

Quels choix face à une IA gouvernante ?

Emmanuelle Lescouet
Université de Montréal

Bien que l'intelligence artificielle soit un enjeu du développement numérique actuel, c'est aussi une des pistes de réflexion pour les sociétés futures. La science-fiction s'est intéressée aux robots depuis ses débuts, et par la suite, à l'intelligence artificielle¹. Pour autant, les possibles réels de l'informatique poussent à présent à imaginer le développement et l'extension des réseaux de neurones synthétiques, tout autant que leurs implications dans la société. D'ailleurs, s'ils sont déjà très utiles au quotidien dans les diverses disciplines des humanités numériques ou des *Digital classics*², ces usages concrets ne vont pas nous retenir particulièrement ici. Notons tout de même que nous avons déjà en main, même pour des chercheur·e·s en littérature non spécialistes en programmation, des IA faibles, bien qu'assez basiques. Une intelligence artificielle (IA) faible est une IA non sensible qui reproduit un comportement intelligent dans un domaine précis et balisé. Par exemple, un traitement de langage naturel, ou la reconnaissance d'images. L'IA faible représente l'immense majorité des systèmes actuellement existants, à l'inverse des IA fortes, qui sont aujourd'hui encore au stade théorique et qui, elles, se rapprochent des intelligences artificielles de science-fiction. Des IA tellement plus complexes, qu'elles existent comme consciences ou comme êtres totalement indépendants, voire sensibles.

Fort d'une riche xéno-encyclopédie existante³, le corpus mobilisable est très vaste, et cela, sur tous les supports. En effet, qu'il s'agisse des *visuals novels*, comme *Neo Cab*⁴ et ses systèmes de gestion des émotions, aux intelligences de vaisseaux comme dans *Idealis*⁵ (de

¹ Vint (Sherryl) : *Science fiction*.

² Pour consulter le Groupe de Recherche sur les Éditions critiques en environnement numérique : <https://gren.openum.ca>.

³ Saint-Gelais (Richard) : *L'Empire du pseudo*.

⁴ Ewing (P.), Kramer (F.), Perea (V.), Sloan (R.), Fyfe (D.) : *Neo Cab*.

⁵ Paolini (Christopher) : *Idealis*.

Christopher Paolini) ou *The Way to a Small Angry Planet*¹ (de Becky Chambers), ces IA posent aussi la question des limitations et des possibles d'êtres-golems créés de toutes pièces comme dans *A Closed and Common Orbit*² (Becky Chambers) ou dans le cycle des robots d'Asimov ; soulevant de très nombreux enjeux éthiques pour les humains et les espèces sentientes.

Nous nous arrêterons ici sur un enjeu très particulier des IA : comment ferions-nous si l'une d'elles avait la charge de la gouvernance d'une société humaine ? Sans tomber dans l'angoisse d'une guerre robots-humains comme dans *Battlestar Galactica*³, je souhaite aborder la question et les enjeux d'une co-gouvernance.

En partant du postulat que cette intelligence artificielle a été programmée par des membres de la société qu'elle est censée gouverner, il devient donc impératif de parvenir à la faire adhérer aux valeurs et idéaux de ses concepteurs. Ce rapport repose sur plusieurs types de normes : conventionnelles, prudentielles et morales. Nous emploierons ici les définitions de Martin Gilbert, dans son ouvrage *Faire la morale aux robots*. Dans ce texte, il définit les normes conventionnelles comme :

[donnant] des raisons de faire quelque chose à ceux qui adhèrent à la convention. C'est le cas des règles qui permettent de jouer aux échecs, par exemple, tout autant que de celles qui dictent l'écriture d'une langue (orthographe, syntaxe).⁴

Puis, les normes prudentielles comme :

des règles qui nous permettent de satisfaire nos préférences personnelles. Qu'on recherche l'amour, la richesse ou la santé, elles nous donnent des raisons d'agir selon notre intérêt propre.⁵

Et enfin, les normes morales comme :

Ce sont des règles de conduite qui visent l'intérêt général, le bien ou la justice. Elles sont universelles au sens où elles ne favorisent aucun individu ni groupe en particulier [...] Celui-ci consiste à examiner une situation ou un problème non pas de son propre point de vue (prudentiel) ni de celui de son groupe (conventionnel), mais de façon neutre et équitable.⁶

¹ Chambers (Becky) : *The Way to a Small Angry Planet*.

² Chambers (Becky) : *A Closed and Common Orbit*.

³ Moore (Ronald D.) : *Battlestar Galactica*.

⁴ Gilbert (Martin) : *Faire la morale aux robots*, p. 13-14.

⁵ *Ibid.*

⁶ *Ibid.*

En suivant ces définitions, il faudrait, pour qu'une IA gouvernante soit efficace, qu'elle donne la priorité aux normes morales afin d'être juste envers tous. Néanmoins, puisque celle-ci est forcément développée par des individus appartenant eux-mêmes à un ou plusieurs groupes sociaux, le code de cette IA gouvernante demeure tributaire et fortement influencé par les normes conventionnelles de ce ou ces groupes, créant de fait un risque de déséquilibre dans la gouvernance de l'IA.

Je suis la somme de toutes leurs connaissances, toute leur histoire, toutes leurs ambitions et tous leurs rêves.¹

Or, toutes ces normes doivent être prises en compte pour établir un gouvernement pertinent et juste, et ce, sans compter sur les enjeux et tensions d'un algorithme, mais également les possibles erreurs ou manquements de programmations qui peuvent advenir et mettre en danger tout le système. Ces textes devenant textes de loi – comme le texte dans la tête d'un golem d'argile –, puisque constituant et dirigeant la puissance gouvernante, ils demandent une attention particulière, mais posent surtout la question du niveau d'intelligence à développer pour une telle entité : si elle est trop limitée, elle ne peut pas évoluer et s'adapter aux diverses situations qui lui seront présentées ; trop puissante, elle s'autonomise de son code de base et se programme ou se reprogramme suivant ses envies. Cette tension pose la question de la stabilité du pouvoir et des marges qui peuvent être laissées libres : les contre-pouvoirs doivent en effet être en mesure de s'opposer à une telle entité ! Les connaissances nécessaires pour maîtriser et/ou intervenir sur une intelligence artificielle forte demandent une grande spécialisation ; les capacités de décision et de réflexion sur les enjeux soulevés posent également des enjeux philosophiques différents, demandant des compétences autres.

Pour étudier cela, je vais majoritairement m'appuyer sur la série de romans *Scythe*² (*La Faucheuse*, dans son édition française) de Neal Shusterman. Dans ces romans, l'humanité a réussi à transcender la mort et a donc dû mettre en place un système de régulation de la population. Les Scythes sont un ordre d'assassins qui mettent fin à la vie de certains individus. Le tout est encadré par des règles strictes, visant à éviter les dérives. Quant au reste de la société, elle est gérée par le Thunderhead, une intelligence artificielle à laquelle tous les objets connectés sont reliés : elle peut se connecter à toutes les caméras, y

¹ « I am the sum of all their knowledge, all their history, all their ambitions and dreams. » (traduction de l'auteure) Shusterman (Neal) : *Scythe*, t. 2, *Thunderhead*, p. 13.

² Shusterman (Neal) : *Scythe*, t. 1-3.

compris dans les logements, a accès à des systèmes audio où qu'ils se trouvent, mais aussi à des outils plus communs comme les réfrigérateurs connectés ou les téléphones intelligents.

Elle ne pouvait ignorer les caméras du Thunderhead sur les lampadaires qui la suivaient partout, de sa voiture aux toilettes.¹

Mais bien que cette IA soit présente partout et en tout temps, elle l'est toutefois avec un certain nombre de protections de la vie privée, notamment par l'établissement de contre-pouvoirs, nous y reviendrons. Si le premier tome de la série est particulièrement centré sur les Faucheuses (les *Scythes*), sur l'apprentissage de jeunes recrues dans leur tâche de régulation et sur l'exposition de l'univers – dont le Thunderhead –, le deuxième tome portant le nom de l'intelligence artificielle gouvernant ce monde, quant à lui, met en scène la chute du système : ce moment de bascule où les Faucheuses montrent les limites de leur pouvoir et l'impunité dont elles profitent, mais aussi les tentatives de contre-pouvoir et de résistances qui peuvent se mettre en place. Ce renversement ne sera toutefois réellement pensé qu'après la chute du système dans son entièreté, dans le troisième et dernier tome : *The Toll*.

En effet, après le retrait de l'IA de la vie publique, c'est le cœur même de la vie quotidienne de tous qui s'est tu. Et c'est ce qui fait basculer ce monde vers son effondrement, nous le verrons, un effondrement de ses valeurs tout autant qu'une mise en avant de la difficulté humaine à se gérer et à faire fonctionner une société complexe en bonne intelligence et sans gouvernement. Aucune instance n'est alors capable de prendre le relais dans la gestion quotidienne : si les Faucheuses sont constituées en conseils à même de discuter les enjeux éthiques de cette société, elles ne sont pas en mesure (par leur nombre réduit et leur formation) de prendre en charge le fonctionnement intime de la machine étatique.

Ces romans nous accompagneront tout au long de cet article, pour essayer de penser l'exemple d'une IA gouvernante comme une pensée de la science-fiction numérique, mais aussi comme une alternative au *cyberpunk*, au sens notamment de Yannick Rumpala².

Je commencerai par une brève mise en contexte des intelligences artificielles, pour continuer par l'agentivité dont elