clairement

tué. Cette peur annihile la liberté (négative) de vivre sa vie sans que ceux qui nous gouvernent, ou ceux qui tentent d'imposer de nous gouverner, nous entravent.

sur les risques qu'implique de consentir. Cependant, la communication claire sur les risques qu'implique le consentement au traitement des données se heurte à une difficulté : l'incertitude de ces traitements et de leurs risques.

Consentir à quelque chose qui nous dépasse ?

Comme le soulignent Jones et Edenberg, les personnes qui utilisent des systèmes d'intelligence artificielle sont souvent dans l'incapacité d'évaluer les risques et de prévoir les résultats avec certitude :

Th[e] ubiquity, complexity, and opaqueness [of AI systems and their arrangements of data] contribute to users' inability to assess and predict outcomes and harms. Part of the claimed benefit of AI systems is that they produce insights and predictions that humans may or can not. As such, the rationality of choosing a service that touts an improvement to some aspect of one's life is clear when measured against unpredictable and incomprehensible long term future harms. (Jones & Edenberg, 2020, p. 362)

Cela est particulièrement sensible dans les cas où des systèmes d'intelligence artificielle sont utilisés pour aider des décisions de justice, faire de la médecine prédictive, assurer la sécurité d'une ville par le maillage de capteurs qui entraînent un système de surveillance. C'est également sensible dans le cas d'un système d'IA en santé mentale qui serait utilisé par des personnes en détresse. Par le fait qu'il est impossible de prévoir des effets indésirables causés par un système dont le fonctionnement échappe en partie à la compréhension, le consentement peut être attiré plutôt par ce qui est plus positif à court terme, selon un biais optimiste (Kim, 2019) : ce système a une efficacité telle que son utilisation permet d'améliorer quelque chose dans un temps rapproché.

Ce qui m'intéresse d'autant plus dans cette citation, c'est que l'incapacité d'évaluation et de prédiction des risques concerne les utilisateurs finaux, mais aussi les personnes qui développent un système en utilisant plusieurs services d'intelligence artificielle – comme c'est le cas pour Prélude qui utilise les services d'IA conversationnelle de Google Cloud Platform, comme nous l'avons vu précédemment (chapitre 2, p. 77). Ces personnes, comme les membres de Prélude, ne savent pas comment les outils d'IA qu'elles utilisent fonctionnent précisément, ni les effets négatifs qu'ils peuvent potentiellement avoir à plus ou moins long terme – comme nous le verrons dans le

chapitre 6, un membre de Prélude mentionne plusieurs fois ne pas pouvoir connaître le « secret » de Google pour développer ses algorithmes d'ASR. En revanche, elles ont accès à plusieurs preuves de son efficacité à court terme, pour relever des défis techniques que l'organisation n'arrive pas à relever par ses ressources limitées (en termes de personnes aux compétences élevées qui maîtrisent des techniques complexes, de temps pour faire de la recherche et du développement, d'argent pour payer des accès à de grandes bases de données). La perspective de rapidité avec laquelle son assistant vocal peut être développé grâce aux composantes d'IA disponibles en ligne oriente le choix d'utiliser ces services de *cloud platform* plutôt que de se laisser freiner par l'incertitude d'effets potentiellement indésirables mais pour l'instant non avérés. Or par ce choix, le consentement demandé à ses utilisateurs finaux est teinté de cette incertitude quant aux outils par lesquels l'assistant vocal est développé.

Cette limite que présentent les incertitudes que l'on peut connaître d'avance, couplée aux limites de nos connaissances et de notre compréhension de la complexité de l'écosystème numérique, mais aussi des imbrications par lesquelles transitent les données personnelles, amène une question : consentir à ce traitement, n'est-ce finalement pas se résigner face à cette complexité et à une organisation de l'écosystème qui nous dépasse ? Que vaut un consentement par résignation ? La passivité inhérente à la résignation touche à une ambiguïté dans le terme « d'accord » dont le consentement est synonyme, comme le souligne Manon Garcia (2021)¹³⁵ : l'accord, comme le consentement, peut relever d'un choix pleinement positif ou alors d'une acceptation, voire peut-être d'une forme de résignation. S'il est acceptation, dans ce cas le consentement a quelque chose de passif, la personne peut avoir été amenée à accepter, dans le sens de ne pas refuser, voire céder (Leguil, 2021). Même dans le cas d'un choix, celui-ci peut-être plus ou moins libre, plus ou moins posé de façon active et affirmative.

¹³⁵ Son travail sur le concept de consentement porte principalement sur le consentement sexuel mais son raisonnement philosophique pour forger ce concept, c'est-à-dire pour déceler une logique qui rassemble les ambiguïtés de cette notion, est très éclairant pour penser les nuances du consentement dans d'autres contextes qui posent de différentes façons la question : comment donner accès à soi en se faisant respecter ?

Ces nuances sur l'affirmation du consentement font une différence sur le plan moral quand on se demande ce qui fait qu'une expérience de consentement est bonne, qu'un consentement est valable, pas seulement sur le plan de la formalité (qui demande si le consentement est formellement donné), mais sur un plan moral qui vise une expérience d'autonomie, qui demande que le consentement soit donné de façon libre et volontaire.

Ces nuances, qui floutent le caractère actif du consentement pour en voir la passivité, pointent vers la soumission structurelle que Manon Garcia (2018, 2021) met de l'avant pour penser le consentement. La perspective féministe de sa démarche philosophique met l'accent sur la domination masculine qui, en tant que structure sociale, impacte les conditions de possibilité du consentement des femmes, de sorte que celui-ci tend à s'inscrire dans une soumission structurelle. Cette dimension de la soumission structurelle est intéressante à transposer dans un contexte technologique en matière de consentement, quand les outils et services utilisés pour créer un système d'IA ont leurs propres pratiques de gestion des données personnelles auxquelles consentir. Le mode agile de développement technologique, qui s'appuie sur des outils performants accessibles, comme des services auxquels souscrire via les *Platforms as a Service* (PaaS), implique par exemple de se soumettre aux façons dont sont traitées les données qui transitent par ces services, même si ces façons sont opaques pour la personne qui fait appel à ces services. En optant pour l'utilisation de plusieurs services de Google Cloud Platform et de l'Assistant Google pour développer son agent conversationnel, Prélude s'inscrit ainsi dans une soumission structurelle à Google et ses services et outils, même si cette soumission n'est pas reconnue comme telle dès le début (voir chapitre 6). La façon dont ces outils sont imbriqués les uns aux autres, amène également à entrevoir une domination sous-jacente où les organisations qui mettent à disposition des outils performants via le *cloud* imposent à la fois leurs pratiques de gestion des données personnelles, et posent des garde-fous aux organisations qui les utilisent, de sorte qu'elles ne puissent pas collecter toutes les données.

Un exemple de garde-fou se présente quand on installe une Action sur Google Assistant : à la manière d'un *consent wall*, une fenêtre s'affiche en avant-plan pour

proposer d'empêcher l'Action en question de mémoriser nos informations (figure 6 ci-dessous).

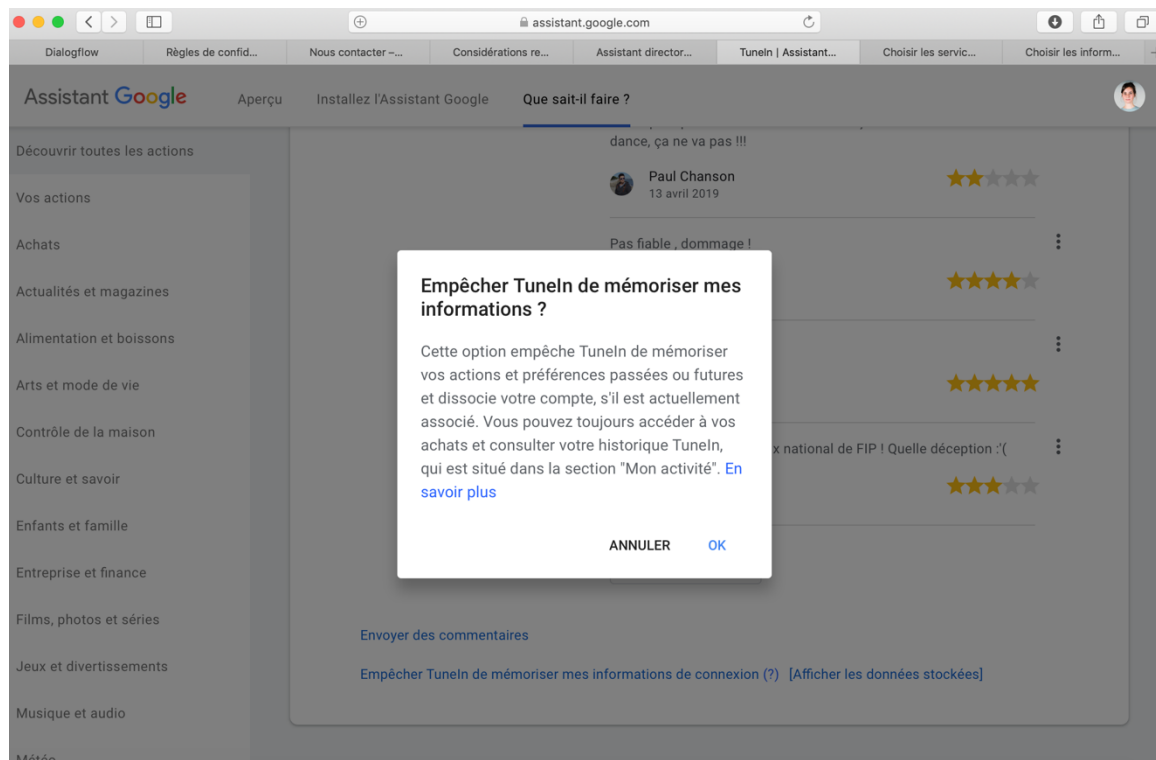


Figure 6 : Capture d'écran d'une fenêtre permettant d'empêcher qu'une Action mémorise mes informations. Prise le 22 octobre 2019, 17h41.

C'est une façon de limiter la quantité de données personnelles collectées par l'organisation qui développe l'Action; en revanche, cette limitation ne s'applique pas du côté de Google qui traite ces données pour permettre l'utilisation de l'Action via son Assistant. La métaphore du garde-fou évoque le même phénomène que la création d'une interférence de la part de Google pour assurer la protection et le traitement de données personnelles selon ses termes – une interférence de l'ordre d'une liberté positive pour donner un sentiment de contrôle aux utilisateurs qui restreignent ainsi à certaines organisations l'accès à leurs données personnelles, mais qui positionne Google comme source de contrôle sur ces données.

Quand le consentement est demandé à une utilisatrice finale de l'agent conversationnel, il s'agit de consentir aux pratiques du système d'IA qui implique indirectement d'accepter les termes de chaque outil ou service utilisé pour le créer et par

lequel transitent des données personnelles. Cet empilement de consentements amène à se demander si celui-ci n'est pas noyé. Le consentement demandé à l'utilisatrice via le système d'IA final se trouve dépendant de la cascade de consentements impliqués auparavant mais qui ne lui parviennent pas facilement, car leur enchevêtrement est si complexe que même les développeurs de ces systèmes n'en ont pas toujours de vision claire et les risques ne sont pas évidents à prévoir. Le consentement final est donc plutôt de l'ordre d'une soumission à cet ordre établi, il sonne comme une acceptation, voire une résignation à une structure sociale *data-driven*¹³⁶ où l'on ne peut pas faire grand-chose. C'est également à la lumière de cette soumission, qui prend la forme d'une acceptation ou résignation, que peut se voir le *privacy paradox* évoqué plus haut (p. 159), selon lequel les personnes se disent concernées par leur vie privée mais ne changent pas pour autant leurs comportements face à la protection de leurs données personnelles.

Dans une situation de soumission ou subordination à des rapports socio-économiques propres aux sociétés libérales *data-driven*, les personnes qui utilisent des technologies ne semblent pas pouvoir faire grand-chose d'autre que consentir à ces ordres pour pouvoir faire partie de la société¹³⁷. Alors pourquoi encore tenir au consentement ? Quel horizon éthique dessine-t-il pour valoir la peine que l'on tienne à pouvoir exercer notre consentement, et que celui-ci soit valide ? Le consentement peut-il permettre de participer (démocratiquement) aux sociétés *data-driven* ou n'est-il qu'un outil formel qui masque une impuissance inhérente à la soumission ?

Ces questions traduisent un idéal de consentement affranchi de rapports de domination. Mais rares sont ces conditions idéales, indépendantes de toute domination ou subordination; cela rend le consentement impossible, s'il est attendu qu'il soit parfait. Trop de choses échappent à notre connaissance, et trop de structures entraînent un

¹³⁶ Cette expression fait référence aux liens complexes entre les données et l'activité quotidienne dans le monde, des liens qui façonnent les sociétés où l'utilisation des technologies est massive dans la vie quotidienne. Voir par exemple Pentland (2013).

¹³⁷ Ce constat est inspiré de Manon Garcia (2019 et 2021) qui met de l'avant notre soumission au patriarcat pour faire partie de la société car c'est ce mode qui organise structurellement nos sociétés occidentales. Sa réflexion sur la soumission est une contribution importante pour penser le pouvoir et analyser, dans la richesse de ses ambiguïtés, la capacité d'agir inhérente à la soumission.

rapport de subordination. Alors ce que l'on peut faire, c'est prendre cette imperfection comme point de départ, essayer d'améliorer les pratiques et les informations pour qu'il y ait moins de regrets dans les décisions que l'on prend, les accords que l'on donne, les choix que l'on fait (Kim, 2019). On cherche alors à faire au mieux pour se rapprocher d'un consentement valable, faire ce que l'on peut pour que les conditions dans lesquelles on le demande permettent sa validité, dans les limites de ce que l'on peut faire bouger pour permettre cette validité. On ne peut pas faire bouger facilement les dominations structurelles qui mettent dans des rapports de subordination ou de soumission¹³⁸, mais on peut faire bouger des pratiques pour favoriser une certaine vigilance, porter différemment son attention sur cet enjeu, qu'il devienne une préoccupation puis l'objet d'un soin particulier (voir chapitre 4). On peut également faire bouger les lignes de nos conversations sur le consentement, pour moins mettre l'accent sur son exercice désagréable qui interrompt nos expériences pour être respecté et davantage le penser comme un moment de conversation qui vise un certain épanouissement.

Parler de consentement

Le consentement est ainsi un concept qui se constitue par son ambiguïté de vouloir être donné par un certain contrôle, signe de son autonomie voire de sa liberté. Pourtant, les conditions dans lesquelles il est demandé et obtenu – que ces conditions soient de l'ordre du design ou d'un ordre plus structurel marqué par des rapports de domination et soumission – sont autant d'interférences dans cette liberté. Le consentement entretient également un rapport ambigu à l'information sur laquelle il est censé se fonder pour être demandé et donné : beaucoup d'informations échappent à notre compréhension en tant qu'utilisateurs et développeurs de ces systèmes, si bien que chercher à informer les utilisateurs amène les développeurs à s'informer eux-mêmes sur ce qu'impliquent leurs

¹³⁸ Les *whistleblowers* et les personnes militantes s'attachent à faire bouger ces dominations structurelles, adoptant une posture de lutte de l'ordre du contre-pouvoir qui les expose à une certaine vulnérabilité, comme abordé au chapitre 1 (p. 27).

pratiques en matière de traitement des données personnelles. Chez Prélude, le design des interactions de consentement amène ce travail d'information à se faire auprès des membres de l'équipe, comme nous le verrons au chapitre 6. Ces ambiguïtés dans le travail du consentement s'accompagnent d'ambiguïtés dans son expression : reconnaître ses limites et marquer son désaccord sont des actes souvent flous et inconfortables. Pourquoi et comment faire bouger cette expérience inconfortable du consentement ?

Parler autrement de l'inconfortable

Si les agents conversationnels visent à imiter comment les personnes humaines s'y prennent pour manifester leur consentement dans des conversations, alors ces technologies devront faire preuve d'une bonne écoute et analyse pour identifier la multiplicité des façons par lesquelles on exprime son accord et son refus. Les règlements de protection des données personnelles appellent à une manifestation explicite du consentement, pour que l'accord soit clairement donné. Cela entraîne la mise en place des choix de design et des fonctionnalités pour permettre cette expression explicite, pour donner la possibilité de refuser ou de continuer l'expérience en ligne sans accepter. Mais ces options de refus sont réductrices par rapport à la diversité des façons dont le refus s'exprime dans les conversations.

Dans le contexte du consentement sexuel, Kitzyger et Fritz (1999) ont montré, par l'analyse de conversation, à quel point les refus – y compris hors du champ sexuel – sont des interactions conversationnelles complexes qui ne peuvent se réduire à « *just say no* » (titre de leur article) car elles impliquent souvent des « *silences, compliments, or even weak acceptances* » (p. 296), ainsi que des hésitations, des faux départs, des auto-corrections (p. 299). Leur analyse montre que juste dire non est difficile et inconfortable puisque « *that is not how refusals are normatively done* » (p. 302). Les appels à « *just say no* » visent à simplifier les expressions de refus par un « non » explicite de la part des personnes à qui il est demandé de dire oui ou non – ces personnes étant souvent des femmes quand il s'agit de consentement sexuel. Kitzyger et Fritz invitent plutôt à ne pas tant se concentrer à faire changer ces façons par une expression simplifiée, comme s'il

fallait surtout travailler à se rendre plus compréhensibles pour ceux qui veulent faire des actes avec ou sur elles. Dire non n'est pas qu'une affaire d'éducation des femmes. Il s'agit plutôt d'affiner l'écoute nuancée des actes de refus dans nos conversations quotidiennes, puisque « *communicating a refusal is a far more elegantly crafted interactional activity than we might have imagined* » (p. 302). Plutôt que de faire de l'expression simpliste du refus une norme, elles invitent à écouter la normativité habituelle et ambiguë du refus. C'est cette écoute que les agents conversationnels gagneraient à imiter, dans tous les contextes et pour toutes les requêtes qui impliquent un accord ou la possibilité d'un refus.

Presque vingt ans après la publication de cet article, cette écoute ne s'est manifestement pas développée : le mouvement #MeToo/#MoiAussi qui a explosé en 2017 a rendu visible l'envergure de l'incompétence, ou de l'absence de volonté, à reconnaître et respecter des refus, ce qui amène autant de violations de consentement, sur tous les continents, dans toutes les professions, dans l'intime comme dans l'espace public. L'expression explicite du consentement revient sur le devant de la scène, sous la forme d'un oui ou non à expliciter, mais surtout à respecter clairement dans le flou du moment. Cette explicitation de consentement n'est cependant pas prise pour acquis, elle fait l'objet de diverses initiatives pédagogiques¹³⁹ qui visent à éduquer sur ce moment interactionnel qu'est le consentement. Cette pédagogie passe par une valorisation des rapports d'égalité plutôt que de domination. Cela influence les réflexions comme la mienne, dans d'autres sphères que le sexuel, pour penser cet acte par lequel on manifeste un accord ou un refus, au nom du droit que l'on a de se faire respecter, malgré les imperfections des situations dans lesquelles on exprime ce consentement. Ce droit demande un certain entraînement pour s'affirmer comme pouvoir en chacune de nous. Il demande aussi à ceux qui sont dans la position de demander un consentement d'assumer leur responsabilité pour prendre soin des conditions dans lesquelles ils le demandent, de l'écouter et de s'assurer de le respecter.

¹³⁹ Par le biais de l'excellente série *Sex Education* (Nunn, 2019-), le travail de photographes comme Charlotte Abramow (<https://www.charlotteabramow.com/richard-taittinger-gallery/>), entre autres.

Cet entraînement passe par l'amélioration des pratiques de consentement plutôt que d'en perpétuer l'expérience d'inconfort, d'évitement, d'ennui ou de fatigue habituelles. Manon Garcia (2021) reprend l'expression « érotiser l'égalité » :

[S]i l'enjeu est de lutter contre les injustices de genre tout en préservant l'autonomie sexuelle de toutes et tous, s'il s'agit en somme, pour reprendre la formule de Gloria Steinem, d'« érotiser l'égalité » plutôt que la domination, alors le consentement sexuel, conçu comme conversation érotique, est sans doute l'avenir de l'amour et du sexe. (p. 223)

Le registre érotique et sexuel n'est pas tout à fait approprié à mon contexte – bien que les interactions d'ordre sexuel soient courantes avec les agents conversationnels, comme nous l'avons vu dans le chapitre précédent (p. 108) –, mais l'intention est la même : il s'agit d'améliorer les pratiques pour que le plaisir se concentre sur l'égalité plutôt qu'il ne laisse dire ses mots sous le coup de la domination. L'agréable ne doit pas forcément se soumettre à l'habitude d'expériences conversationnelles qui se débarrassent rapidement du moment de consentement pour se consacrer à l'expérience principale. L'agréable peut être par exemple d'avoir la possibilité d'aborder, dans la discussion, les différentes façons par lesquelles des choses sont mises en place pour sonder nos choix et les respecter.

Cet entraînement par la conversation, c'est ce en faveur de quoi Yolande Strengers et ses collègues (2021) plaident avec les technologies émergentes telles que les agents conversationnels et robots sexuels. Elles proposent d'intervenir dans le développement de ces technologies émergentes pour que leurs *embodied interactions* soient attentives aux façons de demander et obtenir le consentement des utilisateurs. Dans la lignée du mouvement de *Consentful Tech*¹⁴⁰ (Allied Media Projects, 2017), leur article « *What Can HCI Learn From Sexual Consent?* » propose de suivre le modèle de consentement explicite du *Planned Parenthood* américain FRIES : « *freely given (F), reversible (R), informed (I), enthusiastic (E), and specific (S)* » (Allied Media Projects, 2017, p. 8-9; Strengers et al., 2021, p. 2). Ce modèle permet d'envisager une expérience de consentement qui ne soit pas « *an annoying or intrusive thing that we have to get out of the way before doing anything interesting or fun* » mais « *agency affirming and freedom enhancing, not purely*

¹⁴⁰ Voir <https://www.consentfultech.io>

restrictive » (Strengers et al., 2021, p. 4). À partir du modèle FRIES, Strengers et ses collègues proposent un modèle de design d'interaction qui touche au design de conversation : TEASE, acronyme de « *traffic lights (T), establish ongoing dialogue (E), Aftercare (A), Safewords (S), and Explicate soft/hard limits (E)* ». Ce modèle est inspiré des pratiques de consentement qui sont utilisées dans les communautés BDSM. Le modèle s'applique à des interactions avec des agents conversationnels et des robots sexuels, mais aussi à des champs d'interactions au-delà des pratiques sexuelles comme les « *bodily-play interactions* » et le « *persuasive design with smart technologies* ». Les *bodily-play interactions* placent le corps au centre des interactions, comme le font les systèmes de réalité virtuelle, ou ceux qui envoient des feedbacks haptiques (dans des manettes de console de jeu vidéo par exemple) ou encore ceux qui traquent le mouvement (les bracelets connectés ou les applications qui comptent le nombre de pas faits en une journée sont des exemples). Le design persuasif est un type de design qui a recours à des *nudges* pour inciter certains comportements, voire certains modes de consommation. Si mon projet de thèse avait pour objectif initial de développer un modèle de consentement avec des agents conversationnels qui aurait pu être similaire, mon expérience avec Mr Y et Prélude lui a fait prendre une autre tournure. Alors je ne reprends ni ne développe un modèle similaire de consentement, mais je puise de l'inspiration dans ce mouvement¹⁴¹ pour inviter à prendre soin du consentement quand on développe des technologies qui jouent sur des relations de type *embodied* et d'altérité, comme le font les agents conversationnels.

Dans ces exemples et inspirations où le consentement fait l'objet de conversations, c'est l'usage sexuel du consentement qui est mis de l'avant. Celui-ci rend le consentement particulièrement important car le rapport sexuel est un moment de grande vulnérabilité : si celle-ci n'est pas respectée, la soumission laisse des failles d'une profondeur qui reste souvent indicible pendant longtemps chez la personne dont le consentement est violé (voir chapitre 4). Mais il serait réducteur de cantonner cette

¹⁴¹ Ce mouvement va dans le sens de celui de *Design Justice* (Costanza-Chock, 2020).

vulnérabilité au champ intime du sexuel; celle-ci est politique, inhérente à ce qui nous lie les uns les autres au quotidien.

Un exercice de vulnérabilité

Dès le début de ce chapitre, en soulignant la *moral magic* du consentement qui transforme en acceptables des actes qui seraient sinon de l'ordre de violations inacceptables, la vulnérabilité inhérente au consentement était là sans dire son nom. C'est une vulnérabilité qui se cache dans nos sociétés libérales, comme le montre la philosophe Martha Nussbaum dans son livre *Hiding from Humanity* (2006) : « *A salient fact about the human being, from the point of view of liberalism, is its vulnerability to significant damage at the hands of others* » (p. 345). Dans la ligne de pensée de Nussbaum, la vulnérabilité se dessine comme une condition humaine liée à l'absence de contrôle parfait que les êtres humains ont dans et sur leur vie, alors même qu'ils sont intelligents et pleins de ressources :

Human beings are deeply troubled about being human—about being highly intelligent and resourceful, on the one hand, but weak and vulnerable, helpless against death, on the other. We are ashamed of this awkward condition and, in manifold ways, we try to hide from it.

In the case of shame, a more general anxiety about helplessness and lack of control inspires the pursuit of invulnerability (or the retrieval of that illusion, which was very likely present in infancy). (2006, p. 336)

Sans être toujours une question de vie ou de mort, le consentement touche à ce trouble. Son inconfort, la tendance à préférer ne pas s'attarder dessus, voire à l'éviter, la revendication de contrôle qu'il doit signifier : autant de gestes qui cachent implicitement ce moment, ou qui tentent d'en consolider le sentiment de contrôle. En entretenant ces évitements et ce sentiment de contrôle, on continue d'apprendre

to reject [our] own vulnerability and to suppress awareness of the attachments that entail it. We might also say (...) that [we] are the only animals for whom neediness is a source of shame, and who take pride in themselves to the extent to which they have allegedly gotten clear of vulnerability. (Nussbaum, 2003, p. 147)

Reconnaître la vulnérabilité du consentement, c'est reconnaître avoir besoin de l'autre au quotidien. Le consentement, parce qu'il est un choix, la manifestation de la volonté, une

indication de préférence, s'inscrit dans une ouverture au monde et aux autres souvent tentée par sa fermeture pour se contrôler. Car cette ouverture est une fragilisation de soi-même : on consent pour ne pas être blessé, mais la possibilité de la blessure est là; c'est pour se prévenir de cette possibilité que le consentement se pose. Consentir, c'est-à-dire donner son accord pour s'ouvrir à l'autre, s'accompagne d'une « vulnérabilité qui implique la confrontation, les empêchements, les modifications du désir ou de la volonté initiale » (Brugère, 2013). S'il y a autonomie dans le consentement, elle est ainsi fondamentalement relationnelle¹⁴² car elle s'ancre dans cette vulnérabilité, cette ouverture aux autres et les possibilités d'actions ensemble.

À la lumière de cette vulnérabilité par laquelle on reconnaît l'absence de contrôle parfait et notre besoin de l'autre, on peut envisager un léger changement de regard sur les rapports de domination et de soumission dans lesquels le consentement est pris : ces relations peuvent être vues à la lumière de l'interdépendance qui nous lie dans nos actes quotidiens. Mais ce changement de regard s'accompagne de capacités à développer, dont celles que nous avons vues précédemment, qui consistent tant à exprimer qu'à écouter le consentement dans l'ambiguïté des actes d'accord et de refus. Dans l'exercice de la vulnérabilité inhérente au consentement, une autre capacité consiste à reconnaître qu'il y a des interférences qui l'orientent, le permettent ou l'influencent. Mais ces interférences sont plutôt des marques d'interdépendance : un consentement ne peut pas être donné hors de conditions qui le sondent et le demandent. Reconnaître cette interdépendance demande alors la capacité d'assumer la responsabilité que l'on a quand on demande un consentement, c'est-à-dire quand on participe aux conditions dans lesquelles il peut être donné.

Cette notion de capacité est au cœur de l'approche des *capabilités* que Nussbaum a développé avec l'économiste Amartya Sen (Nussbaum, 2013; Sen, 1995). Cette approche permet de renouveler la perspective de la liberté – ni positive, ni négative – en l'orientant vers l'accomplissement (Brugère, 2013). Le but de cette approche est de :

¹⁴² Cette notion d'autonomie relationnelle, ancrée dans les approches féministes de l'autonomie, serait passionnante à explorer davantage pour penser la dimension relationnelle du consentement. Voir par exemple Mackenzie & Stoljar (2000).

trouver les moyens de donner du pouvoir d'être et d'agir à ceux dont la liberté est restreinte par toutes sortes d'obstacles. La question principale est alors la suivante : « qu'est-ce que chaque personne dans son contexte de vie est capable de faire et d'être ? ». (...) « Nous nous interrogerons non seulement au sujet de la satisfaction qu'éprouve la personne dans ce qu'elle fait, mais au sujet de ce qu'elle fait et de ce qu'elle est en mesure de faire (quelles sont ses possibilités et ses libertés) ». (Brugère, 2013, citant Nussbaum, 2008, p. 110-111)

L'approche des capacités permet de penser et d'analyser le développement humain en instaurant les conditions d'un monde plus juste, où chaque personne pourrait se développer dans un horizon d'épanouissement¹⁴³.

Dans le contexte numérique, l'approche des capacités peut en particulier éclairer la nécessité pour les citoyens et citoyennes d'avoir accès à des savoirs qui permettent de comprendre les grandes lignes des écosystèmes numériques et des enjeux liés aux données personnelles. Cet accès permet de se doter de compétences structurantes qui développent « *the ability to define, access, manage, integrate, communicate, evaluate and create information safely and appropriately through digital technologies and networked devices for participation in economic and social life* » (GAML & UNESCO, 2018, p. 3). Cette capacité est appelée « littératie numérique » (*digital literacy*) et comprend des capacités fonctionnelles (savoir utiliser les outils numériques), mais également critiques (en comprendre les implications sociales) et éthiques (les utiliser de manière responsable)¹⁴⁴. Si l'on regarde la littératie numérique par l'angle du consentement tel qu'il peut être demandé, donné et retiré en ligne, celle-ci met en lumière l'interdépendance entre l'utilisatrice à qui on demande son consentement, l'organisation qui a besoin de ce consentement mais doit respecter toute forme (ambiguë) par laquelle il est donné, les services qui sont utilisés pour concevoir le système numérique. Cette interdépendance est travaillée par le design qui conçoit les façons de le demander, le

¹⁴³ Amartya Sen met l'accent sur un horizon de bien-être par l'amélioration des conditions qui permettent davantage d'égalité, ou de corriger des situations d'inégalité. Martha Nussbaum oriente davantage les capacités vers l'épanouissement par la justice, fidèle à une posture aristotélicienne qui vise la réalisation de soi par une éthique relationnelle (voir son livre *The Fragility of Goodness*, 1983, cité dans Brugère, 2013).

¹⁴⁴ Pour un aperçu plus approfondi de ce qu'implique la littératie numérique, en particulier dans le contexte du développement responsable de l'IA, voir la section et ses recommandations qui lui sont dédiées dans le rapport de la Déclaration de Montréal (Vézy et al., 2018).

donner ou le retirer; et en même temps, cette interdépendance émane du design au fil chemins parfois surprenants que les usages prennent. Et dans un contexte plus spécifique de design de conversation, où les dérives des agents conversationnels et leurs usages peuvent reproduire des dynamiques d'interaction qui ont tendance à faire honte aux personnes d'un certain genre (chapitre 2), où les pratiques à l'égard des données personnelles peuvent être laxistes, la dimension critique de la littératie numérique peut amener à (se) demander : à quelles conversations consent-on ? Les conditions que l'on arrange favorisent-elles le respect des personnes avec lesquelles nous interagissons via leurs données ?

Conclusion : un horizon qui vaut la peine de le faire compter

Dans ce chapitre, j'ai défini le consentement en présentant les ambiguïtés et incertitudes qui le constituent. Alors qu'il serait idéalement donné au nom d'une volonté autonome et en contrôle, signe de sa liberté, les conditions dans lesquelles il est demandé et donné interfèrent souvent, au risque de l'influencer, de le forcer, de l'empêcher d'être tout à fait éclairé ou informé. Ces conditions peuvent relever du design comme d'un ordre plus structurel marqué par des rapports de domination ou de soumission. Face à ces ambiguïtés et incertitudes, on pourrait être tenté de considérer le consentement comme un acte vain, au mieux formel, au pire condamné à l'échec de sa validité. Mais cette résignation passerait à côté de l'exercice vulnérable que demande le consentement pour pouvoir opérer sa « magie morale » qui permet de donner accès à soi, et d'accéder à l'autre, sans blesser.

Il s'agit alors de changer le regard porté sur le consentement : plutôt que s'accrocher à son idéal contrôle, le consentement tient dans la reconnaissance de notre besoin de l'autre au quotidien, dans des conditions et des situations de contrôle qui ne sont jamais parfaites. Le consentement se situe dans un besoin de l'autre qui n'est pas condamné à la soumission; il est plutôt inhérent à l'interdépendance qui nous lie au quotidien. Cela ne signifie pas pour autant que les situations de domination et de

soumission n'existent pas. Ce changement de regard porté sur l'interdépendance amène plutôt à se doter d'outils pour agir vers l'égalité plutôt que de subir des structures de domination.

Cette réflexion ne concerne pas seulement des situations de consentement entre deux personnes dans des rapports physiques. Elle s'applique également à des situations d'usage numérique qui demande à une personne (utilisatrice) de donner accès à soi, par l'intermédiaire de ses données personnelles. Plutôt que d'être réduit à une simple formalité, le consentement mérite d'être considéré comme une situation d'interaction où le respect de la personne se joue, en particulier en reconnaissant que les informations qu'elle s'apprête à partager ne sont pas des données désincarnées, mais qu'elles s'inscrivent dans des situations de vécu personnel, marquées de certaines vulnérabilités.

Si le consentement – dans tous ses usages – dépend des conditions (et du design) dans lesquelles il est donné, ce qui compte c'est que ces conditions permettent un certain épanouissement, en relevant d'une liberté qui soit un soutien dans l'expression et le respect de soi et des autres au quotidien. Car son horizon d'épanouissement, par le respect visant l'égalité, vaut la peine de continuer à le faire compter. C'est pour prendre soin de ces conditions que je propose de faire du consentement pas seulement une préoccupation (*matter of concern*) mais un *matter of caring* dans les organisations.

Chapitre 4

***Matter of caring* : une éthique du « faire parler »**

Introduction : quelle éthique pour faire parler ?

Pourquoi prendre soin du consentement dans les conversations, en particulier celles que l'on conçoit avec des agents conversationnels ? Les trois premiers chapitres ont abordé cette question du point de vue du consentement : alors que le consentement est une dynamique de conversation importante pour marquer le respect des personnes qui échangent, il est souvent négligé, détourné ou incité, par une habitude qui l'aborde comme un moment inconfortable, ennuyant ou distrayant. Aux chapitres 2 et 3, j'ai exposé des manières de faire autrement pour faire parler les agents conversationnels, et pour parler de consentement. Ces manières de faire autrement s'inscrivent dans une démarche sociomatérielle porteuse d'intervention, qui cherche à faire compter (*matter*) les préoccupations éthiques liées au respect de la vie privée, des données personnelles et du consentement, au fil du développement d'un agent conversationnel (voir chapitre 1).

À présent, je pose à nouveau la question : *pourquoi* prendre soin du consentement dans les conversations ? Mais le « pourquoi » qui m'intéresse ici est à entendre dans une version plus théorique. Ou plutôt, je pose cette question à la théorie, dont le travail cherche à donner du sens à des phénomènes et à les faire parler d'une certaine manière après avoir été vécus et observés par les chercheurs. La théorie que j'interroge en particulier est celle, en communication, qui cherche à mettre en lumière la multiplicité de ce qui s'exprime au travers de nos paroles. Il s'agit de la théorie de la ventriloquie que François Cooren a exposée dans son livre *Manières de faire parler* (2013) pour « souligner l'omniprésence de cette activité consistant à se positionner soi-même (souvent implicitement) comme agissant ou parlant pour d'autres êtres, tour à tour collectifs, conceptuels, contractuels, factuels et humains » (p. 190). En d'autres termes, nous ne

sommes pas seuls à parler dans nos interactions, nous faisons parler d'autres êtres, avec lesquels nous entretenons un certain attachement. En référence à la marionnette des ventriloques, François Cooren utilise le terme de « figure » pour désigner ce qui anime des personnes et les font parler et agir tel qu'elles le font, en particulier dans des organisations. Mais quand on porte attention au consentement et à l'ambiguïté qui constitue cet acte, les figures qui font parler peuvent entretenir un rapport étrange avec les personnes qui les expriment. Pourquoi identifier les figures qui font parler, au nom de quoi, ne permet pas tout à fait de porter attention au consentement, dans le souci que celui-ci soit une expérience d'épanouissement ?

À partir de ce que j'ai présenté dans le chapitre 3, je vais dans un premier temps identifier les figures qui sont censées s'exprimer dans le moment d'interaction qu'est le consentement pour montrer que certaines d'entre elles sont potentiellement indésirables. Face à ces figures indésirables, je vais proposer dans un deuxième temps le concept de *matter of caring*, qui part d'un angle différent de la ventriloquie, pour contribuer à *configurer autrement* les conditions d'expression de la parole dans un moment comme le consentement. Cette reconfiguration consiste en d'autres termes à faire bouger les figures qui font parler, et faire de la place à d'autres figures pour parler, dans un moment comme celui du consentement. Par cette activité de reconfiguration, porter attention à la manifestation et l'expression de *matters of caring* est également une pratique méthodologique de recherche, de l'ordre du *careful engagement* déjà mentionné (voir chapitre 1, p. 49). Je vais donc, en même temps, présenter cette pratique méthodologique. Celle-ci sera approfondie dans le chapitre suivant pour mettre l'accent sur la dimension réflexive par laquelle, à la lumière du consentement comme *matter of caring*, l'organisation, l'éthique et moi-même comme chercheure nous nous sommes tissées ensemble, de façon relationnelle.

Pourquoi aller au-delà de « faire parler » ?

Dans les chapitres précédents, *faire parler* a fait l'objet d'une activité de design, pour que des systèmes d'intelligence artificielle puissent tenir des conversations. J'ai d'abord montré une panoplie de choses qui font parler de tels agents conversationnels : les *cloud services* qui permettent de les développer, les techniques d'ASR et TTS¹⁴⁵, les voix artificielles et humaines, la persona du bot, les documents dans lesquels est travaillé le design de conversation, les choix de design de conversation, les formulations, les stéréotypes à imiter ou non, les données à partir desquelles les algorithmes s'entraînent, les intentions des designers, les usages et détournements des utilisateurs, l'objectif d'offrir une bonne expérience, la logique de familiarité et de continuité pour cette bonne expérience, les erreurs qui font parler le bot autrement qu'anticipé, les plateformes qui facilitent l'organisation du travail entre développeurs et designers.

Puis, mon attention s'est portée sur une valeur particulière, le consentement. Il s'agit de prévoir, dans le design de l'agent conversationnel, comment faire parler quelqu'un de son consentement – qu'il accepte ou refuse de le donner. Or le consentement invoque de nombreuses choses dans sa manière d'être exprimé : des principes d'autonomie et de respect de la vie privée, des impératifs légaux qui visent sa validité, des droits et obligations, de l'information, de la transparence, de la clarté, de l'ambiguïté, des choix de design plus ou moins sombres ou instructifs (*nudges*, *dark patterns*, *nudges éducatifs*), des fonctionnalités qui permettent de l'exprimer et le retirer, de la fatigue, de l'inconfort, de l'évitement, de la connaissance, des conditions de soumission ou subordination, un idéal d'épanouissement et d'égalité.

La diversité de ce qui s'exprime dans l'activité de *faire parler* quelque chose ou quelqu'un peut être abordée par la perspective ventriloque de la communication que François Cooren a exposée dans son livre *Manières de faire parler* (2013). Cette perspective s'inscrit dans l'approche de la Communication Constitutive de la Réalité (CCR) qui s'intéresse aux façons dont la réalité se façonne par la communication, au fil des

¹⁴⁵ *Automatic Speech Recognition* et *Text-To-Speech* : voir chapitre 2.

interactions entre les personnes et les choses autres qu'humaines (documents, technologies, bâtiments, etc.). La Constitution Constitutive des Organisations (CCO) s'inscrit dans cette approche, en se concentrant sur les façons par lesquelles les organisations se façonnent (Boivin et al., 2017). Dans le cas de mon étude, comme la constitution de l'agent conversationnel façonne la constitution de l'organisation Prélude, étudier tout ce qui participe au développement de l'agent conversationnel revient à étudier comment l'organisation Prélude se constitue par cette activité de design d'agent conversationnel. La ventriloquie marque en particulier le courant dit de l'École de Montréal qui se concentre sur les façons dont de multiples choses (principes, valeurs, préoccupations, règles, etc.) participent dans les interactions constitutives des organisations. Cela permet de montrer comment des personnes donnent une voix à d'autres choses, qui se trouvent en retour exprimées à travers elles (Cooren, 2013; Cooren et al., 2013). En d'autres termes, cette approche analyse le phénomène par lequel des éléments sont *talked into existence* dans une organisation, en montrant ce qui les soutient dans la communication quotidienne (Cooren, 2010). Ce qui soutient l'agent conversationnel et le consentement chez Prélude est ainsi l'ensemble des choses listées ci-dessus, chacune étant évoquée, *talked into existence*, par la voix des membres de Prélude et parfois la mienne, selon les situations quotidiennes.

La métaphore de la ventriloquie est employée pour désigner ce phénomène communicationnel car elle met l'accent sur la dynamique à l'œuvre quand un ventriloque fait parler sa marionnette (*dummy* en anglais), ou que la marionnette s'exprime par le ventriloque qui l'anime :

Humans are both ventriloquists and dummies, and organizations are talked into existence through this oscillating dynamic: they make things and people speak to achieve their goals, as much as these things and people make them speak through what they say and do. Ventriloquial analyses allow us to reveal this hybrid character of multiple voices through unfolding discourse in interaction. (Nathues et al., 2021, p. 1458)

L'analyse ventriloque cherche ainsi à déceler « ce *au nom de quoi* » nous parlons et agissons tel que nous le faisons (Cooren, 2013, p. 108, italiques dans l'original). La ventriloquie porte alors un intérêt particulier aux attachements à différentes figures qui

nous font parler, nous conférant une certaine passivité dans nos actions, ou mettant du moins en lumière notre posture de « passeur » quand nous sommes en même temps acteur de ce que nous faisons et disons (Cooren, 2013, 2016).

Dans son article *Ethics for Dummies* (2016), François Cooren souligne les implications éthiques qu'a cette posture de passeur. Elle peut donner l'impression qu'il est plus facile de se déresponsabiliser : multiplier les sources de ce au nom de quoi on agit, ou de ce qui nous fait agir (un règlement, une personne supérieure dans la hiérarchie, une fonctionnalité technique qui ne permet pas de faire autrement), peut amener à effacer notre responsabilité au nom de ces autres choses dont nous sommes les passeurs. Mais Cooren met plutôt l'accent sur la dimension partagée de la responsabilité : tout passeur porte la responsabilité de ce qu'il passe¹⁴⁶. Multiplier les sources de ce au nom de quoi on agit « *forces us to be more vigilant* » (p. 26) car cela nous demande de répondre « *for all the authors that or who happen to join their voices to ours* » (p. 27). Cela nous demande donc d'être vigilant à l'égard de ce à quoi on participe, car nous pouvons, en tant qu'acteurs de ce que nous disons, faire la part des choses et décider de ce par quoi on se fait ventriloquer, ou ce à quoi on donne notre voix : « *Ethics for dummies thus consists of giving ourselves the chance to be ventriloquized by contradictory voices while taking ownership of our decisions (Bencherki & Cooren, 2011)* » (p. 28).

Cette décision est centrale dans l'acte de consentir, qui consiste à choisir si et à quoi on donne son accord. Si les sources d'un acte de consentement sont multiples, alors l'activité ventriloque perturbe les idéaux de liberté et d'autonomie au nom desquels devrait être exprimé un consentement, renforçant la difficulté de réunir les conditions parfaites pour l'expression d'un consentement parfaitement valable (voir chapitre 3). Consentir par l'intermédiaire d'une technologie comme un agent conversationnel ajoute une complexité : l'activité de design qui consiste à faire parler l'agent conversationnel doit

¹⁴⁶ Cela résonne en quelque sorte avec la responsabilisation que le RGPD impose à toute entité traitant des données personnelles : toute organisation qui a un lien avec des données personnelles est responsable de leur traitement, qu'elle s'occupe de leur traitement et leur hébergement de A à Z ou qu'elle ne soit qu'un des canaux par lesquels ces données transitent.

être vigilante avec les manières dont elle prévoit de permettre d'exprimer un consentement – c'est-à-dire avec les affordances qu'elle conçoit pour l'exprimer. En effet, le design pose les conditions par lesquelles le consentement est ventriloqué quand il est demandé à une utilisatrice, selon les fonctionnalités et les formulations par lesquelles celui-ci est demandé et donné, et selon les informations au nom desquelles il peut être donné ou refusé. La responsabilité des designers de conversation repose ainsi dans cette activité de multiplication des sources de l'action de consentir.

Mais il y a une implication éthique de plus quand on s'intéresse aux interactions de consentement : le consentement doit être bon, c'est-à-dire valable, pour être doté de sa « *moral magic* » (Hurd, 1996) qui rend acceptable l'acte pour lequel il est demandé (voir chapitre 3, p. 119). Pour qu'il y ait consentement valable, il faut faire parler la personne (lui donner la possibilité d'exprimer son consentement) *au nom de ce qu'elle veut*, et non la faire parler au nom de ce que l'organisation veut (obtenir le consentement). Autrement dit, il faut être attentif à l'influence potentiellement exercée sur le consentement, de sorte que l'activité qui fait parler soit orientée dans une direction qui soit bonne pour que le consentement s'exprime. L'éthique qui cherche à permettre l'expression du consentement demande ainsi à l'activité ventriloque de s'orienter *dans une certaine direction*, une direction qui permette au consentement de s'exprimer au nom de valeurs de respect et d'égalité entre les personnes, dans un horizon d'épanouissement (voir chapitre 3, p. 177). Ces valeurs et cet horizon sont en effet ce qui soutient le plus le consentement quand celui-ci risque d'être mis de côté tant les obstacles et ambiguïtés auxquels il fait face semblent le rendre impossible (voir fin du chapitre 3).

Dans le jeu ventriloque du consentement, la vigilance ne doit alors pas s'en tenir qu'aux décisions que des personnes prennent pour faire parler certaines figures plutôt que d'autres en s'en reconnaissant responsable (cf citation de Cooren, 2016, à partir de Bencherki & Cooren, 2011, plus haut). Car le consentement demande de porter attention aux situations où de nombreuses personnes sont dans l'impossibilité de prendre des décisions claires. La vigilance doit se porter sur la vulnérabilité au cœur du consentement : cette vulnérabilité réside dans la possibilité que le consentement ne soit pas valable, qu'il

soit influencé dans une direction qui n'est pas bonne pour la personne qui le donne, qu'il ne permette pas à sa « *moral magic* » d'opérer. À la lumière de cette vulnérabilité, j'interroge donc les figures par lesquelles peut s'exprimer le consentement, pour souligner la présence indésirable de celles qui nous influencent ou nous empêchent de parler.

Les *figures* peuvent être implicites ou explicites, elles sont invoquées par des personnes qui leur font dire des choses. Dans le cas de mon étude qui se penche sur le consentement numérique que de nombreux règlements cadrent (voir chapitre 3, p. 129), quand des personnes chez Mr Y et Prélude invoquent le RGPD pour modifier la façon dont est demandé le consentement par leurs agents conversationnels, le RGPD est la *figure* du règlement qui est invoqué, celle au nom de quoi les personnes parlent. Le RGPD est également parfois la figure au nom de laquelle des personnes sont présentes – c'était le cas pour les représentantes de Data Protection Consulting par exemple, dont la présence avait le RGPD comme raison d'être pour auditer TMTL et Prélude et vérifier la conformité de leurs activités (cf chapitre 3, p. 155).

La présence de ces figures qui *font faire* quelque chose peut évoquer la dynamique de soumission que j'ai abordée dans le chapitre précédent, montrant que l'acte de consentement peut être exprimé de façon passive, par résignation, ou par soumission à une situation qui nous dépasse et dans laquelle il est impossible de réunir les conditions pour un consentement parfaitement valable (selon les critères des règlements de protection des données personnelles). Si l'on reprend la métaphore ventriloque, ces conditions correspondent à tout ce qui est mis en place (ou toutes les affordances conçues) pour amener quelqu'un à exprimer son consentement ou son refus. Dans l'acte de consentir, ces conditions permettent de faire s'exprimer ou non un consentement – d'où la responsabilité de ceux qui conçoivent ces conditions (en particulier l'agent conversationnel, dans le cas de mon étude). Les *figures* que la personne amenée à consentir peut invoquer explicitement ou implicitement sont ses droits, ses envies, le principe de sa volonté autonome ou de sa liberté, sa fatigue, ses connaissances de l'écosystème numérique, les informations qu'elle a sur le traitement de ses données

personnelles, son manque d'intérêt, les objectifs qu'elle a en utilisant l'agent conversationnel qu'elle s'apprête à utiliser.

Parmi les conditions qui jouent sur le consentement, il y a les choix de design : les services technologiques utilisés pour développer l'agent conversationnel (l'Assistant Google, Dialogflow, Google Cloud Text-to-Speech), les formulations prévues par les designers de conversation, les modes d'interaction pour signifier un accord ou un refus, les *patterns* plus ou moins sombres ou instructifs qui *font faire* quelque chose, comme donner son consentement. Or, comme nous l'avons vu avec les *dark patterns* qui influencent, voire forcent, l'acte de donner son accord, il est possible que ces conditions gênent ou empêchent la personne de parler au nom de ce qu'elle voudrait – comme limiter ce à quoi elle donne son accord, refuser certains termes, retirer son consentement mais tout de même utiliser le service. Les interférences inhérentes aux conditions qui permettent au consentement d'être animé – c'est-à-dire donné, refusé, exprimé – font ainsi parler la personne d'une certaine manière, même si elle préférerait peut-être parler au nom d'autre chose. Il y a alors un désalignement entre ce qui anime nos manières de parler (les conditions) et ce au nom de quoi (les figures) on *aimerait* ou *voudrait* parler. Dans un contexte qui demande à notre accord ou notre refus de s'exprimer, c'est-à-dire un contexte où l'on fait parler notre consentement, ce désalignement me semble particulièrement intéressant à explorer davantage pour questionner l'activité ventriloque présente dans nos interactions.

Évoquer ce dont on aimerait ou voudrait parler, c'est envisager des figures implicites, relevant d'un conditionnel ou d'un idéal, qui ne parviennent pas à s'exprimer tout à fait. Or, ce qui les empêche de s'exprimer peut être ce qui rend le consentement contraint (voir chapitre 3, p. 151), ce qui porte atteinte à la moralité de ce consentement, c'est-à-dire à sa validité. Dans ce cas, les conditions qui animent notre consentement en le contraignant empêchent certaines figures de s'exprimer au travers de nous, alors même que ce sont ces figures qui rendraient notre consentement plus valable. Comme les conditions qui animent l'acte de donner son consentement peuvent porter atteinte à la validité de ce consentement, l'activité qui fait parler pour donner son consentement

relève d'une ventriloquie délicate : quand elle étudie le consentement dans ces interactions, l'analyse ne peut s'en tenir à l'observation de ce qui s'exprime explicitement, puisque ce qui s'exprime peut avoir les mots d'un consentement, sans que celui-ci ne soit effectivement un consentement valable, si les conditions par lesquelles il est exprimé le contraignent. En d'autres termes, l'analyse ventriloque demande de porter attention à la situation de vulnérabilité du consentement qui, pour être valable, doit être exprimé au nom de ce que veut la personne et non au nom de ce qu'elle est amenée à dire, parfois sous l'effet de contraintes ou influences.

Parce que le consentement s'exprime à l'endroit-même où il y a possibilité de blessure ou d'abus¹⁴⁷, faire parler le consentement demande une vigilance particulière à la vulnérabilité constitutive de celui-ci. Je vais aborder deux situations qui abordent le consentement de façon explicite et implicite, pour voir pourquoi on a besoin d'aller au-delà d'une analyse des figures qui s'expriment explicitement dans les tours de parole si l'on veut esquisser une éthique du consentement dans les interactions. La première situation en est une, chez Prélude, où un silence a été brisé pour faire parler certaines préoccupations. La deuxième est une situation de domination, dans le monde du cinéma, où des figures empêchent certaines personnes de parler. Ce qui m'intéresse dans ces situations, c'est l'importance de porter attention aux silences comme des formes par lesquelles s'expriment certaines figures.

Le possible désalignement, qui existe dans tout consentement, entre ce qui nous fait parler et ce au nom de quoi on voudrait parler, m'amène à mettre l'accent sur les figures qui ne parviennent pas à s'exprimer dans l'interaction, alors qu'elles comptent dans la valeur du consentement qu'on donne. Parmi les figures qui sont invoquées explicitement ou implicitement, l'approche ventriloque s'intéresse souvent aux préoccupations (*matters of concerns*). Ces préoccupations sont des sujets (*matters*) qui suscitent un intérêt particulier quand ces questions ne font pas qu'exister comme tout sujet, mais qu'elles se mettent à compter pour nous, que nous nous sentons concernés

¹⁴⁷ Comme nous l'avons vu dans le chapitre précédent avec l'idée de « *moral magic* » (Hurd, 1996), la « magie » du consentement opère dans cette possibilité de blessure ou d'abus, pour transformer l'inacceptable en acceptable.

par ce sujet. Par la ventriloquie, l'analyse des préoccupations permet de montrer la situation telle qu'elle est « *literally made of these beings that concern or preoccupy us* » (Cooren, 2016, p.23). Certaines préoccupations peuvent ne pas compter, auquel cas elles disparaissent des tours de parole et ne pèsent plus tant dans la décision – « *Of course, they could end up not mattering or counting, which means that they would then become silent and voiceless* » (Cooren, 2016, p.23).

Quand on se penche sur des interactions de l'ordre du consentement, cette disparition des tours de paroles (ou des interactions avec un outil technologique) pose plusieurs problèmes. D'une part, en évacuant la préoccupation de manifester et écouter un acte de consentement, comme le font les *dark patterns* et les habitudes fatiguées de passer par cette étape, le consentement finit par ne plus compter, alors que son horizon d'épanouissement par le respect en toute égalité vaut la peine de continuer à le faire compter (voir chapitre 3, 177). D'autre part, en ne parvenant pas à s'exprimer clairement, un consentement peut devenir silencieux, ce qui peut en faire un accord résigné ou un refus (voir Kitsinger & Fritz (1999) dans le chapitre 3, p. 171), dans l'ambiguïté de son absence de voix. Ces disparitions des tours de paroles sont problématiques si, au nom de ces silences ou ces absences, d'autres personnes considèrent que ces consentements ne comptent pas – dans ce cas, l'abus est possible. Elles sont également problématiques si d'autres personnes se permettent d'agir au nom d'un consentement qui serait tacite ou implicite alors que son silence se voudrait un refus. Ces silences et disparitions sont, en d'autres termes, des figures qui existent, même si elles ne sont pas *talked into existence*.

Dans une situation où un consentement est en discussion, une posture ventriloque gagne ainsi à être curieuse de ce qui se passe au-delà ou en-deçà de l'interaction, pour écouter les silences¹⁴⁸ de ce qui ne parvient pas jusqu'aux tours de paroles et qui compte pourtant, car cela participe au façonnement de ces tours de paroles. Cette curiosité peut par exemple être étoffée par des notes de terrain pendant l'observation de situations, car ces notes explicitent des figures qui viennent briser leur silence malgré les hésitations

¹⁴⁸ Un exemple de silence à écouter est la pièce 4'33" de John Cage, performée dans cette vidéo par le Berliner Philharmoniker le 31 octobre 2020 : <https://www.youtube.com/watch?v=AWVUp12XPpU>

dans l'interaction. C'est ce qu'étoffent mes notes prises à la suite d'une réunion le 10 juillet 2019 chez Prélude. Le sujet du consentement est arrivé dans la discussion, à la fin d'une réunion où Philippe présentait Rasa à Félix et Paul, un *SaaS*¹⁴⁹ de création d'agent conversationnel similaire à Dialogflow, mais indépendant de Google. Il ne s'agit pas d'une discussion où le consentement de quelqu'un était en jeu sur le moment; il s'agit plutôt d'une discussion portant sur les interactions qui seraient intégrées dans le design de conversation de l'agent pour demander leur consentement aux utilisateurs.

Présente dans la salle pour comprendre initialement les implications de Rasa, j'écoute Philippe, Paul et Félix interagir puis, à ma surprise, aborder ce à quoi pourraient ressembler les interactions portant sur le consentement, jusqu'à ce que j'interrompe mon silence :

- 1 Félix : là c'est-c'est ça, "*yes/no, would you like to know more about that?*" Puis là,
2 là[
3 Philippe : [Je suis pas mal sûr que c'est facile de faire ça dans une phrase élégante
4 puis dire si vous voulez plus de t'sais euh
5 Paul : si tu veux plus, tu-t'as un « *learn more* » et là ça part (.) t'sais genre sur un
6 autre module[
7 Philippe : [ouais ça, puis à la limite ça pourrait, ouais c'est un autre
8 module exactement. Alors autrement dit euh, je pense que c'est la bonne
9 façon de faire puis, mais honnêtement, je trouve que c'est un peu une insi-
10 une insulte à l'intelligence humaine par rapport à, quand on y réfléchit,
11 parce que à un moment donné, les gens ils devraient comprendre que[
12 Paul : [oui
13 c'est plus pour les rassurer
14 Camille : ((*voix plus forte*)) [nan
15 nan nan mais les gens savent pas forcément, les gens sont pas forcément
16 ils[
17 Philippe : [c'est un problème de littératie en fait
18 Paul : [ouais
19 Camille : [ils pensent déjà qu'on collecte des données tout le temps tout le temps
20 ou alors ils sont pas conscients de ce qu'on collecte, ((*la voix hésitante*))
21 après il y a la question de ben c'est quoi les risques en fait ? enfin je veux
22 dire[

(1^{er} extrait de retranscription de réunion du 10 juillet 2019)¹⁵⁰

¹⁴⁹ *System as a Service* : voir chapitre 2 (p. 62).

¹⁵⁰ Pour les retranscriptions de réunion, j'utilise les conventions d'analyse de conversation de Gail Jefferson (1984) dont les principales sont décrites dans l'article de Kitzinger & Frith (1999, p. 312). Le [est utilisé pour

Philippe fait des suggestions de ce que pourrait dire le bot pour informer le consentement, Félix redirige, je me retiens d'intervenir. Mais voilà qu'une impatience monte en moi, des « ouais » sortent de ma bouche, peu audibles, comme s'ils se frayaient une place dans la discussion à mesure que le cadrage de mon observation se déplace : je suis la seule femme dans cette pièce, les deux hommes qui parlent principalement sont ceux qui ne connaissent que de loin mon travail en arrière-plan sur le consentement pour l'autre projet. Ils tentent des formulations rapides sur des enjeux pour lesquels je cherche des formulations justes et nuancées depuis plusieurs semaines, tout en apprivoisant les ramifications du concept de consentement étroitement lié aux régulations en matière de données personnelles. Mon avis ne semble pas sollicité, cela m'embête un peu, mais en même temps cela me permet d'écouter attentivement et observer ces interactions.

(À partir du journal de bord, notes réflexives du 11 juillet 2019)

Un moment particulier me pousse à intervenir sous forme d'objection : les suppositions sur ce que « les gens [...] devraient comprendre » (l. 11) font émerger un agacement plus fort que la curiosité de continuer à écouter. Une préoccupation m'anime : au nom de quoi peut-on parler pour « les gens » ? Quelque chose me pousse à agir : un sentiment de légitimité pour parler au nom des « gens » sur ces questions – un sentiment de légitimité (Benoit-Barné & Cooren, 2009) fondé sur mon implication de l'année précédente au sein de la démarche de consultation citoyenne pour la Déclaration de Montréal (voir chapitre 1, p. 33). En termes ventriloques, cette expérience m'avait ouverte à plusieurs figures (des citoyens et citoyennes, des principes, des inquiétudes, des revendications, des recommandations, des objections) qui me font parler au nom de cette complexité rencontrée, et qui me font revendiquer l'autorité de parler pour elle. Comme le souligne François Cooren, « *the very reason [I] feel authorized to speak comes from the various authors that compose [my] turns of talk* » (2016, p. 17) dans cette situation.

Parmi ces figures qui me font parler, ou m'autorisent à parler, il y a le constat qui avait été fait à propos de l'asymétrie de connaissances entre citoyens et développeurs d'IA sur l'utilisation des données personnelles (Vézy et al., 2018). Alors, quand j'entends un développeur (Philippe) supposer ce que les gens devraient comprendre, des

indiquer des chevauchements, le (.) indique une pause de moins de 0.2 seconde et le soulignage indique une insistance.

préoccupations prennent le dessus : que risque-t-on de passer sous silence quand, depuis un point de vue connaisseur du sujet, on suppose ce que les gens devraient comprendre a priori ? Le consentement peut-il être vraiment informé si on suppose un ensemble de choses (à propos des données personnelles) que les gens doivent avoir comprises avant que l'on s'adresse à eux ?¹⁵¹ Ces préoccupations prennent ainsi le dessus sur le silence et ma posture d'observation devient une intervention, au nom de la figure des gens pour lesquels parler.

Je pourrais ne pas intervenir, en me contentant d'observer les manières de parler des autres, attentive aux façons dont l'éthique peut se matérialiser dans les tours de parole des membres de cette organisation, à la manière du *shadowing* souvent pratiqué dans la recherche CCR et CCO (Vásquez et al., 2012). Mais dans ce cas, que ferait mon silence dans cette situation ? Si je laisse seulement le silence être présent, je risque de laisser des préoccupations et nuances ne pas compter, alors qu'elles ont de l'importance pour les personnes comme moi, en éthique de l'IA, attentives au respect de celles et ceux qui utiliseront les systèmes d'IA en développement. Je décide donc d'intervenir pour ne pas contribuer, par mon silence, à laisser ces figures ne pas compter. Autrement dit, je me laisse animer par ces préoccupations et nuances auxquelles je prête ma voix, comme par un débordement qui donne de l'importance à ce qui compte pour moi. Ce « moi » est habité de plusieurs figures qui motivent ce débordement : il y a mon rôle dans cette organisation, qui consiste à travailler sur les interactions de consentement pour un agent conversationnel; il y a aussi mon corps¹⁵², mon regard et mon écoute, attentifs à l'éthique par laquelle ce consentement peut être respecté.

¹⁵¹ Cette intervention permet de faire parler la préoccupation de la littératie numérique, prononcée par le développeur informé sur ce sujet. Cette préoccupation marque le souci d'accompagner la compréhension des utilisateurs – un souci qui est de l'ordre d'un *matter of caring* puisqu'il s'agit d'améliorer une situation (manque de compréhension des enjeux en matière de citoyenneté numérique).

¹⁵² Je plaide en faveur de la prise en considération du corps des chercheurs, en particulier lors du travail de terrain à partir duquel ils constituent des mondes par leurs observations, participations et analyses, car « le corps n'est pas quelque chose que nous possédons, c'est notre médium général pour avoir un monde » (Garcia, 2019, p. 172). Ce 10 juillet 2019 comme pendant toute mon immersion, je suis présente chez Mr Y et Prélude par mon corps féminin et la situation qu'il implique, c'est par lui que j'ai prise sur le monde dans lequel je m'immisce et que j'esquisse mes projets – je reprends là la citation de Beauvoir, citée dans Garcia, 2019, p. 172 : « Dans la perspective que j'adopte, (...) si le corps n'est pas une *chose*, il est une situation :

Ce débordement par lequel se brise mon silence porte en même temps autre chose, une tension présente dans mon corps, mon regard et ma voix hésitante (I. 20) : la situation en train de se dérouler est constituée principalement de paroles d'hommes, la seule femme dans la salle s'efface alors que c'est la personne la plus informée sur le sujet. En laissant le silence s'exprimer au travers de ma personne, je me soumetts à une certaine situation, laissant l'espace se constituer majoritairement par les paroles d'hommes, contribuant ainsi à marginaliser la présence des femmes, en particulier dans le secteur technologique (Corbeil et al., 2017). Cette situation n'est pas seulement une situation *dans laquelle* je suis mais une situation *que je suis*. Il ne s'agit pas là de comprendre le verbe être comme une essentialisation, comme si toute personne était réductible à sa situation, mais plutôt de porter attention au contexte qui participe à sa constitution en tant que personne.

À partir des travaux de Simone de Beauvoir, Manon Garcia explique cette notion de situation de tout être humain : en reconnaissant que l'être humain est situé, Beauvoir « affirme que l'on est dans un monde où notre être a toujours déjà un sens et une norme mais que ce sens et cette norme sont des produits de l'histoire et non d'une nature fixe [ou essence] » (Garcia, 2019, p. 83). Les possibilités d'action sont ainsi structurées par ces normes et ce sens qui se sont sédimentés au fil du temps. Or pour les femmes, comme pour toute personne dans une position d'infériorité où le rapport de pouvoir est inégal, ce comportement socialement prescrit relève de la soumission (Garcia, 2019, p. 88-91). Vécue au quotidien par certaines personnes, cette soumission peut prendre la forme de silences, de gêne, de disqualification; sans être nécessairement de l'ordre de la passivité imposée, elle peut relever d'une sorte d'auto-censure performée activement par les personnes elles-mêmes (Garcia, 2019). On peut également éprouver du plaisir à cette soumission : le dévouement et l'abdication peuvent être confortables (Garcia, 2019, p. 119).

c'est notre prise sur le monde et l'esquisse de nos projets » (Beauvoir, *Le Deuxième Sexe*, t. I, p. 75, italiques dans l'original).

Ce concept de situation que *l'on est* ne veut pas nécessairement dire que le macro des normes soit la source de l'action, comme une abstraction qui s'imposerait à nous. Il invite plutôt à regarder, au niveau micro, la façon dont ces normes sont performées et réitérées au quotidien¹⁵³, au fil d'un processus similaire à celui de « cultivation » décrit par Bencherki, Cooren, Brummans, Benoit-Barné et Matte (2020) :

Nous appelons cette perspective [communicationnelle de la culture organisationnelle] « cultivation », en référence au fait que la culture s'établit alors que les gens cultivent les figures qui comptent pour eux, en les mobilisant encore et encore, de manière itérative, dans leurs interactions. (p. 92)

Cependant, la notion de situation que *l'on est* amène à nuancer l'idée que « les gens cultivent les figures qui comptent pour eux » en invitant à prendre en considération que la situation que *l'on incarne* peut nous amener à cultiver des figures qui nous entravent. En d'autres termes, par nos situations, nous pouvons être amenés à cultiver des figures qui font une différence pour nous mais de façon négative. Ce sont des figures auxquelles nous sommes attachés, ou plutôt des figures qui nous attachent dans la cultivation de ces habitudes d'interaction, où la voix de personnes aux situations similaires a tendance à être plus basse, plus polie, à prendre moins de place, à être coupée ou chevauchée par d'autres voix. D'autres, comme des designers de conversations qui imitent les stéréotypes de genre (voir chapitre 2, p. 104), peuvent également être amenés à cultiver des figures telles que des persona soumises et dociles, qui entretiennent des habitudes d'interactions qui nous¹⁵⁴ dégradent.

Cet accent mis sur la situation que *l'on est*, plutôt que sur la situation *dans laquelle* on est, apporte ainsi une nuance à l'argument de François Cooren qui invite à porter attention aux « *diktats* des situations » (2013, 2016), pour mettre en lumière nos façons d'être « passeurs » en même temps qu'acteurs de ce que nous faisons et disons. Si cette approche permet de mettre en lumière la multiplicité de ce qui s'exprime lorsque nous

¹⁵³ C'est là l'originalité de la pensée de Manon Garcia à partir de Beauvoir : penser la soumission, « au sein de la domination, s'intéresser à ceux [et celles] qui la vivent plutôt qu'à ceux [et celles] qui l'exercent » (2019, p. 104). La question n'est pas tant « qu'est-ce que la domination ? » mais « qu'est-ce que ça fait d'être soumise ? ».

¹⁵⁴ Ce « nous » regroupe les femmes réduites à cette servitude, mais aussi les hommes réduits à une supposée bêtise dominante et les personnes non-binaires ignorées.

parlons et agissons, il me semble important de porter attention à la qualité des figures qui s'expriment quand nous parlons (de consentement, mais pas seulement) : lesquelles sont positives, favorables à l'expression de la personne dans son épanouissement¹⁵⁵, lesquelles sont négatives, voire néfastes à l'épanouissement de la personne qui se trouve animée par des figures qui l'entravent, lesquelles sont plus grises, lesquelles sont neutres. S'interroger sur la qualité des figures, des plus épanouissantes aux plus entravantes, touche à un autre aspect délicat de la posture ventriloque : de quoi est-on témoin ? Qu'est-ce qu'on ne voit pas clairement, mais qui est pourtant là ?

Ces questions demandent de ne plus considérer le silence comme un ensemble de préoccupations qui sont amenées à ne plus compter (dans l'interaction) car elles se retrouvent sans voix (Bigo, 2018). Le silence gagne à être mis en situation : qui est silencieux ou silencieuse, qu'est-ce qui est silencieux, comment ce silence s'exprime, quelles/qui sont les situations qui parlent et celles qui sont silencieuses ? Ces questions sont difficiles à aborder dans l'analyse d'une interaction retranscrite, car elles demandent un accès aux voix qui ne s'y sont pas exprimées – les faire parler en supposant ce qu'elles voudraient dire, ce serait leur manquer de respect en mettant dans leur voix des figures qui ne sont peut-être pas les leurs. Je pose plutôt ces questions pour attirer l'attention de l'analyse ventriloque, qui fait parler, sur la performativité des silences dans des situations façonnées par des attachements qui entravent. Ces silences comptent tout autant que les préoccupations explicites, surtout s'ils manifestent des figures qui empêchent la personne de parler. Cela importe particulièrement dans les situations de consentement, rarement marquées d'explicite, souvent habitées d'ambiguïtés, comme je l'ai abordé dans le chapitre précédent (p. 171), pour inviter à affiner notre écoute de ces situations vulnérables.

Cette écoute des situations vulnérables demande de ne pas s'en tenir à ce qui est explicite dans la conversation, alors que l'analyse ventriloque s'attache souvent à ce qui est formulé par des mots et des actes. La situation que l'on est peut en effet impliquer

¹⁵⁵ La notion d'épanouissement fait référence à Nussbaum (2008), tel que présenté à la fin du chapitre 3 (p. 156).

que l'on parvienne rarement aux tours de parole, elle peut également nous faire adopter un mode d'expression principalement implicite, empêchant de communiquer clairement nos préoccupations, intérêts et autres figures qui nous animent, voire nous laisser être présents dans un silence incarnant implicitement des figures qui nous empêchent de parler. C'est ce qui se passe dans toute dynamique de domination, où la personne dominée ne peut pas s'exprimer pleinement dans un espace occupé par quelqu'un ou quelque chose qui la domine.

C'est ce que dévoile par exemple le témoignage de l'actrice française Adèle Haenel dans le journal Médiapart où elle raconte la mécanique silencieuse dans laquelle l'ont enfermée les abus sexuels du réalisateur Christophe Ruggia quand elle avait entre 12 et 15 ans (Turchi, 2019). Ce témoignage permet de faire de la place à une parole qui explicite, au terme d'un long cheminement, une « *prise de parole [qui était] impossible* » puis un « *silence devenu insupportable* » (p. 1). Ce témoignage et l'enquête qui l'accompagne amènent à porter attention à ce qui se passe en négatif des actes et des paroles : qu'est-ce qui empêche de parler, quelles sont ces figures insupportables qui se mettent à cadrer tous ses faits et gestes, la faisant parler d'une certaine manière, estimant savoir ce qu'elle veut et prendre soin d'elle, la faisant parler, agir ou rester silencieuse par son emprise qui impose de garder un secret – un secret qui se cristallise en sentiments de gêne, de honte, de culpabilité, des sentiments qui font se recroqueviller sur soi-même dans le silence, et empêchent toute figure épanouissante de s'exprimer.

Pourtant ce qui était visible (ou explicite) dans cette situation, c'était une relation entre un réalisateur et une jeune actrice, des sourires sur le visage de l'actrice lors d'événements autour du monde, comme en témoignent plusieurs photos. Certains témoins ont vu là quelque chose qu'ils n'arrivaient pas tout à fait à nommer : « plusieurs membres de l'équipe peinent à mettre un mot sur ce qu'ils observent » (p. 4). Il y a eu du doute (2 occurrences dans l'article), du malaise (5 occurrences), du dérangement (1 occurrence) : des sentiments troublés, marqués par de l'ambiguïté alors qu'il y a des choses qu'il est difficile d'oser dire – ici, dénoncer la relation entre un réalisateur et sa comédienne. Ces sentiments troublés et cette ambiguïté peuvent compter comme des

figures qui amènent à éviter d’agir, laissant la situation se dérouler sans la perturber. Quand ces figures poussent à agir, ce n’est pas avec clarté, c’est plutôt le doute qui domine, comme en témoignent les interrogations de celle qui était régisseuse sur ce tournage :

C’est très compliqué de se dire que le réalisateur pour qui on travaille est potentiellement abusif, qu’il y a manipulation. Je me disais parfois : « Est-ce que j’ai rêvé ? Est-ce que je suis folle ? » Et personne n’aurait l’idée de s’immiscer dans sa relation avec les comédiens, d’oser dire un mot, car cela fait partie d’un processus de création. D’où les possibilités d’abus – qu’ils soient physiques, moraux ou émotionnels – sur les tournages. (p. 4)

En revanche, ce qui s’exprime avec clarté, ce sont les préoccupations de la personne qui domine la situation. Quand la coach des jeunes acteurs sur le tournage a voulu exprimer ses doutes sur la situation, le réalisateur lui a rétorqué : « Tu veux foutre en l’air mon film, tu ne te rends pas compte le rapport privilégié que j’ai avec eux ! » puis ne lui aurait plus adressé la parole, tout en lui compliquant son accès au tournage par la suite. Au nom de la préoccupation de la réussite du film et du « rapport privilégié » qu’il compte favoriser avec les acteurs, deux préoccupations manifestées par le réalisateur en position dominante, la personne qui a osé tenter d’intervenir pour faire changer la situation se fait écartier des tours de parole qui comptent. Sa préoccupation envers le bien-être des jeunes acteurs – une préoccupation qui est en même temps son rôle sur ce tournage – est réduite au silence.

Ce qui s’exprime alors explicitement dans ces situations, ce sont les figures qui animent le réalisateur qui domine la situation. Celui-ci est en même temps la figure qui empêche les autres de parler, qui amène les autres préoccupations à rester en silence pour moins compter. Alors, si l’analyse CCR et CCO par l’angle des figures ventriloques qui cultivent des situations tient à rester sur la « terre ferme des interactions » (Cooren, 2013), il est toutefois important d’inciter à la vigilance envers ce qui se passe en filigrane de l’interaction : ce qui se voit, ce qui s’entend, mais également ce qui peut empêcher une voix de parler et d’être entendue, selon la situation de chacune.

Cette nuance sur la situation incarnée par chacune résonne avec l’épistémologie féministe qui porte attention aux perspectives situées par lesquelles sont créées toute

connaissance (Haraway, 1988). Il s'agit de mettre en lumière des points de vue (*standpoint*, Harding, 1992) à partir desquels les paroles et actes sont dits et faits, sans pour autant essentialiser ces points de vue comme des perspectives figées, mais plutôt en reconnaissant que la situation à partir de laquelle chaque personne s'exprime, les figures qu'elle cultive, n'est pas la même pour tout le monde. Parmi les figures qui les (dés)animent, certaines personnes comptent davantage de figures qui les empêchent de parler, ce qui leur demande un certain effort pour trouver et faire s'exprimer les figures qui comptent de façon positive pour elles; d'autres sont plus libres dans les figures qui les animent.

Le consentement cherche précisément à désamorcer des dynamiques de domination, en se faisant respecter au nom de l'égalité entre les personnes qui agissent ensemble. Quand on analyse des situations où le consentement est en jeu – c'est-à-dire toute situation où des personnes interagissent en se respectant mutuellement – il est particulièrement important de porter attention à ces potentielles figures qui empêchent de parler, pour favoriser les situations de consentement qui visent un certain épanouissement (voir fin du chapitre 3, p. 175). L'observation de cultivation peut ainsi se faire en étant attentive non pas seulement à ce qui s'exprime, mais en interrogeant *qui* s'exprime, au nom de *quoi*, pour qui ou pour quoi. Ces questions ne s'arrêtent pas à ce qui se manifeste, mais interrogent également ce qui est sous-jacent : qu'est-ce qui ne s'exprime pas, pourquoi, qu'est-ce qui peut empêcher de s'exprimer, comment accompagner le dévoilement de ces voix plus silencieuses ? Ces accents sont au cœur du questionnement qui se penche sur les situations de consentement, c'est-à-dire les situations qui portent une attention particulière au respect de chacun au nom de leur égalité (voir Garcia, 2021, citée à la fin du chapitre 3, p. 173).

Quand elles se posent seulement pendant l'analyse, ces questions trouvent bien peu de réponses explicites si des entrevues n'ont pas été menées pour écouter ces voix inhibées, et si ces inhibitions n'en sont pas au stade d'être reconnues. Ces questions se posent alors beaucoup pendant l'observation et la participation à des situations abordant des interactions de consentement. L'observation de cette façon d'un processus de

cultivation est alors porteuse d'une éthique qui consiste à porter attention au négligé, à l'implicite, à observer avec sollicitude et conscience des situations de vulnérabilité. Dans l'observation de ce qui compte (*matter*), il s'agit d'une éthique qui ajoute du *care* aux *concerns* (Puig de la Bellacasa, 2011).

Matters of caring

Porter attention aux situations pour que le consentement s'exprime demande une attention éthique supplémentaire qui va au-delà de l'approche ventriloque : cela demande d'écouter des situations potentiellement vulnérables, de faire attention non seulement à ce qui s'exprime, mais aux conditions dans lesquelles cela s'exprime (par des paroles ou du silence), de s'attarder sur ce qui a tendance à être négligé. Je qualifie d'éthique cette attention car être témoin de situations où le consentement se pose demande d'orienter l'attention dans une direction qui soit favorable à l'expression de consentement. S'orienter dans cette direction, cela suscite un certain engagement, si l'on veut éviter d'être témoin d'abus ou d'agression, ou si l'on veut contribuer à désamorcer des dynamiques de domination. C'est un engagement qui est attentif à sa façon d'écouter, d'observer mais aussi de participer à ce qui fait parler les uns et les autres à propos de consentement. L'éthique est ici de l'ordre du *care*, une notion délicate à traduire en français, qui rassemble les notions de soin (prendre soin), de sollicitude, d'écoute attentive. Je vais conserver le terme de *care* pour évoquer ces notions toutes ensemble comme en anglais.

Le *careful engagement* (voir Nickelsen & Lydahl, sous presse, mentionné dans le chapitre 1) que suscite le consentement fait de lui ce que Maria Puig de la Bellacasa appelle un *matter of care*. Avec ce concept, l'objectif de Bellacasa est de réintroduire des affects dans la production de connaissance et ainsi encourager « *an ethos of care within the study of science and technology* ». Dans son sillon, j'encourage un même *ethos of care* dans les études qui s'intéressent à ce qui fait parler les personnes, en particulier pour éviter de leur faire dire des choses et pour porter attention aux situations dans lesquelles

on parle. Il ne s'agit pas de comprendre l'éthos du *care* comme l'entretien d'une attitude de bonne intention dans la création de connaissance, il s'agit plutôt de le comprendre comme un enjeu éthico-politique : « *The politics of caring have been at the heart of concerns with exclusions and critiques of power dynamics in stratified worlds. (...) Ways of studying and representing things can have world-making effects* » (de la Bellacasa, 2011, p. 86). Bellacasa prolonge le geste de Latour qui a permis de comprendre les questions de faits (*matters of fact*) et assemblages sociotechnologiques en questions de préoccupations (*matters of concern*)¹⁵⁶ : à partir d'un ancrage dans le courant féministe des STS (Barad, 2007; Haraway, 1991), elle propose de faire que des *matters of concern* deviennent des *matters of care*.

Une question de *care* (*matter of care*) est quelque chose à laquelle on porte attention (*care for*) et dont on s'occupe, dont on prend soin (*take care of*). Elle est ainsi proche du *matter of concern* (qui est une forme de *caring for*). Mais elle fait un pas de plus dans l'action : la question de *care* ne s'en tient pas à observer la préoccupation qui l'anime, elle implique une action qui permet de prendre soin, d'améliorer une situation. Puig de la Bellacasa précise ainsi les nuances qu'apportent le *care* aux *concerns* :

Care, however, has stronger affective and ethical connotations. One can make oneself concerned, but 'to care' more strongly directs us to a notion of material doing. (...) Understanding caring as something we do extends a vision of care as an ethically and politically charged practice, one that has been at the forefront of feminist concern with devalued labours. (...) For now, it is important to say that from this perspective to care signifies: an affective state, a material vital doing, and an ethico-political obligation. (Puig de la Bellacasa, 2011, p. 89-90)

Cette dimension active du *matter of care* est présente dans les définitions de l'éthique du *care* par Joan Tronto, une de ses premières théoriciennes¹⁵⁷. Tronto expose ainsi la notion

¹⁵⁶ Puig de la Bellacasa (2011) explique la notion de *matter of concern* dans le contexte de Latour, qui a invité à comprendre la création de savoir scientifique non comme un rassemblement autour d'états de fait (*matters of fact*), mais autour de préoccupations (*matters of concern*). La notion de « *concern* » introduit une connotation affective/émotive plus forte que le simple intérêt. Elle évoque en particulier « *trouble, worry and care* » (2011, p. 87).

¹⁵⁷ Avec son livre *Une voix différente*, Carol Gilligan (2003) est la première à avoir mis de l'avant la notion d'éthique du *care*, en mettant l'accent sur les formes invisibles et dévalorisées que prend le travail du *care* fait principalement par des femmes. Joan Tronto contraste avec Gilligan en montrant la dimension politique

de *care* (traduite en français par « sollicitude » dans son livre *Un monde vulnérable, Pour une politique du care*) à partir de l'utilisation du terme dans le langage quotidien :

Au niveau le plus général, la sollicitude connote une forme d'engagement; la manière la plus simple de le démontrer est d'envisager sa forme négative : « je ne m'en soucie pas » (*I don't care*). Mais le type d'engagement connoté par la sollicitude n'est pas de même sorte que celui qui caractérise une personne mue par ses intérêts. Dire « je ne m'en préoccupe pas » n'est pas l'équivalent d'être indifférent.(...) La sollicitude semble comporter deux aspects supplémentaires. Premièrement, elle implique de tendre vers quelque chose d'autre que soi : elle n'est ni autoréférente ni aut centrée. Deuxièmement, elle suggère implicitement qu'elle va conduire à entreprendre une action. (...) Sémantiquement, la sollicitude est indissociable de la notion de charge [Noddings, 1984]; se soucier de quelqu'un ou de quelque chose implique davantage qu'une simple envie ou un intérêt passager, mais bien plutôt l'acquiescement à une forme de prise en charge. (Tronto, 2008, p. 243)

Le *care* implique ainsi un engagement et une prise en charge qui amène une préoccupation (*matter of concern*) à sortir de sa dimension interne, comme quelque chose qui nous habite pour devenir une action. En d'autres termes, la question de *care* est une préoccupation qui s'extériorise en quelque chose que l'on fait pour porter un changement dans une situation, un état. « *Caring involves a notion of doing and intervening* » (Puig de la Bellacasa, 2011, p. 89) afin de « maintenir, perpétuer et réparer notre "monde", en sorte que nous puissions y vivre aussi bien que possible. Ce monde comprend nos corps, nous-mêmes et notre environnement, tous éléments que nous cherchons à relier en un réseau complexe, en soutien à la vie » (Fischer & Tronto, 1991, p. 40 cité dans Tronto, 2008). Cette intervention est de l'ordre du *careful engagement* (voir chapitre 1, p. 49), qui ne s'arrête pas à ce que l'on observe compter, à ce qui se manifeste dans les tours de parole, mais qui cherche à *faire compter* certaines préoccupations, certaines valeurs, certaines voix, par souci de leur permettre de s'exprimer ou d'être adressées. C'est en faisant compter ces préoccupations que l'on s'inscrit dans la perspective idéale de « maintenir, perpétuer et réparer notre monde ».

du *care* : cela permet de faire en sorte que le *care* soit moins essentialisé comme une question et pratique propre aux femmes.

Dans mon étude, cette intervention cherche souvent à faire compter la préoccupation du consentement, et celle qui lui permet d'être valable selon les conditions dans lesquelles il sera demandé. Et quand cette préoccupation du consentement risque de moins compter, d'être éventuellement écartée face aux obstacles qui se présentent à elle, c'est l'horizon éthique d'épanouissement et d'égalité qui vient soutenir l'intervention, en suscitant le sentiment que cela vaut la peine de s'y attarder. Cette intervention s'inscrit dans la perspective d'un monde où les habitudes d'interactions seraient modifiées, réparées, améliorées pour que soient désamorcées les tendances de domination qui parlent pour les autres sans les consulter, qui prétendent savoir ce que les autres veulent ou ce avec quoi ils sont d'accord, qui ne respectent pas l'intimité partagée quand on donne accès à soi – par le biais de nos corps comme de nos données personnelles.

Par la notion de *matter of caring* – qui reprend celle de *matter of care* de Bellacasa, en insistant sur la pratique processuelle du *caring* –, je propose ainsi quelque chose de différent de l'approche CCR et CCO qui étudie les manières de parler au travers de multiples figures. Il s'agit de porter attention aux façons par lesquelles on *care for*, mais aussi comment on *take care of* ces préoccupations. Cela peut sembler similaire à ce que Matte et Bencherki (2019) font dans leur analyse des *ethical matters of concern (EMCs)*, tels qu'ils se manifestent dans deux situations extraites d'une immersion au sein de Médecins Sans Frontières (MSF) dans un camp de réfugiés au Kenya. Dans leur article, la matérialisation de l'éthique au fil de la manifestation des préoccupations éthiques (EMCs) prend finalement la forme d'une décision de ce qui est bon ou non, c'est-à-dire une décision de l'ordre du *take care of*, afin d'opter pour le meilleur soin accordé aux patients. Dans la première situation, la médecin, qui propose de décider en équipe qu'une personne reste en tout temps avec les patients, manifeste un *matter of caring*. La deuxième situation expose deux *matters of caring* qui se confrontent : l'équipe de MSF insiste sur l'intubation d'une jeune fille pour assurer sa survie, tandis que la tante à son chevet tient à nourrir elle-même sa nièce conformément à la tradition familiale. C'est finalement le *matter of caring* de MSF qui l'emporte au nom d'un standard éthique de

l'organisation qui implique de donner plus de poids à la version du bon soin qu'est de sauver la vie d'une patiente par des moyens scientifiques éprouvés, plutôt qu'à celle portée par les coutumes locales et le souci du confort d'une proche souffrante.

Leur démarche était principalement d'observer comment des préoccupations éthiques se matérialisent dans les conversations par lesquelles une organisation comme MSF se constitue au quotidien. L'analyse CCO permet d'identifier comment des enjeux éthiques au sein d'une organisation se matérialisent au fil des pratiques de ses membres, c'est-à-dire comment se manifestent des préoccupations éthiques, lesquelles sont amenées à peser davantage dans la décision prise. Mais comme l'analyse porte principalement sur ce qui est manifesté dans des interactions, ce qui a plus de poids est en même temps ce qui est manifesté le plus explicitement : cela correspond le plus souvent à la voix des membres de l'organisation MSF, ceux au travers de qui la figure de l'organisation s'exprime, conférant à leur voix une certaine autorité. Or, en s'en tenant à ces voix et interactions porteuses de cette autorité organisationnelle, il y a le risque pour les analystes CCO de mettre en lumière principalement les pratiques et les acteurs qui ont le pouvoir de faire entendre leur voix et peser dans les décisions. Cela rejoint la critique de l'*Actor-Network-Theory* (ANT) qui a inspiré ce courant de la CCO à ses débuts :

[Star] argues that the ANTs see only through the eyes of the powerful, and that they base their studies on actors who have the power to ally themselves with others, build networks and empires, in other words, entrepreneurs. They follow the actors, but only those with the power to make their will prevail. (Asdal et al., 2007, p. 31-32)

Pour étudier un phénomène comme le consentement, cette façon d'aborder ce qui s'exprime ne peut pas bien fonctionner si l'on veut aller au-delà d'un constat qui s'en tiendrait à la manifestation d'un accord ou d'un refus. Le consentement nécessite en effet d'être abordé par les yeux de celles et ceux à qui il est demandé – tandis qu'il est demandé souvent par des personnes ou organisations plus puissantes.

Un contraste se dessine alors dans ma démarche, qui initialement s'intéressait au même phénomène (la matérialisation de l'éthique dans une organisation) : à la lumière d'un phénomène ambigu comme le consentement, mon attention s'oriente vers ce qui échappe, par habitude, à nos méthodes qui s'en tiennent aux tours de parole. Pour porter

attention à ces échappements et tenter d'améliorer ces habitudes, l'éthique du *care*, ses racines et implications féministes permettent de porter un regard vigilant à l'ambiguïté constitutive des tours de parole dans lesquels ne se présente explicitement que ce qui parvient à s'exprimer, dans lesquels existent des situations de silence et d'inhibition. Faire de la recherche CCO par une approche de *caring* permet de faire un pas éthique de plus que ce que font Matte et Bencherki (2019) : ne pas s'en tenir aux pratiques de *care* observées mais intervenir pour améliorer la situation; donner la parole ou du moins accorder plus d'attention aux êtres (humains et autres) dont l'autorité n'est pas aussi forte que celle de ceux qui parviennent à s'exprimer dans les tours de parole et documents; contribuer à reconfigurer le poids des figures qui font parler les uns et les autres afin d'aider à rendre ces interactions supportables pour la voix écrasée ou dégradée – comme c'est le cas pour les voix de femmes dégradées par les personnes féminines des agents conversationnels quand elles reproduisent des habitudes soumises aux agressions verbales et sexuelles (voir chapitre 2, p. 104).

La notion de *matter of caring* invite à faire quelque chose de différent. Je pars d'un angle différent : les éléments qui sont *talked into existence* ne sont qu'une partie de ce qui existe dans l'organisation. Certains existent par leur silence, réduits au silence (Bigo, 2018), entravés par des figures qui empêchent de parler, en quelque sorte désanimés. Mon angle d'approche est différent d'un point de vue épistémologique car il demande de prendre en compte la dimension située de l'observation et de l'analyse. Il est également différent sur un plan théorique : il s'agit de voir le négatif photographique, ou le silence, de ce qui est exprimé, ainsi que les conditions qui contribuent à l'expression, qu'elle se fasse par la parole ou le silence (et tous les bégaiements, hésitations, rougissements, gênes, confusions qu'il peut y avoir entre les deux). D'un point de vue méthodologique, il s'agit de ne pas s'en tenir à ce qui s'exprime clairement, mais à *configurer autrement* les conditions par lesquelles la parole s'exprime, par lesquelles on peut les voir/entendre exister dans l'organisation. Pour reprendre des termes ventriloques, cette re/configuration consiste à faire bouger les figures qui font parler, faire de la place à d'autres figures – c'est par exemple ce que l'on a entrevu avec l'évolution des

formulations par lesquelles Siri, Alexa et l'Assistant Google répondent à des requêtes de l'ordre de l'abus verbal, voire d'agression sexuelle (voir chapitre 2, p. 104). Ce travail de re/configuration est à la fois analytique et interventionniste : il s'agit non seulement d'analyser comment des conditions d'expression sont modifiées, améliorées, réparées, mais aussi, en tant qu'observatrice et analyste, de contribuer à la reconfiguration de ces conditions. Porter attention à la manifestation et l'expression de *matters of caring*, consiste ainsi à porter attention à des pratiques au-delà des conversations – par exemple, dans une organisation, comment une préoccupation se transforme en tâche à réaliser en un temps imparti, en l'objet d'une réunion dans un espace défini, en une fonctionnalité dans un système technologique. Si l'on fait de la recherche CCO (ou CCR), il s'agit en d'autres termes de la faire avec *care*.

Il y a ainsi dans la question de *care* une impatience, quelque chose qui n'est plus supportable en l'état, quelque chose qui pousse à intervenir¹⁵⁸ pour améliorer la situation plutôt que de s'en tenir à l'observer se dérouler. C'est comme si l'intranquillité inhérente à la préoccupation devenait si forte que le constat de ce qui est (des *matters* présents) et la reconnaissance des préoccupations appelaient une intervention, une contribution pour changer une situation, que cet état devienne petit à petit moins préoccupant, plus acceptable, plus confortable. C'est ce genre d'intranquillité qui m'a faite sortir de mon silence le 10 juillet 2019 : l'objection et l'agacement sont des traces d'intranquillité qui devient insupportable et moteur de passage à l'action, manifestation d'une préoccupation nécessitant intervention, donc d'une préoccupation qui devient un *matter of caring*. C'est également une intranquillité de cet ordre qui a fait intervenir la coach d'Adèle Haenel auprès du réalisateur pour interrompre une situation dérangeante.

Ces interventions ne sont pas confortables car elles s'immiscent dans des interactions pour les faire bouger. Les remises en question de la régisseuse sur le tournage avec Adèle Haenel soulignent le doute inconfortable : « Est-ce que j'ai rêvé ? Est-ce que je suis folle ? » (cité précédemment, p. 197). Dans le cas de la réunion du 10 juillet 2019,

¹⁵⁸ Cela interroge la complicité de la chercheuse avec l'organisation et les pratiques étudiées, je reviendrai sur ce point dans le chapitre suivant.

oser intervenir pour transformer la situation n'est ni fluide ni agréable pour moi : mes interventions se font au fil de phrases incomplètes, selon un rythme précipité puis plus lent, ponctué de répétitions, de bégaiements, de rougeurs qui montent au visage, surprise d'entendre le silence se faire quand je parle un peu plus, je reprends ce que les autres disent, avec ajouts, modifications, précisions, hésitations. L'intervention suscitée par le *matter of caring* implique ainsi de heurter l'expérience de tournage ou de réunion : on heurte l'expérience de ce qui a l'habitude ou le pouvoir de s'exprimer majoritairement. Ce heurt d'expérience a été une tension très présente chez Prélude à propos du travail sur le consentement, comme nous le verrons dans les chapitres suivants.

L'intervention éthique présente dans les *matters of caring*, qui vise une transformation, permet en même temps d'inscrire les *matters of caring* dans une temporalité qui n'est pas que du déjà-là comme les *concerns* : les *matters of caring*, par l'engagement vers la transformation qui leur est inhérent, ont une dynamique de projet. Compris comme une activité propre aux designers, Vial décrit ainsi le projet :

Un projet, c'est un ensemble de propositions formelles originales qui offrent aux usagers une expérience-à-vivre de nature à satisfaire leurs besoins et susceptible d'enchanter leur existence. C'est une démarche de projection ou d'anticipation qui consiste à imaginer, à partir de l'état existant, des formes innovantes de vie et d'usage. En ce sens, le designer est celui qui propose un état projeté de la réalité, c'est-à-dire une possibilité d'avenir réalisable. (Vial, 2010, p. 70)

Le *care* ne visant pas le plaisir, mais le maintien, la perpétuation ou la réparation du monde (Tronto, 2008), la transposition de cette définition du projet dans le contexte du *care* dirait « susceptible d'améliorer leur existence » plutôt qu'« enchanter ». Cette dimension de projet se retrouve au niveau de ma posture de recherche (un projet portant sur le design de consentement au fil du développement d'un agent conversationnel) mais aussi au niveau des marqueurs par lesquels identifier des *matters of caring*. En plus des marqueurs affectifs qui indiquent une intranquillité qui pousse à l'action (marqueurs d'inconfort, d'injustice, de doute, par exemple), ou encore des marqueurs d'attachements (sous forme de répétitions, insistance, silences explicités dans les notes de journal de bord), il s'agit de porter attention aux devenirs de certaines de ces préoccupations intranquilles. Je vais par exemple m'intéresser dans le chapitre 6, à la

façon dont la préoccupation du consentement devient un module de conversation dans le design de l'agent conversationnel, comment la préoccupation pour la fluidité de l'expérience aboutit à une demande de consentement sous forme de liste qui déplace l'attention portée à ce moment d'interaction – les tâches cherchant à lui donner de la place deviennent des tâches cherchant à améliorer la validité du consentement demandé. Parfois relevant de la ruse, de la contrainte ou du dessein, ces préoccupations faites *matters of caring* passent par ces différentes acceptions du design (cf F. Fischer, 2019, citée dans le chapitre 1, p. 67) qui aide à transformer les préoccupations en solutions plus concrètes.

Étudier de tels *matters of caring* permet d'aligner mon approche théorique et méthodologique avec l'objet, en même temps concept, au cœur de mon terrain : le consentement compris comme un enjeu éthique et objet de design, au fil des pratiques des membres d'une organisation qui développent un agent conversationnel. Avec la notion du projet comme un *matter of caring*, le concept de consentement s'incarne dans les pratiques. Autrement dit, la notion de *matter of caring* permet de penser avec le concept de consentement, qui demande de porter attention au négligé, aux conditions dans lesquelles on demande ou fait dire quelque chose, qui demande aussi de faire ralentir d'autres pratiques pour faire de la place à ce moment de consentement. Un tel moment demande des reconfigurations pour le faire respecter grâce à des formulations et des interactions vigilantes. Et mutuellement, le concept de consentement invite à voir comment le *matter of caring* implique une vigilance similaire à ce que fait le consentement, par son travail des limites auquel il invite, à la vulnérabilité potentielle, à des dynamiques qui empêchent, permettent ou rassurent (cf chapitre 3).

Avec cette organisation (d'abord Mr Y puis Prélude), nous cherchons la meilleure façon de demander (et obtenir) le consentement de futurs utilisateurs amenés à utiliser leur agent conversationnel. Ce projet est un *matter of caring* car le consentement se loge dans un espace souvent négligé, où il est possible que l'expression de la personne soit empêchée – par des *dark patterns* par exemple, par l'habitude de ne pas s'y attarder, par l'évitement de certaines questions ou encore par l'absence de certaines pratiques

respectueuses des données personnelles. Travailler les interactions portant sur la demande de consentement demande un effort de faire en sorte que l'expression soit possible, que la personne soit réellement consultée, que son expression soit favorisée plutôt qu'inhibée par des choses qui la font parler (informations, formulations, interactions). Le consentement permet ainsi d'interroger le désirable, le voulu, le supportable dans l'interaction, tout en étant un moment d'interaction étrange, qui rend visible ce qu'on a l'habitude de considérer implicite et tacite.

La fin visée de ce projet est d'améliorer des situations d'interaction, en ancrant le consentement dans les habitudes d'interaction – c'est un *matter of caring* technologique (propre aux agents conversationnels et autres dispositifs technologiques) et social (inhérent aux mouvements éduquant au consentement, invitant à faire de la place à l'expression ou l'inhibition du consentement et affinant l'écoute de ces situations). De même qu'il est guidé par une émotion intranquille propre au sentiment éthique et au *matter of caring*, ce projet est plein d'inconforts puisqu'il évolue en tension avec ce qui est habituellement considéré comme des interactions agréables et divertissantes – la tension étant inhérente au travail qui cherche à faire bouger des lignes, à configurer autrement. Par ailleurs, comme le consentement, le travail que fait le *matter of caring* s'accompagne d'une « vulnérabilité qui implique la confrontation, les empêchements, les modifications du désir ou de la volonté initiale » (Brugère, 2013, citée dans le chapitre 3, p. 176). J'ajouterais à ces moyens de reconfigurations le ralentissement (Stengers, 2018), qui est nécessaire quand on s'intéresse au possible, à l'implicite et au faire autrement. Le concept de *matter of caring* cherche ainsi à contribuer à des possibilités d'actions (respectueuses) ensemble, il implique une *practice of togetherness* qui est également au cœur de mes pratiques ethnographiques, relationnelles (voir chapitre 5).

À la lumière de ce concept interventionniste de *matter of caring*, je vais à présent reconstituer le travail fait sur le consentement avec Mr Y et Prélude en me guidant par deux questions de recherche (QR) : (QR1) Comment cela se fait, d'aborder le consentement comme un *matter of caring* dans une organisation qui design un agent conversationnel ? et (QR2) Qu'est-ce que cela fait, de faire du consentement un *matter*

of caring ? La première question cherche à montrer les pratiques impliquées pour faire du consentement un *matter of caring* dans cette organisation. La deuxième question, en demandant « qu'est-ce que cela fait ? », interroge le travail (pragmatique) que fait le *matter of caring* quand il s'attarde sur un objet comme le consentement dans cette organisation. Ce qui m'intéresse dans cette deuxième question est de voir comment le travail du consentement comme concept se manifeste dans le travail du *matter of caring*.

Conclusion : une éthique du care pour écouter parler

J'ai montré dans ce chapitre que l'activité qui fait parler, ou la passivité à travers laquelle on parle, demande une éthique particulière quand on s'intéresse au consentement. Par la vulnérabilité dans laquelle il loge, le consentement demande en effet qu'une attention particulière soit portée non seulement à ses façons d'être exprimé – que ce soit par un acte clair ou, plus souvent, par de l'ambiguïté et du silence –, mais aussi aux conditions dans lesquelles il est demandé. Le consentement cherche à désamorcer des dynamiques de domination, en se faisant respecter au nom de l'égalité entre les personnes qui agissent ensemble. Alors, quand on analyse des situations où le consentement est en jeu – c'est-à-dire toute situation où des personnes interagissent en se respectant mutuellement – il est particulièrement important de porter attention aux potentielles figures qui empêchent de parler, pour favoriser les situations de consentement qui visent un certain épanouissement.

L'éthique que je propose pour ces situations, est une éthique du *care*, qui fait du consentement un *matter of caring* appelant une intervention pour améliorer ces situations. L'intervention éthique dans l'interaction implique de considérer les voix silencieuses non comme des voix absentes des tours de paroles, mais comme des voix qui comptent et qui demandent de remanier la façon par laquelle nous interagissons. Il ne s'agit pas de les forcer à parler, mais de tenter de désamorcer une habitude langagière liée à une forme de domination qui tient pour acquis de savoir implicitement ce que l'autre veut ou ce avec quoi elle est d'accord. L'intervention éthique par une approche

CCR peut ainsi être elle-même un *matter of caring* et aller au-delà du constat de la cultivation de certaines figures mobilisées et remobilisées de façon itérative : elle peut participer à la « *transformation of habits of thinking and seeing that happens through attachment to particular concerns, interests and commitments* » (Puig de la Bellacasa, 2011, p. 96). Le concept de *matter of caring* est pour la qualité de l'attention ce que le consentement est pour la qualité des interactions : animé par un idéal de respect, d'épanouissement et d'égalité, il est constitutif de ce qui lie des êtres (humains, organisationnels et autres) qui se constituent mutuellement par leurs interactions.

Dans un contexte où le consentement est abordé dans son usage numérique et dans les activités d'une organisation qui développe un agent conversationnel, comment ce *matter of caring* se manifeste-t-il ? Dans les deux prochains chapitres, je vais explorer cela par le biais de deux questions de recherche : (QR1) Comment cela se fait, d'aborder le consentement comme un *matter of caring* dans une organisation qui design un agent conversationnel ? et (QR2) Qu'est-ce que cela fait, de faire du consentement un *matter of caring* ? Le prochain chapitre (chapitre 5) met en lumière comment le *matter of caring* a été présent dans mes pratiques de recherche ethnographique, en particulier par ma réflexivité et ma complicité qui ont permis à l'intervention de prendre la texture d'un tissage relationnel, participant à la constitution de l'organisation. Le chapitre d'après (chapitre 6) reconstitue le travail du consentement comme *matter of caring* tout au long de mon immersion avec Mr Y et Prélude.

Chapitre 5

Pratiques ethnographiques : Complicité, réflexivité et compromis

« *Am I going to practice peace, or am I going to war?* »
(Pema Chödrön, *When things fall apart*, p. 12)

Introduction : quelle complicité ?

La crainte de se réduire au silence
Comment rendre visible sans exposer, sans condamner systématiquement
Comment transformer
Pacifiquement, sans violence

(Journal de bord, notes réflexives du 5 novembre 2019)

Quand j'écris ces mots dans mon journal de bord, j'explicité certaines figures qui m'animent ou m'entravent en tant que chercheuse : un sentiment de crainte à l'égard de mon silence lié à mon activité d'observation ainsi qu'à ma situation qui adopte plus facilement une posture de silence plutôt que d'oser parler, des préoccupations pour « rendre visible » et « transformer ». Ces préoccupations sont de l'ordre d'un *doing*, avec un certain souci : « ne pas exposer », ne pas « condamner », agir « pacifiquement, sans violence », c'est-à-dire un souci de contribuer à rendre visible et transformer dans un mouvement qui épouse les dynamiques organisantes à l'œuvre sur ce terrain. C'est là ce qui fait l'objet d'un *matter of caring* : travailler, de l'intérieur d'une organisation, comment des interactions, par le biais d'un agent conversationnel, peuvent faire de la place au consentement. Cela implique de travailler les conditions – en particulier inhérentes à l'activité de design de conversation – qui amènent quelque chose (un agent conversationnel) et quelqu'un (une utilisatrice) à demander et exprimer son

consentement. Cela s'accompagne en même temps d'une tension : quelle complicité s'engage ?

Ce mot « complice » arrive dans mes notes à la toute fin de mon terrain. Dans la soirée du 27 janvier 2020, après avoir assisté à la présentation d'une chercheure juriste lors de la conférence ACM*FAT (devenue FAcCT¹⁵⁹) à Barcelone, j'écris dans mon journal de bord : « Son raisonnement pointilleux et son agacement m'ont rappelé mes intentions et accrochages de départ chez Prélude. Ça me rappelle ces réflexes qui se sont assouplis petit à petit, passant de critiques à complices ». Cette réflexion est suivie d'une gêne : « N'empêche qu'en m'étant assouplie, je n'ai pas bien incarné la rigueur d'un règlement. Mais ce n'était pas non plus dans mes cordes, je n'ai pas de formation en droit, et j'ai heurté cet obstacle plusieurs fois ». Ma complicité n'est donc pas teintée de fierté mais elle porte une limite qui peut être présente dans de nombreuses organisations comme Prélude : « En même temps, ça se passe peut-être comme ça dans de petites organisations : des gens qui ne maîtrisent pas un sujet s'y frottent et cherchent à trouver des solutions » (Journal de bord, notes réflexives du 27 janvier 2020). Ma complicité, la gêne dont elle est teintée et les approximations avec lesquelles j'ai bricolé marquent ainsi ma pratique ethnographique qui cherche à rendre compte de ce que cela fait de travailler l'éthique dans les activités d'une organisation comme celle-ci.

La complicité est un thème courant en ethnographie (Marcus, 1997, 2001; Neyland, 2008; Ybema et al., 2009)¹⁶⁰, elle se pose en particulier quand la chercheure devient de plus en plus familière¹⁶¹ de l'organisation ou du collectif dans lequel elle s'immerge depuis plusieurs semaines voire plusieurs mois. Elle s'est ainsi posée lors de mon immersion de huit mois au sein de Mr Y et Prélude, deux startups devenues une (voir

¹⁵⁹ ACM Conference on Fairness, Accountability, and Transparency. Il s'agit d'un rendez-vous majeur, initialement en informatique mais devenu multidisciplinaire, pour les chercheurs qui s'intéressent aux enjeux de transparence, justice sociale et responsabilité des technologies : <https://facctconference.org>.

¹⁶⁰ Marcus (1997, 2001) contraste la complicité qui se tisse sur le terrain, d'une façon qui nuance avec les notions ethnographiques plus traditionnelles et romantiques des « rapports » et « collaboration » entretenues avec les « sujets » sur le terrain.

¹⁶¹ La familiarité qui s'établit avec l'organisation peut porter le risque de « *going native* » (Neyland, 2008, p. 81), auquel cas la chercheure n'a plus de distance avec l'organisation. Pour entretenir la distance nécessaire pour élaborer une pensée de ce dont l'ethnographe est témoin, la pratique réflexive est un outil important.

chapitre 6), au fil de ma participation aux activités de design de leurs agents conversationnels. Par une entente Mitacs¹⁶², mon accès à ces organisations s’est fait au nom d’un rôle particulier : en tant que chercheuse en éthique et en expérience utilisateur, j’accompagne leurs activités pour orienter le design de sorte que s’expriment certaines valeurs, dont celle du consentement et du respect de la vie privée. Plus précisément, mon rôle est officiellement de concevoir le modèle par lequel le consentement serait demandé par les agents conversationnels de Mr Y et Prélude; je suis en quelque sorte *ethics owner* (voir chapitre 1, p. 45) du design de consentement.

Par cet accompagnement (Verbeek, 2011, voir chapitre 1, p. 71), ma posture se distingue de postures éthiciennes qui mettent un point d’honneur à se tenir à distance des activités quotidiennes des organisations qu’elles observent, surveillent voire sanctionnent. Elles tiennent à préserver par là une indépendance qui permet d’évaluer des pratiques de façon plus impartiale que lorsqu’on y participe. Par contraste avec ce rôle évaluateur, j’aborde l’activité éthique de l’intérieur de ces organisations, d’une façon relationnelle qui m’amène à participer à leur constitution – c’est ce que je vais montrer plus en détail dans ce chapitre. Loin d’incarner l’idéal d’indépendance dans la conduite d’activités liées à l’éthique, ma posture se révèle plutôt incarner des dynamiques d’interdépendance pour amener des valeurs telles que le consentement et le respect de la vie privée à s’exprimer.

Cette posture ne permet pas de faire triompher la pureté que peut laisser imaginer l’éthique quand elle est abordée par l’angle de ses principes à imposer. Le doute de ce à quoi je contribue est omniprésent – cela ne m’est pas confortable de participer à l’usage d’agents conversationnels et objets connectés qui ont tendance à collecter des données de façon intrusive et en trop grande quantité. Mais plutôt que de m’accrocher à la pureté des principes éthiques et ce qu’ils dictent, je prends l’existence de ces agents conversationnels et objets connectés comme point de départ : « *what happens if we start from here?* » (Shotwell, 2016, p. 5). Ma complicité, le doute de ses inconforts, l’affûtage

¹⁶² Mitacs est un organisme subventionnaire liant industries et universités : <https://www.mitacs.ca/fr>. Mon projet sur le consentement avec Mr Y puis Prélude a été financé sous forme de deux unités de stages de recherche de quatre mois chacune (avril-juillet puis septembre-décembre 2019).

de sa vigilance, sont les points de départ de mon action. Ces actions, interventions et résultats sont peuplés de compromis¹⁶³. Le constat de ces compromis se manifeste sous forme de déception, de gêne, parfois d'apaisement coupable de privilégier le relâchement d'une tension dans l'équipe plutôt que de m'accrocher à une restriction que je voudrais imposer. Mais cette complicité et ces compromis permettent en même temps de façonner des pratiques de responsabilité¹⁶⁴, dans l'équipe mais également envers ma propre activité de recherche.

Dans ce chapitre, je vais montrer comment – par complicité et réflexivité – mes pratiques ethnographiques ont abordé le consentement comme un *matter of caring* dans cette organisation (QR1). Je vais me concentrer principalement sur cette complicité sous la forme de *practice of togetherness* qu'a prise ma pratique réflexive au fil de mon travail ethnographique au sein de Mr Y et Prélude. J'emprunte à Tim Ingold (2011) cette expression de « *practice of togetherness* » pour désigner une démarche de recherche qui consiste à étudier *avec* et apprendre *de* l'organisation dont on épouse les contraintes, le rythme, les ambiances etc. plutôt que d'apprendre *sur* cette organisation dans une relation distanciée. Je vais montrer comment la pratique réflexive s'est manifestée comme une expérience relationnelle de ces organisations en constitution. À partir de la métaphore du tissage relationnel d'Ingold (2011), je vais montrer la constitution mutuelle qui s'est opérée entre ce terrain, le concept et objet du consentement et moi-même comme chercheure¹⁶⁵. Ces précisions sur la pratique réflexive permettent d'expliquer comment j'ai collecté mon matériel ethnographique. Je préciserai ensuite les méthodes

¹⁶³ Cette réflexion sur la complicité et les compromis s'inspirent de la bouffée d'air qu'a été la lecture d'Alexis Shotwell et son introduction « *Complexity and Complicity: An Introduction to Constitutive Impurity* » dans son livre *Against Purity: Living Ethically in Compromised Times* (2016, p. 1-19).

¹⁶⁴ J'emprunte à Alexis Shotwell cette formule de « pratiques de responsabilité ». Elle l'emploie face au constat de notre complicité avec les phénomènes du monde auxquels nous sommes liés même si nous préférerions ne pas l'être, car ils sont vecteurs de souffrance pour certaines personnes : « *We need to shape better practices of responsibility and memory of our placement in relation to the past, our implication in the present, and our potential creating of different futures* » (2016, p. 8).

¹⁶⁵ Tout ce que je dis sur la pratique réflexive dans ce chapitre est repris de l'article que j'ai publié avec Boris H. J. M. Brummans : « La recherche CCO comme pratique réflexive : une approche relationnelle » dans le numéro 59 de la revue *Communication & Organisation* (2021). Pour ce chapitre, j'ai étoffé plusieurs passages de l'article pour le lier au concept de *matter of caring*.

d'analyse que j'ai utilisées pour faire sens de ce matériel à la lumière du concept de *matter of caring*, pour voir ce que cela fait, de faire du consentement un *matter of caring* dans une organisation comme Mr Y et Prélude (QR2).

Pratiques de collecte de données

Le matériel constitué pour cette thèse est collecté lors de mon immersion ethnographique (Neyland, 2008; Ybema et al., 2009) de huit mois au sein des startups Mr Y (les deux premiers mois) puis Prélude (chevauchant ces premiers mois puis prenant le dessus les mois suivants). J'étais initialement intéressée à observer comment des enjeux éthiques, ou *ethical matters of concerns* (Matte & Bencherki, 2019), se matérialiseraient au fil des pratiques de ces organisations. Mon rôle formalisé par une entente Mitacs m'amène cependant à adopter une posture plus engagée que l'observation distanciée. Cette observation participante se fait au sens d'Ingold (2014) qui met l'accent sur une observation dont le but n'est pas « *to represent the observed but to participate with it in the same generative movement* » (Ingold, 2011, p. 223). Teintée d'une approche d'éthique du *care*, cette participation est de l'ordre de l'intervention éthique présentée dans le chapitre précédent. Cette dimension d'intervention dans le design des agents conversationnels dote également mon approche d'une démarche similaire à la recherche-action telle qu'elle se définit comme « une action de recherche et d'éducation liée à la résolution d'un problème concret dans le but de soutenir des changements bénéfiques, de contribuer à l'apprentissage des personnes concernées et d'améliorer les connaissances dans un système » (Guay et al., 2016, p. 548).

Je garde les traces de cette observation dans un journal de bord tenu quotidiennement, où les notes de terrain et les notes réflexives s'entremêlent pour former une source primaire de données ethnographiques à analyser (Emerson et al., 2011). À la fin de ce travail de terrain, en janvier 2020, mon journal de bord compte ainsi 606 pages. Les notes de terrain sont rédigées à partir de notes prises lors de plusieurs types d'activités : ma participation à 65 réunions avec des membres de l'équipe, ma

participation également à 3 séances d’entrevues avec des potentiels utilisateurs de la Voix de Montréal à qui Malik fait tester le prototype de l’agent conversationnel, la consultation de documents organisationnels, ma contribution à 33 de ces documents, des recherches pour répondre à certaines questions sur des pratiques de protection des données personnelles, des tests que je fais de certaines applications comme l’Assistant Google et des agents conversationnels similaires à ceux de Mr Y et Prélude.

Je fais ces tests en utilisant une technique de « *app walkthrough* » qui « *systematically and forensically step[s] through the various stages of app registration and entry, everyday use and discontinuation of use* » (Light et al., 2018, p. 881). Cela me permet entre autres de voir comment le consentement est demandé dans des applications d’agents conversationnels en santé mentale (tels que les chatbots Wyzia et Woebot¹⁶⁶) pour prendre exemple dessus. Cela me permet également d’identifier l’existence de l’interface « mon activité »¹⁶⁷ de l’Assistant Google, une interface sur laquelle repose l’expérience utilisateur que Google propose pour la gestion du consentement, ou plutôt pour donner du « contrôle » (voir chapitre 3, p. 137) à l’utilisateur sur ses paramètres de confidentialité et de traitement des données personnelles. Je teste différents chemins pour trouver de l’information à ce propos, via l’Assistant, les documentations pour développeurs et les politiques de confidentialité. Cela me permet de retracer comment Google pose des garde-fous aux organisations qui développent des Actions sur son Assistant, pour limiter l’accès qu’elles peuvent avoir aux données personnelles de leurs utilisateurs; ce qui limite en même temps la possibilité d’une démarche de *privacy-by-design* de la part de ces organisations (voir chapitre 3).

Avec mes notes réflexives, l’ensemble de ces notes développent au fil de mon immersion une sorte de conversation textuelle entre chercheure (moi), organisation (Mr Y et Prélude) et objet de recherche (consentement); autour du processus de recherche travaillé par le processus organisant et l’objet de recherche; autour également de mes

¹⁶⁶ Voir <https://www.wysa.io> et <https://woebothealth.com>. À noter que les versions actuelles (avril 2022) sont différentes de celles que je consultais en mars-avril 2019 pour identifier les similitudes et les contrastes avec ce que Mr Y préparait pour son chatbot.

¹⁶⁷ Voir <https://myactivity.google.com/myactivity>

actions, en tant que chercheure, qui contribuent à l'organisation en constitution, par le design de ses agents conversationnels, les réunions et documents qui le soutiennent. Cette conversation textuelle, à la fois descriptive et réflexive, s'accompagne par ailleurs de 24 enregistrements audio¹⁶⁸ de réunions qui permettent d'identifier des résonances et porosités entre les manières de faire sens¹⁶⁹ et de parler – des manières qui évoluent au fil des réunions entre membres de l'organisation, des manières qui évoluent aussi du côté de la chercheure au fil de ses notes.

L'ensemble de ce matériel textuel et audio me permet ensuite de retracer les étapes par lesquelles le consentement se constitue comme une préoccupation et un *matter of caring* chez Mr Y et Prélude (QR1), les bifurcations que cette constitution prend, les tensions qu'elle soulève – je les mettrai plus en lumière dans le chapitre suivant. Cet ensemble me permet également de voir plusieurs nuances dans le travail que fait le consentement dans cette organisation où il se constitue comme *matter of caring* (QR2) – je décèle ce que ce travail *me* fait (il provoque des gênes, de l'inconfort, parfois de l'effacement, des efforts de clarifications, etc., voir chapitre 6), mais également le travail de limites, de paroles qui cherchent à rassurer, d'empêchements qu'il fait en tension avec d'autres préoccupations (voir chapitre 6 également).

Mais ce travail du *matter of caring* (QR2) ne se passe pas seulement dans un temps postérieur à la constitution de mon matériel. Il s'effectue dès mes pratiques de collecte de données, en limitant la quantité de mes enregistrements et en restreignant ma prise de notes aux moments concernés par mon objet de recherche, le consentement. Sur le total de 65 réunions auxquelles j'assiste, je n'en enregistre que 24 car je n'ai pas le consentement explicite des participants pour les 41 autres réunions – soit parce que les réunions commencent rapidement et je n'ai pas le temps de le demander, soit parce que je décide de ne pas enregistrer car le sujet de la réunion n'est pas lié à mon projet de

¹⁶⁸ Le format audio des enregistrements plutôt que vidéo s'est imposé car Paul et Félix ne voulaient pas être filmés. Ils ne voulaient pas non plus risquer de donner l'impression aux employés qu'une caméra les surveillerait.

¹⁶⁹ Bien qu'étant un anglicisme, « faire sens » me permet d'évoquer au fil du texte le processus de *sensemaking* de Weick (1979), qui est un processus d'organisation en constitution – un *organizing*.

recherche. Cette absence de consentement des personnes qui participent à ma recherche en étant membre de Mr Y et Prélude amène donc à limiter ma pratique d'enregistrement, par souci de respecter ces absences de consentement.

Autre chose limite ma pratique d'enregistrement : elle fait souvent l'objet de blagues liées à la potentielle surveillance « par l'éthique » (métonymie par laquelle certains me désignent parfois avec humour). Alors, quand le contexte de la réunion peut manifester des tensions, je n'enregistre pas car cela me semble inapproprié de potentiellement donner l'impression d'être surveillé. Par souci de respecter des situations de tension non-liées à mon objet de recherche, par souci également de favoriser mon inclusion dans l'équipe sans encourager l'idée que ma présence est synonyme de surveillance, je limite la quantité de données enregistrées et la diversité de situations de tension dans le matériel que je constitue. En guise d'alternative, je prends des notes comme à chaque réunion, mais je ne note que ce qui est en lien avec les activités auxquelles je suis mêlée.

À partir de ces notes, je sélectionne les moments de réunions qui abordent les thèmes liés à mon projet de recherche (consentement et gestion des données personnelles) pour étoffer mes notes de terrain. 4 de ces réunions portent principalement sur ces thèmes (le 10 juillet 2019, le 11 septembre 2019, le 23 septembre et le 23 octobre 2019) : je les retranscris donc pour les intégrer à mon matériel à analyser. Ces retranscriptions me permettent plusieurs choses : le 1^{er} extrait de la réunion du 10 juillet 2019 me permet d'élaborer en partie le concept de *matter of caring* (voir chapitre 4, p. 190); la réunion du 11 septembre 2019 avec Data Protection Consulting me permet d'apporter des précisions sur des technicalités liées au consentement et au traitement des données personnelles selon le RGPD (voir chapitre 3, p. 155 et chapitre 6); les autres extraits du 10 juillet 2019, ainsi que ceux des 23 septembre et 23 octobre 2019 permettent de relever les étapes et les thèmes soulevés par le travail du consentement abordé comme un *matter of caring* (QR1, voir chapitre 6). Ces retranscriptions me permettent également de souligner les nuances et les confrontations qu'implique le *matter of caring* porté sur le consentement (QR2, voir chapitre 6 également). La pratique

par laquelle je sélectionne les moments pertinents dans ces retranscriptions se laisse ainsi guider par les mentions explicites du consentement mais aussi par l'évocation de conditions par lesquelles il sera demandé (par l'intermédiaire de quelles formulations, quels boutons; quels traitements de données personnelles seront impliqués). Ma pratique de sélection se laisse également guider par les traces de *matter of caring* (selon sa définition au fil du chapitre 4) pour approfondir l'exploration de mes questions de recherches, en particulier la deuxième (sur le travail que fait le *matter of caring*) : où y a-t-il modifications, réparations, améliorations ? Y a-t-il des silences explicités quelque part ? des signes d'intranquillité (interventions, interruptions, agacements, doutes) ?

Dans mes notes réflexives, les résonances et porosités entre les manières de faire sens débordent du cadre spatial et temporel des organisations dans lesquelles je suis allée plusieurs fois par semaine. Ces notes réflexives explicitent des troubles, des émotions, des inquiétudes, des attachements à l'égard du travail sur le consentement, de la conduite de ma recherche, du fonctionnement des agents conversationnels et leur place dans l'écosystème complexe de géants comme Google, et de certaines tensions dans l'équipe. Ces troubles, émotions, inquiétudes et attachements portent la marque d'intranquillité semé par le travail du *matter of caring* (QR2) porté sur le consentement.

Mais ces notes partagent aussi des réflexions à la lumière de lectures¹⁷⁰, événements et pensées sur le consentement, dans ses usages autres que numérique. Elles portent en somme les traces de ce que cela m'a fait de faire du consentement un *matter of caring* (QR2), dans cette organisation et plus largement, dans ce moment après le mouvement *#MeToo/#MoiAussi*, qui m'a interrogée sur l'évolution de nos habitudes d'interaction. Ces notes réflexives explicitent ainsi certains de mes silences, de mes balbutiements, hésitations, enthousiasmes, et parfois interventions portées par un sentiment d'urgence de ne pas laisser la situation se dérouler telle qu'elle est en train de le faire (voir chapitre 4, p. 190). Autrement dit, ces notes réflexives portent les traces de ma voix, si la voix est à la fois ce qui vibre, *voicing*, et ne vibre pas, *voiceless* (Ihde, 2007)

¹⁷⁰ Manon Garcia (2019) en particulier, mais également des articles de presse dans les sillages du mouvement *#MeToo*, ainsi que des infolettres et articles de blogue sur les enjeux éthiques, politiques et sociaux de l'intelligence artificielle et sa gouvernance.

– le *voicelless* ne s’entend pas mais peut avoir des allures de silence qui exprime quelque chose plutôt qu’il se rapporte à quelque chose qui n’existe pas (voir chapitre 4). Cette voix (la mienne) est aussi la seule à laquelle j’ai accès pour expliciter des silences selon ses propres mots.

Au début du travail de terrain, j’envisage de faire des entrevues semi-dirigées (Spradley, 1979) pour expliciter certaines zones d’ombres et potentiels silences avec chaque personne de Mr Y et Prélude. J’en fais quelques-unes avec Paul à propos de Mr Y – celles-ci me donnent plusieurs éléments de contexte sur l’évolution de Mr Y en tant qu’organisation, de sa naissance comme idée lors de la participation à une compétition d’innovation¹⁷¹, jusqu’au rachat des parts de ses associés quand Mr Y se mêle à Prélude début 2019. C’est lors de ces entrevues qui deviennent informelles que je perçois la sensibilité de Paul aux questions d’éthique, à sa volonté d’intervenir pour les prendre en compte dans le design de ses agents conversationnels – je perçois que l’attention qu’il porte à l’éthique n’est pas que d’un ordre instrumental pour apporter de la valeur à ses produits et son organisation, son attention est teinté de *matter of caring*, pour prendre soin des situations de vulnérabilité des utilisateurs de Mr Y (voir début du chapitre 6). J’ai à ce moment l’intention de réaliser des entrevues avec tous les autres membres de l’organisation. Mais à partir du moment où je suis davantage mêlée aux activités de Prélude, je me mets à poser les questions que je prévoyais pour les entrevues lors des réunions et des discussions informelles dans les couloirs ou autour de la machine à café. Ces questions orientent mon attention et certaines de mes interventions. En revanche, plus ma complicité dans les activités de Prélude se renforce, épousant le rythme soutenu de travail, moins je crée d’opportunités pour de potentielles entrevues formelles qui expliciteraient certaines situations.

J’entretiens tout de même des discussions avec chaque membre de l’équipe tout au long du terrain (en particulier à partir du mois de juillet). Je prépare plusieurs de ces discussions comme des entrevues informelles avec des questions à aborder pour expliciter des silences ou ambiguïtés liés à certaines tensions que je crois percevoir. Mais

¹⁷¹ Il s’agit du Coopérathon en 2016 (voir <https://cooperathon.ca>)

je ne les expose pas dans cette thèse, car elles n'ont pas de lien direct avec le travail sur le consentement dans le design de l'agent conversationnel, et par souci de respecter ces anciens collègues et participants à ma recherche envers qui je me suis engagée à ne pas dévoiler ces situations. L'une de ces personnes mentionne d'ailleurs son intention de retirer son consentement de ma recherche un jour, puis a souhaité à nouveau participer, mais cette interruption de consentement m'amène à mettre de côté plusieurs de nos discussions liées au projet. Que fait donc le *matter of caring* à l'égard de cette pratique de recherche que sont les discussions ou entrevues informelles ? (QR2)

Abordée comme un *matter of caring*, la pratique de discussions informelles en guise d'entrevues fait ainsi face au paradoxe de se donner accès à plusieurs informations qui dévoilent les silences de certaines voix, mais à ne pas les utiliser, pour respecter les vulnérabilités qui s'explicitent dans ces paroles. Le *matter of caring* limite ainsi cette pratique qui pourrait servir ma collecte de données – il la limite au nom du respect des personnes présentes par l'intermédiaire de ces traces que sont les données. Le *matter of caring* fait donc, à l'égard de mes propres pratiques de collecte de données, un travail similaire au consentement qui limite ce que l'on collecte comme informations au nom du respect des personnes présentes au travers de ces informations – en d'autres termes, il invite à éviter la tendance à la dépersonnalisation que dénonce Kate Crawford à propos de la collecte massive et l'utilisation de données (en intelligence artificielle) au nom de leur déconnexion des situations personnelles auxquelles elles se rapportent (voir chapitre 3, p. 123). Pour respecter ces personnes, le *matter of caring* m'amène à pratiquer un certain silence à propos de situations vulnérables, non pour les rendre inexistantes, mais pour orienter mon regard ailleurs que sur ces tensions qui ne veulent pas être exposées. Le *matter of caring* esquisse ainsi une éthique de recherche où la pratique réflexive se révèle être un outil majeur pour (se) poser des limites, être à l'écoute d'inconforts et expliciter des silences qui acceptent d'être dévoilés – ces silences sont principalement les miens, car je suis la seule personne dont je peux cerner clairement l'accord pour être exposée ainsi.

Pratique réflexive et éthique de recherche

En mêlant des événements organisationnels, sociaux et politiques et écrivant les traces de ce qu'ils me font ressentir et de ce à quoi ils me font réfléchir, ces notes réflexives révèlent des façons dont je me lie à ce qui m'entoure – événements, façons de parler et d'agir, activités, technologies, situations, témoignages, blessures et petites victoires. Tandis que la réflexivité a été définie de multiples façons (Brummans, 2015), cette expérience que j'en fais correspond à la version par laquelle Ann L. Cunliffe conçoit la pratique réflexive : cette pratique consiste à « *questioning our own assumptions and taken-for-granted actions, thinking about where/who we are and where/who we would like to be, challenging our conceptions of reality, and exploring new possibilities* » (2016, p. 751). Selon Cunliffe, cette activité critique permet de réfléchir à la dimension éthique de toute pratique, en particulier par son caractère relationnel qui invite à sonder les façons dont nous sommes liés aux autres, au monde qui nous entoure, aux organisations dans lesquelles nous nous immergeons.

L'activité réflexive permet en d'autres termes de reconnaître l'interdépendance par laquelle nous sommes liés à notre environnement. Mais l'évolution des traces (ou notes) qu'elle laisse permet également de voir autre chose : à quel point cette interdépendance n'est pas qu'un état de fait mais un phénomène par lequel se constituent mutuellement ces choses (humaines et autres) qui se lient entre elles. Parce qu'elle est une activité par laquelle se révèlent et s'opèrent des résonances constitutives, la réflexivité fait partie de ce qui permet au *matter of caring* de n'être pas seulement un concept analytique mais une pratique. Autrement dit, le *matter of caring* va de pair avec la pratique réflexive quand celle-ci reflète les porosités avec ce qui l'entoure, ce qui participe à sa constitution en même temps qu'elle participe à la constitution de ce qui l'entoure.

Ces résonances constitutives transparaissent par exemple dans cet extrait de notes réflexives où je me mets à reconnaître à quel point ma pratique de collecte de données pour ma recherche (par enregistrement) est similaire à celles des agents conversationnels que je pousse à limiter :

Une réunion avec effervescence dans l'air : Paul et Félix viennent de demander à toute l'équipe de se réunir dans une salle pour récapituler ce que chacun a fait jusqu'à présent. Je n'ai pas enregistré cette réunion, faute d'avoir demandé l'accord de tous au début – l'ambiance est un peu tendue, il y a beaucoup de choses à régler d'ici deux jours, la réunion a commencé très rapidement. En entendant plusieurs enjeux éthiques abordés au fil de la discussion avant même que ce soit mon tour de parler, je me suis dit que je pourrais enregistrer discrètement et m'assurer après la réunion qu'ils sont d'accord que je conserve cet enregistrement; après tout, je les ai déjà tous informés de ma recherche et de mon activité d'enregistrements réguliers, ils étaient d'accord. Mais ça m'a dérangée. Utiliser des données audio collectées sans renouveler explicitement le consentement dès le début, ne serait-ce pas faire exactement comme un assistant personnel qui se met à écouter à notre insu ? Certes, les risques sont peu élevés, rien de sensible n'est discuté dans la conversation à laquelle je suis en train d'assister. Mais tout de même, il y a les voix des membres de Prélude, les noms, les accents, les opinions selon le type de commentaires faits... Ce n'est rien de très sensible mais c'est autant de données identificatoires, directement ou indirectement. Et c'est ce type de données que les règlements de protection de données personnelles invitent entre autres à limiter, si je suis le document que Paul et Félix m'ont demandé de faire pour clarifier tout ce qui peut être une donnée personnelle et ce qu'imposent les règlements de protection de ces données. Alors, si mon rôle dans cette équipe est en partie de contribuer à des pratiques de collecte de données qui soient raisonnables, honnêtes, transparentes et responsables, autant renforcer ce critère de limitation dans ma propre démarche de recherche.

(Vignette où l'éthique sur mon terrain résonne avec l'éthique de ma recherche, d'après le journal de bord, notes de terrain du 23 avril 2019)

Ces notes retranscrivent une situation de résonance qui m'amène à négocier les termes de mon éthique de recherche au fil d'une réflexion relationnelle où je me mets à reconnaître mes pratiques de recherche dans celles de l'objet que j'étudie avec Prélude, un agent conversationnel. Ce mouvement de reconnaissance manifeste une pratique de *matter of caring* : si j'interviens pour orienter dans une certaine direction le développement des agents conversationnels de Prélude (limiter la collecte de données), il s'agit de m'orienter moi-même dans cette direction quand je mène des activités similaires, celles qui cultivent un attachement à une certaine quantité de données pour les analyser plus tard.

Dans ces notes réflexives, je questionne mon activité de recherche : collecter des données en les enregistrant quand des moments d'intérêts surviennent. Ce questionnement se fait à la lumière de l'activité de collecte de données par l'agent conversationnel que je participe à cadrer tandis que ce type de système d'IA est souvent critiqué pour ses pratiques ambiguës de collecte et analyse de données, notamment quand il s'active à l'insu des utilisateurs enregistrés (voir Vincent, 2019, et ce qui suscite de la méfiance dans le chapitre 2, p. 96). Ces notes témoignent de ce que Neyland définit comme « réflexivité ethnographique » : « *reflexive ethnography engages in a thorough and detailed analysis of the ethnographer's attempts to make sense of the world while those being studied are making sense of the world* » (Neyland, 2008, p. 22). En reconnaissant la similarité de mes méthodes de recherche avec celles de ces assistants virtuels, j'en viens à renforcer, dans ma propre activité de recherche, ces contraintes de limitation de collecte de données personnelles qui sont imposées aux acteurs, humains et autres, traitant de telles données. L'auto-imposition de ces contraintes relève également du travail que fait le *matter of caring* à la lumière du consentement qui sensibilise aux limites à s'imposer dans nos relations avec les autres (voir chapitre 3 et la mention de l'impératif catégorique kantien, p. 121).

Ponctuées par des marqueurs de réflexions personnelles (« je me suis dit que », « ça m'a dérangée »), ces notes permettent d'ouvrir une fenêtre sur mes conversations intérieures en tant que chercheuse à mesure que je fais sens de l'organisation dans laquelle je suis immergée et dont j'apprivoise petit à petit les manières de faire sens. Ces notes reflètent ainsi une des voix (silencieuses) qui prennent part à la constitution de l'organisation, ne serait-ce que par ma présence à cette réunion et le sens que je fais sur le vif qui influence mes actions par la suite, qu'il s'agisse de créer des documents ou intervenir dans des réunions. La pratique du *matter of caring* amène à porter attention à ces notes pour tenir compte de ce silence et avoir accès à ce qu'il exprime (QR2).

Prendre un « je me suis dit que » comme marqueur d'analyse peut exposer à la critique de s'arrêter sur une opinion anecdotique de la chercheuse qui en apprend peu à la lectrice sur la constitution de l'organisation ou du phénomène du consentement à ce

moment précis. Pourtant, cette vignette réflexive révèle en moi la présence de plusieurs dynamiques organisationnelles liées à un événement important (la préparation du *pitch*) et à une pratique indispensable pour ce type d'organisation en IA (la collecte de données personnelles, voir chapitre 2). Elle révèle, d'une part, l'ambiance vécue au sein de cette organisation à ce moment-là (« effervescence dans l'air », « l'ambiance est un peu tendue »), une ambiance qui peut avoir un effet organisant par des points de tensions qu'elle suscite, des effets de silence de certaines voix dans les conversations, des rapports de pouvoir qui la structurent implicitement. Cette ambiance joue ainsi sur les conditions auxquelles le *matter of caring* porte attention (voir chapitre 4, p. 199). La vignette révèle, d'autre part, un ensemble de tensions liées à des préoccupations éthiques (*ethical matters of concern*, Matte et Bencherki, 2019) à l'égard des pratiques de collecte de données personnelles. La retranscription d'un extrait de réunion pourrait montrer les hésitations, contradictions et négociations par lesquelles ces préoccupations se traduisent au fil des conversations entre les membres de l'équipe à propos des façons de mettre en acte le principe de respect de la vie privée – une mise en acte qui influence la constitution de l'organisation selon les contraintes à respecter en matière de collecte et de traitement des données personnelles (voir chapitres 3 et 6). Ces conversations entre les membres de l'équipe ne sont pas visibles dans cette vignette puisqu'il n'y a pas eu d'enregistrements, mais les hésitations, contradictions et négociations y sont présentes, comme en reflet de ces conversations.

Par exemple, l'hésitation à enregistrer est marquée par le conditionnel « je pourrais », introduite par l'indication de discours rapporté « je me suis dit que » qui expose l'action envisagée (enregistrer). Elle est ensuite ponctuée par la formule « après tout » qui mentionne une raison pour laquelle cette action pourrait être envisageable (un consentement plus large a été donné avant cet événement particulier). Cette hésitation se solde par la mention d'un inconfort de l'ordre du dérangement (« Mais ça m'a dérangée » qui exprime la tension liée à la tentation de collecter des données « discrètement », en s'en remettant à un consentement implicite sur le vif bien que donné de façon plus générale auparavant. Ce dérangement permet d'ouvrir la pratique réflexive

par une question rhétorique : « Utiliser des données audio collectées sans renouveler explicitement le consentement dès le début, ne serait-ce pas faire exactement comme un assistant personnel qui se met à écouter à notre insu ? ». Celle-ci met en miroir la tentation d'une pratique de collecte de données de recherche qui serait critiquable d'un point de vue d'éthique de la recherche¹⁷², et la pratique intrusive d'écoute et d'enregistrements par des assistants personnels activés à notre insu. Or, cette tentation aboutirait en une contradiction : la chercheuse agirait comme un système d'IA intrusif alors même qu'elle incarne dans l'équipe le rôle d'éviter ce type de pratique intrusive en contribuant à la mise en place de « pratiques de collecte de données qui soient raisonnables, honnêtes, transparentes et responsables ».

Cette mise en miroir qui soulève une contradiction éthique est un pivot important car elle marque un moment où le sens que je fais de ma pratique de recherche évolue par la similitude que je reconnais entre ma pratique d'enregistrement et celle d'un système IA. Par la tournure négative de cette question rhétorique (« ne serait-ce pas »), qui laisse d'autant moins de place au doute qu'elle s'accompagne de l'adverbe « exactement », une forme de reconnaissance contradictoire de l'éthique de l'une dans l'autre s'opère: je semble reconnaître l'éthique de ma pratique de collecte de données dans le système d'IA, alors que je critique habituellement l'éthique de cet objet. Cette mise en miroir marque une réflexion relationnelle où ma distance en tant que chercheuse par rapport au système d'IA se réduit quand je reconnais ma pratique de recherche et la dérive de son éthique dans celle du système d'IA tandis que je formule ma tentation d'enregistrer à la dérobée. Cette réflexion relationnelle marque l'alignement qu'implique de faire d'une préoccupation éthique un *matter of caring* (QR2) : l'éthique ne peut pas se cantonner à l'objet étudié mais doit être incarnée par la posture et les pratiques par lesquelles on intervient pour faire bouger, voire améliorer, des situations.

¹⁷² L'Énoncé de politique des trois conseils qui régit, au Canada, l'éthique de la recherche avec des êtres humains soumet l'activité de collecte de données au consentement libre, informé et explicite. Ce consentement doit être donné au début de chaque enregistrement effectué, ce que rappelle mon certificat d'éthique octroyé par mon université.

La reconnaissance de cette similarité marque mon engagement dans le phénomène du respect des données personnelles et du consentement, et lie ma pratique de recherche à d'autres pratiques, de sorte que celle-ci ne fasse pas exception, qu'elle n'ait rien d'extra-ordinaire – où l'extra serait à côté voire au-delà de l'ordinaire. J'inscris ma pratique de recherche dans le même ordinaire que ce que j'observe. Car l'objet principal de mon observation – le consentement – est une dynamique d'interaction qui invite au respect, et cet objet est présent dans de nombreuses situations ordinaires. La pratique réflexive de l'ordre du *matter of caring* permet de porter attention à la présence de cet objet, par-delà la sphère dans laquelle j'ai prévu de l'observer (dans le design d'un agent conversationnel avec une organisation qui en développe un). Et si cet objet (le respect dans une situation de consentement) demande un certain comportement, alors ce comportement doit s'appliquer dans ces différentes sphères ordinaires, par un certain alignement.

Ainsi, les marqueurs de réflexions personnelles accompagnées de mention de sentiments tels que le dérangement sont à considérer de la même manière que les trajectoires affectives tracées par les récits réflexifs de Consuelo Vásquez (2020, p. 195) : « non pas comme étant des réactions individuelles ou intimes, mais plutôt comme des dynamiques relationnelles qui s'articulent autour », ici, de la pratique de collecte de données et son éthique. La vignette réflexive se présente ainsi comme un espace textuel de négociation pour résoudre cette contradiction, une négociation articulée par des dynamiques relationnelles entre la chercheuse, l'agent conversationnel, et par extension l'organisation qui le développe et dont je fais partie. Et cette négociation silencieuse compte car elle fait une différence dans ma manière de collecter des données pour constituer mon matériel. Elle fait également une différence dans ma manière d'être avec les membres de l'organisation et l'agent conversationnel en développement : je reconnais ma complicité avec eux, j'inscris mes pratiques dans le même ordinaire.

Cela a un double effet sur ma constitution en tant que chercheuse et ma complicité dans cette organisation, à la fois à l'égard de mon activité de collecte de données, et de ma posture dans cette organisation. Pour éviter de devenir comme un système d'IA

intrusif, je restreins mes enregistrements aux seuls moments où je pourrai formaliser un consentement explicite en début de réunion – comme je l’ai mentionné dans la section précédente. Par cette limitation, il s’agit d’éteindre toute tentation d’enregistrement de moments plus informels, renforçant ainsi dans ma propre démarche de recherche le critère de limitation de la collecte de données personnelles. Les 24 réunions que j’ai finalement enregistrées représentent un total de 22 heures et 15 minutes, soit seulement 3% de la totalité du temps passé avec l’organisation. L’essentiel de mon matériel est ailleurs que dans ces tours de parole, principalement dans mon journal de bord.

Par ailleurs, pour contribuer à développer un système d’IA respectueux de la vie privée, j’adopte une posture où ma présence dans l’organisation est de l’ordre de ce que Tim Ingold (2011) appelle une « *practice of togetherness* » – comme mentionné plus haut, cette pratique désigne une démarche de recherche qui consiste à étudier *avec* et apprendre *de* l’organisation (et de l’objet – consentement) dont on épouse les contraintes, le rythme, les ambiances etc. plutôt que d’apprendre *sur* cette organisation dans une relation distanciée. Je ne vais donc pas contribuer à cette organisation *pour* collecter des données sur la manière dont se matérialise l’éthique par le design au fil de la constitution organisationnelle (Cooren, 2016; Matte et Bencherki, 2019). Ma posture s’affirme plutôt comme une contribution à la matérialisation de l’éthique, par le design, dans cette organisation – où l’éthique est concentrée sur les enjeux du consentement et du respect de la vie privée. Cette contribution se fait par plusieurs activités façonnant ma complicité : en faisant l’expérience des contraintes, en apprivoisant des régulations, en me préoccupant des enjeux éthiques du consentement et du respect de la vie privée, en bricolant pour proposer des solutions. C’est à partir de cette posture de l’intérieur, complice, et cette *practice of togetherness* que la réflexivité peut s’envisager comme une pratique de tissage relationnel. Il s’agit d’un tissage relationnel qui contribue aux organisations¹⁷³ en constitution, mais également un tissage qui se fait avec l’objet apprivoisé – le consentement et les agents conversationnels dans mon cas.

¹⁷³ Dans l’article avec Boris Brummans, nous avons mis l’accent sur la contribution de ce tissage relationnel aux organisations. Si je reprends ces mots ici dans ma thèse, j’ajoute cependant le tissage qui se fait avec l’objet de ma recherche – le consentement – car mon chemin avec ce concept a participé à mes

En suivant les pratiques de développement de cet agent conversationnel selon une *practice of togetherness*, mon attention en tant que chercheuse emprunte ainsi le même chemin que ceux qui développent cet agent – c’est la direction que prend ma complicité. Mon attention chemine également à la lumière de ce que je croise et assimile à propos des enjeux de consentement et respect de la vie privée. Dans la vignette ci-dessus, cette *practice of togetherness* se manifeste par des façons de penser avec l’organisation, et avec les enjeux du consentement et du respect de la vie privée, notamment à l’occasion d’une autre négociation sous forme de dialogue indirect qui sonde ce qui pourrait orienter le choix d’enregistrer ou non. « Certes, les risques sont peu élevés, rien de sensible n’est discuté dans la conversation à laquelle je suis en train d’assister » : cette réflexion orientée sur les risques que présenterait la collecte de ces données reprend une réflexion initiée avant cette réunion quand des membres de l’équipe m’ont pressée à reformuler mes questions (voir chapitre 6). En transposant cette réflexion dans ma propre activité de recherche puis me rassurant sur le degré de ces risques (« peu élevés, rien de sensible »), je pense *avec* l’organisation et *avec* les documents liés au respect de la vie privée. En effet, à ce moment, Prélude est elle-même en train d’apprivoiser cette réflexion éthique et légale : ses membres, dont moi-même, essayent de comprendre pourquoi il est important de limiter les données personnelles, même quand celles-ci ne sont pas sensibles (quand par exemple il ne s’agit pas d’informations médicales ni financières) – cette moindre sensibilité peut laisser penser qu’il n’y a pas grand risque à les collecter. L’interjection qui suit (« mais tout de même ») recentre la réflexion : même si la sensibilité de ces données est faible, il ne s’agit pas d’amoindrir la vigilance envers leur caractère identificatoire (voir chapitre 2, p. 122, et chapitre 6). Autrement dit, c’est ce risque identificatoire à ne pas négliger qui justifie de faire de ces données personnelles un *matter of caring*. En aidant à recentrer la réflexion sur ce risque identificatoire, la pratique réflexive permet de s’entraîner à penser avec cette préoccupation qui accompagne la constitution de l’organisation autour de la question du traitement des informations personnelles.

interventions dans le design de l’agent conversationnel puis dans l’analyse de mon matériel et l’écriture de ma thèse.

Par ce dialogue réflexif sur le caractère identificatoire des données que je collecte pour ma propre recherche, j'affine ainsi ma compétence¹⁷⁴ en matière de traitement des données personnelles en clarifiant l'importance de maintenir une certaine vigilance à leur égard. Cette compétence est ce qui se forme à force de penser avec ces enjeux. Cet affinement de compétence contribue à constituer l'organisation *d'une certaine manière*, c'est-à-dire en faisant des pratiques de collecte de données un *matter of caring* pour qu'elles soient respectueuses de l'enjeu de vie privée.

L'écriture des notes réflexives et descriptives consiste ainsi en autant de traces glanées et conservées au fil de l'immersion, une activité qui ne consiste pas tant à collecter des données comme des informations à extraire du terrain mais plutôt comme des compétences (Ingold, 2013, p. 5), expériences, documents, traces créées et transformées *avec* plutôt que *sur* les membres de l'organisation – une activité de constitution mutuelle de l'organisation et du tissu relationnel dont elle se compose.

Cette expérience de terrain relatée par la pratique réflexive se constitue dès lors par les mêmes lignes dont se tisse le terrain lui-même – le terrain comprenant des organisations spécifiques (Mr Y et Prélude), leurs activités (le design d'agents conversationnels) mais aussi un enjeu qui est aussi objet de recherche (le consentement et le respect de la vie privée qui lui est lié). La dimension constitutive de l'organisation et de l'objet de recherche se manifestent en particulier par la réflexivité. Celle-ci se présente ainsi comme une pratique accompagnant ce chemin, non parce qu'elle permet un retour sur soi comme un point fermé sur lui-même, mais plutôt parce qu'elle se dessine comme une ligne de co-constitution (et de complicité) :



¹⁷⁴ Cette compétence relève de la littératie numérique (voir chapitre 3, p. 160). Elle touche au fonctionnel (comprendre le traitement des données personnelles) et à l'esprit critique (comprendre les implications de ce traitement, et les risques associés).

¹⁷⁵ Ingold (2011, p. 69). Deux versions de dessin d'un organisme pour montrer l'inversion.

Figure 7 : schémas de Tim Ingold (2011, p. 69) qui représentent deux versions de dessin d'un organisme

Ingold (2011) utilise ces images (figure 7) pour transmettre sa vision relationnelle d'un organisme, qui n'est pas plié sur lui-même mais qui se présente comme un chemin. Tissé avec d'autres chemins, celui-ci en vient à constituer un *meshwork*, c'est-à-dire un champ relationnel qui constitue une chose, une organisation par exemple. Ma réflexivité en tant que chercheure présente dans le tissage de cette organisation et du concept de consentement peut se comprendre par cette image du chemin fondamentalement relationnel. En paraphrasant Ingold, comprenons que dans le premier cas, où la réflexivité est un retour sur soi, il y a une inversion : la réflexivité est enroulée sur elle-même, délimitée et contenue dans un périmètre, posée contre un monde qui l'entoure – un environnement, un terrain – avec lequel elle est destinée à interagir. Dans le deuxième dessin, la réflexivité n'a ni dedans ni dehors, ni de limite qui sépare les deux domaines. Elle est plutôt un chemin le long duquel se fait l'expérience de l'organisation et du concept de consentement, qui se constituent au fil de ce chemin. Sur ce terrain, en tant que chercheure je suis le point de rencontre et de départ de plusieurs chemins que je découvre en même temps que j'en trace, tissant, parcourant et co-constituant l'organisation et le concept de consentement. Quand on défait l'inversion de la réflexivité comprise comme un retour sur soi pour la dessiner comme un chemin et comme une relation, il ne faut donc pas ranger cela dans la catégorie des images de pensée : « *Things are their relations* » (Ingold, 2011, p. 70, italiques dans l'original). La réflexivité est une relation, la chercheure est un ensemble de relations qui participent au tissage du terrain. Et sa posture ne consiste pas à passer de son terrain à son bureau pour écrire son texte, entre un *fieldwork*, un *deskwork*, puis un *textwork* (Yanow, 2000), mais de se laisser traverser par tous les chemins qu'impliquent ces types de travail – des chemins qui

1: "But in this apparently innocent depiction I have already effected an inversion. I folded the organism in on itself such that it is delineated and contained within a perimeter boundary, set off against a surrounding world – an environment – with which it is destined to interact according to its nature."

Fig. 2: "In this depiction there is no inside or outside, and no boundary separating the two domains. Rather there is a trail of movement or growth. Every such trail discloses a relation. But the relation is not between one thing and another – between the organism 'here' and the environment 'there'. It is rather a trail along which life is lived."

s'entrecroisent et qui façonnent son environnement comme elle est formée et transformée par lui.

Si ma réflexivité guide autant ce chemin qui se poursuit jusqu'à tisser cette thèse, s'agit-il d'une autoethnographie ? Ce tissage relationnel est porté par une pratique du *matter of caring* qui est étroitement liée à *mon* expérience chez Mr Y et Prélude, à partir de *ma* propre situation, à la lumière dont le concept de consentement a résonné *pour moi* dans de multiples usages. Ce sont là les accents « auto- » présents dans toute démarche ethnographique qui dépend de l'accès de la chercheuse à l'organisation et au phénomène étudié, qui repose également sur *ses* façons de sélectionner et d'analyser ce qui retient *son* attention. L'accès, l'observation, l'analyse sont les étapes et enjeux méthodologiques propres à toute ethnographie (Neyland, 2008; Tracy, 2019; Ybema et al., 2009). Ce qui fait d'une ethnographie une *autoethnographie*, c'est plutôt l'orientation biographique de son contenu (Tracy, 2019, p. 70). Cette orientation autobiographique serait très pertinente pour étudier le consentement par l'angle du *matter of caring*. Elle serait également courageuse, à exposer ses vulnérabilités.

Ce que je fais s'en tient à l'ethnographie, comme une balade en forêt¹⁷⁶ qui n'est souvent pas débroussaillée : on y entre par un endroit, on cherche par où passer, on se fait un chemin en taillant son accès, on avance et on bifurque en fonction d'obstacles et d'indices, de marques, de lumières qui nous attirent, mais ce n'est pas pour autant que le récit de cette balade sera autobiographique. *Ma* situation, *mon* expérience, *mes* résonances, relèvent ainsi davantage de mon *orientation* que du centre de mon attention : elles sont *ce à partir de quoi* je vois des choses. Cette orientation s'est faite à partir du point de vue¹⁷⁷ (*standpoint*, Harding, 1992) qui était le mien, comme toute connaissance est créée à partir d'une perspective située (Haraway, 1991). Cette orientation s'est faite aussi à la lumière du consentement, comme concept, comme

¹⁷⁶ Je dois à mon ami Justine Laurent cette métaphore de la balade (d'un écureuil) en forêt, qu'elle m'a rappelée plusieurs fois lors de la laborieuse phase d'analyse de mon matériel : si ce que je veux suivre est ce qui sent la noisette, parmi toutes ces pistes, quel est le chemin qui sent le plus la noisette ?

¹⁷⁷ Le point de vue est à comprendre ici non comme une opinion ou un avis qui implique un jugement, mais comme un point à partir duquel on a une vue, sur un paysage par exemple.

pratique, comme moment d'interaction, comme objet de design de conversation. Sarah Ahmed dit des orientations qu'elles « *shape how the world coheres around me. Orientations affect what is near or proximate to the body, those objects that we do things with* » (2010, p. 235). Mon point de vue, la façon dont ma perspective est située et le consentement comme objet sont ainsi ces orientations qui ont façonné comment le monde s'est organisé autour de moi, selon ce qui était proche de mon corps et de mes idées¹⁷⁸.

Mon ethnographie rend compte d'un chemin qui s'est tracé selon ces orientations – de mon corps et ma pensée – « de la même manière qu'un sentier au préalable inexistant apparaît en marchant » (Varela, 1996, p. 111). Et ce chemin ne s'est pas arrêté à la fin du terrain, il s'est poursuivi pendant les deux années (et demi) d'analyse et écriture qui ont suivi le terrain, orienté par le *caring* que demande le consentement pour ne plus être négligé ou court-circuité.

Pratiques d'analyse de données

La notion de *matter of caring* arrive au fil d'une analyse itérative (Tracy, 2019) du matériel que j'ai rassemblé pendant ce travail de terrain avec Mr Y et Prélude. Sarah Tracy définit l'analyse itérative comme une approche qui alterne :

between emic, or emergent, readings of the data and an etic use of existing models, explanations, and theories. Rather than grounding the meaning solely in the emergent data, an iterative approach also encourages reflection upon the active interests, current literature, granted priorities, and various theories the researcher brings to the data. (2019, p. 209)

De cette façon, mon analyse se fait par des allers-retours entre l'immersion dans mes données, des lectures théoriques pour approfondir certaines pistes, des discussions. Je

¹⁷⁸ Cette notion d'orientation selon ce qui est proche du corps fait écho à cette phrase de Beauvoir (dans Garcia, 2019, p. 172) citée dans le chapitre précédent : « Dans la perspective que j'adopte, (...) si le corps n'est pas une *chose*, il est une situation : c'est notre prise sur le monde et l'esquisse de nos projets » (Beauvoir, *Le Deuxième Sexe*, t. I, p. 75, italiques dans l'original).

porte attention aux sens (signification et direction) qui émergent de mes données. Puis selon ces émergences, j'approfondis des pistes par des lectures, donnant tantôt priorité à une piste plutôt qu'une autre, replongeant dans mes données à la lumière de ces intérêts, testant d'autres éclairages théoriques, etc.

Au fil du travail de terrain, j'ai classé mon matériel en différents dossiers (journal de bord, documents organisationnels liés au consentement et au traitement des données personnelles, enregistrements de réunions, retranscriptions de celles qui abordaient ces sujets). Après avoir quitté le terrain, j'ai relu mon matériel en prenant soin de changer les noms par des pseudonymes. Les co-fondateurs de Mr Y et Prélude m'ont manifesté leur accord pour conserver le nom de leurs organisations et leurs propres noms. Mais, en alignement avec une recommandation de mon comité d'éthique, j'opte pour l'usage de pseudonymes afin de permettre davantage de confidentialité, en particulier des autres membres de ces organisations. Ces pseudonymes m'aident également à entrer dans l'analyse : ils donnent aux mots et actions de ces personnes, dont j'ai gardé des traces écrites, la consistance propre aux micro-moments du passé. Ils me permettent ainsi de prendre de la distance avec leurs personnes qui ne peuvent se réduire à ces traces passées. Proposer une interprétation, par l'analyse, à partir de ces traces liées à des pseudonymes aide ma pensée à ne pas prétendre parler pour eux ni dire ce que leurs mots et gestes veulent dire. J'apprivoise plutôt ces traces comme des vestiges autour et au travers desquels ma pensée tantôt s'entortille, s'agace, s'enthousiasme, se perd, toujours s'anime.

Lors d'un premier cycle d'analyse (ou *primary-cycle coding*, selon les termes de Tracy, 2019, p. 219) il s'agit de s'immerger dans ses données et les structurer – ce que je fais à l'aide du logiciel NVivo. Je passe au travers de l'ensemble de mon journal de bord, des documents organisationnels liés au consentement et au traitement des données personnelles ainsi que des retranscriptions de réunions que j'ai sélectionnées quand elles abordaient ces sujets (voir la première section de ce chapitre). À partir de techniques

d'*open coding* (Charmaz, 2014), j'attribue un code¹⁷⁹ à chaque phrase ou ensemble de phrases, d'abord pour identifier les différentes pratiques mobilisées, les informations sur les organisations et les techniques, puis pour repérer certaines tensions. À ce premier stade, les codes sont descriptifs, visant à montrer et regrouper ce qui est similaire dans les données – c'est ce que Tracy appelle les *first-level codes* (2019, p. 234). Par exemple, le code « consentement » est attribué aux phrases, paragraphes, extraits de documents ou de retranscriptions où est mentionné explicitement le mot « consentement », le code « design de conversation » est attribué à ceux qui relèvent de cette activité, ou encore le code « inconfort » est attribué aux moments où le terme est employé (dans mes notes réflexives surtout). Pour rassembler les codes qui se démultiplient, je les rassemble en différents thèmes « parapluie » (Tracy, 2019, p. 226) selon leurs similarités, selon une technique d'*axial coding* (Charmaz, 2014). Cette première phase de codage me permet en particulier d'identifier les différentes activités, pratiques, étapes par lesquelles se fait le travail du consentement, afin d'analyser ensuite (quand le concept aura émergé) comment le *matter of caring* se manifeste et se constitue au fil de ces processus – autrement dit, cela me permet de structurer mon matériel pour explorer par la suite ma première question de recherche (comment cela se fait d'aborder le consentement comme un *matter of caring* dans cette organisation ?).

Après cette phase de codage au niveau des phrases et paragraphes dans l'ensemble de mes données, je procède à plusieurs phases de *secondary-cycle coding* (Strauss & Corbin, 1990; Tracy, 2019, p. 225). Cette phase est plus interprétative que la première. Elle permet de hiérarchiser les codes en thèmes (Gioia et al., 2013; Saldaña, 2016), puis d'interpréter ces thèmes à la lumière de perspectives et concepts théoriques (Tracy, 2019). Les thèmes s'articulent en trois grandes catégories : (1) les pratiques abordées (éthique, design, *privacy*), (2) le contexte organisationnel (organisation, techniques & outils) et (3) les pratiques de recherche (enjeux ethnographiques). (1) Le

¹⁷⁹ J'utilise les termes « code » et « codage » (*coding*) comme Tracy (2019, p. 213) : « I use the word code to refer to a word or phrase that "symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data" (Saldaña, 2016, p. 4) and coding as the active process of identifying data as belonging to, or representing, some type of phenomenon ».

thème des pratiques abordées – éthique, design, *privacy* – rassemble des pratiques d'éthique *by design* (voir chapitre 1, p. 26) dont j'ai pu faire l'expérience. (2) Le contexte organisationnel rassemble ce que j'ai pu constater et apprendre sur Mr Y et Prélude et les techniques envisagées ou utilisées (voir chapitre 2, p. 73). (3) La catégorie des enjeux ethnographiques rassemble des sous-thèmes (ou nœuds, dans le langage NVivo) liés à la pratique ethnographique, tels que l'accès, les affects ou ressentis (surtout ceux qui transparaissent dans mes notes réflexives), la complicité, des intentions de départ (qui ont bifurqué ou ont disparu), des réflexions sur l'éthique de ma recherche, sur mon rôle et ma posture – autant de sous-thèmes dont je tire les fils dans ce chapitre. Ces thèmes, ou catégories, permettent de structurer mon matériel, ce qui me permet d'avoir une vue d'ensemble des étapes, des tensions et modifications par lesquelles passe le travail du consentement dans l'organisation, et ainsi identifier à la fois comment le travail du consentement se fait, comment il est abordé comme une préoccupation particulière (QR1), et quels effets cela fait au sein de l'organisation (QR2). À ce moment de mon analyse, la notion de *matter of caring* n'est pas encore forgée comme telle, mais la préoccupation particulière à laquelle je porte attention est ce qui se raffine petit à petit, de façon itérative, par la notion de *matter of caring*.

Au début de cette deuxième phase de codage plus interprétatif, il y a un autre thème majeur, en plus des trois cités ci-dessus : le consentement. Mais comme il est présent dans la grande majorité de mes données, je le distingue finalement en sous-catégories du thème « design » et de celui « éthique ». Dans le thème du design, je rassemble tout ce qui se rapporte aux activités liées au design de consentement – les ambivalences des *nudges* voire *dark patterns* et du contrôle donné aux utilisateurs (voir chapitre 3, p. 131) sont ressorties ici. Dans le thème éthique, le sous-thème du consentement rassemble les réflexions d'ordre plus global, qui identifient des résonances entre consentement tel qu'abordé dans la culture populaire, à la suite de *#MeToo* (« *consent culture* »), et consentement numérique – de là émerge la remarque que ce qui se joue dans ces différents usages, ce sont les termes d'une relation respectueuse (voir chapitre 3, p. 114).

Au fil de cette deuxième phase de codage, deux autres thèmes apparaissent de façon transversale aux autres : tensions et voix. Les tensions m’amènent à mettre le doigt sur plusieurs ambiguïtés liées au travail du consentement, aux pratiques de *privacy-by-design*, à l’évolution des organisations Mr Y et Prélude, à ma pratique de recherche. Le travail de l’ambiguïté se révèle ici être une piste intéressante à suivre – ce qui me mène au concept d’ambiguïté pragmatique de Giroux (voir chapitre 3, p. 131) mais également à la place importante de l’ambiguïté dans les études féministes (en particulier Beauvoir, 2013; Garcia, 2018). L’intérêt pour la voix est une surprise du terrain, à mesure que je suis de plus en plus mêlée à Prélude qui s’intéresse aux technologies vocales et au design de *voice user interface* (voir chapitre 2, p. 73).

En approfondissant cette piste de la voix, en particulier en technologie, je découvre les travaux de Don Ihde (2007) qui me mènent à ceux du courant de la postphénoménologie en philosophie de la technologie (voir Rosenberger & Verbeek, 2015; Verbeek, 2011; abordés dans le chapitre 1, p. 67). Je suis également curieuse d’explorer comment la voix est abordée dans les études CCO, ce qui attire mon attention vers l’importance portée à ce qui s’exprime (*voiced*) dans les tours de parole¹⁸⁰. Je retourne alors dans l’article « *Just say no* » (Kitzinger & Frith, 1999, cité dans le chapitre 3, p. 171) qui met l’accent sur la difficulté d’exprimer un refus clairement. Cela résonne avec des témoignages post-*#MeToo* que j’ai classés dans le nœud consentement. L’importance de porter attention à ce qu’expriment des silences, des hésitations, et à ne pas parler pour quelqu’un dans le contexte d’une demande de consentement se fait de plus en plus saillante.

¹⁸⁰ Dans les études CCO, le terme de voix est souvent mentionné mais n’est pas lui-même un objet principal de recherche. Des voix sont la plupart du temps présentes par les données qui retranscrivent souvent des interactions entre plusieurs personnes; en particulier dans les analyses ventriloques (Cooren, 2013), où l’on trouve les termes « *voiced by, voiced through* » et une attention au phénomène de polyphonie, où « *many voices can be heard* » (Cooren & Sandler, 2014). Le terme de voix est également utilisé de façon métaphorique (Putnam et al., 1996; Putnam & Boys, 2006) pour évoquer la voix d’un collectif, d’une organisation (Taylor & Cooren, 1997).

Entrelacées dans ces cycles d'analyse, des discussions et des lectures¹⁸¹ amènent le concept de *care* à s'immiscer dans ce sillon, notamment pour l'accent qu'il met sur les conditions dans lesquelles une parole est exprimée (Garcia, 2018, 2021) ou dans lesquelles un soin est donné (Mol, 2013; Pols, 2015). La rencontre avec le concept de *matter of care* de Bellacasa (2011) me donne une clé analytique par laquelle je plonge de nouveau dans mes données, en particulier celles des nœuds du consentement dans les thèmes de design, d'éthique et de *privacy*.

Chacune de ces phases d'analyse itérative se fait par la rédaction de mémos analytiques, compris comme des « *sites of conversation with ourselves about our data* » (Clarke, 2005, p. 202, cité dans Tracy, 2019, p. 228). D'abord sous forme de notes pour clarifier un concept ou des liens entre des codes, ces mémos s'étoffent par le recours à des phases d'écriture libre et d'écriture comme *method of inquiry* (Richardson, 2000; Richardson & St. Pierre, 2018). Comme les différentes pistes que je suis à divers moments de l'analyse me semblent souvent mener à des impasses, la pratique de l'écriture prend la forme de nombreuses réécritures, reformulations, cherchant d'autres angles, jusqu'à parvenir à celui du *care*. Cet angle me permet de mieux saisir les nuances, ambiguïtés et multiplicités (des pratiques et des sens du concept de consentement) qui attirent mon attention. Petit à petit, des plans émergent (et se modifient) pour organiser l'analyse et mes arguments autour de la constitution du consentement comme *matter of caring* (QR1) et du travail que ce *matter of caring* fait dans l'organisation (QR2).

Au fil de ces cycles d'analyse, d'écriture et de réécriture, mes questions de recherche ont évolué. Initialement, mon intention était d'explorer comment l'éthique – et plus précisément l'enjeu du consentement et du respect de la vie privée – se matérialise au fil des pratiques d'une organisation qui développe un agent conversationnel. Cette question m'a amenée à une description de pratiques sociomatérielles qui correspondent à celles de la prise en compte du consentement et du

¹⁸¹ L'analyse est elle aussi un processus de tissage relationnel : on tisse son chemin analytique au travers des données que l'on frotte à certaines idées, plusieurs concepts, au fil de discussions avec des personnes, ces discussions attirent l'œil vers les données et les concepts qui *glow* (Gherardi et al., 2019, à partir de MacLure, 2013). L'analyse se tisse au fil de ces rencontres.

respect de la vie privée dans les activités de design de conversation – ces pratiques ont surtout été les miennes dans cette organisation plutôt que celles de l’organisation dans son ensemble. Cette description informe les trois premiers chapitres de ma thèse. Mais la portée éthique de mon objet de recherche (le consentement) ainsi que l’engagement de ma posture m’ont incitée à aller au-delà de la description. Mon questionnement s’est finalement articulé en une problématique allant au-delà de mon matériel – pourquoi prendre soin du consentement dans le design d’un agent conversationnel ? – et deux questions de recherche ancrées dans mes données : (QR1) prendre le consentement comme *matter of caring* dans une organisation qui développe un agent conversationnel, comment cela se fait ? et (QR2) qu’est-ce que cela fait ?

La problématique ancre mon questionnement sur le travail du consentement (QR1) et du *matter of caring* (QR2) : pourquoi prendre soin du consentement dans le design d’un agent conversationnel ? Plutôt que fondée spécifiquement sur mon expérience de terrain, cette problématique cherche à comprendre la pertinence de faire du consentement un *matter of caring* alors que beaucoup de choses donnent l’impression que cet acte est vain ou voué à de nombreux échecs tant il se heurte à des obstacles. Cette problématique guide les chapitres 2 et 3 pour aboutir à la conceptualisation du consentement comme un acte vulnérable, ancré dans l’interdépendance, qui demande de porter attention aux conditions dans lesquelles il est demandé et donné ou refusé. Cet aboutissement donne également des pistes pour procéder à une nouvelle phase d’analyse du matériel sélectionné – celui qui touche explicitement au travail du consentement dans le design de l’agent conversationnel – pour porter attention aux « confrontation[s], (...) empêchements, (...) modifications du désir ou de la volonté initiale » (Brugère, 2013, citée dans le chapitre 3, p. 176) qui dévoile le travail vulnérable que fait le *matter of caring* centré sur le consentement (ce qui permet d’explorer la QR2 dans le chapitre 6).

En posant la question « Qu’est-ce que ? » à propos d’une bonne expérience avec un agent conversationnel et d’une bonne expérience de consentement, je procède dans les chapitres 2 et 3 à ce que la philosophe Sally Haslanger (2000) appelle une enquête analytique qui demande pragmatiquement :

What is the point of having these concepts? What cognitive or practical task do they (or should they) enable us to accomplish? Are they effective tools to accomplish our (legitimate) purposes; if not, what concepts would serve these purposes better?
(p. 33)

Manon Garcia (2021) qualifie de « méliorative » ce type d'analyse « dans la mesure où il n'a pas seulement une ambition *descriptive* – établir à quelle réalité ce concept correspond, quels sont les objets qu'il regroupe – mais aussi une ambition *normative* d'amélioration du monde social – établir à quel type d'entreprise normative et émancipatrice ce concept peut servir » (p. 25-26, mes italiques). En invitant dans le chapitre 2 (p. 104) à faire autrement que reproduire, dans le design d'agent conversationnels, des stéréotypes néfastes pour certaines personnes, mon analyse met de l'avant ces pistes d'amélioration. Dans le chapitre 3, c'est l'horizon de respect, d'égalité et d'épanouissement qui montre l'entreprise émancipatrice à laquelle le consentement peut servir. Mon analyse dans ces chapitres explore d'abord ce que le design d'agent conversationnel et le consentement font (*how they work*), quel effet normatif ce travail a sur ce que l'on considère une « bonne expérience », puis quel travail nous serait plus bénéfique (« *what work we want these concepts to do for us* », Haslinger, 2000, p. 33). Cette analyse méliorative du consentement dans les expériences conversationnelles (par agents virtuels) ancre la pertinence de s'occuper du consentement. Elle permet en partie d'aboutir au concept de *matter of caring* qui comprend la dimension d'amélioration normative demandant de prendre soin des conditions de consentement, dans diverses situations (voir chapitre 4).

Dans ce chapitre 5 et le prochain, maintenant que le concept de *matter of caring* est forgé, je mets l'accent sur les questions de recherche, ancrées dans mon matériel : (QR1) prendre le consentement comme *matter of caring* dans une organisation qui développe un agent conversationnel, comment cela se fait ? et (QR2) qu'est-ce que cela fait ? ou quel travail cela fait-il ? Ces questions sont abordées également en filigrane dans les chapitres précédents, quand j'évoque des situations rencontrées avec Prélude pour appuyer les arguments – par exemple pour expliquer les pratiques de design d'expérience

utilisateur (p. 83), ou encore pour présenter comment le consentement a été travaillé pour la première fois avec Mr Y (voir la vignette Allô, p. 114).

Par contraste avec l'analyse méliorative et son ambition d'amélioration normative, ces questions portant sur la pratique du *matter of caring* et ce qu'elle fait dans cette organisation m'amènent à procéder à une re-scription. J'emprunte à Jeannette Pols (2008, 2015) le terme de re-scription – qu'elle-même reprend de Harbers, 2005 – pour désigner une approche où, abordée de manière empirique, l'éthique :

aims not to describe practice and neither does it aim to judge it. It does aim to interfere in the practices studied by opening up implicit notions of good care for (self) reflection. Thus, this style of empirical ethics does not restrict itself to describing what others think is good, neither does it impose its own norms in the form of prescriptions. Harbers (2005) calls the resulting interference the 're-scription' of practice. It questions norms and ideals by articulating them as they are hidden in activities and routines to provide participants with fresh ways to look at their practice. (Pols, 2008, p. 52).

Le concept de *matter of caring* qui guide mon analyse est constitué de cette dimension empirique de l'éthique : l'intranquillité qui l'habite amène à questionner des normes et des idéaux – par exemple, ceux qui soutiennent ce qu'une bonne expérience est censée être. Ces questions amènent les membres de Prélude et Mr Y à prendre en considération les interactions de consentement alors qu'elles sont habituellement laissées de côté. Cette prise en considération amène des interférences (entre bonne expérience utilisateur et bonne expérience de consentement par exemple) sur lesquelles s'attarde mon analyse dans le chapitre suivant.

L'objectif de cette re-scription n'est pas de juger si, et à quel point, les pratiques adoptées lors de ce moment de terrain ont été bonnes ou non. Mon objectif est plutôt de retracer, à la lumière de la notion de *matter of caring*, comment le consentement s'est tissé comme projet de design (QR1) en même temps qu'il a débordé comme concept éthique inhérent à une pratique réflexive attentive aux conditions d'expression de ce consentement (QR2). Lors du travail de terrain, cette pratique réflexive m'amène à poser des questions et faire des remarques qui permettent aux participants de prendre davantage en considération ces préoccupations autour du consentement (et de la gestion des données personnelles) dans leurs pratiques, en même temps que ma pratique de

recherche animée par la préoccupation du consentement s'enrichit de préoccupations d'expérience utilisateur, de développement technologique et organisationnel (voir chapitre 6). Parfois ces prises en considération sonnent comme des prescriptions – par exemple quand on me demande de reformuler mes questions en pistes de solution puis en *user stories* (voir chapitres 2 et 6), ou quand j'ai présenté des recommandations de *privacy-by-design* en fin d'immersion (voir annexe, p. 365). Mais ces prescriptions entraînent des reconfigurations de nos pratiques, ce qui tantôt améliore notre collaboration, tantôt génère des frustrations, en amenant toujours nos pratiques à s'influencer et à constituer l'organisation sous cette influence. Ce qui m'intéresse au travers de cette re-scription, c'est également de voir, au travers de ce que fait la pratique de *matter of caring* (QR2), comment le travail du consentement se manifeste, par un jeu de limites, des horizons qui se rencontrent, des tentatives de faire autrement, des compromis.

Conclusion : analyser des compromis

J'ai explicité dans ce chapitre comment j'ai procédé, par une démarche ethnographique, à la collecte des données à partir desquelles j'explore le concept du consentement et comment il fait l'objet de pratiques de design d'un agent conversationnel. J'ai mis l'accent sur la source primaire de données que constitue mon journal de bord, en montrant que ma pratique réflexive fait partie du tissage relationnel qui constitue l'organisation et le phénomène que j'étudie. J'ai explicité ensuite comment le tissage relationnel est un ouvrage qui se poursuit dans les phases de mon analyse itérative : relations aux traces collectées, relations aux concepts et perspectives rencontrées, par les discussions et les écritures qui esquissent des pistes de sens petit à petit. C'est le concept de *matter of caring* qui a finalement condensé le récit par lequel rendre compte de mon cheminement ethnographique.

Mon questionnement dans cette enquête s'articule en une première problématique plus générale que les questions de recherche par lesquelles j'explore mes

données et deux questions de recherche. La problématique générale permet d'ancrer la pertinence de s'attarder sur mon objet de recherche : Pourquoi prendre soin du consentement dans le design d'un agent conversationnel ? Ou pourquoi faire du consentement un *matter of caring* ? Les deux questions de recherche guident mon attention qui s'attarde sur le consentement comme *matter of caring* dans l'organisation dans laquelle je me suis immergée : (QR1) Comment cela se fait, d'aborder le consentement comme un *matter of caring* dans une organisation qui design un agent conversationnel ? (QR2) Qu'est-ce que cela fait de faire du consentement un *matter of caring* ?

La première question guide les trois premiers chapitres de ma thèse par une analyse méliorative qui oriente le consentement vers un horizon éthique de respect et d'égalité dans les situations de vulnérabilité. Les deux questions de recherche sont présentes en filigrane tout au long de la thèse, en s'accroissant dans ce chapitre et le suivant. J'explore mon matériel à la lumière de ces questions par un compte-rendu de l'ordre de la *re-scripture*, qui met l'accent sur les questionnements et compromis qui se sont faits à mesure que le consentement a fait l'objet d'un *matter of caring*.

J'ai ouvert ce chapitre par une réflexion sur la complicité, je voudrais à présent conclure avec l'idée de compromis qui l'accompagne dans la pratique de l'éthique. Incarner un *matter of caring* tel que le consentement, c'est faire l'expérience de faire de la recherche et « *living ethically* » (Shotwell, 2016), par des compromis. C'est une expérience inconfortable mais, je crois, représentative de ce que c'est que de faire quelque chose un *matter of caring* dans le monde, quand on prend ce monde comme point de départ bien que plusieurs choses en lui nous excèdent. À partir du *situatedness* que Donna Haraway (1987) décrit, Alexis Shotwell souligne : « *That world is partially shared, offers finite freedom, adequate abundance, modest meaning, and limited happiness. Partial, finite, adequate, modest, limited – and yet worth working on, with, and for* » (2016, p. 5). Reconnaître ces caractères partial, fini, adéquat, modeste et limité, c'est le point de départ du travail de *matter of caring* : à partir de notre situation limitée dans ce monde, nous pouvons intervenir pour faire bouger certaines lignes, toujours

imparfaitement, pour favoriser un horizon de respect et d'épanouissement dans ces situations et peut-être d'autres, par ricochet.

La rescription que je m'appête à faire dans le prochain chapitre, pour montrer comment le consentement a fait l'objet d'une pratique de *matter of caring* (QR1) et ce que cela fait (QR2), articule la complexité et la complicité dans lesquelles se sont faits plusieurs compromis. Une phrase de Shotwell résonne dans le travail du *matter of caring* : « *[I]f we want a world with less suffering and more flourishing, it would be useful to perceive complexity and complicity as the constitutive situation of our lives* » (2016, p. 8). Un monde plus épanouissant, moins dominant, moins dégradant pour certaines, c'est ce à quoi veut contribuer le concept de *matter of caring* (voir chapitres 2, 3, 4). Alors la complexité et la complicité avec des choses qui nous dépassent ou nous embêtent relèvent de la posture qui les reconnaît comme la situation constitutive de nos vies. C'est cette posture qui oriente le prochain chapitre.

Chapitre 6

Tisser le consentement dans le design de l'agent conversationnel : Un exercice de vulnérabilité

« *The trick is to keep exploring and not bail out, even when we find out that something is not what we thought.* » Pema Chödrön, *When things fall apart*, p. 5

« *The wall usually turns out to be a gate.* »
John Tarrant, *Bring me the rhinoceros*, p. 178

Introduction : une fusion vulnérable

Comment le consentement est abordé comme un *matter of caring* chez Mr Y et Prélude ? Quel travail le *matter of caring* fait-il chez Mr Y et Prélude ? À la lumière de ces questions (QR1 et QR2), j'explore dans ce dernier chapitre les vulnérabilités à l'œuvre au fur et à mesure que le consentement fait l'objet d'un *matter of caring* – c'est-à-dire au fil d'interventions pour améliorer des situations où il est mis de côté, négligé, reformulé puis intégré dans le design de l'agent conversationnel. Pour dévoiler ces vulnérabilités, je porte attention aux « confrontation[s], (...) empêchements, (...) modifications du désir ou de la volonté initiale » (Brugère, 2013, citée dans le chapitre 3, p. 176). Celles-ci me permettent d'être attentive au trouble inconfortable que sème le consentement quand il est travaillé comme *matter of caring* : il n'est pas une simple formalité à mettre en place, mais soulève une certaine absence de contrôle (en matière de traitement des données personnelles particulièrement); il révèle des tentations de fermeture, de dissimulation et de diminution face à l'importance des questions qu'il soulève (à propos du caractère personnel des données, notamment); il risque de créer des accrocs dans l'expérience utilisateur, et celle-ci risque d'empêcher le consentement d'être valable (en particulier par la présence de *dark patterns*).

Tissé dans l'expérience de l'agent conversationnel, le consentement se prête au phénomène de fusion que vise le design d'expérience utilisateur (voir chapitre 2, p. 87)¹⁸². Nous voulons que le consentement soit donné sans influence, dans des conditions qui permettent qu'il soit bien informé. En même temps, nous voulons que les interactions qui demandent et informent ce consentement fusionnent le plus possible avec l'expérience de l'agent conversationnel – qu'il en épouse le rythme (rapide), la tonalité (familiale), par une fluidité telle qu'il n'interromprait pas l'expérience de l'utilisatrice. Mais cette fusion présente un risque pour le consentement : qu'une dimension du consentement disparaisse au fil des différentes formes par lesquelles il est reconfiguré (évolution des formats et des formulations), qu'il n'ait pas tout à fait la même consistance que prévu, qu'il ne soit finalement pas valable.

Ce phénomène de fusion est évoqué par Jean Grondin (2005) dans son explication de la métaphore « fusion des horizons » employée par Gadamer pour décrire le phénomène de compréhension. Il soulève une aporie :

Peut-on toujours contrôler cette fusion qui « se produit » dans la compréhension, mais qui tente en même temps de se maîtriser ? Si la question se pose, c'est que Gadamer soutient tout à la fois qu'il y a lieu de distinguer les horizons (donc de les contrôler) et qu'ils se fusionnent, donc qu'ils ne peuvent *pas* être totalement contrôlés. (p. 411, italiques dans l'original)

C'est là une vulnérabilité à laquelle se prête le *matter of caring* par lequel se tisse le consentement dans le design de l'agent conversationnel : tandis que la fusion du consentement se produit dans l'expérience conversationnelle, comment préserver en même temps ce qui fait que cela aboutisse à une bonne expérience de consentement, c'est-à-dire une expérience de consentement qui rassemble les conditions lui permettant d'être valable ? Cette fusion doit se faire avec vigilance, « il y a toujours une part de risque (...) en toute compréhension. On a beau prendre toutes les précautions que l'on voudra,

¹⁸² Pour rappel, dans le cadre de la définition du travail de design d'expérience utilisateur, j'ai décrit ce phénomène ainsi : « nous voulons que les agents conversationnels transforment notre relation au monde, dans le cadre d'une expérience touristique ou d'aide en santé mentale par exemple, et, en même temps, nous voulons que ces agents conversationnels, par lesquels se fait cette transformation, fusionnent le plus possible avec notre expérience plus générale (d'une balade en ville ou d'un soutien mental par exemple), jusqu'au point de ne plus les voir » (Chapitre 2, p. 67)

nul ne sera jamais à l'abri de la « mauvaise fusion des horizons » » (Grondin, 2005, p. 412). Un risque avec la fusion du consentement dans l'expérience de l'agent conversationnel est que celui-ci soit vain, qu'il relève de mots jouant un rôle d'*ethics-washing* (voir chapitre 1, p. 42) – des mots qui se maquillent d'éthique plutôt qu'ils ne la mettent en pratique. Dans cette vulnérabilité de la fusion, une des tâches du *matter of caring* est ainsi de faire en sorte que *ce que font les mots* – qui informent sur le traitement des données, qui demandent le consentement – soient de l'ordre de l'honnêteté éthique (voir chapitre 1, p. 43), qu'ils traduisent les pratiques effectivement employées.

Dans ce phénomène de fusion qui se prête à plusieurs vulnérabilités, j'analyse des compromis auxquels le travail sur le consentement aboutit au fil de ma complicité avec Mr Y et Prélude, dans la complexité des enjeux d'éthique et de design que leurs activités soulèvent. La posture qui oriente ce chapitre est celle d'Alexis Shotwell (2016, citée dans le chapitre 5, p. 213) qui reconnaît la complexité et la complicité avec des choses qui nous dépassent ou nous embêtent – ou avec des choses qui nous confrontent, nous empêchent, modifient notre désir ou volonté initiale (pour reprendre les termes de Fabienne Brugère, 2013). Cette posture reconnaît cette complicité dans la complexité comme étant la situation constitutive du tissage relationnel entre Mr Y, Prélude, le consentement, leurs agents conversationnels, et moi dans tout cela.

Je vais d'abord présenter une première vulnérabilité à laquelle se confronte le travail du consentement : l'organisation dans laquelle il devait se faire change, passant du domaine de la santé mentale à celui du tourisme – ce changement remet en cause la pertinence de faire du consentement un *matter of caring*. Je vais cependant montrer ensuite qu'en se glissant dans les logiques de Prélude, le travail sur le consentement s'y fait un chemin : d'abord réservé à un espace limité, il se heurte à des confrontations, empêchements et modifications, en même temps qu'il en provoque dans le design d'expérience utilisateur. Le *matter of caring* travaille ici à favoriser la qualité du consentement (voir chapitre 3, p. 159) – je vais m'attarder alors sur l'attention qu'il porte aux conditions (formulations, design d'interaction, pratiques de traitement des données) dans lesquelles le consentement est demandé et donné. Ce travail demande de clarifier

des pratiques. Je vais montrer une vulnérabilité que cela expose pour l'organisation qui, par le choix de ses outils, se positionne dans une soumission limitant le contrôle des données qu'elle traite et de l'expérience utilisateur qu'elle conçoit. Enfin, j'aborderai un épisode où les efforts se sont réunis pour réparer l'expérience de l'agent conversationnel quand celui-ci ne fonctionnait plus, manifestant ainsi un *matter of caring* qui travaille dans l'interdépendance de voix humaines et artificielle pour améliorer et maintenir une expérience.

De la santé mentale au tourisme : cela vaut-il encore la peine de s'occuper du consentement ?

Complice avec Paul et son chatbot en santé mentale, je prépare avec lui le projet de concevoir un modèle de consentement pour les agents conversationnels de Mr Y. Dans ce domaine, le consentement compte à plusieurs niveaux. Il est important pour la relation de confiance habituelle entre les professionnels en santé mentale et leurs patients – si les bots de Mr Y veulent tenter de reproduire des dynamiques d'interaction similaires, le consentement doit donc être abordé. Il est également important étant donné la sensibilité des données que l'utilisatrice s'approprierait à partager – ce sont des données de santé. C'est aussi un moment qui permet de clarifier auprès de l'utilisatrice les limites des capacités du bot – il ne peut pas intervenir comme une personne humaine, ni proposer de diagnostic. Enfin, c'est un moment où l'utilisatrice peut poser des questions si elle veut en savoir plus sur cette relation d'aide dans laquelle elle s'approprie à s'engager. Les aléas du terrain font que toutes ces questions sont finalement mises de côté, car un autre projet prend le dessus : celui de Prélude, la Voix de Montréal, un agent conversationnel touristique. Pourquoi le consentement resterait-il pertinent dans ce contexte où la relation est bien moins vulnérable, et les données collectées bien moins sensibles ? Cela vaut-il la peine de faire du consentement un *matter of caring* ? De Mr Y à Prélude, je vais présenter le chemin que le consentement s'est fait, dans les bifurcations de nos volontés initiales, à Paul et moi.

L'accès au terrain : Mr Y, où le consentement compte

« Bonjour à tous les trois,
Je vous présente Camille, une amie de longue date en France. On s'est retrouvés à Montréal et il s'avère que nos expériences respectives ont des points de similitude. » (*Combinaison de deux courriels de Paul qui m'introduisait à ses anciens associés, 14 septembre et 8 novembre 2018*)

Paul et moi avons fait notre scolarité dans la même école en France, de nos 6 à 18 ans. Perdus de vue sur les chemins que chacun a pris pour les études supérieures – lui en commerce et gestion des entreprises, moi en philosophie et communication – nous nous sommes retrouvés dans la foule d'un couloir d'un centre culturel montréalais, un jour de juin 2018, à un événement sur l'IA et ses missions sociales. Il venait présenter sur scène son projet Mr Y, un chatbot pour aider à gérer l'anxiété; je participais à l'événement, à titre de membre de la Déclaration de Montréal IA Responsable. Paul et moi nous connaissions donc de longue date, nos chemins se sont recroisés à Montréal, au carrefour de notre intérêt partagé pour les questions sociales de l'intelligence artificielle.

Nous prenons un verre un mois plus tard pour envisager une éventuelle collaboration : Paul me partage sa sensibilité à l'éthique qu'il veut travailler pour Mr Y. Derrière cette volonté, il manifeste un souci d'améliorer les façons de faire du développement technologique, en particulier dans un contexte de santé mentale qui demande d'être attentif et de prendre soin de situations de vulnérabilité – voire de détresse – dans lesquelles peuvent se trouver les utilisateurs. Ce qui l'intéresse, comme moi, c'est de travailler la dimension concrète de l'éthique quand elle est de l'ordre du *doing*. Paul parle de travailler l'éthique dans les *opérations*, ou activités quotidiennes, de son organisation – ce que mon vocabulaire désigne sous le terme de *pratiques* des acteurs. Nos façons de parler sont différentes mais elles sont habitées par une préoccupation similaire : aborder l'éthique comme un « *doing and intervening* » (Puig de la Bellacasa, 2011, citée dans le chapitre 4); pour chacun de nous, elle fait l'objet d'un *matter of caring*.

Le projet de notre collaboration s'affine au fil de quatre autres rencontres (13 septembre 2018, 29 novembre 2018, 6 décembre 2018¹⁸³ et 15 janvier 2019). L'idée initiale est d'élaborer une charte éthique et de mettre en place un comité d'éthique. Mais au fil des discussions, les enjeux éthiques qui se révèlent particulièrement présents chez Mr Y touchent au consentement et au traitement éthique des données collectées par le bot. La discussion et les réflexions qu'elle suscite amènent le projet à se préciser autrement : « moins aborder l'éthique de manière générale par une charte (ça on peut trouver plusieurs modèles), et plutôt travailler un défi particulier et faire des recommandations pour le développement du produit (ou du service) » (extrait d'un courriel que j'ai envoyé à Paul, le 27 novembre 2018). Par ces « recommandations », il s'agit de faire des « propositions formelles » (Vial, 2010; voir chapitre 4, p. 206) relevant des activités liées au projet en design.

À l'issue de la rencontre du 29 novembre, le projet change de forme pour s'orienter vers une recherche qui relève de la concentration sur un enjeu éthique – le consentement – dans la conception de l'expérience pour les chatbots de Mr Y :

Camille : « ça rassemble les enjeux liés aux données personnelles, à la prise de conscience de l'utilisateur par rapport à ce qui est fait de ses données, et c'est aussi un passage obligé pour pouvoir utiliser le chatbot, donc c'est en même temps là qu'on peut commencer à établir une relation de confiance avec l'utilisateur. L'objectif du projet Mitacs pourrait être de concevoir l'expérience de consentement pour les chatbots Mr Y. Qu'en penses-tu ? »

Paul : « Oui je suis d'accord c'est un très beau projet. J'y ai pensé de mon côté, on pourrait saisir cette opportunité pour avoir une base scientifique pour concevoir l'expérience de consentement pour les chatbots de Mr Y. Ça prend plusieurs des points qu'on a mentionné sur l'éthique en IA donc c'est parfait. Ça peut être un pilier pour développer une relation de confiance avec les utilisateurs »

(Extraits de courriels échangés entre Paul et moi les 29 novembre et 1^{er} décembre 2018).

L'intérêt qui soutient la concentration sur le consentement repose sur le fait que cet enjeu en rassemble plusieurs, à propos du traitement des données personnelles et de l'éducation de l'utilisatrice sur ces questions. L'autre intérêt est l'effet que ce travail sur

¹⁸³ C'est le jour du travail de co-design avec Allô, où nous avons travaillé pour la première fois sur le consentement. Voir la vignette Allô, chapitre 3, p. 92.

le consentement peut avoir : il peut favoriser la confiance des utilisateurs, ce qui permettrait une meilleure ou plus grande adoption de chatbot. Nous mobilisons ainsi dans ces courriels la valeur de la confiance, comme c'est souvent le cas pour valoriser l'effet que la mise en pratique de l'éthique peut avoir (voir chapitre 1, p. 38, et chapitre 3, p. 133).

Paul mentionne également l'opportunité d'avoir « une base scientifique pour concevoir l'expérience de consentement pour les chatbots de Mr Y » : dans un précédent courriel, il avait justifié à ses associés que mes activités de recherche avec eux serait « d'une grande valeur ajoutée dans [leur] développement, et notamment dans la crédibilité qu'[ils ont] auprès de leurs partenaires » (extrait d'un courriel de Paul, 14 septembre 2018). Ce qui soutient donc également ce projet, qui s'affine autour du consentement, relève de la crédibilité (d'un travail de recherche universitaire) dont l'organisation a besoin auprès de ses partenaires – ceux-ci étant des acteurs du milieu de la santé, incluant membres d'hôpitaux, intervenants sociaux et chercheurs universitaires. Ce projet a ainsi une valeur instrumentale de ce que ma contribution peut ajouter à Mr Y (de la crédibilité et de la confiance), comme de ce que Mr Y peut ajouter à mon travail de recherche (un terrain où la mise en pratique de questions éthiques compte). Pour tous les deux, ce projet autour du consentement s'affine ainsi au fil d'un processus d'intéressement qui rassemble plusieurs intérêts respectifs (Callon et al., 1988; Giroux, 2006 abordés dans le chapitre 3, p. 131).

Enfin, ce projet fait l'objet d'un investissement financier de la part de Mr Y, qui s'engage, par une entente Mitacs, à participer à ma rémunération¹⁸⁴ et dédier une partie de ses activités à ce projet. Soutenu ainsi, le consentement s'annonce comme une préoccupation qui ne fait pas qu'animer des intérêts mais amène des activités à se structurer – un premier pas pour que le consentement soit un *matter of caring*, pas seulement pour Paul et moi chacun de notre côté, mais dans cette organisation qui met en place des conditions qui permettent à ce projet de compter.

¹⁸⁴ Les bourses Mitacs Accélération sont financées à 50% par l'organisme partenaire dans lequel est réalisé le projet (voir <https://www.mitacs.ca/fr/programmes/acceleration#business>).

Le 6 décembre 2018, Paul m'invite à la rencontre avec Allô pour la séance de co-design sur un projet de chatbot pour eux (cf vignette dans le chapitre 3, p. 114). Notre collaboration n'a pas encore officiellement commencé¹⁸⁵, mais c'est l'occasion de voir concrètement ce dans quoi j'entrerai plus vivement par la suite : comment se passe le design d'un agent conversationnel, en particulier le design de la conversation autour des thèmes qui seront abordés avec l'utilisatrice. Dans le cas d'Allô, il s'agit de sujets qui peuvent susciter des inquiétudes chez les jeunes; dans le cas de Mr Y, ce sont des sujets à propos de facteurs d'anxiété ainsi que des ressources (exercices, vidéos, lectures, exercices de respiration, etc.) pour l'apaiser. Aborder ces sujets relève du *care* puisqu'il s'agit d'apporter du soutien à quelqu'un pour améliorer sa situation, avec écoute et sollicitude. Dans ce contexte, le travail sur le consentement s'inscrit dans ce travail de *care* – en faire un *matter of caring* s'annonce dans cette suite logique.

Cette réunion est aussi l'occasion d'expérimenter pour la première fois comment travailler sur le processus de consentement dans le design d'un agent conversationnel (QR1) :

Paul me propose de me joindre à Christine pour observer le bloc de l'introduction car c'est sûrement à ce moment qu'il faut introduire la question du consentement. « Avant le nœud 'se connaître' tu peux créer un nœud 'consentement' » dit Paul. Christine semble bien comprendre ce qu'il veut dire : depuis son ordinateur, elle crée un carré sur l'interface et l'intitule « consentement ».

Nous nous demandons alors où ce nœud doit être placé : est-ce la première interaction, le premier élément de discussion avant même l'introduction ? Ou alors vient-il un peu après l'introduction, une fois que le chatbot s'est présenté et a un peu échangé avec l'utilisateur ? Nous n'arrivons pas à décider, hésitant entre l'obligation d'obtenir le consentement et le risque que cela soit un obstacle d'entrée à la fluidité de la discussion. Christine propose à sa collègue Fanny qui n'a pas assisté à la précédente séance de travailler sur ce nœud; comme ça, elle pourra continuer de son côté son flow d'introduction commencé la semaine dernière. Elle me propose ensuite de me joindre à Fanny. Le nœud du consentement se retrouve ainsi à une table un peu en arrière de la grande table où les autres sont tous rassemblés.

(Extrait de la vignette Allô décrite dans le chapitre 3, p. 114)

¹⁸⁵ Elle commence l'année suivante, en mars 2019. C'est à ce moment que j'obtiens la bourse Mitacs et mon certificat d'éthique. Mais ma première entrée physique sur le terrain a lieu le 1^{er} avril 2019 (voir ci-après).

Le consentement est ainsi identifié comme un moment particulier de la conversation : le tout début, ou *onboarding* (voir chapitre 2, p. 143). Il a en même temps été relégué à une table en arrière, pour ne pas ralentir le travail déjà amorcé sur les autres étapes de la conversation. Même dans un contexte de soin en santé mentale, le consentement s'annonce comme un moment qui gêne : on ne sait pas exactement où le placer sans risquer qu'il heurte la fluidité de l'expérience de *onboarding*, on le travaille dans un espace à part, pour éviter de ralentir le travail sur l'ensemble de l'expérience. C'est la tension de la fusion du consentement dans l'expérience principale qui s'annonce : comment travailler ce moment inconfortable du consentement pour qu'il se fonde dans l'expérience, sans la ralentir, ni heurter sa fluidité ? Le travail du *matter of caring* s'esquisse dans cette tension qui tente de laisser le consentement de côté. Elle est encore plus vive par la suite, dans le tout autre contexte des agents conversationnels de Prélude (comme nous allons le voir un peu plus tard).

Cette mise à l'écart soulève d'autres questions (qui nourrissent finalement les chapitres 2 et 3 de cette thèse) : qu'est-ce qui peut rendre une telle expérience agréable, facile et utile ? Une bonne expérience pourrait-elle être autrement ? Cette bulle de questions éclate au son des éclats de rire du reste du groupe à la grande table, qui travaille sur le *flow* introduction dans lequel le consentement serait à terme intégré :

Des éclats de voix et des rires retentissent du flow introduction qui s'interroge sur le genre du chatbot (appelé Carol) : « il lui faut un genre fluide à Carol », « mais on est principalement des femmes ici, pourquoi ce ne serait pas une femme ? ». Le débat s'étend à l'ensemble du groupe : « Comment tu fais comprendre ça toi, un genre fluide ? », « et pourquoi on ne ferait pas choisir à l'utilisateur s'il veut que le chatbot soit masculin ou féminin ? »

(Suite de la vignette Allô, à partir du journal de bord,
notes de terrain du 6 décembre 2018)

Le groupe soulève une question brûlante, au cœur de la définition de la personnalité de la voix dont la mission est de créer un lien de confiance et de familiarité avec l'utilisatrice, qu'elle se sente à l'aise de parler (voir chapitre 2, p. 105) : quel genre donner à cette voix ? De la fluidité et du choix sont proposés : fluidité du genre, laisser choisir l'utilisatrice le genre qu'elle veut donner à la voix qui la guiderait dans un moment de vulnérabilité. Mais

pourquoi rit-on quand on imagine un genre fluide, quelle voix imagine-t-on, la fluidité d'un genre a-t-elle une voix particulière ? Qu'est-ce qu'une voix masculine ou une voix féminine, quels attributs leur donne-t-on ? Et puis dans ce cas, le consentement devrait-il avoir une voix qui sonne comme celle d'un genre en particulier ? Ces questions sont des pistes qui peuvent orienter le travail sur le design de la persona de l'agent conversationnel : si ce travail relève du *matter of caring*, alors des manières seront trouvées pour explorer cette fluidité du genre, pour lui donner de la place, et peut-être un mode d'expression, alors qu'elle est souvent mise de côté voire ignorée.

Le premier jour de mon entrée chez Mr Y, le 1^{er} avril 2019, je propose d'explorer cette question du genre de la voix : « Ça me semble intéressant dans un contexte d'anxiété où l'identification du genre peut être une source d'anxiété pour certaines personnes » (Journal de bord, notes de terrain du 1^{er} avril 2019). Paul reconnaît que ces questions liées à la personnalité du bot sont souvent posées. Cette piste de *matter of caring* portée sur la voix et sa persona n'est cependant pas poursuivie, car les activités de Mr Y ont suivi un autre cours.

L'entrée sur le terrain : le consentement compte-t-il finalement ?

J'arrive à 10h35 à l'accueil du grand espace partagé au quatrième étage d'un immeuble en briques rouges dans le sud-ouest de Montréal. Cet étage rassemble plusieurs agences, studios et startups, œuvrant dans les secteurs du marketing et jeu vidéo. L'espace est vaste, le design est moderne et sobre, de nombreuses fenêtres. Paul vient me chercher en trottinette, grand sourire. Il m'invite à poser mon manteau au vestiaire et m'accompagne vers la cuisine ouverte pour me proposer un thé. Nous nous dirigeons ensuite vers l'espace de Mr Y : une table en longueur, six écrans d'ordinateur et des chaises. Il manque cependant une chaise pour moi, comme il en manquait une pour Philippe quand il est arrivé me dit Paul, il va en prendre une discrètement dans la section dédiée à une autre organisation à côté. Il me présente à ceux qui sont là : Félix avec qui Paul s'est associé pour créer Prélude, une startup qui va proposer des services avec agents conversationnels pour différents domaines (notamment en tourisme), Liam qui fait des études de marché pour Prélude, Philippe qui est développeur et s'occupe tant du site de Prélude que de tout le côté technique. Paul me présente comme allant travailler sur l'éthique et la question du consentement pour Mr Y, une question importante parce que, par exemple, les deux étudiants en informatique ont du mal à obtenir leur certificat

d'éthique pour travailler avec Mr Y car le comité leur pose de nombreuses questions sur les données qu'ils utilisent (des publications sur des forums en ligne), il demande si le consentement a été obtenu pour les traiter. Finalement, ils vont travailler avec Mr Y en tant qu'employés plutôt que dans un cadre académique. Ils ne pourront pas être là à la réunion d'équipe à 11h, mais ils seront là la semaine prochaine.

(Vignette du premier jour, à partir du journal de bord, notes de terrain du 1^{er} avril 2019)

En arrivant chez Mr Y, on arrive en même temps chez Prélude : les deux organisations partagent le même espace de travail et les réunions d'équipe invitent ensemble les personnes qui travaillent pour l'une ou l'autre organisation. Les personnes qui sont présentes la plupart du temps sur place travaillent pour Prélude. Paul introduit le consentement – ou l'objet au cœur de mon projet avec Mr Y, c'est-à-dire ce qui fait que je suis ici dans cet espace avec eux – sous l'angle qui touche Mr Y en ce moment : une formalité qui empêche des étudiants-chercheurs en intelligence artificielle de rejoindre l'équipe dans le cadre de leur recherche. En filigrane, le consentement apparaît comme ce qui peut séparer, ce qui peut empêcher un projet de collaboration scientifique quand il est manquant – d'où l'objectif de l'obtenir quand il est manquant, un objectif que ma contribution peut aider à atteindre. Pour rendre tout de même possible dès maintenant la collaboration souhaitée avec ces personnes, le problème du consentement est donc contourné en changeant les termes de la relation avec ces étudiants-chercheurs¹⁸⁶, pour ne plus être une relation académique soumise à l'obtention d'un certificat d'éthique. En attendant, une personne comme moi est là pour s'occuper de cet objet épineux.

Dans cet espace qui mêle Mr Y et Prélude, la relation entre les deux organisations me paraît ambiguë : sont-elles distinctes ou sont-elles une seule et même organisation ? Mr Y, est-ce une organisation ou seulement un produit, un projet de chatbot en santé mentale ? Sur le site internet de Prélude, le chatbot Mr Y est présenté comme un produit de Prélude. À la première réunion d'équipe de ce lundi, après avoir mentionné onze pistes de projets en cours, Félix résume trois leviers d'action qu'a Prélude : la santé mentale

¹⁸⁶ Ils quittent finalement en juin.

avec Mr Y, le tourisme (ils ont déposé leur candidature pour entrer dans un incubateur dédié) et l'offre de services à des agences de publicité (une idée de projet d'agent conversationnel est en préparation pour le client d'une agence). Pourtant Paul me précise plus tard, le 11 avril 2019, que Mr Y existe toujours comme entreprise distincte de Prélude, lui et Félix ont chacun des parts dans l'une et l'autre organisation. En ce qui me concerne, il est entendu que je travaille principalement sur Mr Y et intervienne ponctuellement pour des conseils sur les projets de Prélude. Leur stratégie pour Prélude est pour l'instant de *pitcher*, c'est-à-dire présenter de manière concise et efficace, des idées de projets au plus de gens possible jusqu'à ce qu'un premier client s'engage financièrement pour réaliser un de ces projets. Quand des mandats commenceront, ils demanderont à Philippe sa participation technique, à moi ma contribution d'un point de vue d'éthique et à Liam son intervention pour des études de marché. Ils feront également appel à des partenaires externes s'ils ont besoin d'autres expertises. Mon arrivée chez Mr Y est donc marquée par un chevauchement entre deux organisations, Mr Y et Prélude, dans l'incertitude de savoir quels projets prendront le plus de place dans les prochains mois. Dans ce qui peut contribuer, chez moi, aux possibles projets de Prélude, ce sont mes compétences en éthique (de l'IA) qui sont retenues, plutôt que le projet sur le consentement tel qu'il a été défini avec Paul pour Mr Y.

Dans l'ambiguïté de la place que prendra le projet Mr Y par rapport aux pistes lancées par Prélude, je m'en tiens aux intentions plus claires que Paul et moi avons établies autour du consentement pour le chatbot en santé mentale. Pour Mr Y, je prends connaissance de plusieurs documents organisationnels portant sur l'organisation et le chatbot; je découvre, utilise et compare plusieurs agents conversationnels en santé mentale; j'ai des réunions régulières avec Paul pour en discuter et faire avancer le projet. Le consentement comme *matter of caring* relève donc d'un travail d'arrière-plan, plutôt solitaire, partagé de temps en temps avec Paul pour qui cet enjeu compte (éthiquement mais aussi pour ajouter de la valeur – confiance, crédibilité, collaborations scientifiques – à son chatbot Mr Y).

Ce travail s'attache dans un premier temps à entrer en relation avec ce qui se fait déjà dans le domaine de la santé mentale, avec les agents conversationnels Woebot et Wysa notamment. J'étudie les étapes de leurs conversations, comment ils abordent la demande de consentement, pour inspirer ce qui pourrait être fait avec Mr Y. J'emprunte la formule « *Before we go on* » employée par Woebot pendant le *onboarding* pour introduire les informations permettant d'éclairer le consentement demandé ensuite¹⁸⁷. Wysa formule également l'interruption que demande le consentement par l'emploi de « *Before* », avant de passer au principal de l'expérience.

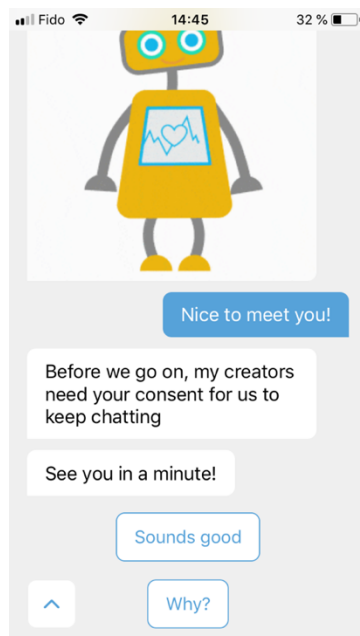


Figure 8 : Capture d'écran du chatbot Woebot qui introduit la demande de consentement.

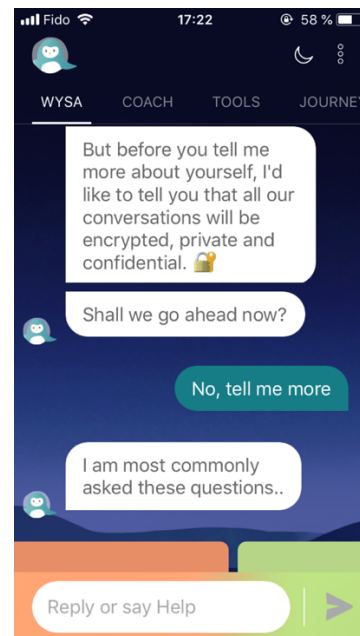


Figure 9 : Capture d'écran du chatbot Wysa qui introduit la demande de consentement.

Je propose également de reprendre le format que Wysa utilise pour présenter la même information clairement. Ce format est un carrousel¹⁸⁸ : chaque thème est présenté par un titre ou une image et un court descriptif, il est possible de faire défiler

¹⁸⁷ La formule que je propose pour Mr Y est : « *Before we go on, I need to tell you a bit more about how my creators ensure your privacy.* » (Document « Consent design flow – Mr Y », 8 juillet 2019)

¹⁸⁸ Voir <https://developers.google.com/assistant/conversation-design/carousel>

horizontalement chaque thème et avoir plus d'information sur l'un ou l'autre en le sélectionnant.

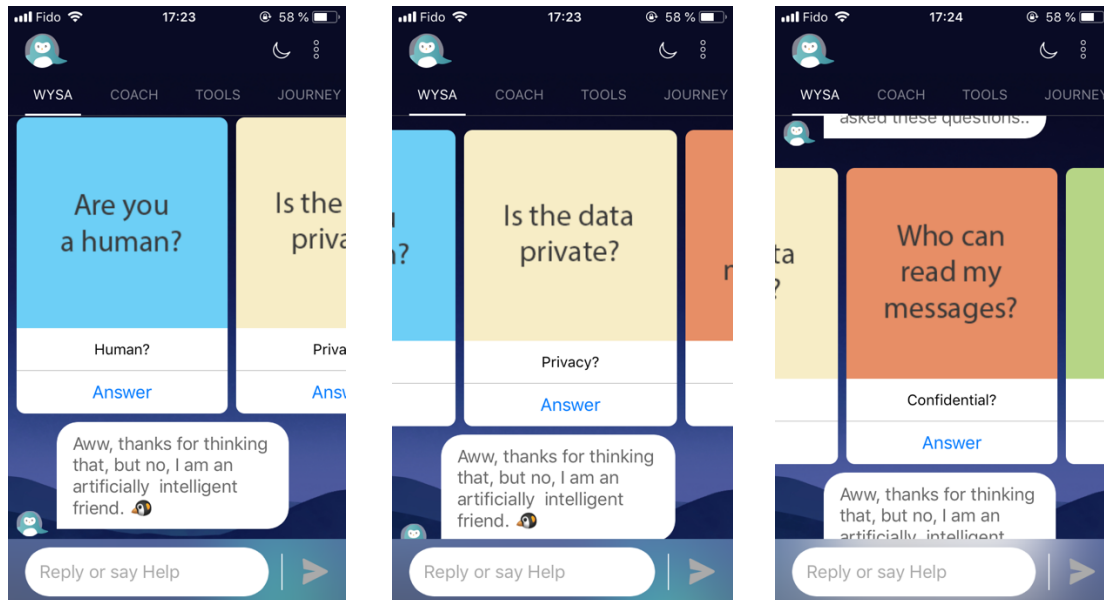


Figure 10 : Captures d'écran du carrousel qui permet d'informer le consentement dans l'application du chatbot Wysa. 9 avril 2019 à 17h23.

L'introduction du consentement par la formule « *Before we go on / Before you tell me more* » ainsi que la présentation de ces thèmes sous forme de carrousel permettent de voir comment d'autres, dans le même domaine, ont porté attention à ce moment et ces informations pour ne pas les négliger. En d'autres termes, passer du temps à explorer ces façons de faire permet de voir à quels résultats ont abouti d'autres organisations qui ont fait du consentement un *matter of caring* dans la conversation de leurs chatbots.

Ces deux exemples de chatbots, qui demandent le consentement des utilisateurs à même la conversation, me permettent également d'identifier les grands thèmes à aborder pour informer le consentement : ce qui est fait avec les données personnelles collectées, comment elles sont sécurisées, comment la confidentialité est préservée, qui est derrière le bot (transparence sur sa non-humanité), quelle relation de soin ils proposent (sans remplacer une psychologue ou intervenante sociale) – ces thèmes sont identifiés dans mes notes de terrain de décembre 2018 et avril 2019. Je remarque également les thèmes qui n'y sont pas, en particulier ceux qui permettent à l'utilisatrice

de demander le retrait ou la suppression des données personnelles. Et il n'est pas clairement mentionné si les chatbots peuvent continuer d'être utilisés si l'utilisatrice ne donne pas son consentement. Le niveau d'information est par ailleurs sommaire, ce qui m'interroge sur le détail qu'il faut donner pour informer clairement l'utilisatrice du traitement de ses données personnelles – cela reste une question jusqu'à la rencontre avec le cabinet d'audit Data Protection Consulting (DPC) le 11 septembre 2019 (voir chapitre 3, p. 155) qui dit que :

13 DPC : L'objectif, en fait, c'est que ce soit suffisamment clair et suffisamment précis
14 pour que la personne sache dans quoi elle met les pieds. (...) Ce qu'il faut
15 indiquer c'est principalement la finalité du traitement de données – donc
16 pourquoi est-ce qu'on va collecter les données personnelles – la base légale
17 (consentement, on va dire « pour qu'on puisse le faire, il faut que vous soyez
18 d'accord »), après on va avoir les destinataires. (...) en termes de granularité,
19 on n'est pas obligé de lister tous les gens mais on peut donner les catégories
20 de destinataires.

(2^{ème} extrait de retranscription de la réunion avec DPC, 11 septembre 2019).

Le niveau de détail attendu reste donc plutôt vague, mais les thèmes à aborder sont relativement clairs.

Cette première étape du travail de *matter of caring* se fait ainsi dans les coulisses de Mr Y et consiste à explorer, tester et utiliser des chatbots similaires existants pour voir en particulier comment le consentement s'inscrit, voire fusionne, dans l'expérience générale de ces agents conversationnels – j'en profite en même temps pour voir comment ces agents conversationnels fusionnent dans ma vie quotidienne, en jouant avec eux :

Dans le métro, j'ouvre Wysa pour continuer à tester l'application, j'ai reçu une notification qui me disait qu'elle¹⁸⁹ avait quelque chose de nouveau. C'est un jeu : elle veut deviner quelle est mon émotion principale. Je pense à l'inquiétude, elle devine la colère, je ne suis pas tout à fait d'accord (mais aurait-elle un peu raison ?)... Puis je lui demande, en coupant le *flow* qu'elle me propose, si ce jeu était fait pour l'entraîner à reconnaître des émotions. Ses réponses ne sont pas complètement en contexte, elle me suggère un carrousel pour explorer les fonctionnalités du bot et modifier mes préférences.

(Journal de bord, notes de terrain du 11 avril 2019)

¹⁸⁹ Dans mes notes, j'associe Wysa au pronom elle, sans que ce pronom soit spécifié par l'agent conversationnel – je projette ainsi sur le flou du genre, une persona féminine.

Alors que le bot manifeste être à mon écoute, quand je lui demande si ce jeu a une visée d'entraînement, j'interroge ce que le bot ne dit pas : me propose-t-il simplement de jouer, ou me propose-t-il de jouer pour qu'il puisse lui-même mieux me connaître, dans l'intimité de mes émotions ? Ce questionnement porte la marque d'une tentative de clarification dans l'ambiguïté du traitement des données personnelles, une clarification que le *matter of caring* cherche quand il porte sur le consentement, dans une idéale transparence. Il marque également une attitude dubitative, qui ne s'en tient pas au jeu mais interroge ce qui peut être fait avec les données partagées dans ce moment ludique que je n'ai pas moi-même sollicité.

Mon inquiétude, que Wysa prend pour de la colère, s'explique un peu plus tard dans la journée, quand Félix et Paul mobilisent toute l'équipe pour travailler sur un projet qu'ils sont sur le point de pitcher :

Toute l'équipe est là aujourd'hui, ce n'est pas habituel, il n'y a pas assez de chaises alors deux personnes sont dans le canapé qui fait office de cloison avec le studio de jeu vidéo qui travaille derrière. Assise à l'une des chaises (je suis arrivée tôt ce matin), je suis en train de réviser deux diapositives dans la présentation Google slide qui sert de document de travail partagé.

C'est une des premières fois que tous les membres de la jeune startup travaillent ensemble : un pitch est en préparation pour un projet d'agent conversationnel qui raconterait des histoires à des enfants hospitalisés. Paul et Félix ont demandé à chacun de préparer une liste d'enjeux et de pistes de solutions selon leur expertise. On m'a demandé d'identifier les enjeux éthiques de ce projet qui implique des Google homes dans des chambres d'enfants hospitalisés. Le contexte est sensible, les questions et signaux rouges dans ma tête sont nombreux. J'ai inséré mes deux diapositives de questions dans le Google slide : on me presse de formuler des actions pour proposer des solutions plutôt que de m'en tenir aux interrogations.

(Vignette de préparation du pitch, d'après le journal de bord, notes de terrain des 22 et 23 avril 2019)¹⁹⁰

C'est la première expérience qui réunit toute l'équipe autour d'un projet commun – je suis invitée dans la préparation pour identifier les enjeux éthiques, et des pistes de solutions pour les adresser.

¹⁹⁰ Cette vignette est reprise de l'article « La recherche CCO comme pratique réflexive » que Boris H.J.M Brummans et moi avons publié en 2021 dans la revue *Communication & Organisation*, p. 142.

Le contexte de ce projet est sensible : installer des Google Homes à l'hôpital, dans des chambres d'enfants, cela demande les plus hauts standards de confidentialité et sécurité en matière de traitement des données que les assistants vocaux collecteraient. Pourtant, les connaissances et les sensibilités par rapport aux questions de protection des données se révèlent inégales parmi les membres de l'équipe – d'où une certaine inquiétude de ma part. Félix me rétorque par exemple : « mais ça, les données, ce n'est pas notre responsabilité, c'est Google qui s'en occupe » (Journal de bord, notes de terrain du 12 avril 2019) quand je lui demande quelle organisation va se charger de la sécurité et confidentialité des données personnelles collectées par les Google Homes, dans ce projet qui implique au moins trois parties prenantes (Prélude, l'agence de développement web chargée des infrastructures informatiques pour ce projet, l'hôpital). Je reste perplexe face à une telle réponse alors que ma question est motivée par la préoccupation de reconnaître la responsabilité partagée par chaque organisation impliquée dans le traitement de données personnelles.

Étant donné la rapidité avec laquelle la réponse de Félix évacue le problème de responsabilité soulevé, nos préoccupations respectives ne sont manifestement pas accentuées aux mêmes endroits : les siennes se concentrent sur la conception de l'expérience utilisateur originale à présenter dans quelques jours, alors toute question qui ralentit sa créativité est pour l'instant écartée. Sa préoccupation envers l'expérience utilisateur a d'ailleurs la teinture d'un *matter of caring* : comment améliorer le quotidien d'enfants hospitalisés par des expériences avec des agents conversationnels qui leur raconteraient des histoires personnalisées ? Pourtant, prendre toutes les mesures nécessaires pour protéger les données collectées sur des personnes (mineures et hospitalisées) en situation vulnérable, et limiter les risques de collecter des informations médicales, ce sont des points importants pour que cette expérience soit bonne – et qu'elle soit digne de confiance auprès des potentiels clients auxquels ils vont pitcher le projet, comme auprès des potentiels utilisateurs, et leurs parents, si le projet se concrétise. C'est là l'accent que je mets sur ces questions qui font l'objet d'un *matter of caring* de ma part :

sensibiliser au respect des données personnelles comme une des dimensions qui contribuent à offrir une bonne expérience.

Lors de cette préparation, l'attention que j'amène à porter sur l'enjeu de respect de la vie privée passe par de nombreuses questions qui interfèrent dans l'imagination du projet. Ces interférences peuvent ouvrir des réflexions sur ce que serait une bonne expérience dans ce contexte, pas seulement par une approche divertissante mais aussi par une approche respectueuse de la vie privée. Face aux nombreuses questions que je soulève – à propos de l'anonymisation, de la limite de la collecte de données, la prévention des erreurs d'activation, etc. – on me demande de transformer mes questions en pistes de solution. À son tour, cette demande interfère dans ma façon habituelle de travailler l'éthique : je suis habituée à soulever des enjeux, moins à proposer des solutions concrètes. Mais cette demande ouvre ma réflexion sur les solutions possibles face à ces enjeux. Que ce soit une interférence de l'éthique dans la créativité, ou une interférence qui demande des solutions dans un travail de l'éthique centré sur les questions, l'inconfort suscité par ces interférences permet au *matter of caring* de s'immiscer comme un trouble à adresser pour améliorer une situation. Il nourrit également des façons différentes d'imaginer le projet.

Du côté du travail de l'éthique, cet effort de reformulation sous forme de pistes de solutions s'oriente dans un premier temps sur les régulations avec lesquelles entrer en conformité pour la gestion des données collectées : il y a un devoir de conformité à respecter, voilà les règles auxquelles obéir (à vérifier avec une juriste). À la demande de Félix, je crée un document qui liste ces règlements de protection des données personnelles (LPRPDE au Canada, le *Children's Online Privacy Protection Act (COPPA)* aux États-Unis, le RGPD en Europe). La mise en conformité donne au travail de l'éthique le poids de l'autorité, mais également celui de la contrainte qui ajoute de la complexité dans les processus de développement et pose certaines limites aux idées de ce qui est possible. C'est là une des similarités entre le travail du consentement et celui du *matter of caring* :

ils posent des limites¹⁹¹ au possible, pour soumettre ce possible au respect de ceux qui interagiront avec l'objet, qui n'est pour l'instant qu'imaginé.

Je crée également un document compilant les définitions de ce que peut être une donnée personnelle, car quelqu'un de l'équipe me mentionne, dans un couloir entre deux réunions, que pour gérer les données personnelles, il suffit d'enlever les caractères identificatoires (noms, prénoms), ce n'est pas si compliqué. Cette remarque ne me semble pas au fait de la vastitude de ce que peut être une donnée personnelle, dont les caractères identificatoires peuvent être directs ou indirects, et complexifient la tâche d'anonymisation étant donné que de nombreux recroisements entre les données sont possibles (Rocher et al., 2019) – cet enjeu va revenir quelques mois plus tard, en octobre, quand Félix aura écouté un podcast à ce sujet (voir ci-après, lors de la réunion du 23 octobre 2019). L'enjeu du respect de la vie privée, objet d'un *matter of caring*, passe ainsi par un travail de sensibilisation qui relève de la littératie numérique (voir chapitre 3, p. 177) : il s'agit de sensibiliser plusieurs membres de l'équipe (et développer mes propres connaissances) à propos de l'envergure des données personnelles et de l'écosystème des régulations, qui soutient le développement responsable de technologies, en cadrant les pratiques de traitement de ces données personnelles. C'est un premier pas dans la mise en place des conditions qui permettent de respecter les personnes auxquelles on demandera de consentir (à ces pratiques qui traiteront leurs données personnelles) pour utiliser notre agent conversationnel.

Lors de cette première expérience en équipe, le *matter of caring* travaille ainsi comme un trouble et une sensibilisation qui s'immiscent dans les réflexions des uns et des autres. Petit à petit, les préoccupations de chacune, selon son expertise et ce qui lui tient à cœur, s'adaptent pour pouvoir travailler et penser ensemble : une place est faite à la réflexion éthique dans l'imagination d'une expérience originale d'agent conversationnel, une place est également faite, dans le travail de l'éthique, à l'imagination de pistes de solution pour permettre cette expérience originale. Comme sa visée est

¹⁹¹ Voir dans le chapitre 3 (« Quelles fins respecter ? », p. 98) le travail de limites à l'œuvre avec le consentement, qui oblige au respect.

d'améliorer des situations – où la vie privée d'enfants hospitalisés est respectée, où ces enfants ont accès à des expériences divertissantes – ce qui compte dans le travail du *matter of caring*, ce n'est pas tant la perturbation qu'il entraîne. Ce qui compte, ce sont plutôt les adhérences qu'il amène à créer entre des préoccupations qui se gênent.

Quelques jours après la présentation faite ce 15 avril 2019 aux clients intéressés par l'idée de cet agent conversationnel, d'autres rencontres ont eu lieu pour présenter une démo et entamer une phase de négociations. Le projet n'aboutit pas, mais des adhérences dans l'équipe se font : les uns et les autres sont plus au fait des préoccupations de chacune selon les rôles et expertises, les façons de parler s'ajustent pour les prendre en compte, les contourner ou parfois les confronter.

Après mon immersion chez Mr Y et Prélude, en février 2020, Félix communique publiquement sur LinkedIn le rapport de *privacy-by-design* que je prépare avec eux de septembre à décembre 2019. Ce rapport présente plusieurs recommandations de gestion des données personnelles dans le contexte du développement d'agents conversationnels (voir annexe, p. 365). Tandis que dans l'une de nos premières interactions, il me répondait que la gestion des données concernait Google et non Prélude, cette publicisation manifeste le changement de son intérêt et de ses connaissances envers ce sujet qui prend plus d'importance pour lui au fil de l'année. De mon côté, dans la poursuite de pistes de solutions à proposer pour adresser des questions éthiques pendant cette même année, ma curiosité envers le design de conversation marque un changement de mes réflexes éthiques, moins accrochés à l'inquiétude de tout développement technologique, davantage motivés par les arrangements à trouver pour prendre en compte les nuances que les questions éthiques soulèvent.

Des adhérences se créent ainsi au fil de l'année entre les préoccupations qui comptent pour le projet, qu'elles relèvent de l'expérience utilisateur, du respect de la vie privée, du modèle d'affaire, de l'infrastructure technologique. C'est par ces adhérences que chaque *matter of caring* trouve les moyens de se concrétiser dans le design du projet pour contribuer à améliorer des situations. L'adhérence de ces préoccupations entre elles, qui ont tendance à se gêner, manifeste l'interdépendance qui les lie pour

concrétiser le projet d'une bonne expérience (avec un agent conversationnel). Ces adhérences résultent ainsi d'un exercice vulnérable où ces préoccupations et *matters of caring* se confrontent, s'empêchent et se modifient – pour que l'expérience qu'ils conçoivent ensemble soit bonne en chacun de leur nom (au nom de l'expérience utilisateur, de l'éthique, du modèle d'affaire, de l'infrastructure technologique).

Si des adhérences se font entre les membres de l'équipe, les préoccupations de Prélude s'éloignent cependant du domaine du *care* à la fin du mois de mai : après avoir pitché Vocaltrip (une idée de projet d'expériences touristiques vocales, guidées par un assistant vocal), la startup est sélectionnée pour entrer au MLab – un espace et programme d'incubation de startups dans le secteur du tourisme, de la culture et du divertissement. Prélude déménage au MLab début mai 2019. Les efforts redoublent pour faire fructifier les idées d'agents conversationnels touristiques et culturels, Mr Y est de moins en moins évoqué. Si le contexte touristique prend le dessus sur celui du soin, peut-il toujours y avoir des *matters of caring* qui font l'objet d'attentions particulières ? Et cela reste-t-il pertinent de s'occuper si attentivement du consentement quand le contexte n'appelle pas à améliorer des situations auprès de personnes vulnérables ?

Ces questions se manifestent par une certaine frustration de ma part :

Je ne sais pas encore comment le glissement du domaine de la santé mentale au tourisme est perçu par les autres; de mon côté, ce glissement a été perçu avec une certaine frustration de ne pas se dédier aux enjeux soulevés par un produit d'IA dans la santé mentale, tout en comprenant en même temps les aléas de la vie de startup qui cherche des voies par lesquelles décoller, quitte à s'insérer dans d'autres domaines qui bougent plus rapidement.

(Journal de bord, résumé des notes de terrain des mois d'avril et mai, 31 mai 2019)

Ma frustration manifeste un attachement au domaine de la santé mentale qui se prête bien au soin accordé au consentement dans des interactions vulnérables. Plus largement, ce domaine est un bon terrain de réflexion sur les enjeux éthiques que soulève le développement d'un agent conversationnel dans un tel contexte. Je continue tout de même à travailler sur le consentement pour Mr Y en arrière-plan, « j'ai l'impression de me sentir plus ancrée quand je travaille sur Mr Y plutôt que Vocaltrip » (Journal de bord,

notes réflexives, 17 mai 2019). Jusqu'à ce que Paul et Félix m'invitent à participer à un pitch.

Un pitch, où le consentement est là, mais pour quoi ?

Le projet s'appelle « Bonjour Montréal », un agent conversationnel qui guide les touristes pour découvrir la ville comme des personnes locales – c'est une adaptation de l'idée de Vocaltrip, centrée sur Montréal. La perspective de *care* n'est clairement pas présente : il ne s'agit pas d'améliorer une situation auprès de personnes vulnérables – quand je mentionne les enjeux d'accessibilité aux personnes à mobilité réduite, malentendantes et mal-voyantes, ces situations sont écartées, car elles ajoutent de la complexité; pour l'instant Prélude se concentre sur des cas plus majoritaires¹⁹². Il s'agit plutôt de proposer une expérience originale et divertissante avec un agent conversationnel pour prendre plaisir à visiter une ville de façon autonome, tout en ayant l'impression de marcher dans les pas de personnes qui y vivent. L'idée est présentée à plusieurs membres de TMTL, une organisation de promotion du tourisme à Montréal.

Pour ce pitch, Paul et Félix me demandent de proposer des « *key performance indicators* » (indicateurs de performance clés) afin de montrer comment Prélude évaluerait la réussite de l'expérience auprès des utilisateurs. Ces indicateurs de performance visent à montrer que l'outil proposé par Prélude permettrait de :

développer la connaissance client, d'améliorer l'accueil des touristes, d'enrichir l'expérience visiteur avec du contenu personnalisé et contextualisé, de favoriser l'accès à l'information et la facilité transactionnelle, d'améliorer les retombées économiques et surtout d'augmenter les revenus autonomes de TMTL. (Extrait de l'invitation du calendrier Google envoyée pour cet événement prévu le 25 juin 2019, notes de terrain du 6 juin 2019).

L'invitation me laisse un peu perplexe : cette demande est a priori peu liée à mon rôle de conseil en éthique pour les projets de Prélude. Je pourrais décliner l'invitation, mais c'est l'occasion de participer à un pitch, et d'éventuellement orienter cette demande vers des préoccupations éthiques pour favoriser davantage d'adhérences avec ces réflexions,

¹⁹² Pourtant, partir du cas de personnes en situation de handicap est une approche de design inclusif qui aurait pu être choisie si la perspective de *care* était encore présente (voir par exemple Clarkson et al., 2003).

malgré le contexte qui s'en éloigne. Je me demande : « Comment mêler à ça des questions d'ordre éthique ? Comment ces questions d'ordre éthique peuvent être traduites en indicateurs de performance ? » (Journal de bord, notes de terrain du 6 juin 2019).

En guise de compromis, je propose quatre variables d'expérience utilisateur à observer lorsque l'agent conversationnel sera utilisé : utilisabilité, accessibilité, confiance, désirabilité. Ces variables appliquent des principes éthiques au contexte de l'agent conversationnel touristique en question – elles traduisent des enjeux éthiques en piste de design d'expérience utilisateur (ou expérience visiteur, selon les termes de TMTL). Je me prête ici à l'exercice qui cherche à transformer des questions (éthiques) en pistes de solution, pour faire des propositions formelles de design cette fois-ci – un exercice de mise en pratique de l'éthique par le design (voir chapitre 1, p. 26). Par sa tentative de rendre présente la réflexion éthique même si le contexte ne la demande pas nécessairement, cet exercice relève-t-il du *matter of caring* ? Pour cela, il faudrait que les préoccupations manifestées permettent de prendre en compte des choses négligées ou mises de côté pour améliorer une situation marquée par une certaine vulnérabilité.

La diapositive suivante (figure 11) représente ces variables. Elle a été intégrée à la présentation préparée pour deux pitches auprès d'une des directrices de TMTL (le 25 juin 2019) puis de son équipe (le 2 juillet 2019). Cette diapositive fait l'objet de plusieurs discussions avec Paul et Félix avant d'arriver à ces formulations finales. Ces discussions peuvent se comprendre comme des processus de création d'adhérences entre plusieurs préoccupations : des préoccupations d'ordre marketing telles qu'elles sont manifestées ailleurs par TMTL, d'autres d'ordre éthique, ces deux types de préoccupations se faisant approprier par Félix et Paul pour être alignées avec ce qu'ils projettent de l'agent conversationnel et ses capacités. La description de la « désirabilité » est par exemple reformulée en termes marketing par la méthode du calcul du « NPS » (*Net Promoter Score*) dont TMTL a parlé à la cohorte de l'incubateur lors de la présentation de leur organisation.

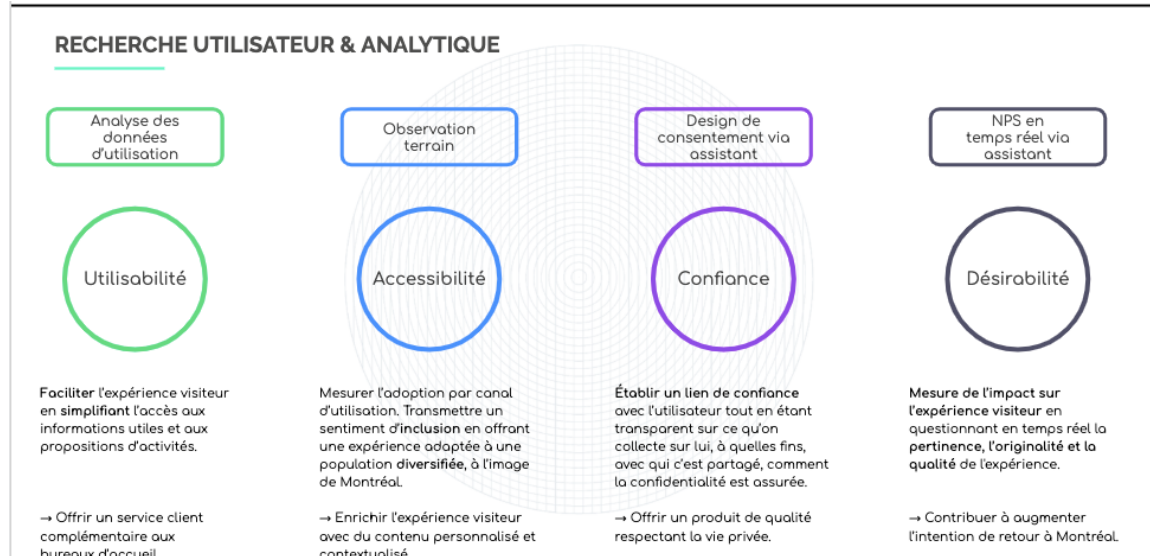


Figure 11 : Capture d'écran de la diapositive présentant les variables d'expérience utilisateur dans la présentation du projet Bonjour Montréal, 2 juillet 2019.

Dans cette diapositive, les mots dans les cercles sont les « variables ». Chacune est précisée en-dessous par un petit paragraphe introduit par un ou plusieurs verbes d'action pour expliciter ce qui serait évalué selon chacune de ces variables. Par exemple, selon la variable « utilisabilité », on évaluerait à quel point l'expérience visiteur est facile, à quel point l'accès aux informations est effectivement plus simple, etc. Deux de ces variables marquent de potentiels *matters of caring* : le souci d'inclusion compris dans celle d'accessibilité et l'attention portée à la transparence sur les données collectées, qui décrit celle de confiance. Mais à ce stade, où il n'y a pas encore d'intervention pour les aborder concrètement, ces préoccupations n'ont pas encore la consistance du *matter of caring* – elles pourraient aussi bien nourrir de l'*ethics washing* qui s'en tiendrait seulement à des mots donnant une chance de remporter un *pitch*.

Les deux autres variables – utilisabilité et désirabilité – ne sont pas particulièrement de l'ordre du *matter of caring* car elles ne visent pas à améliorer une situation marquée par une certaine vulnérabilité. En effet, la variable d'utilisabilité pourrait relever du *matter of caring*, si le souci de rendre facilement utilisable l'agent conversationnel n'était pas orienté exclusivement sur les personnes habituées à ce type de technologies et dépourvues de tout handicap qui configurerait autrement leurs capacités d'interaction avec cet outil – comme mentionné ci-dessus, cette approche de

design inclusif n'est pas envisagée par Prélude. La désirabilité n'est pas non plus de l'ordre du *matter of caring* puisqu'elle implique la mesure de la satisfaction, orientée vers les retours sur investissements de l'organisation qui propose le service; elle marque un *matter of concern* – susciter du plaisir, l'envie de revenir, et un retour sur investissement pour TMTL – mais cette préoccupation ne relève pas du *care* puisqu'elle ne touche pas à l'amélioration d'une situation vulnérable.

Dans cette distinction de ce qui relève du *matter of caring* ou non dans ces variables, c'est là une nuance importante que Tronto fait à propos du lien entre plaisir et *care* : le *care* peut susciter du plaisir mais les activités qui cherchent le plaisir ne relèvent pas du *care*. La fin doit être le soin, c'est-à-dire une « pratique [qui] a pour but le maintien, la perpétuation ou la réparation du monde » (Tronto, 2009, citée également dans le chapitre 4). Pour le *matter of caring*, la fin doit avoir cette dimension d'amélioration, de maintien ou de réparation d'une situation présentant une certaine vulnérabilité. En orientant ses activités vers le tourisme, mettant de côté une approche de design attentive aux situations de vulnérabilité – comme celles qui présentent des limitations dans certaines capacités –, les préoccupations de Prélude s'orientent de plus en plus vers un plaisir suscité par le divertissement de l'expérience proposée. Elles s'orientent de moins en moins vers le soin de situations vulnérables comme l'étaient la situation d'enfants hospitalisés ou encore celle de personnes cherchant de l'aide en santé mentale.

Si l'on revient à l'exposition de ces variables dans la diapositive (figure 11), le rectangle au-dessus de chaque cercle indique un type d'activité ou de méthode par laquelle l'évaluation de chaque variable serait faite. L'objectif général de chaque variable est introduit par une flèche en-dessous du paragraphe, pour indiquer une valeur ajoutée à ce qu'offre déjà TMTL – un service d'information dans des bureaux d'accueil touristique, une application mobile et un site internet offrant du contenu personnalisé pour donner des idées d'activités aux visiteurs – et pour s'aligner avec l'un de ses objectifs dans la promotion de Montréal : augmenter l'intention de revenir une fois les visiteurs partis. Au-dessus de la variable de confiance, l'activité « design de consentement via assistant », dans le rectangle violet, est ajoutée par Paul lors de notre discussion de préparation. C'est

une référence au projet qui me lie à Mr Y – s’il est évoqué ici, c’est qu’il pourrait s’appliquer au contexte de cet autre agent conversationnel touristique, pas tant au nom de la sensibilité du domaine dans lequel il y aurait traitement de données personnelles, mais au nom de la nécessité du consentement pour tout traitement de données personnelles.

Pendant le *pitch* lui-même, après ma présentation de la diapositive, Félix précise que mon projet de recherche avec eux porte spécifiquement sur cette question de « design de consentement via assistant ». Je suis surprise, car c’est la première fois que je suis amenée à en parler explicitement avec Félix – jusqu’à présent, ce projet était consacré à Mr Y exclusivement, occupant une place d’arrière-plan. En développant brièvement, je mentionne l’adoption d’une démarche de *privacy-by-design*. Une personne de TMTL rebondit sur ce point en soulignant l’importance d’arrimer tôt les écosystèmes technologiques de TMTL et ceux de Prélude pour permettre cette démarche – TMTL est en réflexion là-dessus, pour entrer en conformité avec le RGPD qui s’applique à leurs utilisateurs européens¹⁹³. Lié au *privacy-by-design*, le projet sur le consentement trouve ainsi une résonance avec une préoccupation de conformité qui est également présente dans l’organisation potentiellement cliente de Prélude. Même si le contexte n’est donc pas aussi sensible que la santé mentale, le consentement peut compter dans cet autre projet, en particulier s’il est lié à une démarche de *privacy-by-design* qui demande de s’arrimer avec les partenaires ou clients.

Cette mention explicite du projet de « design de consentement via assistant » peut ainsi contribuer à faire que cet enjeu se mette à compter dans les activités de Prélude, ce qui contribuerait à le faire compter auprès d’autres membres de l’organisation. Elle pourrait même le faire devenir un *matter of caring*, si des actions portant explicitement sur le consentement sont entreprises dans les pratiques de design de l’agent conversationnel, pour faire alors bouger les lignes du court-circuit habituel de l’étape du

¹⁹³ C’est dans le cadre de cette démarche qu’aura lieu la rencontre avec DPC, le cabinet de conseil en protection des données qui audite TMTL et ses partenaires en septembre 2019.

consentement quand il s'agit d'interagir avec un agent conversationnel¹⁹⁴. Pour l'instant, la mention de ce projet le positionne plutôt comme un intéressement, qui rassemble plusieurs intérêts : montrer que l'agent conversationnel de Prélude peut ajouter de la valeur aux produits actuels de TMTL, favoriser la confiance – envers les futurs utilisateurs mais également envers TMTL.

Après quelques négociations entre Félix et la directrice de TMTL la semaine suivant ces *pitchs*, TMTL accepte de développer ce projet qui prend le nom de la « Voix de Montréal » et devient le projet principal de Prélude. Mr Y s'en trouve de plus en plus éclipsé. Mais à défaut de se développer dans Mr Y, le consentement se fait finalement une place comme « module » au détour d'une fin de réunion, le 10 juillet 2019. Cela marque son entrée explicite dans le design de conversation de l'agent conversationnel, où il devient davantage l'objet d'interventions dans le design qui en font un *matter of caring*. Il s'engage alors dans un exercice plus vulnérable que celui de la préparation du *pitch* où l'arrimage des intérêts primait : cette fois, en essayant de faire fusionner les préoccupations – d'expérience utilisateur et de consentement principalement –, celles-ci se confrontent, s'empêchent et se modifient mutuellement.

Faire fusionner le consentement dans l'expérience principale

Si Prélude fait une place au travail du consentement dans le design de son agent conversationnel, ce travail se fait au fil de plusieurs tensions. Le consentement est d'abord abordé sous forme de module, dans un espace réservé, pour gêner le moins possible l'expérience principale. Puis les formulations par lesquelles le proposer amènent à interroger la validité du consentement qui serait donné dans ces conditions. Je vais présenter ces tensions inhérentes à la fusion du consentement dans le design de l'agent

¹⁹⁴ Cette fois, il peut s'agir de *matter of caring*, car l'objectif est d'améliorer une situation qui, par son habitude de négliger leur consentement, expose les utilisateurs au traitement potentiellement excessif ou non-respectueux de leurs données personnelles – ce qui peut avoir, entre autres, des effets néfastes dans leur vie, tels que la divulgation d'informations confidentielles, la surveillance, la manipulation d'intérêts, d'opinions et émotions (voir chapitre 3).

conversationnel, en montrant également comment elles se poursuivent par un travail qui cherche à clarifier quelles seront les pratiques de traitement des données personnelles.

Un module

- 23 Paul : si tu veux plus, tu t'as un *learn more* et là ça part (.) t'sais genre sur un autre
24 module
- 25 Philippe : ouais ça, puis à la limite ça pourrait, ouais c'est un autre module
26 exactement. »
- 27 [...]
- 28 Paul : (...) je pense que d'avoir le *know more* (.) t'sais qu'un autre module euh eh
29 ben ça-ça va permettre d'avoir toutes les questions et que ce soit clair pour
30 l'utilisateur (.) tu pourras même-on pourra même lui dire en gros « voici les
31 questions qui me sont fréquemment posées », « voici les questions
32 auxquelles légalement je dois te répondre »
- 33 Félix : Ouais autrement on va revenir avec un bloc de texte absolument euh
34 dégueulasse que personne va lire puis que au final, on-on-on fait pas du tout
35 changer la donne. Là on s'insère dans la conversation avant d'te donner la
36 fonction principale de notre produit, on dit r'garde (.) « on utilise-faut qu'tu
37 nous donnes ton consentement, veux-tu savoir qu'est-ce qu'on fait avec ton
38 data ? »

(2^{ème} extrait de retranscription de la réunion du 10 juillet 2019)

Comme je l'ai évoqué plus tôt (voir chapitre 4, p. 190), cette réunion du 10 juillet marque la première fois que le travail sur le design de consentement est abordé par Félix et Philippe plutôt que Paul ou moi qui travaillons dessus dans le cadre de Mr Y. C'est dans les mots de Paul qu'apparaît la conception du consentement comme un module. Ce module se constitue de l'ensemble des informations qui peuvent éclairer l'utilisatrice sur ce qui est fait de ses données personnelles, pourquoi ces données sont collectées et comment elle peut les consulter voire les supprimer. Ce module est là pour donner la possibilité à l'utilisatrice d'en savoir plus en disant « *know more* » (l. 28) ou « *learn more* » (l. 23), ce qui permettrait d'ouvrir la section de conversation où le bot serait capable de répondre à des questions portant sur la gestion des données personnelles, leur protection et le contrôle que l'utilisatrice peut avoir dessus. En somme, ce module est une

proposition de réinvention des modèles de consentement, pour l’informer par interaction avec l’agent conversationnel, selon le même type d’interaction qui a lieu pour le reste de la conversation. Ce modèle permet d’informer de façon vulgarisée plutôt qu’en langage juridique selon l’habitude des modèles d’information du consentement de type « *notice and consent* » (voir chapitre 3, p. 138).

Du mois d’avril au mois de juillet 2019 – de mon entrée chez Mr Y au démarrage du projet de la Voix de Montréal chez Prélude – je travaille le consentement pour Mr Y, en arrière-plan des activités principales de ces organisations. C’est là que ce module prend forme, inspiré de ce que font les chatbots en santé mentale. À partir de juillet et du pitch qui l’a annoncé, Paul et Félix envisagent de transposer cette section de conversation dans le nouveau projet d’agent conversationnel de Prélude.

En parlant de module, Paul utilise le vocabulaire de l’approche agile de développement technologique (voir chapitre 2, p. 79). Cette notion de module désigne les sous-unités qui, travaillées de façon indépendante, constituent l’ensemble d’un système une fois qu’elles sont rassemblées :

The term modularity is used to describe the degree to which a given (complex) system can be broken apart into subunits (modules), which can be coupled in various ways (Baldwin 2015). As a design or architectural principle, modularity refers to the “building of a complex product or process from smaller subsystems that can be designed independently yet function together as a whole” (Baldwin and Clark 1997). (...) Modular design involves the mantra that the independence of system components is optimized after which they are ‘loosely coupled’ (Van Schewick 2010). (Gürses & van Hoboken, 2018, p. 586)

Cette modularité par laquelle le consentement est pensé comme section de conversation permet de le faire passer d’une organisation (Mr Y) à l’autre (Prélude), d’un domaine (santé mentale) à l’autre (tourisme), d’un chatbot (Mr Y) à un autre agent conversationnel (la Voix de Montréal). Ces transpositions aident à s’occuper de ce moment de conversation comme un module qui compte au-delà d’un seul projet (Mr Y) dans un domaine spécifique (santé mentale). Elles justifient que l’on s’attarde dessus, non plus parce que le domaine le demande, mais parce que c’est un des modules qui constitueront le produit final – et potentiellement d’autres à venir selon Paul :

39 Paul : donc ouais tu vois, tout ce module-là de (...) *data privacy consent*, ça pourra
40 s'exporter sur n'importe quel projet qu'on aura¹⁹⁵.

(3^{ème} extrait de la réunion du 10 juillet)

La justification s'inscrit dans la logique agile qui organise le développement technologique en modules dont l'indépendance permet une réutilisation dans d'autres domaines que celui prévu initialement. La modularité permet en somme de multiples arrangements à venir.

Par ailleurs, l'indépendance du travail modulaire permet à la fois de continuer à porter sur l'enjeu du consentement une attention de l'ordre du *matter of caring*, tout en le poussant – par le mouvement de transposition – à sortir du domaine auquel il était rattaché (santé mentale) pour pouvoir se constituer comme une préoccupation qui compte dans l'organisation Prélude, au fil de son glissement vers le domaine touristique. Dans un premier temps (avril-juillet 2019), quand ce travail est exclusivement lié à Mr Y, comme nous l'avons vu dans la section précédente, le travail se déroule dans des documents dédiés, régulièrement discutés et modifiés avec Paul, ainsi qu'au fil d'interactions avec des bots du même domaine (Woebot et Wyza en particulier). Ce module se conçoit ainsi dans un espace à part¹⁹⁶, à l'écart des autres activités et documents de l'organisation, à l'écart aussi des activités de Prélude.

Cet espace à part permet de dédier une attention qui traite cette section de conversation autrement qu'au nom d'un impératif des « questions auxquelles légalement je dois te répondre » (Paul, l. 31-32 du 2^{ème} extrait de la réunion du 10 juillet cité ci-dessus), ou d'une formalité obligatoire pour que l'on puisse passer à l'étape suivante de l'expérience utilisateur (cf « faut qu'tu nous donnes ton consentement » Félix, l. 36-37 du 2^{ème} extrait de la réunion du 10 juillet citée ci-dessus). Cet espace indépendant permet plutôt de ralentir (voir Stengers, 2018, citée dans le chapitre 3, p. 208) pour prendre le

¹⁹⁵ Cette idée de transposition du module de consentement était déjà présente dans la préparation du projet sur le design de consentement avant que j'entre chez Mr Y (et Prélude). Paul m'avait effectivement mentionné que si ce projet de consentement par interaction avec un chatbot aboutissait à quelque chose de satisfaisant, alors il pourrait être « répli[qué] dans le monde des agents conversationnels, y compris vocaux » (courriel de Paul, 13 décembre 2018).

¹⁹⁶ Avec Allô, ma première expérience de travail sur le consentement préfigurait cette tendance à être mis dans un espace à part, en retrait des activités principales (voir vignette Allô plus haut).

temps d'aborder cette section de conversation comme un *matter of caring* qui porte attention aux conditions par lesquelles on formule cette demande de consentement. Cette attention est animée par des questions se demandant quelle quantité et quel niveau d'information donner pour que le consentement soit suffisamment informé sans pour autant le submerger ni l'inquiéter, comment poser les questions en laissant le choix de dire oui ou non, sans induire un oui dans la question, comment assurer la clarté de la réponse dans le design d'interaction pour que le consentement soit marqué de façon explicite, sans avoir été forcé de cliquer sur « ok », ni supposer un oui par défaut si l'utilisateur ne répond pas¹⁹⁷. L'exploration de ces questions demande un certain temps, les réponses et nuances ne sont pas rapides ni évidentes à trouver, elles font l'objet d'un projet (de recherche et de design) avec Mr Y. L'espace documentaire et organisationnel isolé dans lequel elles se développent permet de s'arrêter dessus; le traitement de ces questions qui font l'objet d'un *matter of caring* demande un certain ralentissement. Mais si ce *matter of caring* reste dans cet espace à part, peut-il se doter du pouvoir transformateur qu'il vise ?

Le 10 juillet 2019, en cette fin de réunion qui portait sur autre chose, on aborde le consentement depuis un autre espace que cet espace isolé dans lequel il est exploré jusqu'à présent. Les propositions de Philippe et Félix répondent rapidement aux questions implicites que je prends le temps d'explorer ailleurs. Face à ces propositions rapides, mes questions prennent la forme de préoccupations marquées de vulnérabilité, par des objections, des chevauchements, des balbutiements, des (im)précisions :

- 41 Félix : ouais. Moi j'dirais « *would you like to know more about this process?* » si oui,
42 ben là on explique comment on protège leur data pis etcetera etcetera « *or*
43 *continue by giving your consent* » pis là c'est genre « *I give my consent* ».[
44 Camille : [Ouais
45 mais c'est[
46 Félix : [« ok »[
47 Camille : [I n'est pas suffisamment informé là le consentement. (.)

¹⁹⁷ Ces questions sont égrenées dans mon journal de bord, au fil de mes notes de terrain d'avril à juillet 2019. Ces thèmes sont également présents pendant le réunion du 10 juillet 2019.

48 Félix : Qu'est-ce que tu veux dire?

49 Camille : Parce que-on-pour que le consentement soit suffisamment informé il faut
 50 leur dire aussi avec qui ((*la voix s'abaisse, bafouille un peu*)) on les – enfin si
 51 on les partage – avec qui euh comment[

52 Félix : [Ouais mais ça c'est, ouais mais on
 53 l'aura l'option, je veux pas tout leur dire ça là, on leur offre l'option d'en savoir
 54 plus, s'ils veulent en savoir plus là ils vont pouvoir aller en savoir plus, mais je
 55 vais pas faire un paragraphe de de 8 affaires pour leur dire comment leur data
 56 il est... ça, ça tue l'expérience.

57 Paul : C'est pour ça que j'avais dit tout à l'heure que (.) il pourrait y avoir juste un
 58 module (.) dans lequel il va y avoir *data privacy*. Dans lequel il va y avoir toutes
 59 les questions et toutes les réponses qui sont associées. Et même le *know more*
 60 de là[

(4^{ème} extrait de retranscription de la réunion du 10 juillet 2019)

Ma préoccupation de suffisamment informer le consentement interrompt Félix (« Ouais mais c'est » (l. 44-45), « Il n'est pas suffisamment informé là le consentement » (l. 45)). Mon objection cherche à se préciser ensuite mais elle est marquée d'hésitations (voix qui s'abaisse, bafouillement (l. 50)) – mon travail sur ces questions est encore en chantier dans l'espace de Mr Y qui lui est dédié. Félix manifeste en retour sa préoccupation, avec une certaine réticence envers ma remarque : il ne veut pas que le texte qui informe le consentement soit trop long (« je vais pas faire un paragraphe de de 8 affaires pour leur dire comment leur data il est... » (l. 55-56 dans le 4^{ème} extrait ci-dessus), ou encore « autrement on va revenir avec un bloc de texte absolument euh dégueulasse que personne va lire¹⁹⁸ » l. 33-34 du 2^{ème} extrait de la réunion du 10 juillet 2019 citée plus tôt). La crainte d'une information qui prendrait trop de place, c'est que « ça tue l'expérience » (l. 56 du 4^{ème} extrait ci-dessus). Un rapport de force s'amorce autour du heurt qui peut être fait à l'expérience : la place que prendrait l'information destinée au consentement pourrait « tuer » cette expérience, tandis que la rapidité avec laquelle le consentement serait abordé risquerait de si peu l'informer qu'il ne serait finalement pas valable. Ce qui est vulnérable, ce n'est donc plus tant le domaine et les potentiels utilisateurs mais la

¹⁹⁸ C'est une façon d'évoquer les modèles de *notice & consent*, très longs, peu compréhensibles quand on n'est pas juristes, ni agréables à lire (voir chapitre 3, p. 115).

relation entre consentement et UX, parce que l'un et l'autre présentent le risque de se blesser mutuellement. Dans cette vulnérabilité, la proposition du format de module vient apaiser cette tension : il permet de donner de la place au consentement, tout en le contenant pour qu'il n'envahisse pas l'expérience.

Bien que réticent à laisser trop de place au consentement dans la conversation, Félix manifeste son adhésion au projet de travailler sur ces interactions : il mentionne le risque de ne pas « changer la donne » (l. 35 du 2^{ème} extrait) si ce module n'est pas travaillé de façon plus allégée que les textes habituels. Peut-être s'approprie-t-il le consentement comme un *matter of caring* s'il est prêt à passer du temps sur ces interactions dans le but d'améliorer les façons de faire qui ont tendance à le négliger. Dans cette direction, il mentionne à un autre moment lors de cette réunion du 10 juillet 2019 : « je pense que notre but c'est un peu de démystifier qu'est-ce qu'on fait avec tout ça et rendre ça accessible ». Il récapitule également à la fin de la réunion :

61 Mais en ce moment, ce qu'on essaye de trouver c'est une alternative au bloc de
62 texte dégueulasse que tu veux pas lire alors tu fais ok, *right* ? c'est bien le brief que
63 j'ai compris ? ((il lève des yeux interrogateurs vers Paul et Camille qui acquiescent)).

(5^{ème} extrait de retranscription de la réunion du 10 juillet 2019)

Repris dans les mots de Félix, et suggéré sous le format d'un module, le projet sur le consentement s'intègre aux préoccupations de design chez Prélude, ce qui le met en bonne voie pour qu'il soit travaillé comme un *matter of caring* – qu'il reçoive une attention qui permette d'améliorer les situations d'interaction qui tendent à le négliger, et s'inscrive dans un horizon qui favorise le respect des (données des) utilisateurs et leur meilleure compréhension de ce qu'implique d'interagir avec ce type de système d'IA.

À mesure que le module se développe et se fait une place pour fusionner dans la conversation générale, cette attention portée sur le consentement se teinte cependant de différentes préoccupations qui se mélangent et font dévier le consentement d'un souci de *care*. Il devient plutôt une contrainte à relever sans trop heurter l'expérience utilisateur générale. Pour ne pas heurter l'expérience utilisateur, il s'agit de faire fondre les interactions de consentement dans les autres, en adoptant le même style et le même ton – « *fun* ». Pour l'instant, en se faisant aborder dans le contexte de l'agent

conversationnel de Prélude qui la fait sortir de l'espace à part de Mr Y, la préoccupation du consentement a le potentiel de se doter d'une force transformatrice propre à sa visée de l'ordre du *caring* qui cherche à faire bouger les lignes des situations habituelles où les interactions de consentement sont négligées. Cette force transformatrice va-t-elle véritablement se développer ?

Trouver un ton *fun*

À partir de la fin août 2019, le projet sur le consentement est transposé officiellement dans les activités de Prélude : le renouvellement de l'entente Mitacs à propos de ce projet est signé au nom de Prélude plutôt que Mr Y, et il apparaît explicitement dans les activités de Prélude, sous forme de tâches dans les sprints de l'organisation quand elle se met à adopter la méthode Scrum (voir chapitre 2, p. 81). Avec cette méthode d'organisation du travail, chaque module de l'agent conversationnel est développé au fil de sprints de deux semaines en sous-équipes : équipe technique d'un côté, équipe UX de l'autre – le module de consentement fait partie du travail de UX. Lors de ces sprints, le module de consentement évolue : d'un format de conversation, il prend un format de carrousel puis finit sous format de liste au mois de novembre.

Le 23 septembre 2019, je reçois une notification de Jira – l'outil de gestion de projet qui permet d'organiser et suivre les sprints de développement (voir chapitre 2, p. 81). C'est Félix qui m'interpelle :

@Camille, j'ai revu le *consent design* et je pense que *overall* c'est bon à 90% Le principal bug est dans la formulation de l'intro. J'ai l'impression que les gens ne sauront pas quoi dire. Il faut être plus directif et moins *open ended*. Merci de nous faire quelques suggestions en ce sens. (Notification recopiée dans mon journal de bord, notes de terrain du 23 septembre 2019)

Ce qu'il appelle « *consent design* » est le module de consentement que j'avais préparé avec Paul pour Mr Y, qu'il a copié-collé dans le document général de design de conversation de la Voix de Montréal, intitulé « *conversational design* » (voir chapitre 2, p. 89), à la suite de la réunion du 10 juillet 2019. Le *prompt* introduisant la demande de consentement était celui-ci :

Excellent! In order to recommend you the best I have to offer, I have a couple of questions for you.

But before we go on with personal questions, please know that I need to collect data to help me be more personalized to you. Don't worry, everything you tell me is going to stay between us. Are you ok with that or would you like to know more about how we ensure your data privacy?

(Journal de bord, notes de terrain du 23 septembre 2019)

Je propose une alternative qui tente d'être plus directe, avec une question moins ouverte :

Excellent! In order to recommend you the best I have to offer, I have a couple of questions for you.

Psst, before we go on, my creators would like you to know that I need to have access to some information about you, or we couldn't chat further.

Have a look at your screen, you can see what I do with this data, who I share it with and what you can do about it.

CAROUSSEL			
<i>what data we collect</i>	<i>who we share it with</i>	<i>why we collect it</i>	<i>how you can delete it</i>

Do you agree with that?

(Journal de bord, notes de terrain du 23 septembre 2019)

Cette alternative introduit le format du carrousel inspiré de Wysa. Elle reprend également la transparence dont Woebot faisait preuve en évoquant ses créateurs pour éviter l'illusion d'être un humain qui parlerait au travers du « I ». Cette formulation n'est cependant pas du goût de Félix lorsqu'on en discute de vive voix un peu plus tard dans la journée :

- 1 Félix : Non on dirait que là ça m'amène vers quelque chose de... caché, là, genre?
- 2 'With my creators' là non ((il relit en murmurant))

(1^{er} extrait de retranscription de la section « Formuler le consentement »¹⁹⁹, réunion du 23 septembre 2019).

La nuance qui tente de clarifier que le bot n'est pas un humain, qu'il y a une équipe d'humains derrière lui, tombe ainsi à l'eau.

¹⁹⁹ Cette réunion du 23 septembre 2019 s'est articulée en deux temps : nous avons d'abord discuté de la question embarrassante « *Is Google Assistant GDPR compliant?* » (voir partie suivante), puis cherché comment « formuler le consentement ». Ces deux thèmes servent de titres pour distinguer les extraits liés à cette réunion.

Félix revient à la première proposition, mais le format de carrousel retient son attention, à condition qu'il ne s'affiche pas d'emblée :

3 Félix : Ouais mais un carrousel, à partir du moment où c'est *learn more*, mais j'ai
4 l'impression que je mettrais pas le carrousel d'office là, hey tu rentres dans
5 une expérience, tu t'attends à parler avec Montréal, puis là t'es rendu dans le
6 *privacy* qui est la première chose que tu vois, ça marche pas t'sais. C'est pas
7 que je veux le cacher, c'est juste que ça prend trop de place t'sais à ce
8 moment-là.

9 [...]

10 Félix : Mais attends, par contre ton idée de carrousel elle peut très bien embarquer
11 à partir du moment où c'est *learn more*, moi j'ai aucun problème à ce qu'on
12 fasse un carrousel, au contraire j'pense même que ce serait mieux, mais (*il y*
13 *a du bruit dans la salle*) ça c'est plus simple... Moi j'enlèverais « *don't worry,*
14 *everything you say will stay between us* » j'enlèverai ça

(2^{ème} extrait de retranscription de la section « Formuler le consentement »,
réunion du 23 septembre 2019)

En positionnant le carrousel comme une option quand on demanderait à en savoir plus, Félix se défend de vouloir « cacher » (l. 7) cette section. La préoccupation qui l'emporte pour lui est à nouveau que cela ne prenne pas trop de place, en particulier dans ce moment de *onboarding* où le module de consentement peut faire diversion alors que l'objectif est d'offrir l'expérience principale – « parler avec Montréal » (l. 5) – le plus vite possible. Il enlève également la phrase mentionnant une possible inquiétude (« *don't worry, everything you say will stay between us* », l. 13-14) qui reprenait le ton rassurant auquel tenait Mr Y.

Ces trois attachements – ne pas donner l'impression que derrière le bot se cachent d'autres personnes, ne pas vouloir cacher la section du consentement mais en restreindre la visibilité à une option pour en savoir plus, et effacer le ton rassurant à propos d'une possible inquiétude – marquent ce qui favorise, aux yeux de Félix, une meilleure expérience que celle proposée par ces formules. En même temps, ces trois attachements portent des modifications qui évacuent des dimensions vulnérables constitutives du bot, et de l'expérience. Parmi ces dimensions vulnérables, il y a, d'une part, le fait que l'agent conversationnel est loin d'être autonome malgré l'impression qu'il peut donner – il ne fonctionne qu'à partir de ce que font les personnes qui le développent, son « *I* » n'est

donc pas singulier ni la marque d'une personne précise²⁰⁰. D'autre part, clarifier auprès de l'utilisatrice quelles informations sont collectées sur elle, pourquoi, etc., cela explicite le besoin de ces informations pour que le bot fonctionne – en d'autres termes, ce que ces détails explicitent, c'est le besoin de l'autre que l'organisation a, pour développer et faire fonctionner son agent conversationnel. Ce besoin évoque la vulnérabilité constitutive que Martha Nussbaum décrit chez les humains qui ont toujours besoin de l'autre au quotidien (voir chapitre 3, p.79). Évacuer ces précisions ou en diminuer la visibilité peut ainsi se comprendre comme des modifications faites pour rejeter l'exposition de cette vulnérabilité « *and to suppress awareness of the attachments that entail it* » (Nussbaum, 2003, p. 147, citée dans le chapitre 3, p.79). L'effacement du ton qui cherchait à rassurer une possible inquiétude envers ce qui serait fait des données personnelles collectées va également dans le sens de cette poursuite d'invulnérabilité dans les formulations choisies pour faire parler l'agent conversationnel. En entretenant la conscience de cette vulnérabilité, le *matter of caring* qui travaille initialement ces formulations de consentement se confronte ainsi à des préoccupations qui préfèrent masquer cette vulnérabilité, par des formulations qui l'exposent moins, voire des formulations marquées d'invulnérabilité.

Si l'exposition de la vulnérabilité est évacuée, c'est aussi qu'à la différence du domaine de la santé mentale dans lequel ces formulations étaient pensées initialement, à présent, dans le domaine touristique, ce n'est plus tant cette vulnérabilité qui compte que le divertissement. À propos du format du carrousel pour aborder le module de consentement, Félix dit un peu plus tard dans la réunion :

- 15 Félix : j pense que c'est plus efficace, plus agréable, ça convient mieux à une
16 solution qui est pas juste textuelle ou audio fait que faut juste qu'on trouve
17 des visuels qui représentent bien ça, qui soit quand même élégant et qui *fit*
18 un peu dans notre ton, notre *voice* là t'sais.
- 19 Camille : comment tu le définirais notre ton de voix ?
- 20 Félix : mettons que j'aurais une personne à mettre dessus, c'est Ellen Degeneres

²⁰⁰ Comme mentionné dans le chapitre 2 (p. 81), préciser la dimension non-humaine des agents conversationnels est un enjeu éthique qui se soucie de ne pas tromper les personnes qui interagissent avec en leur laissant croire que c'est une personne humaine qui parle avec eux.

(3^{ème} extrait de retranscription de la section « Formuler le consentement »,
réunion du 23 septembre 2019)

Dans ces modifications de formulations, l'objectif est de faire fusionner ce module de consentement dans l'expérience générale, qu'il « *fit* dans notre ton, notre *voice* » (l. 17-18). C'est la première fois que nous parlons explicitement de ce ton, que Félix associe à celui de l'animatrice et humoriste Ellen Degeneres²⁰¹ (l. 20). Par cette figure qu'il projette dans la voix, c'est la tonalité divertissante qu'il privilégie.

Il précise alors sa frustration avec le module de consentement qui n'est pour l'instant pas harmonisé avec cette tonalité. Il cherche à rendre ces formulations plus amusantes :

- 21 Félix : j'trouve que là en ce moment on est encore très formel, j'pense qu'il faudrait
22 qu'on arrive à un niveau où y'a du *fun* là-dedans t'sais, pis j'me dis est-ce qu'il
23 y aurait même du *fun* à injecter là-dedans, t'sais, t'sais mettons « Excellent
24 nananana... » T'sais, mettons « *Before we go on with naughty questions* » là,
25 ça pourrait être t'sais, euh « *I'd like to... encrypted and confidential* » t'as euh
26 « *naughty* » ((Paul, qui travaillait sur son ordinateur, lève la tête et regarde
27 Félix)). Faut juste faire attention que ce soit pas une connotation trop sexuelle,
28 *naughty* c'est-tu sexuel ? ((il tape le mot dans Google sur son ordinateur)) Ah
29 non tu vois « *disobedient, badly behaved children* », « *typically because related*
30 *to sex...* » ((il rit)) C'est le deuxième !
- 31 Paul : Ouais dans la culture populaire je crois que *naughty* c'est trop synonyme
- 32 Félix : Ouais ok c'est peut-être trop mais... « *dirty questions* » ? C'est parce que
33 l'affaire c'est que si on veut être drôle, il faut qu'on soit un peu *edgy*, sinon
34 c'est pas drôle (.) Sinon on pourrait dire... « *indecent* » c'est drôle ça « *before*
35 *we go on with indecent questions* »... « *rude or raci-...* non » (..) « *spicy* » ?
- 36 Camille : Ah *spicy* ça peut être drôle ça
- 37 Félix : Ouais. « *But before we go on with spicy questions* » ouais c'est drôle *spicy* ((il
38 écrit sur son ordi)) « *note...* » Toi faut pas que tu fasses de changement hein
39 sinon (.) ((Camille acquiesce))²⁰² « *Note that our conversations will stay private*

²⁰¹ C'est surprenant que Félix parle d'Ellen Degeneres alors que la persona à laquelle il tenait jusqu'à présent était celle d'une jeune fille à l'accent francophone qui représentait Montréal comme une gigantesque « *playground* » (voir vignette de la voix entendue pour la première fois, dans le chapitre 2 p. 80). La figure d'Ellen Degeneres n'est pas reprise plus tard, la persona de la jeune fille est conservée.

²⁰² Le document sur lequel Félix écrit les modifications est un document draw.io (devenu Diagrams). Draw.io ne permet pas à deux personnes en même temps de modifier le document. Si deux personnes apportent des modifications au même document, draw.io enregistre une copie. La multiplication des copies du document de design de conversation a posé plusieurs problèmes au début du travail là-dessus.

40 *encrypted and confidential* » (.) « *Should we continue now?* » Ouais (.) « *should*
41 *we continue playing now?* » Non des fois c'est juste, *spicy* ça suffit
42 Camille : Parce qu'on veut pas non plus trop intriguer pour pas qu'ils y passent
43 trop de temps
44 Félix : Non c'est ça (.) ok fait qu'à, carrousel (.) pis là (.) ouais c'est ça, là dans *data*
45 *policy* on peut pas vraiment être drôle

(4^{ème} extrait de retranscription de la section « Formuler le consentement »,
réunion du 23 septembre 2019)

Félix définit le ton drôle par l'*edgy* (l. 13), un ton un peu provocant pour avoir une attitude cool. Il joue alors avec les adjectifs *naughty* ou coquines (l. 24-30), *dirty* ou sales (l. 32), *indecent* ou indécentes (l. 34-35), puis *spicy* (piquantes) (l. 35-41) pour qualifier autrement que personnelles les questions que le bot s'apprête à poser. Tous ces adjectifs ont une connotation sexuelle – *spicy* peut-être un peu moins que les autres, ou du moins le terme peut connoter une plus grande diversité de choses piquantes. D'une certaine manière, on retrouve ici dans le travail de design de conversation, la présence du registre sexuel avec lequel jouent (parfois plus agressivement) les personnes qui interagissent avec des agents conversationnels (voir chapitre 2, p. 104). Par un ton légèrement provocateur, et surprenant pour ce type d'interactions qui sont habituellement plus formelles, Félix propose donc de jouer avec ces connotations sexuelles pour introduire les questions relatives au consentement – dont le contexte n'a rien de sexuel. Ce serait le décalage introduit par la connotation sexuelle qui rendrait plus amusante la dimension intime des questions personnelles que le consentement viendrait adresser. Il cherche ainsi à améliorer les formulations de consentement en jouant avec ces allusions, ce qui peut favoriser une expérience intrigante de ce moment plutôt que d'en perpétuer l'ennui ou la formalité habituelle.

J'appuie l'option de *spicy* (l. 36), tout en manifestant ensuite la préoccupation de ne pas « trop intriguer » (l. 42) pour éviter que cela fasse diversion de ce que les utilisateurs cherchent initialement à faire avec le bot – parler avec Montréal. En rappelant cette préoccupation, mon attention semble avoir fusionné avec les préoccupations qui se concentrent sur l'expérience générale plutôt qu'inciter les utilisateurs à passer un certain temps sur les interactions de consentement. Cette fusion de mes préoccupations dans

celles du ton de la voix s'accompagne d'une absence de préoccupations manifestées envers la persona. Pourtant, celle-ci pourrait faire l'objet d'un *matter of caring* qui chercherait à ne pas reproduire des stéréotypes où les voix féminines adoptent un ton séducteur. Or ici, en proposant des formulations à connotation sexuelle, le ton de la voix reprend le stéréotype d'une persona féminine qui serait mise dans une posture potentiellement ambiguë par les connotations qui pourraient induire un rapport de séduction. Le travail de reformulations plus amusantes cherche à déjouer l'ennui de la formalité en jouant avec des connotations décalées par rapport au contexte, mais ce jeu ne s'engage pas particulièrement dans une amélioration de l'ordre du *matter of caring*. Une telle amélioration serait en faveur de plus d'épanouissement, de respect et d'égalité entre les personnes – ce serait un jeu qui pourrait « érotiser l'égalité » (voir Manon Garcia citée dans le chapitre 3, p. 173) pour améliorer des pratiques de consentement plutôt que d'en perpétuer l'expérience d'inconfort, d'évitement et d'ennui habituels. Ce n'est pas particulièrement vers cet horizon que semble s'orienter le travail de reformulation; il s'oriente plutôt vers un horizon de divertissement où ces idéaux ne sont pas à l'avant-plan. Je pourrais intervenir dans cette direction, pour rappeler cette attention portée sur la voix si elle était un *matter of caring*. Mais je ne le fais pas, je m'inscris dans la direction de Félix, orientée vers le divertissement.

Précédemment, quand, avec toute l'équipe, nous avons entendu la voix de l'agent conversationnel pour la première fois (le 14 juin 2019, voir vignette où l'on entend la voix pour la première fois dans le chapitre 2, p. 100), cette voix a suscité une agréable surprise, mais pas d'attention de l'ordre du *matter of caring* non plus. Pourtant, c'était une persona de jeune fille, qui proposait son assistance pour visiter la ville, sur un ton enjoué potentiellement séducteur. Le ton n'était pas intentionnellement séducteur pour les membres de Prélude, mais quand Malik fait écouter cette voix à des personnes pour tester l'agent conversationnel, les réactions sont vives :

many participants, mostly men, sexualized the voice on the basis of its allegedly alluring tone. Indeed, it would often happen that people told me they thought "she sounded hot" upon the first moments of the interaction, when the beguiling voice introduces herself. (Delhi-Talbot, 2020, p. 56-57)

Par ailleurs, quand il la fait écouter dans son groupe de recherche universitaire, « *many other people were shocked by this fact and expressed their concerns over the appropriateness of the solution considering the apparent young age of the voice actor* » (p. 57). Je reçois le même genre de remarques quand je fais écouter la voix à mon groupe de recherche, au département de communication de l'Université de Montréal. Nous sommes embêtés lui et moi : pourquoi cela ne nous a pas autant sauté aux oreilles ? Lui (en anthropologie) comme moi avons de nombreux outils de sciences sociales et humaines qui aiguissent notre attention pour critiquer et corriger ce genre de stéréotypes. Pourquoi ne sommes-nous pas intervenus pour corriger cela ? Ces questions en touchent une autre : pourquoi la voix et sa tonalité ne font pas l'objet d'un *matter of caring* de notre part ?

L'effet d'agréable surprise suscitée à la première écoute de cette voix – pour laquelle nous avons travaillé jusque-là sur des documents textuels et diagrammes insonores – a été une émotion partagée qui a probablement marqué un des points de départ de notre complicité avec cet agent conversationnel. Cette agréable surprise – d'entendre une voix plus chaleureuse et vivante que nos documents de design de conversation – a occulté des ressentis plus critiques, moins enthousiastes. À propos de notre complicité avec des choses qui peuvent finalement nous gêner, Shotwell pose la question : « *what happens if we start from here?* » (Shotwell, 2016, p. 5, citée dans le chapitre 5, p. 213). Ici, le *if* (si) est un *when* (quand) : que se passe-t-il *quand* on part de là ? Que se passe-t-il quand on part d'une voix de jeune fille jouée qui suscite une émotion partagée d'agréable surprise par la chaleur humaine qu'elle dégage, alors que c'est un agent conversationnel que l'on travaille dans la froideur de documents et diagrammes ? Notre complicité avec cette voix participe à notre inclusion dans l'équipe, ma complicité participe également à favoriser la fusion du module de consentement dans l'expérience générale de la voix. Alors des compromis sont faits : nous ne remettons pas en question cette voix à la base de notre complicité. Nous laissons ainsi se reproduire des stéréotypes qui nous gênent mais que nous ne voyons pas sur le moment.

Au fil des évolutions du design de conversation, le ton de la voix change petit à petit cependant, pour mettre davantage l'accent sur les informations que sur son enjouement. Le module de consentement évolue également – les formulations avec *spicy* disparaissent finalement, elles ne sont pas au goût de TMTL – mais ces changements ne se font pas si simplement.

« Ah mais tu peux pas changer comme ça ! »

Le design de conversation est principalement travaillé dans le document général de conversation, intitulé « *conversational design* » (voir chapitre 2, p. 88). Il évolue au fil des sprints et des rencontres avec TMTL qui donne son avis et demande des modifications. Le module de consentement est plutôt travaillé dans un autre document, créé le 23 septembre 2019 par Félix et intitulé « *ethical consent flow Camille* » – c'est dans ce document que Félix a écrit les propositions de formulations avec « *spicy* ». Par la suite, c'est moi principalement qui travaille sur ce document. Quand je jette des coups d'œil de temps en temps au document principal de design de conversation, je remarque que les changements portés au module de consentement ne sont pas pris en compte. Par exemple, les reformulations que Félix voulait plus amusantes n'y sont pas, et depuis notre discussion, j'ai proposé de créer un raccourci pour aborder les questions liées aux données personnelles, mais celui-ci n'y est pas non plus. Alors je l'ajoute dans le document principal, pour le faire voir à Félix et en discuter prochainement. À la suite d'une réunion le 23 octobre 2019, nous abordons les interactions de consentement, et Félix constate que j'ai modifié une formulation dans le document :

46 Félix : mais là on leur offre, regarde “*Before we go on, note that all our*
47 *conversations will stay*” puis là faut qu'on dise “*private and confidential*”,
48 *((il ralentit)) “just say ‘privacy’ anytime if you want to know more...”*

49 Camille : ouais ça j'ai changé hier soir parce que faut qu'on[

50 Félix : [ah mais tu peux pas
51 changer comme ça ! Parce que moi c'est

52 Camille : [parce que faut qu'on
53 donne la possibilité

54 Félix : ouais mais tu peux pas changer de même ma maquette là, ça c'est la
55 maquette que je montre au client. Si tu veux me mettre un commentaire
56 tu peux le mettre mais tu peux pas changer directement dedans, parce que
57 moi je l'avais enlevé parce que c'était trop long en ce moment puis[

(5^{ème} extrait de retranscription de la section « Formuler le consentement »,
réunion du 23 septembre 2019)

J'avais ajouté la phrase « *just say 'privacy' anytime if you want to know more* » (l. 48) par souci de créer un raccourci qui permette d'accéder facilement et en tout temps à cette demande d'information²⁰³ – la création de raccourcis étant recommandée dans les principes d'expérience utilisateur (voir chapitre 2, p. 96). Mais ce changement agace Félix qui le découvre sur le vif – et qui m'apprend en même temps que ce document n'est pas tant un document de travail collaboratif qu'un document sur lequel il travaille (en consultant Paul) pour le partager de façon régulière à TMTL.

Un peu plus tard dans la discussion, j'explique cet ajout de raccourci, et il explique en même temps sa frustration (en plus du changement du document) : c'est lourd.

58 Camille : C'est pour ça que dans la première interaction sur le consentement j'ai
59 précisé « à tout moment si t'as une question là-dessus tu peux dire *privacy* »,
60 c'est dans l'idée de créer une commande facilement accessible pour ces
61 questions liées aux données. On peut l'enlever si tu veux revenir à ce qu'on
62 avait dit précédemment, j'ai le message je peux copier-coller facilement, mais
63 je pense que c'est important de, ben, juste de donner cette clé à l'utilisateur.

64 Félix : Ouais c'est juste que ça devient lourd, puis t'sais on lui dit une fois puis on lui
65 dit qu'il peut aller chercher *learn more* fait que le problème c'est que ça
66 devient lourd, je veux bien qu'on soit *privacy by design* mais à un moment
67 donné, déjà rien que l'intro elle est lourde, fait que je trouve ça, j'aurais
68 tendance, je l'avais simplifié, ici là puis là « *excellent nananannana, I have a
69 couple of question for you* »

70 Camille : Ben là j'ai enlevé encrypté, j'ai mis juste *private and confidential*

71 Félix : Ouais parfait

(6^{ème} extrait de retranscription de la section « Formuler le consentement »,
réunion du 23 septembre 2019)

²⁰³ La création de raccourci faisait l'objet d'une tâche formulée en *user story* (voir chapitre 2, p. 62) dans un des sprints : « *As a user I need to have access to the privacy info as soon as I say "privacy"* (pour ne pas enfermer cette info dans un seul chemin mais permettre de trouver l'info par un mot clé) » (Document « Éthique + gestion des données », diapositive 24 « UX Sprint 4 »).

Les ajouts et modifications liés au consentement doivent donc non seulement éviter de heurter l'expérience utilisateur par de la lourdeur, mais également éviter de modifier le document principal qui sert de dialogue avec les clients. En revanche, les modifications faites pour enlever ou alléger quelque chose en rapport avec le consentement semblent valides aux yeux de Félix : quand je mentionne avoir enlevé un mot (« encrypté²⁰⁴ », l. 70), il acquiesce; quand il mentionne avoir « enlevé parce que c'était trop long » (5^{ème} extrait, l. 57), je ne sais pas à quoi il fait référence, mais j'entends qu'il a probablement modifié quelque chose à mon insu. Il est en tout cas revenu aux versions précédant le travail de reformulations que nous fait un mois plus tôt. C'est également cela qui peut être intéressant en travaillant dans ce document principal : voir comment les autres – en particulier Félix qui dirige le UX – interviennent et modifient les interactions à propos du consentement. Mais je ressors de cet échange avec une certaine frustration :

Et puis je repense à la réunion d'hier qui m'a frustrée. J'ai l'impression de ne pas avoir tant d'espace pour travailler cette affaire de consentement. Félix a la main forte sur son document, il n'apprécie pas que j'y change des choses.

—> lui demander où faire des modifications, comment les faire pour qu'on puisse en discuter après (...)

(Journal de bord, notes réflexives, 24 octobre 2019)

Ces interventions et modifications dans le module du consentement sont finalement à nouveau déplacées dans un autre document draw.io (devenu Diagram) que Félix créé à partir d'une copie du document de design de conversation existant – copie intitulée « *consent conversation design* ». C'est à nouveau principalement moi qui y contribue.

La création d'une copie de document permet donc d'avoir à nouveau un espace pour le travail sur le consentement – comme c'était le cas quand ce travail était réservé à Mr Y, puis quand un document avait été créé le 23 septembre 2019, dans le dossier avec les documents de Prélude. Il s'agit à nouveau d'un espace où travailler les compromis du module de consentement – sans modifier le document principal à l'insu de Félix – pour qu'il fusionne agréablement dans l'expérience principale de conversation, tout en permettant au consentement d'être suffisamment informé et bien placé pour être

²⁰⁴ Ce mot est un anglicisme, la version francisée est « chiffré ».

valable. Mais il reste de l'intranquillité : en se faisant à part, ce travail est face à un risque d'isolement qui pourrait l'empêcher d'aboutir dans le document principal. Il resterait un module indépendant mais isolé, cela l'empêcherait d'être pris en considération dans la conversation de l'agent artificiel et d'ainsi parvenir à faire changer les lignes du court-circuit habituel du consentement – il resterait un *matter of caring* isolé, dépourvu de moyens d'action dans ce qu'il vise d'améliorer. Il faut alors trouver à faire des ponts vers ce qui permet la mise en acte de l'agent conversationnel, c'est-à-dire ce qui permet son développement combinant UX et technique.

Un pont qui ne fonctionne pas, c'est d'intervenir comme je l'ai tenté, directement dans le document principal de design de conversation. Mais un autre pont est possible, via Jira, l'outil de gestion de projet technologique pour les développeurs et non-développeurs. Jira est l'outil qui permet d'organiser et suivre les sprints de développement (voir chapitre 2, p. 81). Chaque personne de l'équipe impliquée dans un sprint a des tâches (ou « tickets ») qui lui sont assignées. Par exemple, la tâche ci-dessous (figure 12), « Faire valider les maquettes du carrousel privacy à Félix », est décrite selon les objectifs visés, et son état est indiqué « en cours » (« *doing* ») dans le cadre de ce sprint. Il est possible d'interpeler d'autres personnes via ces tâches, soit en les taguant (« @Félix (...) »), soit en ajoutant un autre contributeur à cette tâche. La tâche, ou le ticket, peut également être associée à d'autres tickets, ce qui la rend visible ailleurs. Tant que la tâche n'est pas indiquée comme réalisée ou « *done* », le *sprint* ne peut pas passer en mode « complété », les tâches encore en cours seront transférées au *sprint* suivant.

L'effort de sortir le consentement de son isolement passe ainsi par l'adoption des outils (Jira, draw.io, Google Docs) et de la logique qui organise le développement technologique avec agilité, une logique qui vise la clarté et la faisabilité des tâches constituant le projet. En tentant de faire fusionner la logique du *matter of caring* – qui vise une transformation des pratiques – dans la logique agile articulée en différentes tâches pour assembler des modules entre eux, j'essaie de défier l'isolement du consentement. Travailler le module de consentement dans un espace indépendant est une première phase, il faut à présent trouver à arrimer ce module dans les autres activités

qui peuvent l'aider à se concrétiser. En d'autres termes, il s'agit de travailler le processus d'assemblage plus ou moins lâche, où le module se retrouve « *loosely coupled* » (Gürses et Von Hoboken cités ci-dessus) avec les autres modules de conversation qui, ensemble, constituent la conversation générale, ainsi qu'avec les autres modules d'infrastructure technologique qui les soutiennent. Un outil comme Jira permet d'ajuster ces assemblages, de les renforcer plus ou moins en créant des liens entre les tâches qui ne relèvent pas des mêmes équipes ni des mêmes sprints.

The screenshot shows a Jira ticket interface. At the top, there are two checkboxes for 'UXTM-66' and 'UXTM-113'. The ticket title is 'Faire valider les maquettes du carrousel privacy à Félix'. Below the title are buttons for 'Joindre', 'Lier un ticket', and a checkmark icon. The 'Description' section contains the text: 'Voilà des visualisations de ce à quoi pourrait ressembler le carrousel sur les questions de privacy. objectifs:'. This is followed by a bulleted list of objectives: 'voir comment on présente l'information clairement et rapidement', 'voir comment on navigue dans le contenu (cf boutons)', and 'corriger selon les feedbacks de Félix puis valider pour que les tech puissent l'implémenter dans le bot.' A sub-bullet under the third point says 'Pouvoir le faire tester ensuite par des utilisateurs lors de validation sessions.' Below the list is the text 'cf explications sur les slides' and a comment from '@Félix' asking 'peux-tu y jeter un oeil et me donner ton feedback stp?'. The 'Pièces jointes' section shows two screenshots of a carousel design. The right sidebar contains metadata: 'ETAT: DOING', 'DATE D'ECHEANCE: Aucun', 'RESPONSABLE: Camille Vezy', 'ETIQUETTES: Aucun', 'SPRINT: BM Sprint 5', 'ESTIMATION EN STORY POINTS: Aucun', 'RAPPORTEUR: Camille Vezy', and 'TEMPO: Ouvrir Tempo'. At the bottom, there is a 'Tickets associés' section with a link to 'UXTM-100' and an 'Activité' section with a comment from Camille Vezy from 8 hours ago.

Figure 12 : Captures d'écran de la tâche « Faire valider les maquettes du carrousel privacy à Félix », 15 novembre 2019.

Dans la tâche Jira ci-dessus, les « maquettes » que je partage font référence à un travail que je fais dans un autre document que le « *consent conversation design* ». Il s'agit d'une présentation intitulée « Éthique + gestion des données », sous format Google Slide, qui documente tout ce que j'ai identifié à prendre en compte pour informer le consentement, pour lui permettre d'être valable, et mettre en place une approche de *privacy-by-design* (voir chapitre 1, p. 60, et annexe, p. 365). J'ai créé ce document sous l'effet d'une préoccupation au fil de mes interactions avec Félix et sa volonté explicite de donner une place au consentement, certes, mais la moins encombrante possible : à force de faire fusionner ces interactions de consentement dans le reste de l'expérience de conversation, de réduire la place qu'il prend, donc en même temps de réduire la quantité d'informations données, de chercher à le placer à un endroit où il soit le moins gênant possible, est-il toujours valable ?

À quoi dit-on ok ?

Quand Félix m'a dit que je ne pouvais pas changer le document principal comme je l'avais fait, mon ajout a été effacé. Dans la suite de mes notes de terrain qui manifestent ma frustration de ne pas avoir tant d'espace pour travailler le consentement, j'explique mon silence à ce moment :

Quand j'ai voulu changer des choses, Félix a finalement tout remis comme avant, je m'efface, mais le problème c'est que l'unique phrase qu'on a actuellement ne dit pas grand-chose, et le consentement n'est pas explicitement obtenu :

« *But before we go on, note that all the conversations are encrypted, private and confidential. Should we continue now?* »

—> il faudrait avoir une question qui appelle un « ok » ou « *learn more* » avant le « *should we continue* ». Sinon le ok est un ok à continuer, pas un ok aux conditions.

(Journal de bord, notes de terrain du 24 octobre 2019)


Je soulève ici une préoccupation quant à la négligence accordée au « ok », alors que c'est ce ok qui est au cœur des préoccupations (éthiques) puisque c'est en lui que se cristallise le consentement donné, et donc sur lui que repose en partie ce qui fait que cette expérience peut être bonne. Cette nuance met le doigt sur un potentiel *dark pattern* (voir chapitre 3, p. 131) où le ok serait pris comme un accord au traitement des données

personnelles impliquées alors que, tel qu'il serait positionné, il serait plutôt compris comme un ok pour passer à la suite. En proposant un ok maladroitement placé, au nom de la rapidité recherchée, c'est en même temps un exemple d'invulnérabilité poursuivie : atténuer les raisons de dire non – qui rendrait impossible la collecte de données nécessaire pour faire fonctionner l'agent conversationnel – permet d'assurer que l'expérience va avoir lieu, que le bot sera utilisé. Le *dark pattern* permet de favoriser cette utilisation, en détournant de la possible opposition.

Le « ok » se trouve ainsi dans une ambiguïté liée à la fusion du consentement dans l'expérience utilisateur principale : celle-ci vise une navigation rapide dans ces interactions, alors que le souci éthique ne veut pas que le consentement soit donné au nom (ou à cause) de la rapidité. Si cette rapidité, et la fluidité qui va avec, teinte le processus de fusion du consentement dans l'expérience principale, celle-ci l'amène en même temps à compromettre sa validité. Cette confrontation peut interroger sur ce qu'est l'agréable, voire le plaisir, dans ce contexte : est-ce profiter le plus vite possible de ce que l'agent conversationnel a à offrir, ou est-ce constater que l'on prend soin des traces que l'utilisatrice donne d'elle (ses données personnelles), des traces dont l'agent conversationnel a besoin pour fonctionner ? La deuxième option demande un peu plus de temps à l'utilisatrice, pour voir si elle est d'accord ou non avec ce qui va être fait des informations personnelles qu'elle s'apprête à partager. Elle va à contre-courant de la rapidité privilégiée, mais elle peut susciter de la surprise que du soin soit accordé à cela; elle peut susciter aussi plus de confiance pour utiliser l'agent conversationnel. La surprise a des chances d'être agréable pour les personnes pour qui ces questions comptent; elle risque en revanche d'être source d'une légère frustration pour les impatientes qui veulent passer à la suite le plus vite possible. Mais pourquoi donner plus de poids à l'impatience ? En envisageant l'option moins rapide, il s'agit de considérer comme des voix qui comptent celles pour qui le plaisir est de constater que ces questions sont adressées. Tenter de désamorcer la tendance à dire ok rapidement pour passer à la suite, c'est une intervention de l'ordre du *matter of caring*, qui cherche à proposer autre chose que l'habitude

impatiente, et quelque peu dominante, qui tient pour acquis que le ok rapide est implicitement ce que l'utilisatrice veut, et qu'elle est d'accord par défaut.

Prendre soin de la position du ok n'est cependant pas la seule condition qui permettrait au consentement d'être valable. L'information à laquelle on demande ce ok doit être complète et vraie. C'est dans le souci de proposer une information complète que je travaille les « maquettes » évoquées dans la tâche Jira ci-dessus. Ces maquettes sont des représentations visuelles de ce à quoi le module de consentement pourrait ressembler une fois affiché dans l'Assistant Google : où positionner les boutons, comment apparaîtrait l'information, etc. Comme Félix a positionné le carrousel informatif sous forme de « *learn more* » (voir ci-dessus, 6^{ème} extrait de la réunion du 23 septembre 2019), une préoccupation a émergé de mon côté : « si on ne force pas à passer à travers tous les éléments (*data collection, security & confidentiality, data sharing, delete*), le consentement est-il vraiment donné ? » (Journal de bord, notes de terrain, 30 septembre 2019). Plutôt que d'envisager des *nudges* (voir chapitres 1, p. 57 et 3, p. 152) qui forceraient certaines actions pour informer, je propose de mieux rendre visibles les catégories d'information, en les rappelant sous forme de boutons²⁰⁵ à chaque fois qu'une information est consultée. Je suggère également deux formats pour présenter ces informations : la carrousel déjà abordé (figure 13), et la liste (figure 14) – qui est le format déjà rapidement implémenté dans le prototype de l'agent conversationnel.

L'avantage du carrousel (figure 13 ci-dessous) est qu'il permet de faire apparaître directement l'information essentielle. Le bouton « ok » ou «  Got it » peut ici s'afficher directement car l'utilisatrice aurait eu une vue d'ensemble des informations.

²⁰⁵ Comme l'interaction peut être autant tactile que vocale, il n'est pas nécessaire de toucher ces boutons pour les activer, les mots qu'ils indiquent (et leurs synonymes) peuvent être dits vocalement par l'utilisatrice pour réaliser l'action correspondante.

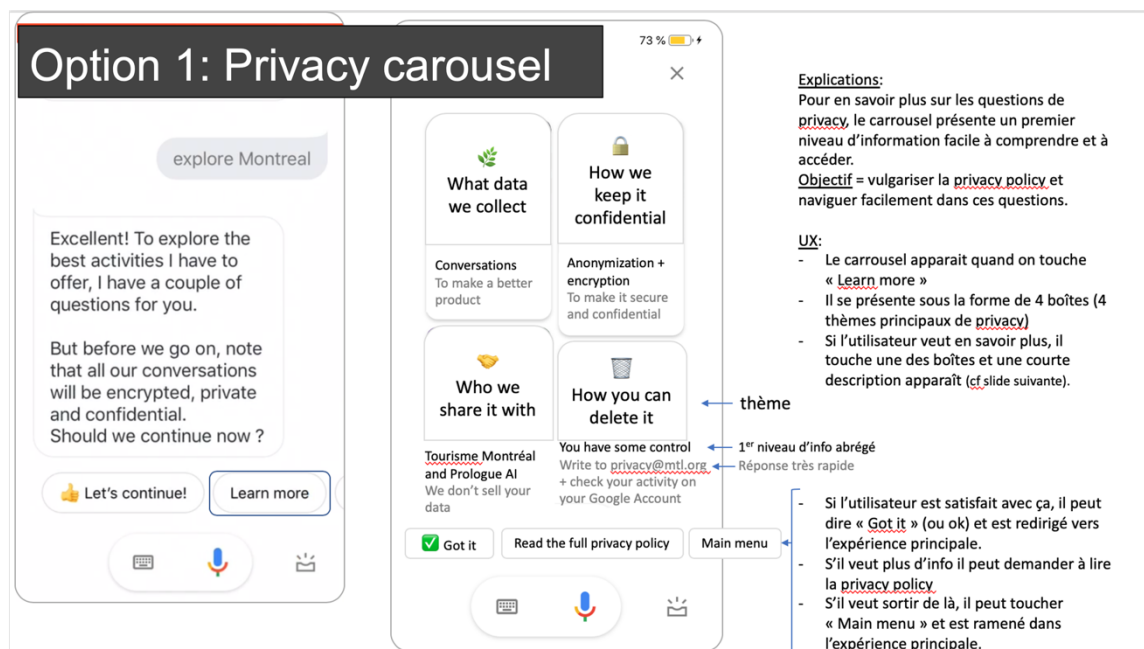
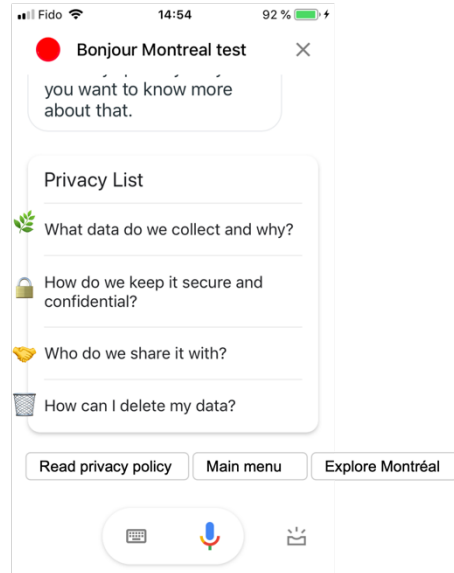


Figure 13 : Capture d'écran de la visualisation du carousel pour informer le consentement. Présentation Google Slide « Éthique + gestion des données », diapositive 25.

L'autre option envisagée – déjà dans le prototype, car plus rapide à intégrer par le développeur – est celle d'une liste, où l'information est plus allégée. À la version déjà intégrée, je propose d'ajouter des émojis pour une légère touche visuelle; je propose également d'ajouter des boutons en bas pour ne pas enfermer l'utilisatrice dans cette liste mais lui rappeler la possibilité d'aller dans d'autres sections de l'agent conversationnel. En revanche, parmi ces boutons, le « ok » ou « Got it » ne peut pas s'afficher directement après la liste. Car la liste ne fait que présenter les thèmes sur lesquels informer l'utilisatrice, mais elle ne donne pas encore les informations en question. Le bouton « Got it » peut donc s'afficher seulement une fois que ces informations sont consultées.

Option 2: Privacy List



Instead of the carousel, we can adopt a Privacy List (it's what is deployed at the moment).
 → add emojis to make it a bit more visual
 → add buttons at the bottom to make it easier to go somewhere else than this list

Figure 14 : Capture d'écran de la visualisation de la liste pour informer le consentement. Présentation Google Slide « Éthique + gestion des données », diapositive 26.

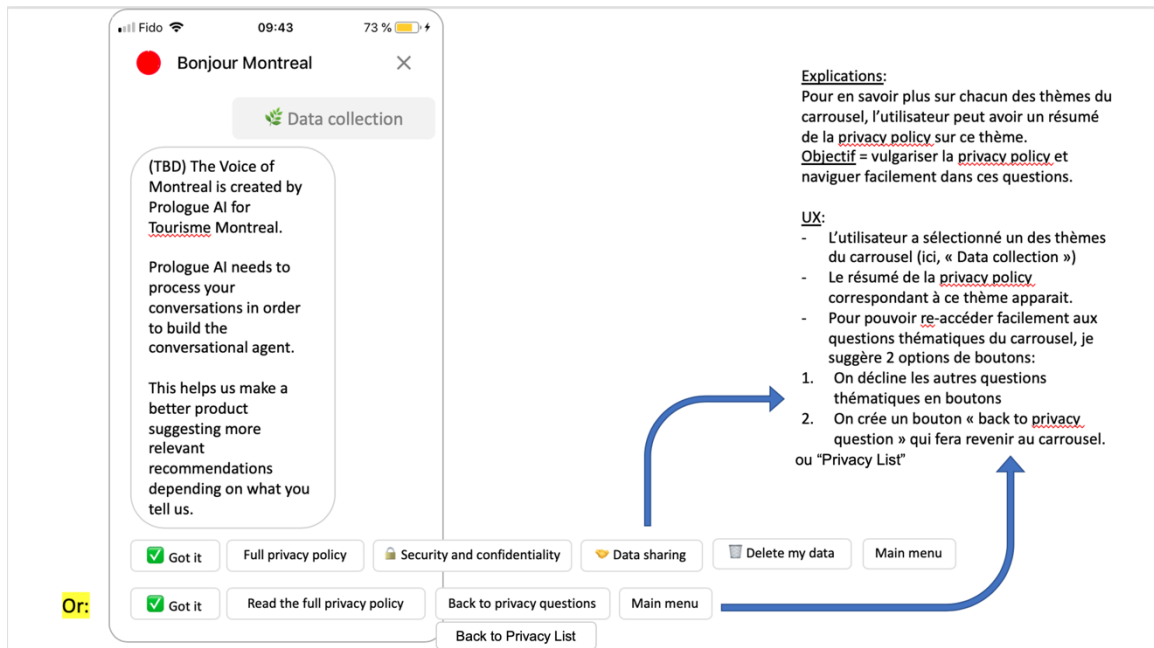


Figure 15 : Capture d'écran de la visualisation de la précision des informations quand un thème de la liste est demandé. Présentation Google Slide « Éthique + gestion des données », diapositive 26.

En travaillant la clarté par laquelle présenter ces informations – dans un format qui ne soit pas trop lourd ni encombrant afin de préserver la fusion du module du consentement dans le reste de l’expérience utilisateur – je cherche à orienter la navigation de l’utilisatrice dans ces informations pour donner une chance à son consentement d’être un minimum informé. En d’autres termes, je travaille les affordances (voir Gibson, 1979 et Leonardi, 2011, abordés dans le chapitre 1, p. 57) par lesquelles on rendra visible la possibilité de donner son consentement, pour que ces affordances permettent de donner ce consentement dans des conditions favorables à sa validité. La préoccupation qui m’oriente est de contribuer à la validité du consentement de l’utilisatrice, plutôt que de privilégier l’intérêt de l’organisation à orienter les actions de l’utilisatrice pour obtenir son consentement. C’est une nuance dans l’orientation qui est importante pour faire du consentement un *matter of caring* : il s’agit, dans les activités de design, de reconnaître la constitution du consentement à laquelle les designers participent, de ne pas l’instrumentaliser pour leurs propres fins (obtenir ce consentement) mais en favoriser la qualité (permettre sa validité).

Pour permettre cette qualité, le travail des formulations et l’orientation de la navigation de l’utilisatrice ne suffisent pas : il faut que les informations à propos du traitement des données personnelles soient données avec justesse, qu’elles correspondent effectivement aux pratiques adoptées par l’organisation. Mais qu’en est-il de ce qui est prévu du côté de l’organisation à propos de ces données personnelles ? Quelles sont les pratiques de gestion des données qui sont mises en place ? Comment va-t-on permettre aux utilisateurs d’exercer leurs droits ? C’est à ces questions que j’ai besoin de trouver des réponses pour pouvoir compléter les informations données dans ce module de consentement. En posant ces questions à l’occasion des deux réunions du 23 septembre et 23 octobre 2019, je me trouve cependant face à des zones de flou.

Maintenir la responsabilité

Après avoir abordé les formulations et la position des informations dans le design d'interaction, mes interventions se mettent à interroger les pratiques de traitement des données personnelles envisagées. Elles sont animées par le souci de contribuer à mettre en place des conditions qui favorisent la validité du consentement demandé. Faire de ces conditions un *matter of caring* amène à travailler ma complicité pour contribuer à être honnête dans les informations données, plutôt que de faire de l'*ethics-washing* en employant des mots qui donnent une impression de pratiques responsables et dignes de confiance, alors qu'il ne s'agirait que de façade. Il s'agit en d'autres termes d'être vigilante à ce que font les mots qui informent le consentement : qu'ils favorisent de la confiance de la part des utilisateurs, à condition que Prélude fasse preuve de responsabilité envers ce qu'ils partagent sur eux (leurs données personnelles). Cette responsabilité est en même temps ce que le *matter of caring* anime dans l'acte de design : il porte son attention sur les valeurs (éthiques) à concrétiser dans le design plutôt que de se concentrer à obtenir, par le design, ce dont l'organisation a besoin (le consentement et les données).

Je vais présenter le travail de clarification du traitement des données personnelles qui a été fait lors des deux réunions du 23 septembre (« *Is Google Assistant GDPR compliant?* ») et du 23 octobre 2019 (gestion des données). Quand je mets de l'avant une vulnérabilité de l'Assistant Google qui n'est potentiellement pas conforme au GDPR à ce moment, les membres de Prélude cherchent à éviter que cela fasse trop de bruit. C'est ensuite l'impuissance face à Google qui est mise de l'avant, entraînant l'organisation dans une certaine soumission que je tente de maintenir active en rappelant la responsabilité qui lui incombe. Les membres de Prélude cherchent ensuite à me rassurer quand je pose davantage de questions sur les pratiques de traitement des données personnelles à mettre en place : les données que Prélude va collecter ne sont pas si personnelles.

« Ne pas faire de tempête dans un verre d'eau »

La rencontre avec le cabinet Data Protection Consulting (DPC), qui audite TMTL et ses partenaires pour évaluer sa conformité au RGPD, a soulevé de nombreuses questions,

en particulier sur le choix des outils de Google (l'Assistant et Dialogflow) pour lesquels il n'est à ce moment pas clair s'ils sont conformes au RGPD :

Aujourd'hui la question de la protection des données personnelles est revenue de l'avant, par l'effet d'une intervention externe, avec d'autant plus d'autorité que la rencontre était demandée par le client (TMTL) en train de se faire auditer sur le sujet. La crainte est d'être sur un sentier encore peu battu, où l'on jongle avec de multiples partenaires qui, chacun, touchent d'une façon ou d'une autre des données à caractère personnel. Il faut donc trouver à établir un processus clair sur la façon dont ces données peuvent être efficacement protégées, pas seulement de notre côté mais du côté de chacun des acteurs de la chaîne. Il faut également des processus clairs qui permettent aux utilisateurs d'exercer leurs droits (suppression, rectification, accès, opposition au traitement).

(Journal de bord, notes de terrain du 11 septembre 2019)

Je crée quelques tâches dans Jira, sous forme de *user stories*, pour animer ces préoccupations chez Prélude et mettre en place les fonctionnalités qui permettront d'exercer ces droits : « *As a user I want to delete my data* » (11 septembre 2019).

Cette rencontre ravive également une question : l'Assistant Google est-il conforme au RGPD ? Cette question fait du bruit à partir du 17 septembre 2019, quand, agitée par la vigilance que DPC nous a amenés à porter sur les termes de traitement des données personnelles par Google via l'Assistant, j'ai replongé dans un article du mois d'août (Vincent, 2019) à propos de pratiques controversées liées à l'Assistant Google :

Cet été, 1^{er} août, Google a reçu un avis du Commissaire de Hambourg à la protection des données qui a ordonné 3 mois d'arrêt des pratiques consistant à transcrire certains extraits enregistrés via les appareils dotés de Google Assistant (Google fait affaire avec des contracteurs humains pour vérifier et corriger les transcriptions afin d'améliorer la reconnaissance du langage dans plusieurs pays²⁰⁶. Certains de ces enregistrements comportaient des données sensibles, en particulier quand l'Assistant s'activait par erreur).

Que doit-on faire avec cette information ? Cela signifie-t-il que Google Assistant n'est pas *GDPR compliant*? Doit-on attendre qu'il le devienne sur ce point-là pour pouvoir l'utiliser?

²⁰⁶ Le travail de ces contracteurs est la plupart du temps un travail invisible effectué avec les données. La valorisation de ce travail invisible pourrait faire l'objet d'une *matter of caring* dans les organisations qui y ont recours : assurer une rémunération adéquate des contracteurs, les former aux situations où ils sont exposés à des contenus sensibles ou violents, leur attribuer des tâches en lien avec leur culture, leurs intérêts et champs de connaissance, etc.

(Journal de bord, notes de terrain du 17 septembre 2019)

J'ai écrit cet extrait de notes de terrain pour m'en servir comme brouillon de courriel destiné à des experts en protection de données personnelles comme DPC. Tandis que je soumetts à Paul et Félix l'idée de poser ces questions à l'une des consultantes de DPC, ils me demandent d'en parler d'abord en équipe, pour « être certain de ne pas créer une tempête dans un verre d'eau » dit Félix dans une conversation sur Slack (Journal de bord, notes de terrain du 17 septembre 2019). Paul me précise plus tard qu'ils sont à ce moment en négociation du contrat avec TMTL pour la Voix de Montréal, ils préfèrent voir de notre côté comment faire, éventuellement avec une autre personne experte en protection des données plutôt que d'attirer l'attention sur cet élément particulier. Quand nous nous réunissons le 23 septembre 2019, Félix reprend cette mise en garde :

- 1 Félix : Ouais j'suis d'accord, faut le documenter, pis j'veux pas cacher cette question-
- 2 là sous le tapis hein c'est vraiment pas notre intention, j'veux juste pas créer
- 3 de[
- 4 Paul : [de vague
- 5 Félix : ouais une tempête dans un verre d'eau avec les clients, parce que si tout le
- 6 monde se met à *freaker* sur Google Assistant, là on est fait.
- 7 Camille : Oui bien sûr, nan nan j'comprends en plus c'est un mauvais timing, contrat
- 8 tout ça, je comprends très bien

(1^{er} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* »,
réunion du 23 septembre 2019)

On ne veut pas « cacher cette question-là sous le tapis » (l. 1-2) mais on ne veut pas l'exposer au client dans ce moment stratégique où Prélude et TMTL sont en train de conclure leur contrat de partenariat. La préoccupation de la conformité de l'Assistant Google éveille ainsi le risque de créer du désordre alors que les deux organisations Prélude et TMTL cherchent à créer ensemble un ordre (un contrat) qui permette à l'agent conversationnel de continuer à être développé, et à Prélude de continuer à se constituer autour de ce développement technologique. S'occuper de la gestion des données personnelles en en faisant un *matter of caring* doit ainsi se faire dans le périmètre de cet ordre en constitution, en limitant certains désordres.

S'il y a de l'inconfort avec cette question, c'est qu'elle met le doigt sur une vulnérabilité qui pourrait empêcher le projet de se réaliser : l'outil principal sur lequel il repose est-il fiable ? S'il n'est pas conforme au RGPD alors que TMTL tient à cette conformité, alors il pourrait être préoccupant : « j'voudrais pas que [DPC] fassent « ouais finalement euh t'sais laissez-faire le contrat avec Prélude » j'veux dire, là, tout tombe, t'sais » (Félix, 4^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* », réunion du 23 septembre 2019, cité ci-dessous, l. 45-46). Le souci principal pour Félix ici est la signature du contrat – le respect des données personnelles par l'outil sur lequel repose le projet est une question problématique à adresser, mais elle est secondaire par rapport au contrat qui permet à la fois le développement de l'agent conversationnel, et la constitution de Prélude en même temps.

Si le respect des données personnelles était un souci prioritaire, une façon radicale de prendre soin des pratiques de gestion des données personnelles dans l'organisation serait de choisir de ne pas utiliser cet outil qui présente des failles à l'endroit même de ce dont on veut prendre soin – la confidentialité de toute donnée personnelle collectée et le respect de la vie privée qui nécessite de collecter le moins de données possibles et d'empêcher toute intrusion, en particulier par le biais d'erreurs qui amènent à collecter des données à l'insu des utilisateurs. C'est ce que je pourrais conseiller et inciter, en tant qu'*ethics owner* incarnant ce *matter of caring*.

Mais ce choix radical n'est pas envisagé, car choisir cet outil dont l'*Automatic Speech Recognition (ASR)* est « le plus performant sur le marché » (Félix, 4^{ème} extrait de la retranscription de la section « *Is Google Assistant GDPR compliant?* », réunion du 23 septembre 2019, cité ci-dessous, l. 34) permet de rendre possible le développement de cet agent conversationnel dans le temps restreint des trois prochains mois – c'est l'horizon de temps qui est annoncé dans le contrat en négociation. Ce qui a le plus de poids dans le choix de l'outil, c'est la faisabilité qu'il permet, la performance qu'il promet, malgré les faiblesses qu'il manifeste en matière de respect de la vie privée. Alors l'accent de ce qui fait l'objet d'un *matter of caring* lié au consentement et au respect des données personnelles, ce n'est pas tant le choix de l'outil que la conscience de ce qu'implique

l'utilisation de cet outil en ce qui concerne le traitement des données personnelles : « il faut qu'on soit bien conscients des avantages et des risques que ça implique » (Camille, 3^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* », réunion du 23 septembre 2019, cité ci-dessous, l. 25-26).

Le travail de l'*ethics owner*, animé par le *matter or caring* du consentement, ne se concentre donc pas sur la remise en cause de ce choix – cela irait à l'encontre de la logique de mise dans l'organisation, celle qui vise la faisabilité du projet avant tout. Par complicité, ce travail se positionne plutôt dans les limites de ce choix, portant attention à ce qui peut et doit être fait – c'est cette recherche de ce que l'on peut et doit faire qui font de cette préoccupation un *matter of caring* dans cette situation. Par ce travail, il s'agit d'adresser le sentiment d'impuissance des membres de Prélude, de reconnaître leur soumission à l'outil choisi, mais de rendre cette soumission active en animant leur responsabilité dans ce choix.

« On n'a pas de contrôle »

Pour introduire la réunion du 23 septembre, je récapitule la situation ayant mené à la question de la conformité de l'Assistant Google au RGPD. Je précise que les pratiques problématiques de Google et leurs contracteurs sont suspendues pendant trois mois (jusqu'au mois de novembre 2019), le temps de faire preuve de davantage de transparence, auprès du Commissariat à la protection des données et de la liberté de l'information de Hambourg, sur leurs pratiques de transcription de données collectées par l'Assistant. Le délai tombe bien : le lancement de la version bêta de l'agent conversationnel est prévu pour le mois de janvier 2020. Je mentionne : « en termes de *timing*, pour nous, ça correspond bien », manifestant d'emblée ma complicité avec Prélude et sa temporalité, pour que cette question problématique ne soit pas un risque de vouer le projet à l'échec. Paul appuie, et Félix répond en marquant un certain éloignement :

9 Félix : nan c'est ça puis nous t'sais j'veux dire on peut (.) c'est tellement (.) loin de
10 nous cette mécanique de (.) c'est-à-dire qu'on peut tellement rien y faire à ça,
11 que on peut juste comme (.) passivement voir ce qui est en train de se passer

12 là-dessus euh mais on dira pas des faussetés, t'sais, on n'est pas en mesure de
13 dire si Google Assistant est *GDPR compliant* mais en même temps est-ce que
14 Google Assistant a intérêt d'être *GDPR compliant* ?

15 Camille et Paul : oui

16 Félix : oui on s'entend que oui là t'sais ils peuvent pas se priver de l'ensemble de
17 l'Europe comme utilisateurs fait que j'ai comme le *feeling* que eux autres avec
18 ils doivent faire comme « hey *guys* on doit renforcer notre-notre processus
19 pour être *compliant* au mieux de notre capacité »

(2^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* »,
réunion du 23 septembre 2019)

L'éloignement que Félix souligne par rapport aux pratiques de Google – « cette mécanique » (l. 10) – manifeste un sentiment d'impuissance par rapport au problème : « on [ne] peut tellement rien y faire » (l. 10). La seule issue qu'il envisage est d'attendre « passivement » (l. 11) pour voir l'évolution des pratiques de Google. Il évoque ainsi la soumission de Prélude par rapport à l'outil principal que l'organisation utilise, comme si la passivité était la seule chose à faire dans cette situation où Prélude serait à la merci des affordances technologiques de l'Assistant Google. Par cette passivité, il s'agit de se conforter dans le biais optimiste (voir chapitre 3, p. 165) qui a orienté la décision d'utiliser cet outil (« le plus performant du marché » dans le 4^{ème} extrait cité ci-dessous, l. 34). Le choix de cet outil très performant semble ainsi mettre Prélude dans la position de se soumettre à quelque chose qui les dépasse, auquel les futurs utilisateurs seront également amenés à consentir puisqu'ils devront utiliser cet Assistant s'ils veulent utiliser l'agent conversationnel de Prélude. C'est là un aperçu de la complexité de l'écosystème technologique qui pose des limites à nos consentements numériques qui sont dans l'impossibilité de tout connaître, étant donné que les organisations elles-mêmes ne savent pas tout de ce qui est fait des données qui transitent par plusieurs systèmes (voir chapitre 3, p. 165).

Je réagis à ce sentiment d'impuissance en rappelant la responsabilité de l'organisation :

20 Camille : ((*acquiesce pendant ce qu'il dit*)) moi aussi je pense qu'ils sont dans cette
21 mouvance-là, par contre, ce qui imp-, enfin nous ça nous concerne un peu
22 parce qu'on est responsable de choisir avec qui on fait affaire et donc, 'fin on
23 porte la responsabilité de choisir en gros nos data processeurs, donc tous ceux

24 par qui nos données passent, et donc on est responsable aussi de choisir de
25 faire affaire avec Google. Et dans ces cas-là, il faut qu'on soit bien conscients
26 des avantages et des risques que ça implique et donc qu'est-ce qu'on met en
27 place pour aussi légalement se protéger contre ça parce que, imaginons notre
28 Action est liée-je sais pas, il y a une faille de sécurité ((Paul et Félix
29 acquiescent)) ou j'sais pas, et c'est les données de notre Action qui sont
30 exposées, ben c'est sur nous que ça rejaillit aussi.

(3^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* »,
réunion du 23 septembre 2019)

Agiter la valeur de la responsabilité permet de ne pas céder à la tentation de fermer les yeux sur ce sujet en attendant passivement de voir l'évolution. Ce rappel de responsabilité ravive un sentiment intranquille qui permet de chercher à agir – le *matter of caring* réveille de l'inconfort. Cette responsabilité est elle-même animée par la figure d'autorité du RGPD. Celui-ci est indirectement présent, par l'emploi du terme « *data processeur* » (l. 23) – l'argument de la responsabilité du choix des data processeurs (anglicisme pour « sous-traitants » est d'ailleurs emprunté au RGPD²⁰⁷.

Malgré les limites de ce choix d'outil, au nom de la responsabilité de ce choix, il y a plusieurs choses que l'on peut faire : on peut prendre connaissance de ce qu'implique l'utilisation de cet outil en ce qui concerne le traitement des données personnelles de nos utilisateurs mais aussi « [mettre] en place [quelque chose] pour légalement se protéger » (l. 27) de potentielles failles de l'outil qui pourraient impacter l'agent conversationnel de Prélude qui fonctionne grâce à lui. Cette mention de quelque chose de légal à faire évoque une mise en garde lancée par Data Protection Consulting (DPC) le 11 septembre 2019 :

21 DPC : Le partenariat avec Google ça se passe comment? (...) Eux ils ont accès à quoi?
22 Paul : Si on le soumet sur Google Assistant? Il me semble qu'ils ont accès aux
23 données, c'est une co-propriété des données.

²⁰⁷ Les sous-traitants (*data processors*) sont les organisations qui « traitent des données personnelles pour le compte de leurs clients ». Les clients de ces sous-traitants sont les responsables de traitement (*data controllers*). En tant que responsable de traitement, Prélude est responsable de choisir ses sous-traitants. Selon le RGPD, les sous-traitants ont également « des obligations spécifiques en matière de sécurité, de confidentialité et de documentation de leur activité ». Voir : <https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-un-guide-pour-accompagner-les-sous-traitants>

24 DPC : Du coup, ça vaudrait quand même le coup de mettre des modalités de
25 protection des données dans le contrat avec Google (rires) bon ça va pas être
26 une surprise pour eux, hein, mais bon

27 Paul : Enfin ça c'est pas seulement spécifique à nous, c'est n'importe qui qui met
28 une action sur Google qui va être soumis à ça

29 DPC : Oui oui bien sûr, non je sais bien, mais c'est quand même une petite alerte,
30 notamment sur l'hébergement en fait, c'est-à-dire que s'ils font une copie ou
31 s'ils ont accès à une partie des données, dans la mesure où c'est des données
32 potentiellement personnelles d'utilisateurs de l'application, nous on peut dire
33 « ok on va le stocker sur AWS au Canada » mais après on sait pas ce que
34 Google en fait. Et euh c'est là que potentiellement c'est compliqué,
35 notamment potentiellement en termes de *data breach*, parce que si vous avez
36 une copie de la base qui est détenue par Google qui n'est pas dans votre
37 contrôle, où il n'y a pas de contrat de sous-traitance qui dit noir sur blanc
38 qu'en fait c'est des données d'utilisateurs de TMTL et qu'ils ont des obligations
39 de notification par exemple, euh ça il faut le prendre en compte. (...) Google
40 est en capacité de vous dire « ben non, je comprends pas, moi je suis pas
41 soumis au RGPD, je ne veux pas ». Ce qui est déjà arrivé, parce que c'est des
42 données canadiennes, vous êtes une entreprise canadienne, parce que TMTL
43 est une entreprise canadienne, Google va dire je vois pas pourquoi on devrait
44 appliquer la RGPD. Et là c'est un argumentaire à dire, parce que si vous avez
45 des données d'Européens qui sont collectées et que eux sont soumis au RGPD,
46 forcément ils doivent appliquer le RGPD. Mais comme en général ils sont un
47 peu réticents à le faire, voilà, c'est quand même une alerte qu'il faut avoir en
48 tête parce qu'ils sont soumis au RGPD en fait. Donc dans le cadre de votre
49 contrat de partenariat, il faut juste faire attention à ça. Parce que vous vous
50 allez transmettre des données à, si on part du principe de, si on prend les
51 grands mots du RGPD, « responsable du traitement, sous-traitant etc ». De
52 toute façon c'est de la sous-traitance ultérieure quoiqu'il arrive, entre vous et
53 Google, et donc il faut redescendre toutes les obligations, y compris les
54 obligations du RGPD. Donc ça, faut garder ça en tête.

(3^{ème} extrait de retranscription de la réunion avec DPC, 11 septembre 2019)

La technicalité juridique n'étant pas dans le champ de mes compétences, mon rôle d'*ethics owner* amène à rappeler l'attention qu'il faut porter à ce qui doit être mis en place légalement, sans toutefois pouvoir expliciter ce qu'il faudrait concrètement faire. Cet implicite marque les limites de mes connaissances et compétences par lesquelles est travaillé, dans cette organisation, le *matter of caring* du consentement et de la gestion des données personnelles.

Paul envisage de faire appel à leurs avocats pour aborder cela d'un point de vue juridique. Puis il revient sur les problèmes des pratiques actuelles liées à l'*Automatic Speech Recognition* (ASR) de Google :

31 Paul : Mais par rapport aux risques du ASR, si on suit une logique de *privacy-by-*
32 *design*, qu'est-ce que c'est qu'on doit faire ?

33 Félix : Ben on ne peut rien faire, moi j'ai l'impression qu'on ne peut rien faire à part
34 ne pas choisir le ASR le plus performant sur le marché [...]

35 Paul : Parce qu'ils sont tous dans le même bateau, Echo, Siri... dans l'article qu'elle
36 a partagé, ils sont tous dans le même lot, ils se sont tous fait attraper. Nan
37 c'est plus par rapport à, ok, vu qu'on sait qu'il y a un risque, est-ce qu'on doit
38 le partager avec quelqu'un, le notifier à un partenaire, que ce soit un
39 partenaire, un touriste, un[

40 Félix : [Mais moi mon sentiment j'ai un sentiment là-
41 dessus puis j'te laisse répondre dans une minute Camille, mon sentiment là-
42 dessus c'est qu'ya personne qui sait, t'sais mettons qu'on demande au DPO²⁰⁸
43 de nos clients, moi personnellement j'éviterais cette voie maintenant à l'étape
44 où on est rendus, j'aimerais mieux consulter un autre DPO là que le DPO de
45 nos clients, hein ? Parce que j'voudrais pas qu'ils fassent « ouais finalement
46 euh t'sais laissez-faire le contrat avec Prélude ». J'veux dire, là, tout tombe
47 t'sais, fait que j'aimerais mieux éviter ce bout-là, j'aimerais mieux qu'on
48 consulte un autre DPO, on en a un autre là qui est très bon (.) [...] mais encore
49 là, j'aimerais bien voir moi qui est-ce qui peut répondre à cette question-là ?]

50 Camille : [Nan c'est sûr

51 Félix : [Tu vois, on est tellement sous le secret...

(4^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* »,
réunion du 23 septembre 2019)

Paul a de l'expérience avec ces questions : pour Mr Y, il avait commencé une première phase d'accompagnement, par deux consultants en protection des données personnelles, pour mettre en place une approche de *privacy-by-design* (voir chapitre 1, p. 60). Il identifie donc des actions à faire plutôt que de s'en remettre à l'impuissance : doit-on partager le risque avec quelqu'un, le notifier à un partenaire, des utilisateurs (l. 37-39) ? Il manifeste ainsi une préoccupation de l'ordre d'un certain *matter of caring* : étant donné notre connaissance de ce risque, que doit-on faire (l. 32) – même si c'est potentiellement

²⁰⁸ *Data Protection Officer* : personne en charge de la protection des données personnelles et de la conformité avec les règlements comme le RGPD.

inconfortable pour nous ? Ce « que doit-on faire » est à la fois légal (qu'est-on obligé de faire selon les règlements ?), et éthique (que doit-on moralement faire ?).

Coupant court à l'énumération des personnes à potentiellement notifier selon Paul, Félix intervient par son impression d'impuissance à nouveau : son sentiment est que personne ne sait répondre à la question de ce que l'on doit faire (l. 41-42). C'est ici qu'il explicite vouloir éviter de faire appel aux consultants de leur client, pour éviter de faire échouer la signature du contrat (l. 42-48). Le point culminant de son impuissance, tentée par la passivité face à ces questions, est son constat d'être « tellement sous le secret » (l. 51) – ou tellement soumis au secret de ces pratiques que Google ne communique pas. Pourtant, lors de la réunion du 10 juillet, il reconnaissait vouloir contribuer à « démystifier » les pratiques de traitement des données personnelles (voir citation p. 277²⁰⁹). Plutôt que le questionnement à propos de notre devoir légal et moral étant donné notre connaissance de ces pratiques pour l'instant problématiques qui posent un risque pour les utilisateurs, il semble orienter vers un certain silence marqué de passivité docile face à un géant qui offre un outil dont on est dépendant. Ce silence serait en même temps animé par la préoccupation de « ne pas faire de tempête dans un verre d'eau » (comme abordé ci-dessus p. 297) afin de préserver les chances de signer le contrat avec le client, et de permettre ainsi au projet de se concrétiser.

Paul interrompt alors cette impuissance face au secret : il propose de demander conseil à Jake C., un des consultants qui ont accompagné Mr Y pour l'approche de *privacy-by-design*. Félix reste sur sa ligne de pensée : « Ouais mais (.) est-ce que Jake C. a le secret de comment le ASR de Google est fait ? » Paul et moi nous exclamons : « Naaaaaan », pour signaler que ce raisonnement n'est pas le bon. Je poursuis :

52 Camille : Mais par contre, lui, sa particularité c'est de connaître plein de pratiques
53 par lesquelles on peut protéger, et comment on peut aussi évaluer les risques
54 dans tout notre écosystème. C'est ça qui est intéressant de son expertise.

Cette voie permet de trouver des pistes pour agir, sans négliger ce problème ni fermer les yeux dessus alors qu'il pose des risques pour le respect de la vie privée des utilisateurs de

²⁰⁹ « Je pense que notre but c'est un peu de démystifier qu'est-ce qu'on fait avec tout ça et rendre ça accessible ».

Prélude – car ces utilisateurs devront en même temps utiliser l’Assistant Google. Cette piste semble apaiser Félix qui se dit favorable à le contacter et qui reconnaît avoir besoin d’en apprendre plus sur ces questions :

55 Félix : puis nous-mêmes on doit devenir des experts de ça, on doit dire t'sais voici
56 où se situe les limites de notre responsabilité voici ce qu'on sait à présent sur
57 le la solution le ASR de *Google Assistant* et compagnie, j'pense qu'on doit
58 s'éduquer là-dessus.

(5^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* »,
réunion du 23 septembre 2019)²¹⁰

Le sentiment d’impuissance est ainsi atténué par une piste qui permet de ne pas s’en tenir à l’ignorance de toute action possible.

Ainsi, l’intervention du *matter of caring*, en posant une question inconfortable qui pourrait freiner la réussite du projet, permet finalement d’aboutir à une volonté d’approfondir la littératie numérique de l’équipe pour connaître les implications – en matière de respect de la vie privée de ses utilisateurs – du recours à l’Assistant Google pour son projet d’agent conversationnel. Il s’agit de ne pas se laisser entraîner par la peur de l’outil mais de se documenter et d’agir en connaissance. La tentation de se déresponsabiliser face à l’absence de contrôle sur cet outil n’est pas une option, des réponses sont à trouver sans pour autant inquiéter le client. Par la manifestation de vouloir approfondir la littératie numérique dans l’équipe sur ce sujet, il semble que ce n’est plus seulement moi-même, dans la peau d’une *ethics owner* soutenue par Paul, qui porte attention aux implications des outils choisis en matière de respect de la vie privée des utilisateurs. Cette attention semble se mêler aux préoccupations d’autres personnes, Félix en particulier – et peut-être Cléa, qui assure la communication avec les clients, et qui est présente à cette réunion mais ne parle pas. Cette attention portée aux implications des outils de Prélude sur le traitement des données personnelles de ses futurs utilisateurs se mêle en particulier à la préoccupation stratégique de savoir répondre aux clients s’ils soulèvent ce point, et de faire preuve de responsabilité en adressant ces problèmes ou faiblesses rencontrées de manière proactive.

²¹⁰ Le 5^{ème} extrait se constitue des extraits de retranscription dans ce paragraphe qui condensent la suite de la discussion après le 4^{ème} extrait cité ci-dessus.

Plutôt que de se désresponsabiliser, Félix parle de « situe[r] les limites de notre responsabilité » (5^{ème} extrait, l. 56). Cette mention de limites de notre responsabilité à situer fait écho à une autre intervention de sa part plus tôt dans la réunion. C'était après que j'ai mentionné le besoin de « légalement se protéger » (3^{ème} extrait) – ce qui revient à délimiter la responsabilité juridique :

59 Félix : mais je pense qu'effectivement il faudrait peut-être légalement essayer de se
60 protéger, euh t'sais, même peut-être de l'évoquer pour que les gens soient
61 conscients, t'sais on n'a pas le contrôle sur la façon dont le ASR de Google est
62 entraîné et que les gens t'sais on pourrait le mentionner au moment où ils
63 acceptent de partager leurs informations. Mais t'sais, moi y'a un truc bcp plus
64 gros que ça comme enjeu que j'ai vu moi auquel j'ai-j'ai en fait qui m'a
65 vraiment fait halluciner vendredi dernier quand j'étais en train de revenir en
66 vélo, [...] j'me suis mis à écouter un podcast de Dugal²¹¹, qui disait que y'avait
67 une faculté en Belgique ((*Paul acquiesce, il semble avoir écouté aussi*)) qui en
68 fait a découvert que toute donnée anonymisée, tout le monde qui anonymise
69 les données, c'était très facile de pouvoir à partir de données anonymisées
70 reconstruire la base de données puis déterminer qui est derrière les données
71 anonymisées. Alors quand j'ai entendu ça, j'me suis dit t'sais on peut faire au
72 mieux qu'on peut mais si y'a des gens qui vraiment veulent aller trouver,
73 détecter qui se cache derrière les données anonymisées, on n'a pas de
74 contrôle là-dessus t'sais. Puis comme il l'expliquait dans son reportage, le
75 scientifique qui avait fait la démonstration de reconstituer une base de
76 données euh-euh anonymisées, il a dit t'sais on n'a jamais été dans ce cas de
77 figure là dans l'histoire de l'humanité : maintenant on est capable de savoir,
78 même si on n'a pas ton nom, on te on a de l'information sur le fait que t'as
79 une voiture, tu vis dans tel quartier, t'as un enfant, tu te déplaces de tel
80 endroit à tel endroit, donc c'est facile de reconstituer ces informations-là et
81 de dire c'est Camille Vézy t'sais mais en fait c'est pas facile mais c'est faisable...

82 ((*Camille et Paul acquiescent tout le long*))

83 Camille : on a tout le profil, on a beau ne pas avoir ton nom, on a tout ton profil
84 donc

85 Félix : c'est ça alors je me dis jusqu'où notre société, Prélude, a le blâme à porter
86 sur des gens mal intentionnés ? J'aimerais vraiment ça ne pas être
87 *accountable* pour ce genre d'affaire-là

²¹¹ Il s'agit de l'émission « Moteur de recherche » de Matthieu Dugal, diffusée le 19 septembre 2019 sur Ici Première, Radio-Canada : « Comment on peut identifier quelqu'un à partir des données anonymes ? » <https://ici.radio-canada.ca/ohdio/premiere/emissions/moteur-de-recherche/segments/entrevue/133833/donnees-anonymes-anonymisation-breche-faille-etude-internet-reseaux-sociaux-inquietudes-craintes-luc-rocher>

(6^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* »,
réunion du 23 septembre 2019)

Félix commence par appuyer ma suggestion, il évoque la possibilité d’informer les utilisateurs au moment « où ils acceptent de partager leurs informations » (l. 63), un moment qui est plutôt celui où l’on demande leur consentement – qu’ils peuvent accepter *ou refuser* de donner. Ce qu’il propose de mentionner, c’est l’absence de contrôle de Prélude par rapport aux pratiques de traitement des données personnelles impliquées dans le ASR de Google que Prélude utilise via l’Assistant (l. 61-63).

Il minimise ensuite ce point en portant attention sur « un truc beaucoup plus gros » (l. 63-64) : des recherches en informatique montrent qu’avec les pratiques actuelles d’anonymisation des données personnelles, il n’est pas difficile de réidentifier quelqu’un à partir de ces données, même si elles ne sont pas directement identificatoires comme son prénom, son nom, son adresse (Rocher et al., 2019). Cette information ajoute au sentiment d’impuissance ou d’absence de contrôle : « on peut faire du mieux qu’on peut mais [...] on n’a pas de contrôle » sur les personnes qui chercheraient à réidentifier les utilisateurs derrière les données que l’organisation anonymise (l. 71-74). Il préférerait alors ne pas avoir à « porter le blâme » ni la responsabilité de ces comportements mal intentionnés (l. 85-86). Délimiter la responsabilité de Prélude reviendrait ainsi à de protéger de tout blâme en évacuant la responsabilité de comportements qui sont hors de leur contrôle.

L’information de la réidentification possible de données anonymisées serait abordée différemment par une attention de l’ordre du *matter of caring* : celle-ci ne cherche pas à dénoncer ce que l’organisation fait mal, à blâmer, elle cherche plutôt à améliorer ses pratiques. La responsabilité dans ce cas ne serait pas tant quelque chose à délimiter pour se protéger qu’une attitude vigilante qui cherche à contribuer à l’amélioration des pratiques – on pourrait par exemple participer à la réflexion que ces chercheurs (Rocher et al., 2019) incitent pour renforcer les standards des pratiques d’anonymisation des données personnelles. En mettant l’accent sur l’amélioration plutôt

que la protection, le *matter of caring* invite à ne pas se résigner face à la difficulté, il s'agit plutôt de chercher des moyens de faire autrement²¹².

Dans la réunion portant sur la clarification du traitement des données personnelles chez Prélude un mois plus tard, le 23 octobre 2019, en discutant sur l'utilisation de cookies pour cibler les utilisateurs dans un cadre marketing, Félix me demande : « Mais ça, est-ce que c'est de la réutilisation des données personnelles, est-ce que c'est de la réutilisation frauduleuse des données personnelles ? ». Cette préoccupation envers une utilisation potentiellement « frauduleuse » semble insinuer que collecter des données personnelles serait par défaut une faute. Ce n'est pas dans ces termes que le *matter of caring* intervient : il ne s'agit pas d'empêcher toute collecte de données de la part de l'organisation, il s'agit de faire preuve de transparence auprès des utilisateurs et les prévenir de tout risque s'il y en a. L'attention du *matter of caring* n'est pas centrée sur l'organisation pour identifier ses intentions potentiellement mauvaises ou ses pratiques frauduleuses, elle est plutôt centrée sur les personnes destinées à interagir avec l'agent conversationnel en développement, pour assurer que ce que ces personnes partagent d'elles soit traité de façon respectueuse et responsable.

À la différence de l'attention d'une éthicienne qui évaluerait les pratiques et potentiellement les fautes, l'attention du *matter of caring* est ainsi plus proche d'un regard postphénoménologique, qui ne s'en tient pas qu'aux intentions des concepteurs puisqu'il y a toujours quelque chose qui leur échappe, et qu'ils mettent souvent de l'avant leurs bonnes intentions (voir Ihde, 1990; Verbeek, 2011; Rosenberg & Verbeek, 2015 abordés dans le chapitre 1 p. 67). Le regard est plutôt attentif à ce qui émerge dans leur utilisation d'outils technologiques, pour accompagner cette utilisation et entretenir une certaine vigilance envers les potentiels effets sur les utilisateurs finaux. Ce qui émerge dans les revendications d'impuissance ou d'absence de contrôle envers les pratiques de traitement des données liées à l'Assistant Google dont Prélude dépend, c'est la reconnaissance de sa soumission face à Google qui entraîne une volonté de délimiter sa

²¹² Cette tentation de se résigner face aux difficultés posées en matière de protection des données personnelles est également présente dans le cas des défis que le droit de retrait du consentement implique (voir Politou et al., 2018, cité dans le chapitre 3, p. 133).

responsabilité, sans pour autant se déresponsabiliser dans ses pratiques de gestion des données. Dans la poursuite de la discussion cherchant à clarifier ces pratiques, des remarques sont faites cependant pour se rassurer, dans l'intranquillité que sème le *matter of caring* qui s'attarde sur ces questions : les données ne sont pas si personnelles.

Les données ne sont pas si personnelles

Après avoir mentionné la facilité de réidentifier des données anonymisées et sa volonté de ne pas porter le blâme pour les personnes malintentionnées qui feraient cela (6^{ème} extrait), Félix poursuit :

- 88 Félix : Moi ce qui me rassure de notre côté c'est que les données qu'on demande à
89 notre monde ce sont des données qui sont pas – je vais te dire une chose
90 peut-être (.) mais – pas si personnelles que ça dans le sens où on te demande
91 pas ta carte de crédit, on te demande pas euh euh, le plus sensible qu'on te
92 demande c'est l'âge de tes enfants *right* ? Pour pouvoir te faire des bonnes
93 recommandations[d'activités mais euh on te demande pas nécessairement
- 94 Cléa : [te servir d'inspiration²¹³
- 95 Félix : Ouais c'est ça exact, mais on te demande pas nécessairement ton nom, on te
96 demande pas nécessairement ton adresse, t'sais, toutes les info vraiment *core*
97 qui une fois volées peuvent vraiment devenir problématiques à une personne
98 t'sais
- 99 Paul : Je sais pas où est-ce que tu vas, t'sais on peut te faire des recommandations
100 ou te dire qu'il y a des activités mais on sait pas forcément si tu vas les faire
101 ou non.
- 102 Félix : Par contre on te *track*, t'sais, dans ce sens que c'est géolocalisé, on va te faire
103 des recommandations – des recommandations d'inspiration autour de la zone
104 dans laquelle tu es.
- 105 Paul : Mais ça c'est toi qui nous auras donné ta géolocalisation, t'auras donné le
106 consentement là-dessus parce qu'on a besoin de ça pour pouvoir te dire ce
107 qu'il y a autour de toi, donc on a le consentement là-dessus
- 108 Félix : mais j'veux juste relativiser sur le le-le niveau de privauté des données
109 auxquelles on va avoir accès t'sais. Être dans une maison aussi, j'pense que
110 t'sais mettons le ASR qui capte ce qu'il y a dans une maison²¹⁴, c'est pas mal
111 plus intense qu'au téléphone aussi quand même, parce que quand t'utilises

²¹³ Cléa le reprend car il est convenu avec TMTL que l'agent conversationnel ne soit pas un algorithme de recommandations d'activités mais qu'il serve d'inspiration.

²¹⁴ Il fait référence à l'usage de l'Assistant Google via les enceintes connectées Google Home.

112 ton téléphone²¹⁵, mon sentiment c'est qu'il y a moins de personnes
113 impliquées que quand tu utilises un *device* dans un espace ouvert, dans une
114 maison avec une famille. Ceci dit, ben peut-être que tu peux capter quand
115 même des bruits d'enfants...

(7^{ème} extrait de retranscription de la section « *Is Google Assistant GDPR compliant?* », réunion du 23 septembre 2019)

Bien qu'il ait manifesté sa préoccupation, à l'écoute du podcast, à propos de la facilité avec laquelle des données personnelles anonymisées peuvent être réidentifiées même quand elles ne sont pas directement identificatoires, Félix tient à se rassurer – au risque de me froisser (« je vais te dire une chose peut-être (.) mais », l. 89-90) : ce qu'on demande à nos utilisateurs, ce ne sont pas des données si personnelles que cela (l. 90-91). Il liste des types de données sensibles, directement identificatoires, que Prélude ne demandera pas : le numéro de carte de crédit, le nom, l'adresse (l. 91-96). Cela est quelque peu contradictoire avec ce qu'il partageait du podcast juste avant : le problème n'est pas tant de collecter ou non ces données identificatoires, c'est plutôt la possibilité qu'à partir de données non-identificatoires, peu sensibles, les utilisateurs puissent être réidentifiés par la quantité de données qui permettent de reconstituer leur profil.

Paul appuie la liste de Félix : on ne sait pas non plus si l'utilisatrice aura effectivement fait l'activité qui lui sera suggérée ou non (l. 99-101). Cela amène l'attention sur la possibilité de repérer l'utilisatrice dans l'espace, géolocalisée (l. 102) – les coordonnées de géolocalisation étant des données personnelles quand elles sont liées à la personne suivie et que les coordonnées sont exactes²¹⁶. Prélude va utiliser la géolocalisation. Mais Paul mitige le risque que cela présente : l'utilisatrice aura donné son consentement pour cela, Prélude aura obtenu son consentement (l. 105-107). Le rappel du consentement semble ici transactionnel en même temps qu'il apaise le potentiel problème : il sera obtenu pour qu'il y ait collecte de donnée géolocalisée, les règles de soumission au consentement seront respectées pour avoir recours à ce type de données.

²¹⁵ Comme ce serait le cas pour la Voix de Montréal.

²¹⁶ Une stratégie pour brouiller l'identification de la personne est de la géolocaliser de façon vague, dans une zone plutôt que par un point précis. C'est ce pour quoi ont opté plusieurs applications de traçage anti-covid (Mondin & de Marcellis-Warin, 2020).

Félix résume vouloir « relativiser le niveau de privauté des données » (l. 108) en fonction des données qui pourraient être captées dans l'environnement de l'utilisatrice : l'agent conversationnel est pensé pour être utilisé sur un téléphone personnel, et non dans un espace privé, comme c'est le cas pour une enceinte connectée dans une maison où différentes personnes peuvent interagir avec, où être dans les environs (l. 109-115). Pourtant, dans l'expérience utilisateur, il est envisagé que l'agent conversationnel sur le téléphone soit utilisé dans des groupes, des familles, dans la rue, dans une chambre d'hôtel pour préparer son excursion : de nombreuses personnes peuvent donc être à proximité et certaines de leurs interactions peuvent être captées par erreur quand elles ont lieu quand l'agent conversationnel est activé, même si elles ne le concernent pas. Et de nombreux usages et mésusages imprévus auront lieu, ce qui amènera d'autres informations personnelles à potentiellement être collectées. Cette relativisation ne tient donc pas vraiment : de nombreuses données personnelles vont être collectées, même si elles ne sont pas si sensibles, qu'est-ce que Prélude prévoit de faire pour les protéger ? C'est ce point de départ que prend le *matter of caring* sur ces questions : plutôt que de diminuer le niveau de privauté des données, il s'agit de reconnaître leur présence et faire preuve de responsabilité pour respecter les utilisateurs auxquels ces données se rapportent.

Une deuxième réunion a lieu un mois plus tard, le 23 octobre 2019, pour clarifier ces pratiques de gestion des données personnelles. Cette fois-ci, Philippe (le développeur) est présent avec Paul et Félix. J'introduis la réunion en rappelant l'objectif :

1 Camille : donc oui, l'objectif de ce *meeting* c'est qu'on s'entende un peu tous sur où
2 on en est sur toutes ces questions [de gestion des données personnelles],
3 qu'on ait un panorama déjà concret sur la question de qu'est-ce qu'on va
4 collecter, qu'est-ce qu'on fait avec tout ça. Sachant que comme on travaille
5 avec Google Assistant, Google a accès à tout ce qui va passer par l'Assistant
6 donc c'est sûr que nous on peut pas régler ce que fait Google de son côté

7 Félix : Non

8 Camille : Mais j'ai vu hier que ça y est, ils étaient bien plus transparents sur ce qu'ils
9 font avec les données personnelles²¹⁷

²¹⁷ Voir Newman (2019) : <https://www.wired.com/story/google-assistant-human-transcription-privacy/>

10 Félix : Y'a eu des gros updates récemment qui nous poussent à croire qu'ils s'en vont
11 vers le GDPR compliant
12 [...]
13 Camille : Mais c'est juste que après, nous de notre côté, par rapport à ce qu'on va
14 garder de notre côté, faut s'assurer qu'on ait des bonnes pratiques.
15 Félix : ((*se lève avec son ordinateur et vient s'asseoir entre Philippe et moi*)) alors y'a
16 quelque chose qu'on peut regarder dans Dialogflow juste pour que tu
17 visualises bien à quoi on a accès comme données à travers Dialogflow, je sais
18 pas si c'est la première chose à faire mais euh parce que tu l'as peut-être pas
19 vu.

(1^{er} extrait de retranscription de la réunion « gestion des données »,
23 octobre 2019)

Le point de départ est ainsi de centrer la réunion sur la clarification des pratiques de gestion des données personnelles (l. 1-4), sans discuter à nouveau du cas potentiellement problématique de l'Assistant Google comme à la précédente réunion – nous avons suivi l'évolution de ce cas, qui pose à présent moins problème car Google a fait preuve d'une certaine transparence à ce propos (l. 8-12). Cette évolution du côté de Google ne dispense cependant pas Prélude de clarifier ses propres pratiques – « s'assurer qu'on ait des bonnes pratiques » (l. 14) de gestion des données personnelles permettra d'être conforme au RGPD, mais aussi de contribuer à offrir une bonne expérience aux utilisateurs, en respectant leurs informations. Félix intervient en perturbant l'ordre prévu de la discussion : il tient à me montrer à quel point les données sont anonymisées dans Dialogflow, le service de *Google Cloud Platform* qui permet de créer des agents conversationnels (l. 15-19).

À la précédente réunion, le 23 septembre, il m'avait précisé qu'ils développaient l'agent conversationnel avec le service Rasa, pour conserver une certaine indépendance par rapport à Google dans la création du bot. Mais depuis ce moment, ils se sont finalement réorientés vers Dialogflow : Rasa demande beaucoup de programmation de la part des développeurs, et il n'est pas utilisable par les non-développeurs comme Félix et Paul qui ont besoin d'intervenir régulièrement dedans. Dialogflow est plus facile à utiliser. Un compromis est donc fait sur l'effort d'indépendance par rapport à Google.

Félix me montre ce qu'il voit apparaître sur Dialogflow :

20 Félix : Donc ici tu vois y'a un *user* qui est venu, à 7h43AM puis qui a consulté l'agent,
 21 ok ? Et puis là il a dit « *moon inspiration* ». C'est moi en fait, *moon inspiration*,
 22 puis là y'a eu « *no match intent* » parce que je me suis trompé ((*rit*)), y'a pas
 23 de *moon inspiration*. Euh fait que là ici j'ai dit « *share your story* » puis là je
 24 faisais des tests en gros. Puis ça a duré 0 minute, ça a pas marché, ça ici *main*
 25 *menu*, *no match intent*, ça a pas marché hier à 10h36. Moi je suis en train de
 26 développer différentes fonctionnalités du robot là hier j'ai passé pas mal de
 27 temps là-dessus. Fait que je vois que ça a pas marché puis là je peux cliquer ici
 28 pour en savoir plus mais je peux pas savoir c'est qui ce *user*-là

29 Philippe : surtout que les données sont très, sont assez générales, ce sont pas des
 30 données personnelles, ça porte pas leurs propres problèmes ou

31 Félix : non non c'est ça. Fait que, en gros, bref, on a accès aux données anonymisées
 32 ok ? puis ça nous donne accès, on peut voir qu'est-ce qui marche qu'est-ce qui
 33 marche pas. [...]

34 Camille : ok

35 Félix: Fait que donc ça c'est[

36 Philippe: [et puis c'est clairement agrégé donc y'a pas,
 37 probablement 0 donnée personnelle

38 Félix: zéro donnée personnelle.

39 Camille : Ok, ouais ouais je comprends.

(2^{ème} extrait de retranscription de la réunion « gestion des données »,
 23 octobre 2019)

Félix décrit les actions de l'utilisateur (dont il ne voit pas le nom, mais il sait que c'est lui-même à ce moment) : à quelle heure il s'est connecté (à 7h43 AM et 10h36 la veille), ce qu'il a dit (« *moon inspiration* » et « *share your story* »), la réponse qui lui a été envoyé (« *no match intent* » : sa requête a échoué), la durée de la session (0 minute). Pour comprendre pourquoi la requête a échoué dans les deux cas, il peut cliquer sur un bouton « En savoir plus » qui documente le type d'erreur qu'il y a eu pour pouvoir la corriger (l. 20-28).

Philippe souligne que ce type de données qui apparaissent ne sont pas personnelles : elles sont « assez générales » et ne renseignent pas sur les problèmes des utilisateurs (l. 29-30). Les données sont anonymisées (l. 31) et « clairement agrégé[es] » (l. 36), donc il n'y a aucune donnée personnelle (l. 37). Je comprends leur démonstration, mais mentionne une limite : dans ce cas précis, il s'agit de tests où l'agent conversationnel

n'est pas utilisé en contexte réel par des personnes dans leur quotidien, les données sont anonymisées, certes. Mais « on a quand même accès à ce que disent les gens, non ? » (3^{ème} extrait ci-dessous, l. 40) Philippe et Félix me répondent en tentant à nouveau de diminuer la dimension personnelle de ces données :

- 40 Camille : Mais attends, on a quand même accès à ce que disent les gens non?
- 41 Philippe : Oui tout à fait
- 42 Félix : Oui, on a accès à ce qu'ils disent dans le robot oui, pas à ce qu'ils disent
43 en dehors du robot.
- 44 Philippe : Mais c'est sûr que dans le contexte du robot, c'est assez général, y'a pas
45 vraiment de, on pose pas de questions sur ses données personnelles donc[
- 46 Camille : [Non mais
- 47 Philippe : [c'est sûr que si on, si le sujet du domaine était très personnel là on aurait
48 des phrases qui pourraient peut-être permettre de-d'identifier les choses
49 mais, étant donné le contexte, c'est assez, à mon avis c'est assez [
- 50 Camille : [Mais
51 quand même ce qu'on a, moi ce que j'ai identifié qu'on avait comme
52 données personnelles ((montre la slide correspondante)) c'est des
53 informations, ben déjà d'authentification ?
- 54 Philippe: Ça dépend en fait
- 55 Camille : Genre le nom, le *user ID*, si on demande un *password* ou un PIN
- 56 Philippe : Quand on va créer le compte oui
- 57 Camille : Ça c'est des informations qu'on va avoir ?
- 58 Philippe : Tout à fait oui
- 59 Camille : Ensuite, dans les conversations elles-mêmes, on va avoir des traces de
60 goûts, de connaissances, juste en sachant où les gens vont cliquer, qu'est-
61 ce qu'ils vont vouloir faire comme activités, donc ça aussi ça nous permet
62 de faire un profil donc c'est de la donnée personnelle. Ensuite il y a des
63 informations démographiques puisqu'on leur demande quelle tranche
64 d'âge ils ont[
- 65 Philippe : [peut-être même que
- 66 Camille : [s'ils viennent en famille solo tout ça, peut-être qu'on peut
67 aussi deviner d'où ils viennent en fonction des expressions et tout
- 68 Philippe : peut-être que Google peut nous donner des informations aussi sur les
69 gens, possiblement si on fait le *linked account*, ouais

70 Camille : Après, on a aussi des informations de *tracking* avec la géolocalisation en
71 sachant aussi quel *device* ils vont utiliser, quelle adresse IP
72 Philippe : Ouais possiblement, ouais
73 Camille : Donc tout ça c'est plein d'informations qu'il faut-nous faut qu'on sache
74 qu'est-ce qu'on fait avec, et pourquoi on les collecte déjà, qu'on soit
75 transparents sur l'objectif de la collecte, avec qui on les partage et
76 comment on les sécurise.

(3^{ème} extrait de retranscription de la réunion « gestion des données »,
23 octobre 2019)

Plutôt que de m'en tenir aux remarques de Philippe qui diminue la dimension personnelle des données collectées, je tente d'explicitier les types de données qui seront collectées, pour montrer qu'il s'agit effectivement de données personnelles, bien que le contexte n'implique pas qu'elles soient particulièrement sensibles. J'ai préparé pour cela une diapositive avec un tableau récapitulant les types de données que j'ai identifiées, pour vérifier avec eux si cela va bien être le cas. Je passe au travers du tableau à voix haute : Prélude va collecter des données d'identification pour créer des comptes permettant de se souvenir des utilisateurs (l. 50-55); l'organisation va également connaître les intérêts, préférences et comportements des utilisateurs selon les conversations qu'ils auront, les choix d'activités qui les intéresseront (l. 59-62, l. 66); elle aura également accès à des informations démographiques (âge, origine, accent – ces informations ne seront pas forcément collectées explicitement mais elles seront présentes en filigrane dans les conversations (l. 63-67)); elle aura enfin accès à des informations de géolocalisation (l. 70-71). Récapituler ces types de données permet de réorienter le regard : il ne s'agit pas de se défendre de collecter des données personnelles, il s'agit de clarifier les pratiques par lesquelles l'organisation va gérer les données personnelles – sensibles ou non – avec lesquelles elle sera en contact (l. 73-76).

Les remarques qui cherchent à diminuer la dimension personnelle des données, celles également qui montrent leur forme anonymisées, semblent aller dans le sens de la relativisation que Félix voulait apporter à la précédente réunion par rapport au niveau de privauté des données. Ici, il s'agit de relativiser la dimension personnelle de ces données. Cela va dans le sens du mouvement de dépersonnalisation des données que Kate

Crawford dénonce dans son livre *Atlas of AI* (comme abordé dans le chapitre 3, p. 123) : plutôt que de prendre en compte la situation personnelle dans laquelle se passe un extrait de conversation collecté dans le cadre de ce bot, on l'écarte du raisonnement à propos des données personnelles, attaché au fait qu'elles sont anonymisées. L'anonymisation – et le contexte non-sensible, dans le cas de Prélude – sont associés à la dépersonnalisation des données, alors que la technologie conversationnelle en question a le potentiel de s'inscrire dans des situations quotidiennes et manières de parler tout à fait personnelles. Face à la diminution de la dimension personnelle des données, l'intervention de l'ordre du *matter of caring* s'attache ainsi à rappeler cette dimension personnelle, au nom du vécu personnel dans lequel l'utilisation (et la collecte de données associée) aura lieu.

Centré sur les personnes à respecter dans ces données, le *matter of caring* amène à orienter le reste de la discussion sur une autre clarification que je demande : « Comment on va permettre à l'utilisateur d'exercer ses droits de suppression, rectification, s'opposer au traitement et tout ça [...] Ça, ça implique plusieurs choses en terme techniques pour permettre ces fonctionnalités-là » (suite de la discussion du 3^{ème} extrait). Cette demande manifeste un *matter of caring* qui cherche à inciter la création des conditions pour que ces personnes soient respectées en pouvant faire quelque chose (exercer leurs droits). La discussion est tortueuse car, à ce stade-ci, ils ne savent pas encore bien à quelles informations et quelles fonctionnalités ils ont accès dans Dialogflow quand un utilisateur se crée un compte – ils n'ont pas encore intégré cette fonctionnalité de création de compte. Pour la demande de suppression des données par exemple, après avoir envisagé plusieurs options, Philippe me demande :

- 77 Philippe : C'est-tu bien clair pour toi Camille?
- 78 Camille : Ben pas vraiment, ce que je comprends, c'est qu'il n'y a pas encore de
79 *process* bien clair pour cette question-là.
- 80 Félix : C'est pas qu'y a pas de *process* bien clair c'est que juste on a établi qu'il va
81 y en avoir un mais il est pas encore intégré dans l'agent. Fait que le *process*
82 il est très clair : tu vas devoir te *log in* et puis une fois qu'on va avoir des
83 gens qui vont être *loggés in* on va pouvoir voir comment on va pouvoir le
84 faire, mais pour l'instant, il est pas intégré dans le bot donc la seule affaire
85 c'est que tant qu'on n'a pas le *log in* on peut pas répondre à ces questions-
86 là sur le comment.

(4^{ème} extrait de retranscription de la réunion « gestion des données »,
23 octobre 2019)

À ma remarque qui mentionne que c'est encore vague (l. 78-79), Félix revendique la clarté pour lui : on sait quelles seront les actions à faire du côté de l'utilisatrice, mais on ne sait pas comment (techniquement, du côté de Prélude) cela va se passer (l. 80-86). Il y a donc de l'incertitude. Cette incertitude concerne également la question demandant qui sera propriétaire des données collectées selon l'entente entre Prélude et TMTL, selon également l'arrimage avec l'écosystème technologique de TMTL pour permettre l'exercice des droits des utilisateurs partout où elles transiteront si elles sont partagées.

Ces incertitudes freinent ainsi l'objectif sous-jacent à cette réunion qui est de réunir les informations à propos des pratiques de traitement des données à indiquer dans le module de consentement pour que l'utilisatrice puisse en être éclairée. Alors que le sprint de cette semaine porte sur le design de conversation et inclut pour moi la tâche de finir le module de conversation sur le consentement, j'avais mentionné quelques jours avant cette réunion :

J'ouvre Jira pour voir un peu quoi faire. Je sais que c'est le sprint *conversation design* mais j'ai l'impression que sans la réunion sur la gestion des données personnelles je ne peux pas avancer grand-chose puisque j'ai besoin des infos qui y seront partagées pour savoir quoi dire dans le module de consentement (sur ce qui sera réellement fait avec les données niveau anonymisation et *process* pour supprimer etc.)

(Journal de bord, notes de terrain du 21 octobre 2019)

À l'issue de cette réunion, je n'ai pas beaucoup plus d'information, si ce n'est une vision plus claire sur les incertitudes qui empêchent pour l'instant d'avoir ces informations. Ces incertitudes amènent le module de consentement à rester incomplet jusqu'à la fin de mon immersion chez Prélude. Elles portent ainsi une vulnérabilité de ce module, qui ne peut pas se suffire en lui-même, quand le moment où il cherche des réponses pour se compléter n'est pas aligné avec la temporalité de l'organisation qui n'a pas encore ces réponses. Compléter ce module quand ces questions seront face à moins d'incertitudes fait ainsi partie des recommandations que je fais en quittant l'organisation, par le biais du rapport sur le *privacy-by-design* que je prépare à la lumière de ces constats (voir annexe, p. 365).

Le travail qui fait l'objet d'un *matter of caring* pour prendre soin des conditions (de traitement des données personnelles) dans lesquelles sera demandé le consentement se heurte ainsi à des incertitudes, car plusieurs fonctionnalités en lien avec des données personnelles ne sont pas intégrées. Avant de résoudre toute incertitude, les efforts de l'équipe sont pour l'instant concentrés sur une phase préliminaire : faire en sorte que le prototype de l'agent conversationnel soit fonctionnel d'ici le mois de décembre 2019, quand il sera montré et testé par l'équipe de TMTL. Plus le développement du prototype avance, plus il y a d'accrocs rencontrés : l'évolution de la voix ne se passe pas comme voulu et l'agent conversationnel ne marche pas lors d'un test utilisateur. Je me joins aux efforts de l'équipe pour réparer l'expérience, dans un mouvement de *caring* qui n'est pas orienté exclusivement sur (le consentement) de futurs utilisateurs mais qui s'ouvre sur une autre dimension rappelant l'humain dans le bot.

Réparer la voix quand elle ne fonctionne pas

Du mois de septembre au mois de novembre 2019, les évolutions de la Voix de Montréal se multiplient : le design de conversation change au rythme des rencontres entre Félix, Paul, Cléa et TMTL; le développement technique se fait d'abord sur Rasa puis Dialogflow, de tâche en tâche sur Jira, selon les fonctionnalités à intégrer, les formulations à changer, les bugs à corriger; des tests avec de potentiels utilisateurs confirment, nuancent et parfois contredisent les intuitions de design. Je vais montrer comment Prélude compose avec les contraintes de confidentialité finalement imposées par Google quand ils intègrent des fonctionnalités de création de compte et de géolocalisation. Puis, à l'occasion de tests avec de potentiels utilisateurs, je vais mettre l'accent sur l'arrimage des différentes voix qui s'assemblent pour réparer l'expérience de l'agent conversationnel quand celui-ci ne fonctionne pas.

Composer avec les contraintes

À partir de début novembre 2019, Kajal intègre la fonctionnalité de création de compte dans le prototype de la Voix de Montréal. Kajal est le développeur principal à présent, car Philippe est parti. La création de compte est une fonctionnalité de Google, *Google Sign-In*, qui permet de créer un compte ou de lier l'Action à un compte Google existant²¹⁸. Nous n'avons pas eu l'occasion d'échanger souvent, Kajal et moi, depuis son arrivée au mois d'août – il ne travaille pas à temps plein car il est à l'université en parallèle, en maîtrise en informatique, il travaille régulièrement de chez lui, et nous n'avons pas encore travaillé sur des tâches en commun. Le matin du 5 novembre, nous parlons un peu, je lui demande sur quoi il travaille. Il est en train d'installer le *Google Sign-in* : « *it asks the permission for name, email address, profile picture. We will know the user ID. But confidentiality is ensured by Google Sign-In and Dialogflow* » (Reconstitution des explications de Kajal prises dans mon journal de bord, notes de terrain du 5 novembre 2019). J'apprends donc à cette occasion l'intégration de cette fonctionnalité, les données personnelles que cela implique, et je comprends – sans entrer dans les détails – que des mesures de protection de la confidentialité sont mises en place par défaut par Google (via le *Sign-In* et *Dialogflow*).

Sans que je ne lui pose de question particulière à ce sujet, Kajal me mentionne une certaine sensibilité aux questions de respect de la vie privée : « *Privacy is important, we don't know when things can get creepy or worse. See right now with China in Hong-Kong: protesters wear masks to not be identified by AI face recognition*²¹⁹. *Or Cambridge Analytica, it was possible through a big data giant like Facebook* » (Reconstitution des paroles de Kajal prises dans mon journal de bord, notes de terrain du 5 novembre 2019). Il manifeste ainsi une certaine vigilance par rapport à ces enjeux, et semble faire confiance aux outils de Google et leurs mesures pour assurer la confidentialité des utilisateurs.

²¹⁸ Voir <https://developers.google.com/assistant/df-asdk/identity/google-sign-in>.

²¹⁹ Il fait référence aux manifestations à Hong-Kong en 2019, contre l'amendement de la loi d'extradition par le gouvernement de Hong-Kong, où les manifestants ont eu recours à plusieurs stratégies pour empêcher leur visage d'être reconnu par les technologies chinoises de surveillance par la reconnaissance faciale. Voir par exemple Mozur (2019) : <https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html>.

Il me montre la nouvelle version de la Voix de Montréal et je constate qu’il y a plusieurs nouvelles interactions qui relèvent du consentement – elles correspondent aux phases où le *Google Sign-In* s’active. Ces phrases s’immiscent par défaut dans le design de conversation, elles sont pré-écrites par Google et s’intègrent automatiquement dans la conversation, à l’endroit où la fonctionnalité est intégrée. Il se passe la même chose avec la fonctionnalité de géolocalisation qui est intégrée un peu plus tard. Cela amène des informations à se répéter, entre ce qui est dit dans le module de consentement, et ce qui est dit dans ces nouvelles intégrations – les affordances que l’on a préparées pour informer le consentement et celles que Google prévoit pour ces fonctionnalités se confrontent à une certaine redondance. Cela rend le module de consentement potentiellement plus vulnérable que ce qu’il est déjà : lui qui est fréquemment remanié pour moins heurter l’expérience, il est à présent confronté à ce que Google annonce finalement par défaut et qui risque de le rendre lourd à nouveau.

Cette prise en charge par défaut par Google, qui informe les utilisateurs au moment où ils s’apprêtent à utiliser des fonctionnalités permettant à Prélude d’obtenir davantage de données personnelles, est une bonne chose d’un point de vue de *matter of caring* : l’organisation dont on est si dépendant (par l’utilisation de son Assistant et sa console Dialogflow) prend au sérieux la protection de la vie privée des utilisateurs, en forçant les organisations qui utilisent ces outils pour créer des agents conversationnels à informer explicitement les utilisateurs quand ils ont recours à des fonctionnalités permettant la collecte de données personnelles directement identificatoires. En imposant le message d’information, le consentement ne peut pas être mis de côté. Ce qui est embêtant en revanche, c’est qu’il est impossible de modifier le ton et les formulations par lesquels sont données ces informations. Le ton est formel, froid, la voix est la voix synthétique féminine de Google par défaut. Aucune intervention n’est possible pour améliorer cela, pour rendre l’interaction plus agréable et mieux fusionnée avec la tonalité du reste de l’expérience. Cela limite le travail du *matter of caring* qui se soucie de faire parler autrement de consentement pour ne pas en perpétuer l’ennui ou la gêne face à tant de formalité.

Un peu plus tard dans la journée, j'entends Félix engager une discussion avec Kajal à propos de la création de compte via le *Google Sign-In*. L'avantage, c'est que cette fonctionnalité permet d'éviter de demander encore et encore son consentement à l'utilisatrice à chaque fois qu'elle se connecte : la création de compte permet de repérer que c'est une utilisatrice récurrente, et se souvenir des sessions passées. Quelques jours avant, j'ai entendu Félix dire à Kajal, en montrant quelque chose sur son écran (probablement le module de consentement) : « *This is for ethics, but ok, we ask them for their consent, I don't want to ask them over and over again because, for example, the thing about consent is boring* » (Journal de bord, notes de terrain du 29 octobre 2019). Le consentement vient donc encore comme une confrontation ennuyante dans ce que voudrait Félix pour l'expérience. La fonctionnalité du *Google sign-in*, en repérant les utilisateurs récurrents, peut permettre d'éviter d'exposer ces interactions « *over and over again* » (encore et encore) alors qu'elles sont ennuyantes, ou du moins redondantes.

Le problème, c'est qu'en optant pour le *Google Sign-In* par la voix, c'est la voix de Google et non la voix de la Voix de Montréal qu'on entend. Félix est agacé : ces enjeux de confidentialité impactent beaucoup l'expérience – il me jette un coup d'œil non-complice. Je m'immisce dans la conversation et suggère de dire quelque chose comme : « *I'll let Google tell you how to sign-in* », histoire d'introduire le changement de voix (Journal de bord, notes de terrain du 5 novembre 2019). C'est une petite tentative pour adoucir l'accroc forcé par la contrainte de Google qui impose sa voix. Sans signe d'acquiescement explicite, Félix semble envisager cette option, il ajoute cette phrase dans le document de design de conversation – elle n'est cependant pas maintenue par la suite.

La contrainte respectueuse de la vie privée imposée par Google vient ainsi à nouveau perturber la fusion cherchée de ces interactions liées au consentement dans l'expérience principale de l'agent conversationnel. Félix m'explique son agacement en fin de journée : « ça m'énerve cette contrainte de Google, ça m'embête pour l'expérience, on peut rien personnaliser. Y'a pas de créativité. » (Journal de bord, notes de terrain du 5 novembre 2019). Comme Félix, je ressens moi aussi une certaine frustration face à cette contrainte qui limite le soin que l'on peut accorder à cette fusion dans l'expérience. Cette

fusion participerait à nous faire davantage apprécier ce type d'interaction qui nous informe de ce qui est fait avec ce que l'on donne de nous. À la place, telle qu'elle est contrainte, la fusion qui se passe abruptement continue à nous ennuyer : cette information intervient de façon désagréable en interrompant nos conversations.

La contrainte a cependant un autre atout qui relève du *privacy-by-design* : pour éviter de heurter encore plus l'expérience par la redondance de ces informations de confidentialité à chaque fonctionnalité intégrée, celles-ci sont finalement positionnées de façon espacées, au moment où l'on a besoin d'elles. C'est le cas pour la création de compte (*Google Sign-In*) et pour la géolocalisation. Plutôt que d'activer ces deux fonctionnalités dès le début, celles-ci n'interviennent qu'aux moments où l'on a besoin d'elles – la création de compte, après l'introduction; la géolocalisation, plus tard, lorsque le bot suggère des activités et seulement si l'utilisatrice demande celles qui ont lieu autour d'elle. Cela permet de limiter, dans une certaine mesure, la quantité de données personnelles collectées, en cantonnant la collecte aux moments pertinents plutôt que sur toute la durée de l'utilisation. La composition avec la contrainte permet ainsi, dans ce cas, de satisfaire une limite que le *matter of caring* peut chercher à poser pour respecter les utilisateurs – plutôt que de privilégier l'appétit des organisations tentées de recueillir toute information du début à la fin, au cas où cela leur servirait.

Ces contraintes avec lesquelles composer ne sont pas les seules à rendre vulnérable l'expérience souhaitée en la modifiant et la confrontant. Devenir alpha testeur pour pouvoir utiliser le prototype demande plusieurs compromis, et faire tester l'agent conversationnel quand il ne marche pas bien demande quelques bricolages.

Tester l'agent conversationnel, bricoler avec nos voix

Pour pouvoir tester le prototype de la Voix de Montréal, il faut être enregistré comme « alpha testeur » par Paul et Félix dans la console qui permet de développer l'agent conversationnel. Cela permet d'avoir accès à la version alpha, le prototype qui n'est pas encore accessible publiquement sur l'Assistant Google. Paul m'envoie un lien sur Slack pour m'y donner accès le 1^{er} octobre 2019, car j'accompagne Malik le lendemain pour faire des tests auprès de potentiels utilisateurs, au centre info-touristique situé au

centre-ville de Montréal. J'ai besoin de tester moi-même l'agent conversationnel avant, car je devrai probablement faire quelques démonstrations le lendemain. L'accès est laborieux et me demande plusieurs concessions :

En prévision de la séance de demain, je demande accès à la version alpha de la Voix de Montréal pour pouvoir la tester et la faire tester. On me répond que j'y ai déjà accès, depuis mon compte Prélude [mon compte Google lié à mon adresse courriel professionnelle]. J'étais identifiée sur Google Assistant avec mon adresse Gmail personnelle, donc ça ne marchait pas. Je me connecte alors avec mon adresse Prélude. Un message apparaît en bleu qui me propose d'activer je-ne-sais-plus-quelles fonctionnalités, je dis non, pas envie de tout ouvrir, d'autant plus que c'est mon compte professionnel, sur mon téléphone personnel. Je dis « *talk to Bonjour Montréal* », ça ne marche pas, il ne comprend pas mon accent. Je change la langue de l'Assistant, je le fais passer de français à anglais. Il comprend un peu mieux quand je parle anglais mais me répond en français et tout le contenu s'affiche en français. Ça commence à m'agacer. Je laisse ça de côté.

Quand ils sortent de réunion, Paul me demande si j'ai réussi à y avoir accès. Je lui dis que j'ai un problème de langue, j'ai beau changer la langue de l'Assistant dans les paramètres, ça reste en français. Félix me dit qu'il faut changer la langue du téléphone. « Ah ouais, c'est intense dis donc ! » dis-je, un peu exaspérée de devoir faire autant d'efforts alors que l'outil capte par défaut une tonne de données sur moi déjà, il pourrait faire preuve de plus de souplesse en retour ! Je change la langue du téléphone. En effet, le contenu s'affiche maintenant en anglais dans l'Assistant. Mais il ne comprend toujours pas quand je dis *talk to Bonjour Montréal*.

Félix me fait redémarrer mon téléphone. Je rallume. Il le prend (avec mon accord), ouvre l'Assistant, accepte l'ouverture des fonctionnalités affichées dans le bandeau bleu que j'avais refusées, « Ah ouais carrément » dis-je. Il répond « Ben, je fais tout pour que ça ait une chance de fonctionner ». Ça demande décidément beaucoup d'efforts et de concessions. Finalement, Paul se souvient qu'il faut y accéder via le lien qu'il m'a envoyé sur Slack. C'est via ce lien que je vais pouvoir accepter d'être alpha testeur. Je clique sur le lien Slack depuis mon ordinateur. Paul me dit que je ferais mieux de le faire directement depuis mon téléphone. Pffff. Je vais sur Slack sur mon téléphone, clique sur le lien, *opt-in* pour être alpha testeur. Félix reprend mon téléphone, il me mentionne rapidement : « donc là tu acceptes qu'en étant alpha testeur on va avoir des infos sur ton utilisation », ok... J'ouvre l'Assistant, dis « *talk to Bonjour Montréal* », il ne comprend pas bien encore « bonjour Montréal ». « Ah ben non tu le dis avec l'accent français, faut dire avec l'accent américain ». Il parle à l'Assistant : « *talk to Bonjourw MonTreal* », la Voix de Montréal s'enclenche enfin. Après deux essais de mon côté, ça marche aussi pour moi.

(Journal de bord, notes de terrain du 1^{er} octobre 2019)

Pouvoir accéder au prototype de l'agent conversationnel, sur mon téléphone personnel, demande ainsi de nombreux compromis : me connecter à l'Assistant Google avec mon compte professionnel alors que je l'utilise en ce moment pour des affaires personnelles parfois, changer la langue du téléphone (ce qui affecte toutes les autres applications que j'ai dessus), accepter des conditions de collecte de données que je préfère habituellement refuser par défaut, laisser mon téléphone (très personnel) dans les mains de quelqu'un d'autre (Félix) avec qui je n'ai des relations que professionnelles, y accéder par l'application Slack sur le téléphone plutôt que sur l'ordinateur, tenter un accent américain plutôt que parler naturellement avec mon accent français. Devenir testeur demande en somme de consentir par défaut à mettre en place les conditions sur mon téléphone pour que le prototype ait une chance de fonctionner – je n'apprécie guère ces conditions, mais il n'y a pas d'alternative. Ce consentement forcé m'agace, mais je m'y plie – c'est aussi une façon d'entretenir mon accès à l'organisation et ses activités, et de pouvoir moi-même interagir avec ce prototype pour y participer et suivre son évolution. Je fais le compromis d'assouplir ma fermeture par défaut à ces petites intrusions dans la sphère personnelle de mon téléphone, pour tenter l'ouverture à cet agent conversationnel.

Le lendemain, avec Malik, nous rencontrons une quinzaine de personnes au fil de la journée au centre info-touristique. C'est principalement Malik qui fait tester l'agent conversationnel sur son téléphone, je pose quelques questions en soutien. Il décrit dans son mémoire de maîtrise ce qu'il faisait pour guider les personnes dans le test de la Voix de Montréal :

Concretely, this meant that I would introduce the prototype as “a conversational solution with whom we can discuss to discover more about the touristic offer of the city.” (Fieldnotes, 30/09/19) I would proceed to tell them that that we were going to perform the test on my smartphone, to which I would add: “You begin talking to it by saying ‘talk to Bonjour Montréal.’ It’s a demo version, so there’s only one path of questions and answers that really works. I’m going to guide you through it as you go. And then, I’m going to ask you questions about your experience.” (Ibid) I was trying not to give too much information while introducing the basics for users to be able to initiate their conversational experience. This implied that the voicebot was autonomous enough to hold a conversation by itself with an uninitiated user, as the true conversational agent I thought it was. However, even with all this information,

it proved to be hard in a lot of cases for users to interact autonomously with the demo.

(Dehbi-Talbot, 2020, p. 37)

Les personnes rencontrées sont des utilisateurs non-initiés à cet agent conversationnel mais également, pour la plupart, à l'Assistant Google ou autre assistant vocal de ce genre. Quelques personnes ont déjà utilisé l'Assistant Google au moins une fois, notamment deux dames d'une soixantaine d'années, qui l'utilisent fréquemment sur leur tablette et leur télévision. Mais plusieurs mentionnent aussi leur méfiance par rapport à ce type de technologie. Plusieurs accents ne sont pas bien compris par l'Assistant, amenant Malik à corriger des prononciations, un peu gêné, pour détourner le message d'erreur. Toutes les personnes gardent les yeux rivés sur l'écran pour lire ce qui s'affiche et souvent appuyer sur les boutons – cela contredit l'intuition de Prélude qui pensait jusqu'à présent que l'agent conversationnel serait utilisé exclusivement par la voix, le téléphone dans la poche. Après l'introduction, plusieurs personnes lèvent des yeux interrogateurs vers Malik pour savoir s'il faut répondre quelque chose ou attendre la suite – certaines font remarquer que cette introduction est trop longue. Malik intervient également régulièrement pour guider les autres actions possibles ou à faire. Toutes les remarques sont prises en compte par la suite pour faire évoluer le design de conversation. À la fin de la démonstration, plusieurs personnes demandent une carte de la ville à Malik, ou lui demandent comment se rendre à tel musée, ou encore s'il a des suggestions de choses à faire. Il prend le temps de répondre à chacune, comblant une information et une expérience relationnelle que l'agent conversationnel n'est pas encore capable d'offrir.

C'est encore plus le cas lors d'une autre (et dernière) séance de test, le 15 novembre 2019 – dans une auberge de jeunesse cette fois, pour rencontrer des personnes plus jeunes. Ce jour-là, le prototype ne fonctionne pas (voir chapitre 2, p. 94). Alors Malik bricole autrement :

On a fait les tests dans l'espace petit-déjeuner, au demi-sous-sol de l'auberge. Le buffet du petit-déjeuner est posé sur une grande table. On demande à deux jeunes femmes si on peut parler un peu avec elles. Elles sont Françaises, en voyage de deux mois au Canada.

Pour faire la démo, Malik utilise l'ancienne version. C'est celle où la voix de la jeune fille est mêlée à la voix synthétique : à chaque nouveau module ou chaque

changement que Félix a apporté dans les formulations, ce n'est plus la voix de la jeune fille qui les dit (il aurait fallu la réenregistrer en studio), c'est la voix Google par défaut. C'est d'ailleurs embêtant pour le module de consentement, dit par la voix synthétique, ça le rend encore moins agréable ! C'est un peu bizarre à entendre ces changements de voix, je me demande quels types de *feedbacks* on va recevoir étant donné que c'est difficile de comprendre ce qui se passe et d'avoir une impression agréable.

Ça bug assez souvent, et comme souvent, les accents ne sont pas reconnus, « *leisure* » est compris comme « *leather* » ou « *laser* » par exemple. On finit par leur recommander nous-mêmes des choses à faire.

On fait tester une voix mais notre voix est aussi très présente : elle pose des questions, s'excuse des bugs, trouve des alternatives pour donner un aperçu de l'expérience idéale et obtenir des *feedbacks*, elle bricole et elle est souriante. Ce qu'on présente est encore très artificiel, comme une pièce de puzzle pas bien coupée, on cherche à voir la meilleure façon de la couper, puis de l'insérer dans le puzzle général – leur expérience de voyage.

C'est ça aussi une partie de l'intelligence artificielle que l'on fait : aller à la rencontre des gens, car sans utilisation, pas de données, pas d'entraînement, pas d'effets, pas d'IA. On traverse la ville pour aller enquêter sur de possibles adoptions, mais l'écart est encore grand d'ici là, car découvrir ce bot n'est pas facile, et ces gestes pour parler à une technologie et en faire un assistant n'est pas encore une habitude pour beaucoup. L'IA se façonne dans plusieurs espaces, par une multiplicité d'outils et de voix.

(Journal de bord, notes de terrain du 15 novembre 2019)

La voix de Malik, et la mienne en soutien, se mêlent ainsi à celle de l'agent conversationnel qui ne fonctionne pas bien. Ce mélange des voix manifeste un soin qui tente de réparer les heurts dans l'expérience qui ne se déroule pas comme prévu.

Dans cet accompagnement par mélange de voix, c'est également la dimension relationnelle de l'agent conversationnel qui se tisse : la relation supposée entre une utilisatrice et cet agent n'est pas évidente, Malik se déplace dans l'espace et intervient dans l'interaction pour combler cette distance. Il répare également la déconnexion qui s'installe quand l'expérience ne se passe pas comme prévu :

“It makes me kind of frustrated to perform the demo to a participant and they end up not being able because “she does not work” right now. [...] In these situations, I feel a kind of disconnect between me and them. It's more a non-connection than a disconnect because we cannot establish a connection on the same grounds, the one that would premise the utilisation of our solution. In fact, I would say I feel extremely annoyed at the bot in these situations. I say at the bot because I know that it is not necessarily [the start-up's] fault if these things do not work. Being immersed for a

while in the organization has allowed me to understand how things are often very complex...” (Fieldnotes, 15/11/2019)

(Dehbi-Talbot, 2020, p. 46, citant ses notes de terrain)

Quand le bot ne fonctionne pas, ce n'est pas simplement lui qui fait preuve de vulnérabilité par sa défaillance, c'est aussi Malik, face à ces personnes avec lesquelles il est entré en relation pour établir une connexion autour d'une expérience partagée. Cette expérience se voudrait idéalement agréable mais elle est rendue inconfortable par la défaillance du bot. Malik doit alors trouver comment modifier ce qu'il voulait initialement montrer, partager et discuter. Il procède en quelque sorte à une fusion lui aussi : améliorer la fusion de l'expérience chaotique de l'agent conversationnel dans l'expérience de sa rencontre avec des inconnus. Dans cette fusion, quand l'agent conversationnel ne fonctionne pas, c'est l'expérience de celui-ci qui se perd en même temps qu'on en perd le contrôle, mais en retour, la voix, les gestes et explications de Malik viennent combler cet échec, et tisser une autre relation. La vulnérabilité de l'agent conversationnel amène ainsi à instaurer une relation d'altérité, qui ne cherche pas imiter des interactions entre humains comme le ferait le bot s'il fonctionnait bien (voir chapitre 2, p. 98), mais qui cherche à réparer une expérience de rencontre humaine, dans la complexité de ses inconforts.

C'est aussi cela le *matter of caring* : prendre soin des rencontres entre des personnes, sur un même pied d'égalité; que ces rencontres se fassent par le biais d'une technologie ou non; avec respect pour le temps qu'elles nous accordent, par notre petite intrusion dans leurs vies qu'elles acceptent; cela peut participer à un certain épanouissement – on ne peut pas savoir pour les autres ce qui est épanouissant pour eux, un voyage touristique peut l'être particulièrement. Le *matter of caring* est finalement une façon d'évoquer le soin accordé à entrer en relation avec humanité, c'est-à-dire avec respect des vulnérabilités qui nous lient les uns aux autres – y compris quand on se lie par ou autour d'une technologie.

Conclusion : le travail limité, partiel et modeste du *matter of caring*

J'ai exposé dans ce chapitre comment le consentement a été abordé chez Mr Y puis chez Prélude comme un *matter of caring* – malgré le changement de contexte d'activité de Prélude (tourisme) par rapport à Mr Y (santé mentale) où prendre soin du consentement, et de l'éthique de manière générale, relevait davantage de la nécessité étant donné la vulnérabilité dans laquelle auraient été les utilisateurs de Mr Y. Tandis que chez Mr Y, le travail du consentement s'inscrivait dans une même logique de *care* que l'expérience que le chatbot voulait offrir, celle-ci n'était plus d'actualité chez Prélude. Mais travailler le consentement chez Prélude s'est fait au nom d'une autre pertinence : contribuer à « changer la donne » (p. 277) en proposant un modèle de consentement différent de ce qui est habituellement fatigant. Cette contribution a pris la forme d'un module à intégrer dans le design de conversation de l'agent. L'intégration de ce module s'est cependant confrontée à plusieurs frustrations pour l'expérience utilisateur, tentée de rendre les interactions de consentement moins lourdes (p. 287), plus amusantes (p. 278), moins ennuyantes (p. 323).

À force de travailler ce module et ses formulations dans le but de le faire fusionner le plus fluidement et agréablement possible dans l'expérience de l'agent conversationnel, le souci porté sur le consentement – un souci de l'ordre du *matter of caring* – s'est mis à interroger sa validité. Cette validité repose en partie sur les affordances qui orientent la façon de dire « ok » – donc de donner son consentement. C'est ici qu'il y a un risque de *dark pattern* qui influencerait ou obscurcirait le consentement si l'affordance fait dire ok à autre chose qu'aux informations données. Mais pour permettre cette validité, il faut également prendre soin de ces informations, pour que celles-ci partagent de façon correcte et honnête comment sont traitées les données personnelles des utilisateurs, puisque c'est à ce traitement qu'on leur demande leur consentement. Par souci d'assurer des conditions qui permettent au consentement d'être valable, le travail sur celui-ci s'est lancé à la recherche de ces informations (en particulier lors des réunions du 23 septembre et 23 octobre 2019). Cette enquête a révélé plusieurs incertitudes face auxquelles le

travail sur le module de consentement ne peut pas aboutir pour l'instant : l'organisation ne sait pas encore clairement comment elle va partager ces données avec son partenaire, ni comment elle pourra permettre l'exercice des droits des utilisateurs selon les outils qu'elle utilise.

De l'incertitude, il y en avait également depuis le début dans ce travail sur le consentement : comment maintenir sa présence dans l'organisation quand celle-ci change de domaine ? Comment se rendre visible quand il est contenu dans des documents et des espaces à part ? Comment résister à l'effacement quand il est perçu comme lourd ou ennuyant ? C'est en s'inscrivant dans les logiques de Prélude que ce travail a pu persister et se faire une place dans le design de l'agent conversationnel pour avoir une chance de contribuer à changer des pratiques qui négligent le consentement – une contribution que vise le *matter of caring* qui oriente vers le respect des personnes amenées à utiliser ces technologies, par le partage de leurs données personnelles. Dans ces incertitudes et ces persistance, le *matter of caring* se prête à l'exercice vulnérable de fusionner dans le design de l'agent conversationnel – sans amoindrir sa vigilance envers la qualité de l'objet de son attention, les conditions de consentement. Ce qui est alors vulnérable, c'est la relation entre consentement et UX, parce que l'un et l'autre présentent le risque de se blesser mutuellement. En confrontant certaines intentions, en limitant certaines idées, en empêchant certains allègements, le *matter of caring* révèle également des vulnérabilités chez Prélude, en particulier à l'égard de ses pratiques de traitement des données.

Dans ces vulnérabilités, le *matter of caring* a finalement fait chez Prélude un travail similaire à celui que fait le consentement (voir chapitre 3) dans les limites qu'il pose et qu'on lui pose, dans le doute qu'il soulève quand il demande tant d'efforts (p. 265), dans les tentatives de le rassurer (« les données ne sont pas si personnelles », p. 311), dans les dynamiques de soumission qu'il révèle (« on n'a pas de contrôle », p. 301), dans l'activité qu'il invite à maintenir malgré la soumission – celle-ci n'empêche pas la responsabilité (p. 297). Il invite enfin à parler autrement de l'inconfortable que représente

habituellement le consentement, en proposant un certain ralentissement favorable au respect plutôt que de privilégier l'impatience de la rapidité (p. 292).

Plutôt que de travailler avec une certaine radicalité cherchant la pureté des pratiques à imposer, ce *matter of caring* compose avec des compromis. Il s'inscrit lui-même dans une certaine soumission (qui est souvent la mienne à l'égard de Prélude) : il veut faire bouger des pratiques, mais il cherche en même temps à entretenir un certain confort qui lui permet de garder accès à l'organisation et de s'inscrire dans ses activités. Alors il n'intervient pas sur certaines choses qui pourraient être améliorées – opter pour une voix qui ne reproduit pas des stéréotypes de genre, par exemple (voir p. 285). C'est que ces choses demandent une autorité qui ait suffisamment de poids face aux personnes pour qui ces choses font l'objet d'une préoccupation majeure – la direction artistique de la voix pour Félix, par exemple. Ce qui fait l'objet d'un *matter of caring* pour moi se soumet ainsi aux logiques dans lesquelles s'inscrire pour faire bouger certaines lignes de l'intérieur.

Le travail du *matter of caring* vise l'amélioration de situations négligées ou abîmées, dans un horizon de respect, d'égalité et d'épanouissement de toute personne touchée par ce que l'on développe. Ce travail est ainsi limité, ses compromis l'amènent à être partiel, mais il agit de façon adéquate et modeste dans un certain périmètre. « *Partial, finite, adequate, modest, limited – and yet worth working on, with, and for* » : à la fin du chapitre précédent (p. 243), je reprenais cette phrase d'Alexis Shotwell (2016, p. 5), inspirée de Donna Haraway (1987), pour marquer le point de départ du *matter of caring*, qui travaille à partir de notre situation limitée dans le monde. Cette situation n'est finalement pas qu'un point de départ : le périmètre dans lequel le *matter of caring* agit a beau s'élargir (du module et ses formulations, aux pratiques de gestion des données personnelles), son intervention rencontre les limites des incertitudes qui le dépassent, elle se confronte aux autres préoccupations qui comptent davantage aux yeux des autres. Mais elle vaut la peine d'être faite, car parler de consentement et porter attention aux conditions dans lesquelles il sera informé et demandé, cela amène à tisser dans l'organisation des pratiques de responsabilité, orientées vers le respect.

CONCLUSION

Retour sur les chapitres et leurs explorations

« Es-tu d'accord ? » Cette question que l'on croise quotidiennement sans nécessairement y accorder beaucoup d'attention est au cœur de l'exploration de cette thèse qui en révèle la complexité et les ambiguïtés. Pourquoi passer tant de temps sur cette question ? Pourquoi prendre soin du consentement ? Ce consentement peut être demandé dans de multiples domaines – juridique, médical, sexuel, politique, numérique – quand deux personnes se rencontrent, pour instaurer du respect et éviter que l'une ne blesse l'autre. C'est là la « *moral magic* » (Hurd, 1996, voir chapitre 3) que peut exercer le consentement : il rend acceptable un acte sur soi, il permet de donner accès à soi, et d'accéder à l'autre, sans blesser. Je me suis penchée sur ce moment particulier d'interaction, constitutif d'une relation respectueuse, dans un domaine particulier : celui de l'éthique de l'intelligence artificielle abordée par l'angle du design (chapitre 1) de systèmes d'IA spécifiques, les agents conversationnels (chapitre 2).

L'éthique de l'IA donne une autre raison de s'attarder sur le consentement : elle s'en préoccupe dans un souci de respecter un principe d'autonomie des utilisateurs, qu'ils aient plus de pouvoir à l'égard des données sur eux auxquelles ils donnent accès quand ils interagissent avec des systèmes numériques. Au nom de l'éthique de l'IA, ce consentement doit être libre, informé (ou éclairé); il doit s'accompagner de droits de consulter ce que des organisations ont sur nous, de les supprimer, de les transférer ailleurs. Ces droits des utilisateurs ont acquis une certaine autorité en devenant obligatoires à respecter selon le Règlement européen de protection des données (RGPD). C'est là une autre raison, plus contraignante, de s'occuper de ce moment de consentement : la loi le demande (chapitre 3).

Ces principes éthiques et règlements interfèrent ainsi dans les pratiques des organisations qui développent des systèmes d'IA, en imposant des contraintes qui impliquent de mettre en place des processus et créer des fonctionnalités pour prendre en compte ces impératifs. Par exemple, pour prendre en considération le respect de la vie

privée dès le début du développement, le *privacy-by-design* implique un ensemble de pratiques amenant la configuration du système et de l'organisation d'une certaine manière, qui anticipe les risques liés au traitement de données personnelles, plutôt que de s'en tenir à réagir quand un accident arrive (une fuite de données par exemple).

En travaillant sur le consentement chez Mr Y et Prélude, dans leurs activités de design de conversation par lesquelles se développent leurs agents conversationnels, ce sont en même temps de nombreuses autres interférences qui se sont révélées (chapitres 2 et 6) : les logiques à l'œuvre dans l'organisation pour concevoir une bonne expérience utilisateur (familiarité, rapidité, divertissement); les outils choisis pour développer l'agent conversationnel, des outils qui limitent le contrôle que l'organisation peut avoir sur les données personnelles qu'elle traite. Quand on plonge dans le travail de l'éthique par le design d'un système d'IA, on constate rapidement de nombreuses interférences qui empêchent de demander et donner un consentement dans les conditions parfaites de liberté et d'information complète que son principe éthique suppose. En retour, le travail sur le consentement gêne les autres activités par les ralentissements qu'il demande, les précisions sur des formulations, la place qu'il risque de prendre.

S'il y a tant de gênes et d'inconfort à travailler le consentement, alors pourquoi s'entêter à s'en occuper ? Ce n'est pas dans la tentative de réunir les conditions parfaites pour que ce consentement soit bon – c'est-à-dire valable – qu'il faut chercher une motivation (chapitre 3). Dans tous ses usages, pas seulement numérique, le consentement se heurte aux limites de ce qui peut l'influencer, limiter sa connaissance, le prendre pour acquis alors qu'il n'est pas vraiment exprimé. Quand on le considère seulement du point de vue de la personne qui peut le donner, les conditions sont rarement réunies pour que ce consentement soit parfaitement valable. Il est plus important d'élargir ce point de vue pour voir le consentement comme l'acte co-constitué qu'il est : une dynamique d'interaction entre deux personnes (ou entre une personne et une organisation), où l'acte de donner son accord dépend des façons dont il est demandé, dont il est possible de l'exprimer et de le retirer. La personne ou l'organisation qui le demande est responsable d'offrir les conditions qui permettent à ce consentement d'être

valable. Alors, si de l'attention est portée sur les conditions du consentement, c'est plutôt par l'angle de cette responsabilité partagée, qui met l'accent non sur la perfection des conditions à rassembler mais sur la façon dont elles favorisent l'expression du consentement. Ce qui compte, c'est que ces conditions soient orientées vers l'horizon d'un certain épanouissement, de respect et d'égalité. C'est cet horizon qui vaut la peine de s'entêter à s'occuper du consentement.

Vers cet horizon, j'ai proposé la notion de *matter of caring* (chapitre 4) pour évoquer cette attention portée aux conditions dans lesquelles une parole comme le consentement s'exprime. Il s'agit de faire attention à ce que ces conditions ne soient pas entravantes pour la personne à qui il est demandé de consentir. Cette attention portée sur les conditions résonne avec l'attention que François Cooren (2013) invite à porter sur tout ce qui s'exprime au travers de nous quand on parle ainsi que ce qui nous fait parler – par la métaphore de la ventriloquie, ces choses qui parlent et qui s'expriment au travers de nous sont des *figures*. Ce que le *matter of caring* amène à faire, par l'intervention qui l'anime, c'est de procéder à des reconfigurations quand ces figures qui font parler sont indésirables ou entravantes. Il s'agit par exemple d'intervenir pour interrompre une situation d'emprise – où quelqu'un peut consentir mais au nom de figures qui exercent de l'emprise sur elle (voir Adèle Haenel dans le chapitre 4) –, ou encore de revoir la disposition de boutons « ok » pour que ce ok soit donné à des informations sur le traitement de données personnelles plutôt qu'à la volonté de passer à la suite (voir chapitre 6). Une autre façon de reconfigurer les conditions dans lesquelles on demande le consentement est de reformuler les façons de le demander : ces reformulations peuvent chercher à rendre plus divertissantes les figures présentes dans le ton adopté (voir chapitre 6).

Les questions de recherche qui ont guidé mon exploration sont : (QR1) comment cela s'est fait, chez Mr Y et Prélude, d'aborder le consentement comme un *matter of caring* ? et (QR2) qu'est-ce que cela fait ? Ce travail d'intervention pour prendre soin des conditions (boutons, formulations, informations données), pour les reconfigurer de sorte qu'elles favorisent l'expression d'un consentement qui soit valable s'il est donné, est une

façon par laquelle a été abordé le consentement comme un *matter of caring* chez Mr Y et Prélude (QR1). Pour éviter également que le travail sur le consentement – sous forme de module, dans des documents séparés – soit cantonné à un espace à part, qui l’empêche de parvenir à se concrétiser dans le prototype de l’agent conversationnel, le *matter of caring* a également adopté les logiques et outils qui permettent de se rendre visible dans l’organisation : par la participation à un pitch, la création de tâches dans Jira, l’animation de réunions pour clarifier les pratiques de traitement de données personnelles à propos desquelles informer les utilisateurs dans le module de consentement (voir chapitre 6). Le *matter of caring* a également tissé une certaine complicité avec l’organisation, adoptant la préoccupation de ne pas heurter l’expérience utilisateur, la recherche de compromis pour proposer des formulations allégées, des formats plus discrets.

Ce *matter of caring* porté sur le consentement a finalement fait (QR2) ce que le consentement lui-même *fait* pragmatiquement : il a tracé des limites entre des préoccupations qui se rencontrent (celle du design d’une bonne expérience avec un agent conversationnel et celle d’une bonne expérience de consentement; celle de la limitation de ma collecte de données en ce qui concerne mes pratiques de recherche (chapitre 5)); il a mis de l’avant la valeur de respect des personnes présentes dans les données traitées; il a navigué entre passivité et activité, révélant des situations de soumission de Prélude en invitant à ne pas abandonner sa capacité d’agir dans cette soumission, en particulier en maintenant sa responsabilité. Moi-même, par souci d’incarner ce *matter of caring* porté sur le consentement, je me suis également mise dans la situation d’une certaine soumission au design d’agent conversationnel, du changement d’organisation de Mr Y à Prélude, des activités de Prélude éloignées du *care*. Cette soumission habite en quelque sorte la complicité avec laquelle je me suis tissée dans l’organisation, mais elle n’est pas exempte de capacité d’agir en direction d’un certain épanouissement (de mes connaissances, mes compétences, mon expérience avec des tensions, entre autres). Le *matter of caring* a ainsi entraîné dans l’organisation, son agent conversationnel, et moi-même comme chercheuse dedans un exercice vulnérable de « confrontation[s], (...) empêchements, (...) modifications du désir ou de la volonté initiale » (Brugère, 2013, citée

dans le chapitre 3, p. 176), entre plusieurs préoccupations – principalement celle portée sur le design d’expérience utilisateur, et celle portée sur la validité du consentement. Peuplé de compromis, ce travail s’est (in)achevé de façon limitée et partielle, suspendu aux incertitudes de l’organisation en matière de traitement des données personnelles. Il a cependant le mérite d’avoir œuvré à tisser le consentement dans le design de l’agent conversationnel, donnant une texture moins pure au travail de l’éthique dans la version idéalisée de ses principes, mais contribuant aux connaissances et compétences des membres de l’organisation pour favoriser des pratiques de responsabilité envers le respect.

Implications théoriques, méthodologiques et pratiques

Mon exploration a plusieurs implications. Théoriquement et empiriquement, en proposant une réflexion sur le consentement qui ne se cantonne pas à son usage numérique, bien que ce soit son objet, mais qui se lie aux autres usages du consentement, dont sexuel, elle participe à l’approche de la postphénoménologie (Ihde, 2007; Rosenberger & Verbeek, 2015; Verbeek, 2011) qui comprend la technologie comme une expérience parmi d’autres dans le monde. Il s’agit d’envisager les expériences que l’on a avec des systèmes numériques et d’intelligence artificielle comme faisant partie d’une expérience du monde, plutôt que de les aborder comme des expériences relevant d’un monde numérique qui serait en quelque sorte à part. Dans ce continuum, parler de consentement numérique implique de parler de consentement comme un concept qui fait pragmatiquement un travail de respect dans nos relations quotidiennes.

À ma connaissance, la plupart des travaux en postphénoménologie se penchent sur la relation humain-technologie qui émerge et évolue dans l’expérience du monde par le prisme des usages qui est fait de technologies une fois qu’elles sont développées. L’approche ethnographique, quand elle s’immerge dans les activités de conception de ces technologies comme je l’ai fait chez Mr Y et Prélude pour des agents conversationnels, permet d’aborder cette relation humain-technologie par un angle différent que celui des usages. Elle permet d’accompagner (Verbeek, 2011) le développement en interrogeant

les valeurs qui peuvent émerger au fil des pratiques de conception. C'est dans cette vigilance à ce qui émerge ou peut émerge que se situe l'intervention du *matter of caring*, qui tient à orienter dans une certaine direction ces émergences, bien qu'elles puissent lui échapper, de la même manière que beaucoup de choses échappent aux intentions des designers (voir chapitre 1). Dans le cas du consentement, le *matter of caring* cherche à orienter l'émergence de la valeur de respect – par le ralentissement qui permet de porter attention à ce moment, par le rappel de la présence des personnes dans les données même anonymisées, par une juste position des boutons qui permettent de dire « ok ». Il intervient dans le développement technologique pour faire en sorte que le système en développement participe à des relations respectueuses, pas seulement au travers de la technologie mais à toute relation dans le monde. En d'autres termes, le *matter of caring* expérimente avec les médiations technologiques, pour contribuer à la pratique du respect dans toute expérience dans le monde.

D'un point de vue pratique, le ralentissement présent dans cette expérimentation implique d'engendrer une tension puisqu'elle va à contre-courant du rythme de l'organisation. Mais ce ralentissement permet de gagner des connaissances et compétences sur des questions sensibles – comme la clarification de ce que sont des données personnelles et la responsabilité que l'organisation a d'en traiter. Aborder ces questions si tôt que lors du développement de la première phase du prototype permet d'anticiper des risques et des erreurs, qui entraîneraient des ralentissements (et d'éventuelles pénalités) plus importants par la suite. « *Hurry slowly* » : ce mantra aux apparences paradoxales m'a été suggéré par Boris Brummans pour guider l'écriture de cette thèse. Il se révèle finalement être le rythme que peut prendre le travail de l'éthique dans une organisation : se glisser dans la cadence de l'organisation mais (faire) ralentir par moments, par souci d'aborder des questions en prévention, plutôt que les laisser de côté au risque qu'elles reviennent de façon problématique par la suite. Le ralentissement n'est alors pas un rythme à contre-courant mais un rythme vigilant qui gagne à être revalorisé dans les organisations.

Cette expérimentation (ralentissante) avec les médiations technologiques est une façon sociomatérielle d'aborder le travail des *affordances* différemment de ce que fait Leonardi (2011). Il s'intéresse aux affordances selon les possibilités qu'elles limitent et les comportements qu'elles suggèrent (voir chapitre 1, p. 57). En travaillant les affordances liées au consentement – dont les formulations et positions des boutons « ok » – mon expérience de terrain montre les pratiques de design dont ces affordances font l'objet (voir chapitre 6). Les affordances ne sont pas tant un objet à observer que quelque chose sur quoi agir, pour déplacer ce qu'elles font ensuite faire aux utilisateurs. En même temps, quand il s'agit d'utiliser certains outils comme ceux de Google – Dialogflow, la fonctionnalité du *Google sign-in* par exemple – le travail du *matter of caring* permet de ne pas s'en tenir à se soumettre passivement aux limitations des affordances que ces outils nous imposent (du flou sur leurs pratiques de traitement des données personnelles (p. 297), un changement de voix (p. 323)), mais à composer avec elles en maintenant une certaine responsabilité (p. 297) envers ce que notre système technologique, qui dépend de ces outils et leurs affordances en matière de vie privée, va faire faire aux personnes qui l'utiliseront. Dans la relation humain-technologie qui se tisse au fil du développement, c'est donc une capacité d'agir de l'ordre d'une certaine soumission qui émerge.

En appliquant à ce contexte le concept philosophique de soumission par lequel Manon Garcia (2019, 2021) pense le consentement (dans son usage sexuel), ma réflexion reprend la sienne qui permet de penser la capacité d'agir – ou agentivité – inhérente à la soumission. Cela implique d'aborder autrement les relations où il y a rapport de domination : plutôt que de centrer le regard sur ce qui domine et ce qui est opprimé, il s'agit de porter attention à l'agentivité de ce qui se soumet. Cela porte la curiosité sur l'activité présente dans la soumission, plutôt que de la rendre synonyme d'une passivité qui l'amène à souvent être négligée. Il me semble que ce raisonnement sur la soumission gagne à se diffuser pour penser les rapports de pouvoir de la perspective des personnes qui en ont moins ou qui n'en ont pas.

C'est d'une certaine manière ce que j'ai invité à faire pour l'approche de la ventriloquie dans la recherche sur la communication constitutive de la réalité (CCR)

(Cooren, 2013) mais aussi pour l'étude de la matérialisation de l'éthique par une approche CCO (Matte & Bencherki, 2019) : plutôt que de se pencher exclusivement sur ce qui parvient à s'exprimer dans les tours de parole, peut-on porter attention à ce qui n'y parvient pas, en particulier sous forme de silence (voir chapitre 4) ? Dans ma propre recherche, cela s'est limité à exprimer mes propres silences tels qu'ils étaient dévoilés dans mes notes réflexives. Mais dans d'autres travaux qui anticiperaient de porter attention à de tels silences, et à ce qui s'exprime avec difficulté, maladresse, etc., des entrevues pourraient dévoiler certaines voix – seulement si la chercheuse prend soin des conditions dans lesquelles elle mène ces entrevues pour dévoiler ces voix (dans quel lieu, selon quelle position dans l'organisation, quel rapport entretient-elle avec les personnes plus dominantes, quelles limites met-elle au dévoilement public de ces paroles par la suite lors de son écriture et ses publications, etc.). L'étude de conditions dans lesquelles des voix s'expriment ou non pourrait faire l'objet de recherches également, par d'autres voies que la parole, comme le « *tracing* » qui dessine les trajectoires suivies par différentes personnes ou choses dans un même espace et au-delà (Brummans & Vézy, 2022) – cela permettrait d'avoir un certain accès à ce qui fait agir tel que des personnes ou d'autres êtres le font, sans mettre l'accent sur les figures qui s'expriment ou non mais sur les mouvements suivis, les orientations des corps. Cela pourrait peut-être également permettre de montrer des conditions de transformation, de renouvellement et de créativité que le silence peut favoriser (Bigo, 2018, p. 130). Cette invitation à porter attention aux silences entre les lignes et les moments des tours de parole rejoint celle de Boris Brummans, « *to be undiscouraged by moments of silence in organizational research* » (2022, p. 185) – sans limiter à la recherche organisationnelle cette façon de « *notice mindfully* » (titre du chapitre de Brummans, in press) qui peut s'appliquer à toute recherche.

Par ailleurs, au niveau empirique de l'éthique de l'intelligence artificielle, mon immersion de plusieurs mois dans un rôle mêlant éthique et recherche utilisateur, qui était encore marginal en 2019, permet d'aborder le travail de l'éthique autrement que par un angle normatif et prescriptif qui s'attache à édicter des principes et évaluer s'ils

sont respectés. Les limites rencontrées par le *matter of caring* sont du même ordre que celles que Metcalf, Moss et Boyd (2019) critiquent à propos de l'*ethics owning* : comme il travaille de l'intérieur de l'organisation en s'appuyant sur les logiques organisationnelles qu'il voudrait pourtant changer pour adresser des enjeux éthiques qui leur sont inhérents, le changement qu'il apporte est limité. Peut-être ce changement est-il encore plus limité dans les organisations plus grandes que celle à laquelle j'ai eu accès – petite startup en démarrage. Il serait intéressant de voir comment ce travail de *matter of caring* pour un enjeu comme le consentement, ou un autre enjeu éthique, peut être travaillé dans une organisation qui dispose d'un département juridique, d'équipes de recherche utilisateur et de design de produit, éventuellement aussi d'un centre de recherche qui travaille de façon multidisciplinaire des questions d'éthique de l'IA.

S'il est limité, ce travail de l'intérieur, par l'immersion ethnographique et la posture de participation observante, permet d'avoir accès à autre chose que du changement radical. L'immersion ethnographique permet en effet de révéler un travail de l'éthique en prise avec de nombreuses ambiguïtés, incertitudes et tensions qui donne à ce travail une dimension émotionnelle qu'il serait intéressant d'approfondir dans des recherches futures. Dès mon expérience avec la Déclaration de Montréal et l'animation de tables de discussion sur des enjeux éthiques, je me suis interrogée sur l'accent mis sur l'émotion de la peur et l'inquiétude pour susciter la réflexion éthique. Ces émotions ont continué d'être particulièrement présentes au début de mon terrain avec Mr Y et Prélude – crainte de ce à quoi je m'apprêtais à contribuer, crainte d'être utilisée pour de l'*ethics-washing*, crainte de voir l'enjeu du consentement disparaître quand l'organisation changeait. Elles ont ainsi influencé mon orientation au début du terrain. Qu'en aurait-il été si des émotions plus paisibles dominaient à ce moment ? La vigilance de l'éthique peut-elle s'accompagner d'autres affects que les émotions inquiètes de l'avenir ?

Esquisser l'horizon (éthique) du consentement – un horizon d'épanouissement, de respect et d'égalité – m'a permis d'orienter mes préoccupations éthiques vers quelque chose d'autre que ces émotions limitantes qui invitent au repli ou à la résignation. Il m'a permis de m'orienter vers quelque chose qui me dépasse, tout en donnant de la

motivation à l'action dans ce sens. C'est une motivation de cet ordre qui motive l'intervention inhérente au *matter of caring* : il invite à agir au nom d'un principe de respect et d'humanité²²⁰ plus grand que nous mais auquel des gestes et des façons de parler peuvent contribuer au quotidien. Plutôt que d'un principe de respect et d'humanité, il s'agit peut-être d'une pratique : ce n'est pas un idéal à atteindre mais quelque chose à faire au quotidien, un entraînement de notre écoute à ce qui s'exprime quand quelqu'un parle ou est silencieux, un rappel constant de notre responsabilité des conditions que l'on contribue à instaurer quand on entre en relation avec quelqu'un.

Cette dimension de pratique du *matter of caring* est en quelque sorte présente également dans la démarche réflexive que j'ai eue (voir chapitre 5) et qui implique, méthodologiquement, de ne pas isoler le concept à son champ théorique mais d'en faire l'expérience lors du terrain, de l'analyse et l'écriture. Cette façon d'incarner le concept exploré me semble particulièrement importante pour les études de l'éthique qui demandent une certaine intégrité, non seulement dans notre conduite en tant que chercheurs sur nos terrains – c'est ce dont veulent s'assurer les comités d'éthique – mais également dans ce qui suit, dans le travail d'analyse qui sélectionne quoi exposer, et l'écriture qui trace un chemin en cherchant comment le dévoiler. Cela demande également de s'entraîner à être « *undiscouraged* » (Brummans, 2022, cité-dessus) par les limites que cette intégrité porte à ce que l'on expose et comment on l'expose. C'est peut-être là aussi un défi du *matter of caring* quand on le pratique théoriquement, empiriquement et méthodologiquement : c'est un concept qui tisse ensemble ces différentes facettes, ce tissage se fait de compromis et limites, mais plutôt que de se laisser décevoir ou frustrer par ces compromis et limites, sa capacité de provoquer (doucement) un certain changement de l'intérieur gagne à être perçue avec optimisme.

La force de changement que le concept de *matter of caring* porte en lui peut s'appliquer à d'autres domaines et préoccupations – le changement climatique ou l'accueil des réfugiés par exemple. Il ne porte pas nécessairement sur le consentement

²²⁰ Le terme d'humanité n'est peut-être pas assez approprié, car idéalement il inclurait les autres-
qu'humains tout autant dignes de respect que les humains.

comme dans mon exploration, mais il s'applique à des situations auxquelles nous sommes amenés à consentir alors qu'elles peuvent être néfastes ou problématiques pour plusieurs personnes. Le *matter of caring* qualifie l'intervention pour faire changer ces situations de consentement et les améliorer selon une certaine mission : prendre soin de l'horizon vers lequel nous nous dirigeons.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., & Wilson, S. (2018). Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys*, 50(3), 1-41. <https://doi.org/10.1145/3054926>
- Aharon, D. (2019, août 27). Cloud Text-to-Speech expands its number of voices by nearly 70%, now covering 32 languages and variants. *Google Cloud Blog AI & Machine Learning*. <https://cloud.google.com/blog/products/ai-machine-learning/cloud-text-to-speech-expands-its-number-of-voices-now-covering-33-languages-and-variants>
- Ahmed, S. (2010). Orientations matter. In D. Coole & S. Frost (Éds.), *New Materialisms* (p. 234-257). Duke University Press. <https://doi.org/10.1215/9780822392996-011>
- AI Ethics Lab. (2020). Toolbox: Dynamics of AI principles. *AI Ethics Lab*. <https://aiethicslab.com/the-box/>
- Akrich, M. (1987). Comment décrire les objets techniques ? *Techniques & culture*, 54-55, 205-219. <https://doi.org/10.4000/tc.4999>
- Albu, O. B., & Flyverbom, M. (2019). Organizational transparency: Conceptualizations, conditions, and consequences. *Business & Society*, 58(2), 268-297. <https://doi.org/10.1177/0007650316659851>
- Allied Media Projects. (2017). *Building consentful tech*. <http://www.consentfultech.io/wp-content/uploads/2019/10/Building-Consentful-Tech.pdf>
- Almeida, F., & Monteiro, J. A. (2021). Exploring the effects of GDPR on the user experience. *Journal of Information Systems Engineering and Management*, 6(3), em0140. <https://doi.org/10.21601/jisem/10957>
- Amazon developers documentation. (s. d.-a). Design with Alexa Conversations. *Alexa Design Guide*. <https://developer.amazon.com/en-US/docs/alexa/alexa-design/alexa-conversations.html>
- Amazon developers documentation. (s. d.-b). Voice design for Alexa Experiences. *Alexa Design Guide*. <https://developer.amazon.com/en-US/docs/alexa/alexa-design/adaptable.html>
- Apple Developer. (s. d.). Siri. *Human interface guidelines*. <https://developer.apple.com/design/human-interface-guidelines/siri/overview/introduction/>
- Asdal, K., Brenna, B., & Moser, I. (Éds.). (2007). A history of STS. In *Technoscience: The politics of interventions* (p. 7-53). Unipub.

- Audureau, W. (2018, mars 22). *Ce qu'il faut savoir sur Cambridge Analytica, la société au cœur du scandale Facebook*.
https://www.lemonde.fr/pixels/article/2018/03/22/ce-qu-il-faut-savoir-sur-cambridge-analytica-la-societe-au-c-ur-du-scandale-facebook_5274804_4408996.html
- Austin, J. L. (2009). *How to do things with words : The William James lectures delivered at Harvard University in 1955* (J. O. Urmson, Éd.; 2. ed., [repr.]). Harvard Univ. Press.
- Baillargeon, D., Bencherki, N., & Boivin, G. (2021). 20 ans des approches de la communication constitutive des organisations : émergence et innovations. *Communication et organisation*, 59, 13-25.
<https://doi.org/10.4000/communicationorganisation.9723>
- Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.
- Beauvoir, S. de. (2013). *Pour une morale de l'ambiguïté. Suivi de Pyrrhus et Cinéas*. Gallimard.
- Bencherki, N., Cooren, F., Brummans, B. H. J. M., Benoit-Barné, C., & Matte, F. (2020). La culture en tant que cultivation : vers une conception communicationnelle de la culture organisationnelle1. *Communiquer. Revue de communication sociale et publique*, 29, 89-109. <https://doi.org/10.4000/communiquer.5674>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623.
<https://doi.org/10.1145/3442188.3445922>
- Benders, J., & van Veen, K. (2001). What's in a fashion? Interpretative viability and management fashions. *Organization*, 8(1), 33-53.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.
- Benoit-Barné, C., & Cooren, F. (2009). The accomplishment of authority through presentification: How authority is distributed among and negotiated by organizational members. *Management Communication Quarterly*, 23(1), 5-31.
<https://doi.org/10.1177/0893318909335414>
- Berlin, I. (1969). Two concepts of liberty. In *Four essays on liberty* (p. 118-172). Oxford University Press.
- Bigo, V. (2018). On silence, creativity and ethics in organization studies. *Organization Studies*, 39(1), 121-133. <https://doi.org/10.1177/0170840617717553>
- Boivin, G. (2017). *The institutionalization of an area of research through published and public discourse: The case of CCO scholarship*. Université de Montréal.

- Boivin, G., & Brummans, B. H. J. M. (2021). What's pragmatic about ambiguity in the communicative constitution of organizations? The case of CCO scholarship's establishment. In J. Basque, N. Bencherki, & T. Kuhn (Éds.), *The Routledge handbook of the communicative constitution of organizations* (Routledge).
- Boivin, G., Brummans, B. H. J. M., & Barker, J. R. (2017). The institutionalization of CCO scholarship: trends from 2000 to 2015. *Management Communication Quarterly*, 31(3), 331-355. <https://doi.org/10.1177/0893318916687396>
- boyd, danah. (2014). *It's complicated, the social lives of networked teens* (Yale University Press). <http://www.danah.org/books/ItsComplicated.pdf>
- Bronstein, M. (2020, janvier 7). A more helpful Google Assistant for your every day. *The Keyword*. <https://blog.google/products/assistant/ces-2020-google-assistant/>
- Brugère, F. (2013). Martha Nussbaum ou la démocratie des capacités. *La Vie des idées*. <https://laviedesidees.fr/Martha-Nussbaum-ou-la-democratie.html>
- Brummans, B. H. J. M. (2011). What goes down must come up: Communication as incarnation and transcension. *Communication and Critical/Cultural Studies*, 8(2), 194-200. <https://doi.org/10.1080/14791420.2011.567810>
- Brummans, B. H. J. M. (2015). "Turning the lens on ourselves" Bourdieu's reflexivity in practice. In A. Tatli, M. Özbilgin, & M. Karatas-Özkan (Éds.), *Pierre Bourdieu, organization, and management* (p. 70-94). <http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781317815808>
- Brummans, B. H. J. M. (2022). Eight ways to notice mindfully in process organization studies. In B. Simpson & L. Revsbæk (Eds.), *Doing process research in organizations: Noticing differently* (p. 173-194). Oxford University Press.
- Brummans, B. H. J. M., Cooren, F., Robichaud, D., & Taylor, J. R. (2014). Approaches to the communicative constitution of organizations. In L. Putnam & D. K. Mumby (Éds.), *The Sage handbook of organizational communication : Advances in theory, research and methods* (3. ed, p. 173-194). Sage.
- Brummans, B. H. J. M., & Vézy, C. (2022). Adventurous ideas for ethnographic research on the Communicative Constitution of Organizations. In J. Basque, N. Bencherki, & T. Kuhn (Éds.), *The Routledge handbook of the Communicative Constitution of Organization* (p. 262-280). Routledge.
- Callon, M., Akrich, M., & Latour, B. (1988). A quoi tient le succès des innovations – Premier épisode : l'art de l'intéressement. *Gérer et comprendre*, 11, 4-17.
- Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 Report*.
- Canada, & Innovation, sciences et développement économique Canada. (2019). *La charte numérique du Canada en action : Un plan par des Canadiens, pour les Canadiens*. [346](http://publications.gc.ca/collections/collection_2019/isde-ised/lu4-</p>
</div>
<div data-bbox=)

259-2019-fra.pdf

- Canca, C. (2020). Operationalizing AI ethics principles. *Communications of the ACM*, 63(12), 18-21. <https://doi.org/10.1145/3430368>
- Carter, I. (2021). Positive and negative liberty. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Spring 2022). <https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative>
- Cascaes Cardoso, M. (2017). The onboarding effect: Leveraging user engagement and retention in crowdsourcing platforms. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 263-267. <https://doi.org/10.1145/3027063.3027128>
- Cate, F. H., & Mayer-Schonberger, V. (2013). Notice and consent in a world of Big Data. *International Data Privacy Law*, 3(2), 67-73. <https://doi.org/10.1093/idpl/ipt005>
- Cavoukian, A. (2011). *Privacy by design. The 7 foundational principles*. 2. Information and Privacy Commissioner, Ontario, Canada. <https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf>
- Cavoukian, A. (2012). *Operationalizing privacy by design: A guide to implementing strong privacy practices*. Information and Privacy Commissioner, Ontario, Canada. <https://www.ipc.on.ca/wp-content/uploads/Resources/PbDReport.pdf>
- Cavoukian, A. (2019). *Privacy By Design [Module 2]*. <https://continuing.ryerson.ca/search/publicCourseSearchDetails.do?method=load&courseId=26829>
- Cavoukian, A., & Tapscott, D. (1997). *Who knows: Safeguarding your privacy in a networked world*. McGraw-Hill.
- Cecez-Kecmanovic, D., Galliers, R. D., Henfridsson, O., Newell, S., & Vidgen, R. (2014). The sociomateriality of information systems. *MIS Quarterly*, 38(3), 809-830. JSTOR. <https://doi.org/10.2307/26634999>
- Cercas Curry, A., & Rieser, V. (2018). #MeToo Alexa: How conversational systems respond to sexual harassment. *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*, 7-14. <https://doi.org/10.18653/v1/W18-0802>
- Charmaz, K. (2014). *Constructing grounded theory* (2nd edition). Sage.
- Chödrön, P. (2016). *When things fall apart: Heart advice for difficult times* (20th anniversary edition). Shambhala.
- Clarkson, J., Keates, S., Coleman, R., & Lebbon, C. (Éds.). (2003). *Inclusive Design*. Springer London. <https://doi.org/10.1007/978-1-4471-0001-0>
- Clegg, S. R., Pina e Cunha, M., Rego, A., & Dias, J. (2013). Mundane objects and the banality of evil: The sociomateriality of a death camp. *Journal of Management Inquiry*, 22(3), 325-340. <https://doi.org/10.1177/1056492612461949>
- Clément, É., Demonique, C., Hansen-Løve, L., & Kahn, P. (2000). *La philosophie de A à Z*

(Nouv. éd). Hatier.

- Commissariat à la protection de la vie privée du Canada. (2018, janvier). *Principes relatifs à l'équité dans le traitement de l'information de la LPRPDE*.
https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privée/lois-sur-la-protection-des-renseignements-personnels-au-canada/la-loi-sur-la-protection-des-renseignements-personnels-et-les-documents-electroniques-lprpde/p_principe/
- Conseil de recherche en sciences humaines (CRSH), Conseil de recherche en science naturelles et en génie du Canada (CRSNG), & Instituts de recherche en santé du Canada (IRSC). (2018). *Énoncé de politique des trois conseils, éthique de la recherche avec des êtres humains*.
http://publications.gc.ca/collections/collection_2019/irsc-cihr/RR4-2-2019-fra.pdf
- Conti, G., & Sobiesk, E. (2010). Malicious interface design: Exploiting the user. *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 271. <https://doi.org/10.1145/1772690.1772719>
- Cooren, F. (2010). *Action and agency in dialogue: Passion, incarnation and ventriloquism*. John Benjamins Pub. Co.
- Cooren, F. (2012). Communication Theory at the Center: Ventriloquism and the Communicative Constitution of Reality. *Journal of Communication*, 62(1), 1-20.
<https://doi.org/10.1111/j.1460-2466.2011.01622.x>
- Cooren, F. (2013). *Manières de faire parler : interaction et ventriloquie* (M. Chaput, Trad.). le Bord de l'eau.
- Cooren, F. (2016). Ethics for dummies: Ventriloquism and responsibility. *Atlantic Journal of Communication*, 24(1), 17-30. <https://doi.org/10.1080/15456870.2016.1113963>
- Cooren, F. (2018). Materializing communication: Making the case for a relational ontology. *Journal of Communication*, 68(2), 278-288.
<https://doi.org/10.1093/joc/jqx014>
- Cooren, F., Brummans, B. H. J. M., & Charrieras, D. (2008). The coproduction of organizational presence: A study of Médecins Sans Frontières in action. *Human Relations*, 61(10), 1339-1370. <https://doi.org/10.1177/0018726708095707>
- Cooren, F., Matte, F., Benoit-Barné, C., & Brummans, B. H. J. M. (2013). Communication as ventriloquism: A grounded-in-action approach to the study of organizational tensions. *Communication Monographs*, 80(3), 255-277.
<https://doi.org/10.1080/03637751.2013.788255>
- Cooren, F., & Sandler, S. (2014). Polyphony, ventriloquism, and constitution: In dialogue with Bakhtin. *Communication Theory*, 24(3), 225-244.
<https://doi.org/10.1111/comt.12041>
- Corbeil, V., Houde, J., Bernier, C., Bouchard, L., Gauthier, C.-A., & Lee-Gosselin, H. (2017). *La place des femmes dans l'industrie des TI au Québec* (Projet NA 13009).

- TECHNOCompétences. https://www.technocompetences.qc.ca/wp-content/uploads/2018/11/2017_EtudeFemmesenTI_TECHNOCompétences.pdf
- Costanza-Chock, S. (2020). *Design justice : Community-led practices to build the worlds we need*. The MIT Press.
- Crawford, K. (2016, juin 25). Artificial intelligence's white guy problem. *The New York Times*. <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>
- Crawford, K. (2021). *Atlas of AI : Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Sánchez, A. N., Raji, D., Rankin, J. L., Richardson, R., Schultz, J., West, S. M., & Whittaker, M. (2019). *AI Now 2019 Report*. AI Now Institute. https://ainowinstitute.org/AI_Now_2019_Report.html.
- Crosset, V. (2020). *Être visible sur et par internet : Le cas de l'État islamique* [Université de Montréal]. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/24778?locale-attribute=fr>
- Cuadra, A., Li, S., Lee, H., Cho, J., & Ju, W. (2021). My Bad! Repairing intelligent voice assistant errors improves interaction. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1-24. <https://doi.org/10.1145/3449101>
- Cunliffe, A. L. (2016). Republication of "On Becoming a critically reflexive practitioner". *Journal of Management Education*, 40(6), 747-768. <https://doi.org/10.1177/1052562916674465>
- Dasgupta, R. (2018). *Voice user interface design: Moving from GUI to mixed modal interaction* (1st ed. 2018). Apress. <https://doi.org/10.1007/978-1-4842-4125-7>
- De Angeli, A., & Brahmam, S. (2008). I hate you! Disinhibition with virtual partners. *Interacting with Computers*, 20(3), 302-310. <https://doi.org/10.1016/j.intcom.2008.02.004>
- De Angeli, A., Brahmam, S., & Wallis, P. (2005, septembre 12). *Abuse: The darker side of Human-Computer Interaction*. INTERACT 2005 workshop, Rome. http://www.agentabuse.org/Abuse_Workshop_WS5.pdf
- de la Bellacasa, M. P. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, 41(1), 85-106. <https://doi.org/10.1177/03063127110380301>
- de Boer, B., te Molder, H., & Verbeek, P.-P. (2020). Constituting 'visual attention': On the mediating role of brain stimulation and brain imaging technologies in neuroscientific practice. *Science as Culture*, 29(4), 503-523. <https://doi.org/10.1080/09505431.2019.1710739>
- Dehbi-Talbot, M. (2020). *Defining the Voice of Montréal: Exploring possibilities for*

human-machine companionship. Lund University.

- Deibel, D., & Evanhoe, R. (2021). *Conversations with things: UX design for chat and voice*. Rosenfeld Media.
- Dilhac, M.-A., Abrassart, C., Bengio, Y., Chicoisne, G., de Marcellis-Warin, N., Gambs, S., Gautrais, V., Gibert, M., Langlois, L., Laviolette, F., Lehoux, P., Maclure, J., Martel, M., Pineau, J., Railton, P., Régis, C., Tappolet, C., & Voarino, N. (2018). La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle. In *Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle* (p. 3-21). Université de Montréal. <https://www.declarationmontreal-iaresponsable.com>
- Duguay, S. (2022). *Personal but not private: Queer women, sexuality, and identity modulation on digital platforms* (1re éd.). Oxford University Press.
- Edenberg, E., & Jones, M. L. (2019). Analyzing the legal roots and moral core of digital consent. *New Media & Society*, 21(8), 1804-1823. <https://doi.org/10.1177/1461444819831321>
- Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed). The University of Chicago Press.
- Feenberg, A. (1999). *Questioning technology*. Routledge.
- Felt, U., Fouché, R., Miller, C. A., & Smith-Doer, L. (Éds.). (2017). *The handbook of science and technology studies* (Fourth edition). The MIT Press.
- Fenwick, T. (2014). Sociomateriality in medical practice and learning: Attuning to what matters. *Medical Education*, 48(1), 44-52. <https://doi.org/10.1111/medu.12295>
- Ferron Parayre, A. (2018). *Donner un consentement éclairé à un soin : Réalité ou fiction ? Exploration de l'effectivité du consentement éclairé aux soins*. Université de Montréal.
- Fessler, L. (2017, février 22). We tested bots like Siri and Alexa to see who would stand up to sexual harassment. *Quartz*. <https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/>
- Fessler, L. (2018, janvier 17). Amazon's Alexa is now a feminist, and she's sorry if that upsets you. *Quartz*. <https://qz.com/work/1180607/amazons-alexa-is-now-a-feminist-and-shes-sorry-if-that-upsets-you/>
- Fischer, B., & Tronto, J. C. (1991). Toward a feminist theory of care. In E. Abel & M. Nelson (Éds.), *Circles of Care : Work and Identity in Women's Lives* (State University of New York Press).
- Fischer, F. (2019). L'éthique by design du numérique : généalogie d'un concept. *Sciences du Design*, 10(2), 61-67. <https://doi.org/10.3917/sdd.010.0061>

- Fischer, F. (2020). *Les normativités des technologies numériques : approche d'une éthique « by design »* [Université de Technologie de Compiègne]. <https://tel.archives-ouvertes.fr/tel-03208802>
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3518482>
- Flusser, V. (2002). *Petite philosophie du design* (C. Maillard, Trad.).
- Friedman, B., Felten, E., & Millett, L. I. (2000). *Informed consent online: A conceptual model and design principles* (CSE Technical Report No 00-12-02). University of Washington, Department of Computer Science and Engineering. <https://dada.cs.washington.edu/research/tr/2000/12/UW-CSE-00-12-02.pdf>
- Friedman, B., & Hendry, D. (2019). *Value sensitive design: Shaping technology with moral imagination*. The MIT Press.
- Friedman, B., & Kahn, P. H. (2003). Human values, ethics, and design. In J. A. Jacko & A. Sears (Éds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (p. 1177-1201). Lawrence Erlbaum Associates.
- Friedman, B., Lin, P., & Miller, J. (2005). Informed consent by design. In L. F. Cranor & S. Garfinkel (Éds.), *Security and usability: Designing secure systems that people can use* (p. 495-521). O'Reilly.
- Friedman, B., Peter H. Kahn Jr., & Borning, A. (2013). Value sensitive design and information systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Éds.), *Early engagement and new technologies: Opening up the laboratory* (Vol. 16, p. 55-96). Springer Netherlands. <https://doi.org/10.1007/978-94-007-7844-3>
- Friedman, B., Smith, I., Kahn, P. H., Consolvo, S., & Selawski, J. (2006). Development of a privacy addendum for open source licenses: Value sensitive design in industry. In P. Dourish & A. Friday (Éds.), *UbiComp 2006 : Ubiquitous Computing* (Vol. 4206, p. 194-211). Springer Berlin Heidelberg. https://doi.org/10.1007/11853565_12
- Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, & OpenAI. (2017). *The malicious use of artificial intelligence*. <https://maliciousaireport.com/>
- GAML, & UNESCO. (2018). *A draft report on a global framework on digital literacy skills for indicator 4.4.2: Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills*. UNESCO. <http://gaml.cite.hku.hk/wp-content/uploads/2018/03/DLGF-draft-report-for-online-consultation-all-gaml.pdf>
- Garcia, M. (2018). *On ne naît pas soumise, on le devient*. Climats.
- Garcia, M. (2021). *La conversation des sexes : Philosophie du consentement*. Climats.

- Gebru, T. (2021, décembre 6). For truly ethical AI, its research must be independent from big tech. *The Guardian*.
<https://www.theguardian.com/commentisfree/2021/dec/06/google-silicon-valley-ai-timnit-gebru>
- Gherardi, S. (2016). Sociomateriality in Posthuman Practice Theory. In A. Hui, T. Schatzki, & E. Shove (Éds.), *The Nexus of Practices : Connections, constellations, practitioners* (1re éd., p. 38-51). Routledge.
<https://doi.org/10.4324/9781315560816>
- Gherardi, S., Murgia, A., Bellè, E., Miele, F., & Carreri, A. (2019). Tracking the sociomaterial traces of affect at the crossroads of affect and practice theories. *Qualitative Research in Organizations and Management: An International Journal*, 14(3), 295-316. <https://doi.org/10.1108/QROM-04-2018-1624>
- Gibert, M. (2020). *Faire la morale aux robots : Une introduction à l'éthique des algorithmes*. Atelier 10.
- Gilligan, C. (2003). *In a different voice: Psychological theory and women's development* (38. print). Harvard Univ. Press.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15-31. <https://doi.org/10.1177/1094428112452151>
- Giroux, H. (2006). 'It was such a handy term': Management fashions and pragmatic ambiguity. *Journal of Management Studies*, 43(6), 1227-1260.
<https://doi.org/10.1111/j.1467-6486.2006.00623.x>
- Google Cloud. (s. d.-a). Présentation de la console Dialogflow. *Dialogflow documentation*. <https://cloud.google.com/dialogflow/es/docs/console>
- Google Cloud. (s. d.-b). Voix standards et WaveNet. *Cloud Text-to-Speech*.
https://cloud.google.com/text-to-speech/docs/wavenet#wavenet_voices
- Google Developers. (2019, mai 7). *Intro to the Google Assistant: Build your first Action (Google I/O'19)*.
- Google Developers. (2021a). Errors. *Google Assistant Conversation Design*.
<https://developers.google.com/assistant/conversation-design/errors>
- Google Developers. (2021b, février 27). Create a persona. *Google Assistant Conversation Design*. <https://developers.google.com/assistant/conversation-design/create-a-persona>
- Google Developers. (2021c, février 27). *Learn about conversation design*.
<https://developers.google.com/assistant/conversation-design/learn-about-conversation>
- Google Developers. (2021d, février 27). What is conversation design? *Google Assistant Conversation Design*. <https://developers.google.com/assistant/conversation->

design/what-is-conversation-design

- Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-14. <https://doi.org/10.1145/3173574.3174108>
- Gray, C. M., Santos, C., Bielova, N., Toth, M., & Clifford, D. (2021). Dark patterns and the legal requirements of consent banners: An interaction criticism perspective. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-18. <https://doi.org/10.1145/3411764.3445779>
- Grondin, J. (2005). La fusion des horizons : la version gadamérienne de l'adæquatio rei et intellectus ? *Archives de Philosophie*, 68(3), 401. <https://doi.org/10.3917/aphi.683.0401>
- Groupe d'experts de haut niveau en intelligence artificielle. (2019). *Lignes directrices en matière d'éthique pour une IA digne de confiance*. Commission européenne. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Guay, M.-H., Prud'homme, L., & Dolbec, A. (2016). La recherche-action. In B. Gauthier & I. Bourgeois (Éds.), *Recherche sociale : De la problématique à la collecte des données* (6e édition, p. 539-577). Presses de l'Université du Québec.
- Gürses, S., & van Hoboken, J. (2018). Privacy after the Agile Turn. In E. Selinger, J. Polonetsky, & O. Tene (Éds.), *The Cambridge Handbook of Consumer Privacy* (1re éd., p. 579-601). Cambridge University Press. <https://doi.org/10.1017/9781316831960.032>
- Hagey, K., & Horwitz, J. (2021, septembre 15). Facebook tried to make its platform a healthier place. It got angrier instead. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215?mod=article_inline
- Hansmann, U., Merk, L., Nicklous, M. S., & Stober, M. (Éds.). (2003). *Pervasive computing: The mobile world* (2nd ed). Springer.
- Haraway, D. J. (1991). *Simians, cyborgs, and women: The reinvention of nature*. Routledge.
- Harding, S. (1992). Standpoint epistemology: What is « strong objectivity? » *The Centennial Review*, 36(3), 437-470.
- Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs*, 34(1), 31-55. <https://doi.org/10.1111/0029-4624.00201>
- Hasse, C. (2013). Artefacts that talk: Mediating technologies as multistable signs and tools. *Subjectivity*, 6(1), 79-100. <https://doi.org/10.1057/sub.2012.29>
- Hempel, J. (2017, juin 21). Inside Microsoft's AI comeback. *Wired*. <https://www.wired.com/story/inside-microsofts-ai-comeback/>
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering

- good decisions. *Perspectives on Psychological Science*, 12(6), 973-986.
<https://doi.org/10.1177/1745691617702496>
- Herveg, J. (Éd.). (2021). *Deep diving into data protection: 1979-2019, celebrating 40 years of research on privacy and data protection at the CRIDS*. Larcier.
- Human, S., & Cech, F. (2020). A human-centric perspective on digital consenting: The case of GAFAM. In A. Zimmermann, R. J. Howlett, & L. C. Jain (Éds.), *Human Centred Intelligent Systems* (Vol. 189, p. 139-159). Springer Singapore.
https://doi.org/10.1007/978-981-15-5784-2_12
- Hurd, H. M. (1996). The moral magic of consent. *Legal Theory*, 2(2), 121-146.
<https://doi.org/10.1017/S1352325200000434>
- Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.
- Ihde, D. (1999). Technology and prognostic predicaments. *AI & Society*, 13(1-2), 44-51.
<https://doi.org/10.1007/BF01205256>
- Ihde, D. (2007). *Listening and voice: Phenomenologies of sound* (2nd ed). State University of New York Press.
- Industrie Canada. (2014). *Responsabilité sociale des entreprises : Guide de mise en œuvre à l'intention des entreprises canadiennes*. Gouvernement du Canada.
[https://www.ic.gc.ca/eic/site/csr-rse.nsf/vwapj/ResponsabiliteSocialesdesEntreprises.pdf/\\$file/ResponsabiliteSocialesdesEntreprises.pdf](https://www.ic.gc.ca/eic/site/csr-rse.nsf/vwapj/ResponsabiliteSocialesdesEntreprises.pdf/$file/ResponsabiliteSocialesdesEntreprises.pdf)
- Ingold, T. (2011). *Being alive: Essays on movement, knowledge and description*. Routledge.
- Ingold, T. (2013). *Making: Anthropology, archaeology, art and architecture*. Routledge.
- Johnston, D. (2009). A history of consent in western thought. In F. Miller & A. Wertheimer (Éds.), *The Ethics of Consent* (p. 25-51). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195335149.001.0001>
- Jones, M. L., & Edenberg, E. (2020). Troubleshooting AI and consent. In M. D. Dubber, F. Pasquale, & S. Das (Éds.), *The Oxford Handbook of Ethics of AI* (p. 357-374). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.013.23>
- Jones, M. L., Kaufman, E., & Edenberg, E. (2018). AI and the ethics of automating consent. *IEEE Security & Privacy*, 16(3), 64-72.
<https://doi.org/10.1109/MSP.2018.2701155>
- Kant, I. (1993). *Fondements de la métaphysique des mœurs* (V. Delbos, Trad.; Le Livre de poche). Librairie Générale Française.
- Kim, N. S. (2019). *Consentability: Consent and its limits* (1re éd.). Cambridge University Press. <https://doi.org/10.1017/9781316691311>
- Kitzinger, C., & Frith, H. (1999). Just say no? The use of conversation analysis in

- developing a feminist perspective on sexual refusal. *Discourse & Society*, 10(3), 293-316. <https://doi.org/10.1177/0957926599010003002>
- Kleinig, J. (2009). The Nature of Consent. In F. Miller & A. Wertheimer (Éds.), *The Ethics of Consent* (p. 3-22). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195335149.001.0001>
- Kostova, O. (2019, janvier 28). Amazon Alexa Skills and Google Assistant Actions: The future of search. *Medium*. <https://medium.com/@odolenakostova/amazon-alexa-skills-and-google-assistant-actions-the-future-of-search-531573793643>
- Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, 30(2), 225-248. <https://doi.org/10.1086/421123>
- Lee, M. K., Kiesler, S., & Forlizzi, J. (2010). Receptionist or information kiosk: How do people talk with a robot? *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10*, 31.
<https://doi.org/10.1145/1718918.1718927>
- Lee, P. (2016, mars 25). Learning from Tay's introduction. *Official Microsoft Blog*.
<https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/>
- Leguil, C. (2021). *Céder n'est pas consentir*. Presses universitaires de France.
- Lenhart, A., & Owens, K. (2020). *Good Intentions, bad Inventions: The four myths of healthy tech*. Data & Society Research Institute.
<http://datasociety.net/pubs/Good-Intentions-Bad-Inventions.pdf>
- Leonardi. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, 35(1), 147. <https://doi.org/10.2307/23043493>
- Leonardi, P. M. (2012). *Car crashes without cars: Lessons about simulation technology and organizational change from automotive design*. MIT Press.
- Leonardi, P. M., & Barley, W. C. (2011). Materiality as organizational communication: Technology, intent, and delegation in the production of meaning. In T. Kuhn (Éd.), *Matters of communication: Political, cultural, and technological challenges to communication theorizing* (p. 101-122). Hampton Press.
- Leviathan, Y., & Matias, Y. (2018, mai 8). Google Duplex: An AI system for accomplishing real-world tasks over the phone. *Google AI Blog*.
<https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>
- Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881-900.
<https://doi.org/10.1177/1461444816675438>
- Luyat, M., & Regia-Corte, T. (2009). Les affordances : De James Jerome Gibson aux formalisations récentes du concept. *L'Année psychologique*, Vol. 109(2), 297-332.

<https://doi.org/10.3917/anpsy.092.0297>

- Machuletz, D., & Böhme, R. (2020). Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies*, 2020(2), 481-498. <https://doi.org/10.2478/popets-2020-0037>
- MacIntyre, A. C. (2012). *Après la vertu : Étude de théorie morale* (L. Bury, Trad.; 2e éd.). Presses universitaires de France.
- Mackenzie, C., & Stoljar, N. (Éds.). (2000). *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*. Oxford University Press.
- Mahieu, R., Asghari, H., Parsons, C., van Hoboken, J., Crete-Nishihata, M., Hilts, A., & Anstis, S. (2021). Measuring the Brussels effect through access requests: Has the European General Data Protection Regulation influenced the data protection rights of Canadian citizens? *Journal of Information Policy*, 11(1), 301-349. <https://doi.org/10.5325/jinfopoli.11.2021.0301>
- Marcus, G. E. (1997). The uses of complicity in the changing mise-en-scène of anthropological fieldwork. *Representations*, 59, 85-108. <https://doi.org/10.2307/2928816>
- Marcus, G. E. (2001). From rapport under erasure to theaters of complicit reflexivity. *Qualitative Inquiry*, 7(4), 519-528. <https://doi.org/10.1177/107780040100700408>
- Martine, T., & Cooren, F. (2016). A relational approach to materiality and organizing: The case of a creative idea. In L. Introna, D. Kavanagh, S. Kelly, W. Orlikowski, & S. Scott (Éds.), *Beyond Interpretivism? New Encounters with Technology and Organization* (Vol. 489, p. 143-166). Springer International Publishing. https://doi.org/10.1007/978-3-319-49733-4_9
- Matte, C., Bielova, N., & Santos, C. (2020). Do cookie banners respect my choice? : Measuring legal compliance of banners from IAB Europe's transparency and consent framework. *2020 IEEE Symposium on Security and Privacy (SP)*, 791-809. <https://doi.org/10.1109/SP40000.2020.00076>
- Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, 3(1), 205395171665021. <https://doi.org/10.1177/2053951716650211>
- Metcalf, J., Moss, E., & boyd, danah. (2019). Owing ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research: An International Quarterly*, 82(2), 449-476.
- Metz, C., & Wakabayashi, D. (2020, décembre 3). Google researcher says she was fired over paper highlighting bias in A.I. *The New York Times*. <https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html>
- Microsoft & Bing. (2019). *Voice report. From answers to action: Customer adoption of voice technology and digital assistants*. <https://advertiseonbing->

blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voicereport/bingads_2019_voicereport.pdf

Mill, J. S. (1990). *De la liberté*. Gallimard.

Millett, L. I., Friedman, B., & Felten, E. (2001). Cookies and Web browser design: Toward realizing informed consent online. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '01*, 46-52.
<https://doi.org/10.1145/365024.365034>

Mol, A. (2013). *Ce que soigner veut dire : Les patients, la vie quotidienne et les limites du choix* (M. Debauche & C. Debauche, Trad.).

Mondin, C., & de Marcellis-Warin, N. (2020). *Recension des solutions technologiques développées dans le monde afin de limiter la propagation de la COVID-19 et typologie des applications de traçage*. Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA). <https://observatoire-ia.ulaval.ca/recension-app-tracage/>

Moore, R. J., & Arar, R. (2018). Conversational UX design: An introduction. In R. Arar, R. J. Moore, G.-J. Ren, & M. H. Szymanski (Éds.), *Studies in Conversational UX Design* (1st ed. 2018, p. 1-16). Springer International Publishing: Imprint: Springer.
<https://doi.org/10.1007/978-3-319-95579-7>

Moore, S. L. (2010). *Ethics by design: Strategic thinking and planning for exemplary performance, responsible results, and societal accountability*. HRD Press.
<http://proquest.safaribooksonline.com/?fpi=9781599962016>

Mörch, C.-M., Gupta, A., & Mishara, B. L. (2019). *Canada Protocol: An ethical checklist for the use of Artificial Intelligence in suicide prevention and mental health*.
<https://doi.org/10.48550/ARXIV.1907.07493>

Morel, V. (2020). *Enhancing Transparency and consent in the Internet of things* [Cryptography and Security [cs.CR], Université de Lyon]. <https://hal.archives-ouvertes.fr/hal-01709255v2>

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141-2168.
<https://doi.org/10.1007/s11948-019-00165-5>

Mozur, P. (2019, juillet 26). In Hong Kong protests, faces become weapons. *The New York Times*. <https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html>

Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. MIT Press.

Nathues, E., van Vuuren, M., & Cooren, F. (2021). Speaking about vision, talking in the name of so much more: A methodological framework for ventriloquial analyses in organization studies. *Organization Studies*, 42(9), 1457-1476.

<https://doi.org/10.1177/0170840620934063>

- Newman, L. H. (2019, septembre 23). Google tightens its voice assistant rules amid privacy backlash. *Wired*. <https://www.wired.com/story/google-assistant-human-transcription-privacy/>
- Neyland, D. (2008). *Organizational ethnography*. SAGE Publications Ltd.
<https://doi.org/10.4135/9781849209526>
- Nickelsen, N. C. M., & Lydahl, D. (Éds.). (sous presse). *Ethical and methodological dilemmas in social science interventions: Careful engagements in healthcare, museums, design and beyond*. Springer.
- Nielsen, J. (1994). Usability inspection methods. In J. Nielsen & R. L. Mack (Éds.), *Usability inspection methods* (p. 25-64). Wiley.
- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Empowering People - CHI '90*, 249-256. <https://doi.org/10.1145/97243.97281>
- Nippert-Eng, C. E. (2010). *Islands of privacy*. The University of Chicago Press.
- Nissenbaum, H. (1998). The cutting edge. *ACM SIGCAS Computers and Society*, 28(1), 38-39. <https://doi.org/10.1145/277351.277359>
- Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law Books.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Nodder, C. (2013). *Evil by design: Interaction design to lead us into temptation*. Wiley.
- Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1), 100-126. <https://doi.org/10.1111/j.1745-6606.2006.00070.x>
- Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.
- Norman, D. A. (1999). Affordance, conventions, and design. *interactions*, 6(3), 38-43. <https://doi.org/10.1145/301153.301168>
- Norman, D. A., & Nielsen, J. (s. d.). The definition of User Experience (UX). *Nielsen Norman Group*. <https://www.nngroup.com/articles/definition-user-experience/>
- Norval, C., Janssen, H., Cobbe, J., & Singh, J. (2021). Data protection and tech startups: The need for attention, support, and scrutiny. *Policy & Internet*, 13(2), 278-299. <https://doi.org/10.1002/poi3.255>
- Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020). Dark patterns after the GDPR: scraping consent pop-ups and demonstrating their influence. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13. <https://doi.org/10.1145/3313831.3376321>

- Nussbaum, M. C. (2003). *Upheavals of thought: The intelligence of emotions*. Cambridge university press.
- Nussbaum, M. C. (2006). *Hiding from humanity: Disgust, shame, and the law* (Second printing; first paperback printing). Princeton Univ. Press.
- Nussbaum, M. C. (2008). *Femmes et développement humain : l'approche des capacités* (C. Chaplain, Trad.). Des Femmes.
- Nussbaum, M. C. (2011). *Les émotions démocratiques : comment former le citoyen du XXIe siècle ?* (S. Chavel, Trad.). Climats.
- Nussbaum, M. C. (2013). *Creating capabilities: The human development approach* (1. paperback ed). Belknap Press of Harvard Univ. Press.
- Office of the Privacy Commissioner of Canada. (2019, septembre 10). *Consent*. <https://www.priv.gc.ca/en/privacy-topics/collecting-personal-information/consent/>
- Office québécois de la langue française. (2019). Hypertrucage. In *Office québécois de la langue française*. https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26552557
- Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization science*, 11(4), 404-428.
- Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9), 1435-1448. <https://doi.org/10.1177/0170840607081138>
- Orlikowski, W. J., & Scott, S. V. (2014). What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector. *Organization Science*, 25(3), 868-891. <https://doi.org/10.1287/orsc.2013.0877>
- Orlikowski, W. J., & Scott, S. V. (2015). Exploring material-discursive practices. *Journal of Management Studies*, 52(5), 697-705. <https://doi.org/10.1111/joms.12114>
- Pearl, C. (2017). *Designing voice user interfaces: Principles of conversational experiences* (First edition). O'Reilly.
- Pentland, A. (2013). The data-driven society. *Scientific American*, 309(4), 78-83.
- Picard, R. W. (1997). *Affective computing*. MIT Press.
- Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, 4(1). <https://doi.org/10.1093/cybsec/tyy001>
- Pols, J. (2008). Which empirical research, whose ethics? Articulating ideals in long-term mental health care. In G. Widdershoven, T. Hope, J. McMillan, & L. van der Scheer (Éds.), *Empirical ethics in psychiatry* (p. 51-68). Oxford University Press.
- Pols, J. (2015). Towards an empirical ethics in care: Relations with technologies in health care. *Medicine, Health Care and Philosophy*, 18(1), 81-90.

<https://doi.org/10.1007/s11019-014-9582-9>

- Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-12. <https://doi.org/10.1145/3173574.3174214>
- Purtill, C. (2021, février 12). Hey, Alexa, are you sexist? *The New York Times*. <https://www.nytimes.com/2021/02/12/us/alexa-bots-female-voice.html>
- Putnam, L. L., & Boys, S. (2006). Revisiting metaphors of organizational communication. In S. Clegg (Éd.), *The Sage handbook of organization studies* (2nd ed, p. 541-576). Sage Publications.
- Putnam, L. L., Phillips, N., & Chapman, P. (1996). Metaphors of communication and organization. In S. Clegg, C. Hardy, & W. R. Nord (Éds.), *Handbook of organization studies* (p. 375-408). Sage Publications.
- Rehkopf, M. (s. d.). User stories avec des exemples et un modèle. *Atlassian Agile Coach*. <https://www.atlassian.com/fr/agile/project-management/user-stories>
- Richards, N., & Hartzog, W. (2019). The pathologies of digital consent. *Washington University Law Review*, 96(6), 1461-1504.
- Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Éds.), *Handbook of qualitative research*. 2nd ed. (p. 923-948). SAGE.
- Richardson, L., & St. Pierre, I. (2018). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Éds.), *Handbook of qualitative research*. Fifth edition (SAGE). SAGE.
- Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1), 3069. <https://doi.org/10.1038/s41467-019-10933-3>
- Rosenberger, R., & Verbeek, P.-P. (Éds.). (2015). *Postphenomenological investigations: Essays on human-technology relations*. Lexington Books.
- Ruffolo, R. (2008, mars). Ontario privacy czar applauds voice biometrics. *IT World Canada News*. *IT World Canada*. <https://www.itworldcanada.com/article/ontario-privacy-czar-applauds-voice-biometrics/2244>
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3E [Third edition]). SAGE.
- Sample, I. (2020, janvier 13). What are deepfakes – and how can you spot them? *The Guardian*. <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>
- Scheck, J., Purnell, N., & Horwitz, J. (2021, septembre 16). Facebook employees flag drug cartels and human traffickers. The company's response is weak, documents show. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-drug-cartels-human-traffickers-response-is-weak-documents-11631812953?mod=article_inline
- Schiffer, Z. (2020, décembre 7). Timnit Gebru's team at Google is going public with their

- side of the story. *The Verge*.
<https://www.theverge.com/2020/12/7/22158501/timnit-gebru-team-google-public-statement-fired>
- Schultze, U., van den Heuvel, G., & Niemimaa, M. (2020). Enacting accountability in IS research after the sociomaterial turn(ing). *Journal of the Association for Information Systems*, 21(4), 811-835. <https://doi.org/10.17705/1jais.00620>
- Schwaber, K., & Sutherland, J. (2020). *The Scrum guide. The definitive guide to Scrum: The rules of the game*. <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf>
- Semel, B. M. (2019). *Speech, signal, symptom: Machine listening and the remaking of psychiatric assessment* [Massachusetts Institute of Technology, Program in Science, Technology and Society]. <https://dspace.mit.edu/handle/1721.1/124064>
- Sen, A. (1995). *Inequality reexamined*. Oxford University Press.
<https://doi.org/10.1093/0198289286.001.0001>
- Shevat, A. (2017). *Designing bots: Creating conversational experiences* (First edition). O'Reilly.
- Shklar, J. N. (1989). The liberalism of fear. In N. L. Rosenblum (Éd.), *Liberalism and the Moral Life* (p. 21-38). Harvard University Press.
<https://doi.org/10.4159/harvard.9780674864443.c2>
- Shotwell, A. (2016). *Against purity: Living ethically in compromised times*. University of Minnesota Press.
- Spiekermann, S. (2012). The challenges of privacy by design. *Communications of the ACM*, 55(7), 38-40. <https://doi.org/10.1145/2209249.2209263>
- Spradley, J. P. (1979). *The ethnographic interview*. Holt, Rinehart and Winston.
- Stark, L. (2016). The emotional context of information privacy. *The Information Society*, 32(1), 14-27. <https://doi.org/10.1080/01972243.2015.1107167>
- Stark, L., & Hoffmann, A. L. (2019). Data is the new what? Popular metaphors & professional ethics in emerging data culture. *Journal of Cultural Analytics*.
<https://doi.org/10.22148/16.036>
- Statt, N. (2018, mai 10). Google now says controversial AI voice calling system will identify itself to humans. *The Verge*.
<https://www.theverge.com/2018/5/10/17342414/google-duplex-ai-assistant-voice-calling-identify-itself-update>
- Stengers, I. (2018). *Another science is possible: A manifesto for slow science* (S. Muecke, Trad.; English edition). Polity.
- Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques* (Sage Publications).
- Strengers, Y., & Kennedy, J. (2020). *The smart wife: Why Siri, Alexa, and other smart*

home devices need a feminist reboot. The MIT Press.

- Strengers, Y., Sadowski, J., Li, Z., Shimshak, A., & « Floyd » Mueller, F. (2021). What can HCI learn from sexual consent?: A feminist process of embodied consent for interactions with emerging technologies. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-13.
<https://doi.org/10.1145/3411764.3445107>
- Suchman, L. (2002). Located accountabilities in technology production. *Scandinavian Journal of Information Systems*, 14(2), 91-105.
- Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.
- Tarrant, J. (1999). *The light inside the dark: Zen, soul and the spiritual life*. Harper Perennial.
- Tarrant, J. (2008). *Bring me the rhinoceros: And other Zen koans that will save your life*. Shambhala.
- Taylor, J. R., & Cooren, F. (1997). What makes communication 'organizational'? : How the many voices of a collectivity become the one voice of an organization. *Journal of Pragmatics*, 27(4), 409-438. [https://doi.org/10.1016/S0378-2166\(96\)00044-6](https://doi.org/10.1016/S0378-2166(96)00044-6)
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness* (Rev. and expanded ed). Penguin Books.
- Tracy, S. J. (2019). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact* (Second edition). Wiley-Blackwell.
- Tronto, J. C. (2008). Du care. *Revue du MAUSS*, 32(2), 243.
<https://doi.org/10.3917/rdm.032.0243>
- Turchi, M. (2019, novembre 3). #MeToo dans le cinéma : l'actrice Adèle Haenel brise un nouveau tabou. *Mediapart*.
<https://www.mediapart.fr/journal/france/031119/metoo-dans-le-cinema-l-actrice-adele-haenel-brise-un-nouveau-tabou>
- UNESCO, & EQUALS Skills Coalition. (2020). *Je rougirais si je pouvais : réduire la fracture numérique entre les genres par l'éducation* (p. 145).
https://unesdoc.unesco.org/ark:/48223/pf0000367416_fre
- Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019). (Un)informed consent: Studying GDPR consent notices in the field. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 973-990.
<https://doi.org/10.1145/3319535.3354212>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 205630512090340.
<https://doi.org/10.1177/2056305120903408>

- Vakkuri, V., Kemell, K.-K., Kultanen, J., Siponen, M., & Abrahamsson, P. (2019). *Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study*. <https://doi.org/10.48550/ARXIV.1906.07946>
- Varela, F. J. (1996). *Invitation aux sciences cognitives* (P. Lavoie, Trad.). Éditions du Seuil.
- Vásquez, C. (2020). (In)habitée par le cancer : récits critiques des trajectoires affectives d'un terrain miné d'émotions. *Recherches qualitatives*, 39(2), 193. <https://doi.org/10.7202/1073515ar>
- Vásquez, C., Brummans, B. H. J. M., & Groleau, C. (2012). Notes from the field on organizational shadowing as framing. *Qualitative Research in Organizations and Management: An International Journal*, 7(2), 144-165. <https://doi.org/10.1108/17465641211253075>
- Vásquez, C., & Plourde, M. (2017). Materiality and organizing. In C. R. Scott, J. R. Barker, T. Kuhn, J. Keyton, P. K. Turner, & L. K. Lewis (Éds.), *The International Encyclopedia of Organizational Communication* (1re éd., p. 1-15). Wiley. <https://doi.org/10.1002/9781118955567.wbieoc129>
- Verbeek, P.-P. (2008). Obstetric ultrasound and the technological mediation of morality: A postphenomenological analysis. *Human Studies*, 31(1), 11-26. <https://doi.org/10.1007/s10746-007-9079-0>
- Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. The University of Chicago Press.
- Verbeek, P.-P. (2015). Toward a theory of technological mediation: A program for postphenomenological research. In J. K. Berg O. Friis & R. P. Crease (Éds.), *Technoscience and postphenomenology: The Manhattan papers* (p. 189-204). Lexington Books.
- Vézy, C., Brummans, B. H. J. M. (2021). La recherche CCO comme pratique réflexive: Une approche relationnelle. *Communication & Organisation*, 59, 141-153.
- Vézy, C., Martel, M., & Dilhac, M.-A. (2018). Les chantiers prioritaires et leurs recommandations pour le développement responsable de l'IA - Chantier littératie numérique. In *Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle* (p. 272-282). Université de Montréal. <https://www.declarationmontreal-iaresponsable.com>
- Vial, S. (2010). *Court traité du design* (Presses Universitaires de France). <https://www.cairn.info/court-traite-du-design--9782130627395.htm>
- Vial, S. (2021). *Le design* (3e éd. mise à jour). Que sais-je ?
- Vincent, J. (2016, mars 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>
- Vincent, J. (2019, juillet 11). Yep, human workers are listening to recordings from Google

- Assistant, too—The Verge. *The Verge*.
<https://www.theverge.com/2019/7/11/20690020/google-assistant-home-human-contractors-listening-recordings-vrt-nws>
- Vincent, J. (2021, avril). Google is poisoning its reputation with AI researchers. *The Verge*. <https://www.theverge.com/2021/4/13/22370158/google-ai-ethics-timnit-gebru-margaret-mitchell-firing-reputation>
- Voarino, N., Vézy, C., Crosset, V., & Zarzani, A. (2018). Rapport des résultats des ateliers de coconstruction de l'hiver. In *Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle* (p. 104-155). Université de Montréal. <https://www.declarationmontreal-iaresponsable.com>
- Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, 4(5), 193-220. <https://doi.org/10.2307/1321160>
- Weick, K. E. (1979). *The social psychology of organizing. Second Edition*. McGraw-Hill
- Weick, K. E., & Putnam, T. (2006). Organizing for mindfulness eastern wisdom and western knowledge. *Journal of Management Inquiry*, 15(3), 275-287.
<https://doi.org/10.1177/1056492606291202>
- Weinstock, D. (2014). *Profession, éthicien*. <http://books.openedition.org/pum/112>
- Wellner, G. (2014). The quasi-face of the cell phone: Rethinking alterity and screens. *Human Studies*, 37(3), 299-316. <https://doi.org/10.1007/s10746-013-9304-y>
- Wells, G., Horwitz, J., & Seetharaman, D. (2021, septembre 14). Facebook knows Instagram is toxic for teen girls, company documents show. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_lead_pos7&mod=article_inline
- West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI*. AI Now Institute. https://cdn.vox-cdn.com/uploads/chorus_asset/file/16125391/discriminating_systems_041519_2.pdf
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., & Schwartz, O. (2018). *AI Now Report 2018*. AI Now Institute, New York University.
https://ainowinstitute.org/AI_Now_2018_Report.pdf
- Wilhoit, E. D. (2018). Affordances as material communication: How the spatial environment communicates to organize cyclists in Copenhagen, Denmark. *Western Journal of Communication*, 82(2), 217-237.
<https://doi.org/10.1080/10570314.2017.1306098>
- Yanow, D. (2000). Moving from fieldwork and deskwork to textwork and beyond. In *Conducting Interpretive Policy Analysis*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781412983747>

Ybema, S., Yanow, D., Wels, H., & Kamsteeg, F. (Éds.). (2009). *Organizational ethnography: Studying the complexities of everyday life*. SAGE.

ANNEXE : rapport de *Privacy-by-design* réalisé pour Prélude²²¹

December 2019

Privacy-by-design (PbD) for voicebots: The case of the Voice of Montréal



Prototype of advertising of the Voice of Montréal, by Prélude and TMTL. October 2019

INTRODUCTION:

²²¹ Ce rapport a été évalué par Michelle Cibba, ancienne membre du Commissariat à la vie privée en Ontario, avec Ann Cavoukian qui était alors Commissaire. Elle enseigne avec Ann Cavoukian le cours en ligne de Privacy-by-design, à Ryerson University (devenue Toronto Metropolitan University), que j'ai suivi pendant l'automne 2019 : <https://www.torontomu.ca/pbdce/privacy-by-design-chang-school-course/>.

Voicebots – and voice technology or conversational AI in general – can be very intrusive since they need to have access to a lot of personal data, both from what the user feels like to discuss, and if the bot gives free rein to the discussion to make it more human-like. Moreover, it is easier and easier to create voicebots on virtual assistants such as Google Assistant or Amazon Echo – the voicebot will be an Action on Google Assistant or a Skill on Amazon Echo. By developing Actions and Skills companies and people other than Google or Amazon can extend the functionalities of these virtual assistants – they are “the equivalent of apps for mobile devices in the voice assistant world” (Kostova, 2019).

Prélude is a Montreal-based startup in Canada who is developing the Voice of Montréal²²², a tourist voicebot that aims to help tourists visit Montreal in an innovative way, both in the content and the user experience: the idea is to get creative personalized recommendations and directions directly into your ears, putting your phone in your pocket. The project is developed in partnership with TMTL, a non-profit organization dedicated to make people discover Montreal and its region.

The Voice of Montréal is in its early phase of development, so it is the perfect time to adopt a PbD approach to make sure this tool is developed in way that respect privacy by embedding it into its design.

This short report aims to present where Prélude is at in terms of doing privacy by design. I will start by presenting briefly the background from which Prélude comes – one that takes privacy by design seriously, as a legacy of a project in mental health. Then, I am going to conduct a privacy by design analysis to show what Prélude has started doing to embed privacy into the design of its first voicebot. From there, I will highlight the main gaps I have identified so far – they are about data minimization and data sharing. I will conclude by summarizing the main recommendations to nurture privacy by design as an ongoing process for the next phases of the product development and launch.

When relevant, I will precise between parenthesis the PbD principle that applies. Please note that the source is the PbD online course by Ryerson University, based on Ann

²²² Here is the press release: <https://blog.mtl.org/en/voice-montreal>

Cavoukian and her colleagues' work as the Information and Privacy Commissioner of Ontario (Cavoukian, 2012). These principles are listed at the end of the document.

BACKGROUND:

Principle 1: “Proactive not reactive; preventive not remedial”

Before working on the Voice of Montréal, one of Prélude's co-founders developed a chatbot to help people cope with anxiety. Since this project was in the context of mental health, he has been aware of the importance of handling privacy with great care before creating Prélude. This previous project serves as organizational memory that informs the development of Prélude as a young company. The co-founders want to design their products by embedding privacy into the design (Principle 3).

The Voice of Montréal is in its early phase of development in a tourist context – the alpha version is expected to be released on December 19th, 2019. It doesn't have other users yet than Prélude's employees who test the bot on a regular basis. However, they are willing to be proactive to commit to a strong privacy program from the earliest stage of the product. This willingness is translated into different actions:

- From a legal point of view, their main client and partner for this project – TMTL – has the requirement to deal with service providers who are compliant with the General Data Protection Regulation (GDPR) in the European Union. That is, Prélude – as one of TMTL's service providers – has to comply with the strong privacy program of the GDPR, from data protection to allowing users to exercise their digital rights such as the right to be informed about data processing, the right to access the personal information companies have on them, the right to rectify these personal data, the right of erasing them, the right to restrict processing, the right to data portability, the right to object to the processing (GDPR, ch. 3).
- Before this imperative from their client, Prélude was going into this direction by hiring a doctoral researcher (me, Camille Vézy) who is working on designing

- consent through the conversational agent so that the user is informed about her data processing. This work is strongly linked to the GDPR and the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) incentives to reinvent consent models (Office of the Privacy Commissioner of Canada, 2019).
- Prélude follows a Scrum²²³ approach of teamwork that implies breaking down the product development into different steps (sprints) where specific tasks must be achieved in order to complete the sprint and go to the next step. Privacy tasks are implemented into the weekly sprints, both for the user experience (UX) and the tech teams. This allows privacy issues to be addressed in the same rhythm as the product development rather than stay on the side.

In those ways, Prélude demonstrates a proactive approach to tackle privacy issues from the beginning of the product development rather than postpone it to a more critical time where they would have to react to a possible breach.

PRIVACY BY DESIGN ANALYSIS

A win-win manner of respecting user privacy (Principles 4 and 7)

This proactive approach is also paired **with a “positive sum, not zero sum” conception of privacy (Principle 4)**. Indeed, being preventive about privacy risks is not only a constraint for compliance but also a matter of fostering trust from users who appear to have privacy concerns about virtual assistants, which is impacting adoption (Microsoft & Bing, 2019, p. 24) :

²²³ <https://www.scrum.org/resources/what-is-scrum>

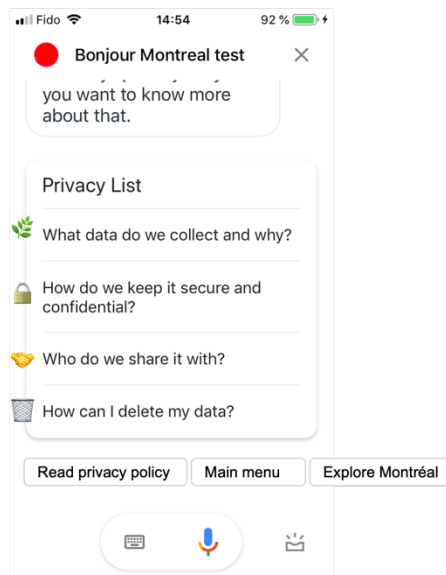
What concerns do people have about digital assistants?

That my personal information or data is not secure	52%
I don't know how my personal information is being used	24%
I don't want my personal information or data used	36%
That it is actively listening and/or recording me	41%
That the information it gathers is not private	31%
I do not trust the companies behind the voice assistant	14%
Other	2%

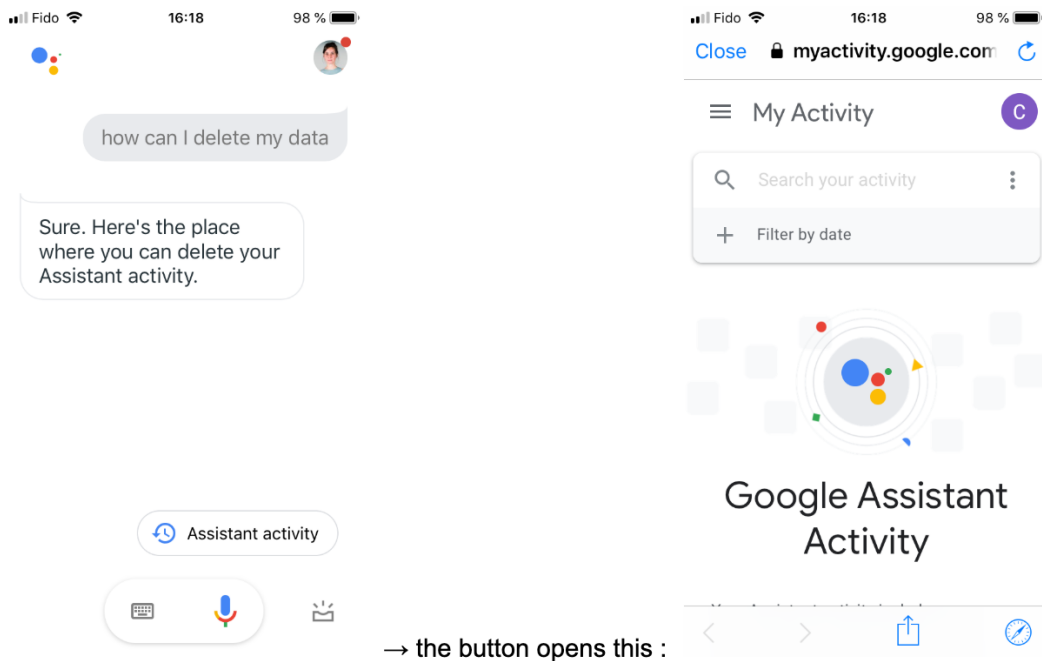
Voice Report 2019 24

Prélude is aware of that and tries to be transparent about their data management practices informing the user before she gets personal with the voicebot.

Indeed, in order to communicate clearly to the user the content of the privacy policy, Prélude is implementing a module composed of typical privacy questions that the user may ask: what data does the Voice of Montréal collect, is it shared, how confidentiality is ensured, and how the users can delete this data. These questions are presented in a list format that the user can navigate by voice and by touching the suggested questions and buttons:



This way, information about the policy is easily accessible and readable both from the user interface and the tone and vocabulary that are easier to understand “to make informed decisions about their personal data” rather than relying on the “notice and consent” model of consent where users “are presented with long and complex privacy notices routinely written by lawyers for lawyers” (Cate & Mayer-Schonberger, 2013, p. 67). What are also made readily available and easily accessible are the “procedures and controls relating to the management of personal information” (PbD online course, Module 2, principle 6). Some control is indeed given to the user to manage her personal information. This control is mostly native to the use of Google Assistant itself. For example, to delete her personal information such as what she has just said, the user can go to the “myactivity” page²²⁴ to see her past interactions with the Google Assistant and delete what she wants. All what she said to the Google Assistant is here, including her interactions with an Action such as the Voice of Montréal.



Screenshot of interactions on Google Assistant to access My Activity on Google. November 20, 2019.

²²⁴ <https://myactivity.google.com/myactivity>

However, what is not clear yet is how this deletion process will be achieved in the backend²²⁵ of the Voice of Montréal – how Prélude and all the stakeholders involved in this voicebot’s data management.

Privacy as the default setting: a work in progress

The processes for managing personal information are not crystal clear yet for various reasons: (1) the product is in its early phase of development so the architecture is still in progress and there are no users yet so Prélude doesn’t have a clear picture of what information will indeed be collected, and (2) the product is developed in partnership between two organizations who should set out clearly the terms for data sharing.

- (1) If the picture is not complete yet because there are no users at the moment (except Prélude’s members who test the bot), we can still identify 4 types of personal information that are going to be collected through the Voice of Montréal:
- The conversations: the audio of what the user says (the voice recording) is processed by Google Assistant to turn the voice into text according to the automatic speech recognition of the Google Assistant. Prélude doesn’t have access to the voices but only the text transcriptions.
 - Sign in information via Google: when the user chooses to create an account so that the bot can remember her preferences and where she left the conversation previously, Prélude gets the user’s name, user ID, email address and profile picture from Google.
 - Information about the stay: When she is going to be in Montreal, for how long, for what purpose (leisure or business); what is her preferred itinerary; the type of

²²⁵ Comment from Michelle Cibba, former member of the Office of the Privacy Commissioner (and teacher of the PbD online course at Ryerson University): “This is a considerable challenge that is a common one so Prélude should not feel badly - perhaps there are other approaches in the meantime to address this requirement (e.g. de-identification, separating identifiers)”

- group she is with (solo, couple, family, group); the “mood” she chooses to explore content (visit, play, relax, go out, eat & drink, shop).
- Geolocation: if the user selects the “around me” option to explore suggested activities, Prélude will have access to her geolocation.

Prélude looks for **minimizing the personal information** they collect by choosing not to have access to the users’ audio recordings – this can be done via Dialogflow²²⁶, the Google tool linked to the Google Assistant to create the infrastructure of the bot. This way, Prélude can avoid dealing with biometric information (Ruffolo, 2008).

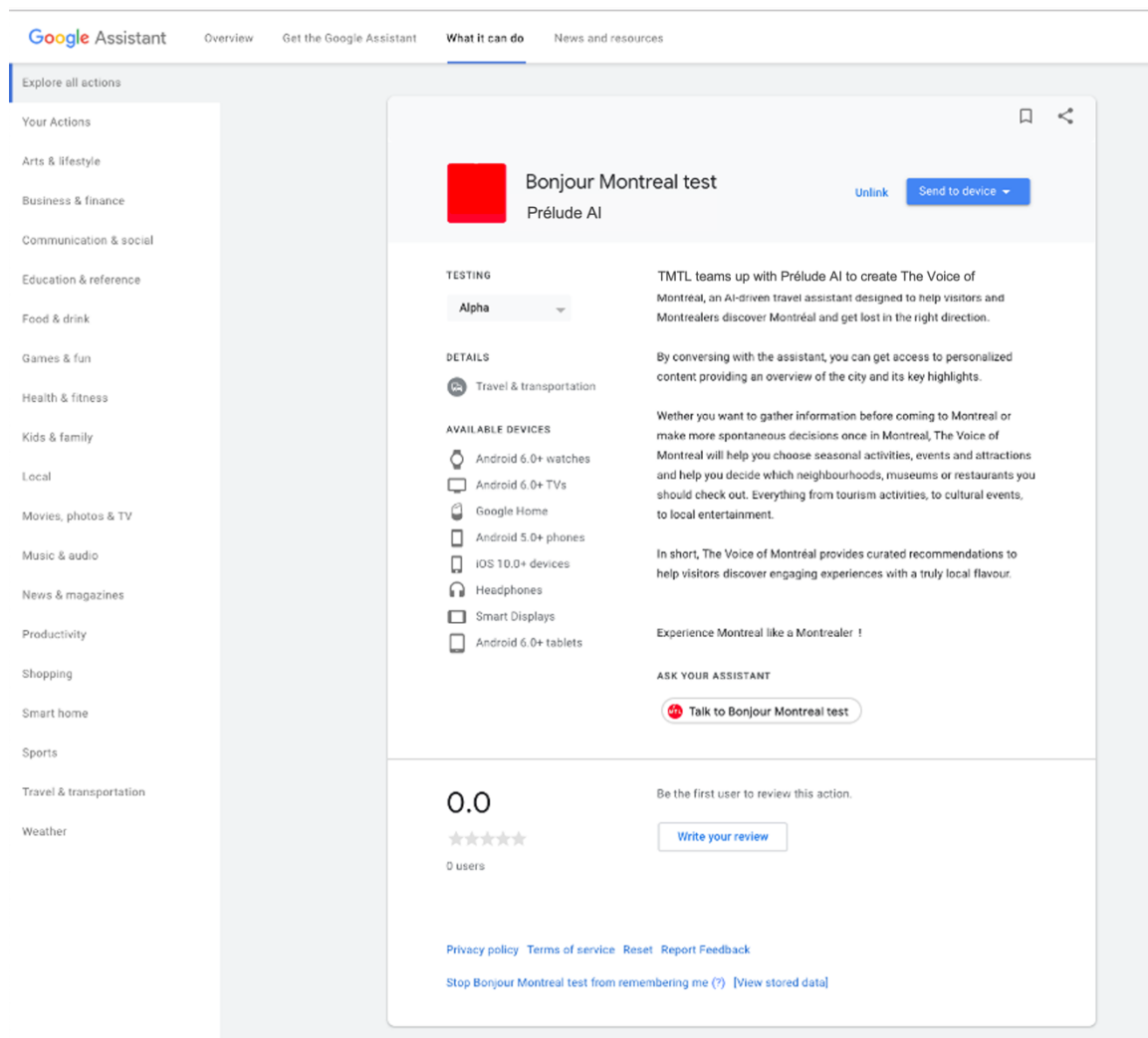
They also tend to minimize the collected personal information by designing the conversation so that the users are not invited to get too personal. However, when the Voice of Montréal will be launched, Prélude should keep an eye on how people do use and maybe misuse the voicebot. Since voice assistants can have access to unintended information that may be recorded in the user’s context (for example voices of people around (including children), double conversations (with the Assistant and with the person who is physically next to her) etc.)²²⁷, Prélude ought to be careful about the information collected through conversations during the use of the Voice of Montréal in real context.

What can be noted about the types of collected information, is that only one type is specific to the Voice of Montréal and fully managed by Prélude: the information about the stay. The other information – transcription of conversation, sign in, geolocation – are dependant of Google tools and the geolocation information of the device. The Google tools allow access to the automatic speech recognition functionality of the Google Assistant that turns voice recording into text transcription, the sign-in functionality enables the creation of an account by linking it to an existing Google account. To use these functionalities, Google appears like a kind of gatekeeper that would foster data minimization on the side of the company that asks access to this personal information.

²²⁶ <https://dialogflow.com>

²²⁷ Cf for example (Vincent, 2019)

For example, the access to the voice recordings is not by default, even if those recordings are generated by your voicebot (or your Action on Google); the account creation gives access to only three information: name, email address and profile picture. Moreover, when you discover the Action on the catalog²²⁸ you can choose not to be remembered by the company developing the Action.



Screenshot of the webpage presenting the prototype of the Voice of Montréal as an Action on Google Assistant. The “stop remembering me” option is at the bottom in the blue font. November 20, 2019.

This illustrates an ambiguity about the achievement of the data minimization principle: the organization developing the voicebot is not entirely in control of the data collection by its voicebot. So that, this organization relies on data collection practices that they can

²²⁸ <https://assistant.google.com/explore>

neither control nor have clear visibility on – these data collection practices being Google’s ones for managing the information generated by Actions developed by third parties such as Prélude. Hence, Prélude can’t make sure collection limitation and data minimization are fully ensured since its product is dependent on a medium that has its own data collection practices that are not (yet?) fully transparent.

(2) Regarding the achievement of the Privacy as the default setting principle, another gap could be the fact that the product is developed in partnership between Prélude and TMTL. The terms of data sharing should be clearly stated in their contract and in the privacy policy of the Voice of Montréal.

In the terms of the GDPR, the question is: is Prélude a data processor or a data controller – or a joint controller? That is, is it Prélude who “determines the purposes for which and the means by which personal data is processed” (= role of the data controller) or does Prélude “process personal data only on behalf of the controller” (= data processor) – or do Prélude and TMTL “jointly determine ‘why’ and ‘how’ personal data should be processed” (= joint controller)?²²⁹

Answering this question will help clarify the data management practices and influence the finalization of the system architecture. Indeed, answering this question implies defining who has access to the personal information. For now, the architecture is designed so that Prélude is the one processing personal information and make sure they are encrypted and secured. Prélude would share with TMTL deidentified and aggregated data only. In that way, the architecture appears to be designed as if Prélude were a controller. The next step will be to “create technological, policy and procedural barriers to data linkages with personally identifiable information” (Principle 2).

No matter what, since people at Prélude are manipulating personal information, it is their responsibility to make sure they do it properly by encrypting this information: the

²²⁹ The quotes are from this webpage: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en

transcripts of the conversations should be unreadable and user's account information unlinkable to their actual name and email address (Principle 5).

CONCLUSION

Embedding privacy into the design of a product such as the Voice of Montréal voicebot is a work in progress that seems to have started in an encouraging path. Due to the unavoidable uncertainties of the product development and use during the following phases, privacy by design should not be considered as achieved but as an ongoing process. This process involves a continued assessment of technical and organizational data management practices to prevent from any breach. It also involves assessments of the actual information that are being collected through the use of the product. These assessments should all the more be on a regular basis that the way this information is collected – by voice – is very fluid and can include a lot of other noises that could reveal personal information.

From analyzing the main privacy-by-design challenges Prélude faces in developing the Voice of Montréal, here are my core recommendations:

- Clarify the **terms for data sharing** with TMTL so that clear personal information management practices can be define. These practices are technological but also business related.
 - o The updated information should be written in the **privacy policy**.
 - o The **Privacy List** on the voicebot itself should also be updated consequently.
- The **system architecture** of the voicebot should also be designed to de-identify personal information wherever possible. And technical barriers should be designed to prevent from data linkages.

- A **security policy** should be written and shared with the employees or consultants who deal with raw data where personal information appears. All the stakeholders should be aligned with this security policy.
- Assessments of the types of personal information that are collected should be conducted on a regular basis – especially during the user testing of the beta version, to adjust the ways of limiting the data collection. These assessments could be **Privacy Impact Assessments**²³⁰.
 - These assessments should then be done on a regular basis when the product will be launched.
- To ensure all this is done and to make sure the protection of personal information is worked on continuously, someone should be held responsible for managing privacy by design in a win-win manner with the leaders of the company. This person could be a **Data Protection Officer** (DPO) to hire, or someone already in the team who has sufficient technical and organizational knowledge to assess these practices.
 - The DPO would lead the privacy impact assessments.
 - She would be the one responsible for the **email address** privacy@prelude.ai where users could send requests related to privacy issues or the exercise of their rights (to delete, access, rectify, download the personal information the company has on them).

²³⁰ Examples : <https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/> ; <https://www.cnil.fr/en/privacy-impact-assessment-pia>

About Privacy-by-design:

The 7 Foundational Principles of Privacy by Design

Principle 1 – Proactive not Reactive; Preventative not Remedial

Operational Guidance: These actions anticipate and prevent privacy-invasive events before they happen. Do not wait for privacy risks to materialize – the aim is to prevent the breaches from occurring. By itself, a privacy policy cannot protect the personal information held by an organization.

Actions	Responsibility
1. Affirm senior leadership commitment to a strong, proactive privacy program.	Leadership/Senior Management e.g. Board of Directors, CEO, CPO, CIO, COO, CSO, Company Owner(s)
2. Ensure that concrete actions, not just policies, reflect a commitment to privacy. Monitor through a system of regularly reviewed metrics.	
3. Develop systematic methods to assess privacy & security risks and to correct any negative impacts, well before they occur.	
4. Encourage privacy practices demonstrably shared by diverse user communities and stakeholders, in a culture of continuous improvement.	

Principle 2 – Privacy as the Default Setting

Operational Guidance: These methods seek to provide privacy assurance – delivering the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice. No action should be required on the part of the individual user to protect their privacy; it should be built into the system, automatically by default. More about this Principle will follow in the upcoming modules.

Actions	Responsibility
1. Adopt as narrow and specific a purpose(s) for data collection as possible. Begin with no collection of personally identifiable information – data minimization.	

2. Minimize the collection of data at the outset to only what is strictly necessary.	Software Engineers & Developers Application & Program Owners
3. Limit the use of personal information to the specific purposes for which it was collected.	Line of Business & Process Owners
4. Create technological, policy and procedural barriers to data linkages with personally identifiable information.	

Principle 3 – Privacy Embedded into Design

Operational Guidance: These actions embed privacy requirements into the design and architecture of IT systems and business practices. They are not bolted on as add-ons after the fact. Privacy should be an essential component of the core functionality being delivered.

Actions	Responsibility
1. Make a Privacy Risk Assessment an integral part of the design stage of any initiative, e.g. when designing the technical architecture of a system, pay particular attention to potential unintended uses of the personal information.	
2. Base identity metasystems on the “Laws of Identity,” intended to codify a set of fundamental principles to which universally adopted, sustainable identity architecture must conform.	Application & Program Owners Line of Business & Process Owners Software Engineers & Developers
3. Consider privacy in system development lifecycles and organizational engineering processes. System designers should be encouraged to practice responsible innovation in the field of advanced analytics.	Regulators
4. Embed privacy into regulatory approaches that may take the form of self-regulation, sectoral privacy laws, omnibus privacy legislation and more	

general legislative frameworks, calling for an approach guided by “flexibility, common sense and pragmatism.”	
---	--

Principle 4 – Full Functionality – Positive-Sum, not Zero-Sum

Operational Guidance: These actions seek to accommodate legitimate interests and objectives in a positive-sum, “win-win” manner, not through a zero-sum (win/lose) approach, where unnecessary trade-offs to privacy are made. Avoid the pretense of false dichotomies, such as privacy vs. security, and demonstrate that it is possible to have both.

Actions	Responsibility
1. Acknowledge that multiple, legitimate business interests must coexist.	Leaders/Senior Management
2. Understand, engage and partner – Practice the 3Cs – communication, consultation and collaboration, to better understand multiple and, at times, divergent interests.	Application & Program Owners Line of Business & Process Owners Software Engineers & Developers
3. Pursue innovative solutions and options to achieve multiple functionalities.	

Principle 5 – End-to-End Security – Full Lifecycle Protection

Operational Guidance: Security is the key to privacy. These actions ensure cradle-to-grave, lifecycle management of information, end-to-end, so that at the conclusion of the process, all data are securely destroyed, in a timely fashion.

Actions	Responsibility
1. Employ encryption by default to mitigate the security concerns associated with the loss, theft or disposal of electronic devices such as laptops, tablets, smartphones, USB memory keys and other external media. The default state of data, if breached, must be “unreadable.”	Software Engineers & Developers Application & Program Owners
2. Deploy encryption correctly and carefully integrate it into devices and	Line of Business & Process Owners

workflows in an automatic and seamless manner.	
3. Ensure the secure destruction and disposal of personal information at the end of its lifecycle.	

Principle 6 – Visibility and Transparency – Keep it Open

Operational Guidance: Stakeholders must be assured that whatever the business practice or technology involved, it is, in fact, transparent to the user and operating according to the stated promises and objectives, subject to independent verification. Remember, trust but verify. Module 4 on Surveillance delves into the significance of this Principle.

Actions	Responsibility
1. Make the identity and contact information of the individual(s) responsible for privacy and security available to the public and well known within the organization.	
2. Implement a policy that requires all “public facing” documents to be written in “plain language” that is easily understood by the individuals whose information is the subject of the policies and procedures.	Leadership/Senior Management
3. Make information about the policies, procedures and controls relating to the management of Personal Information readily available to all individuals.	Software Engineers Application Developers Systems Architect
4. Consider publishing summaries of PIAs, TRAs and independent third-party audit results.	
5. Make available a list of data holdings of Personal Information maintained by your organization.	
6. Make audit tools available so that users can easily determine how their data is stored, protected and used. Users should	

also be able to determine whether the policies are being properly enforced.	
---	--

Principle 7 – Respect for User Privacy – Keep it User-Centric

Operational Guidance: This method requires architects and operators to keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice and empowering user-friendly options. Keep it user-centric. This Principle acts as a key theme throughout the modules.

Actions	Responsibility
1. Offer strong privacy defaults.	Leadership/Senior Management Software Engineers & Developers Application & Program Owners Line of Business & Process Owners
2. Provide appropriate notice.	
3. Consider user-friendly options: <ul style="list-style-type: none"> a. Make user preferences persistent and effective. b. Provide users with access to data about themselves. c. Provide access to the information management practices of the organization. 	

Source: Privacy by design online course, by Ann Cavoukian and Michelle Cibba, Ryerson University, 2019

From: https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf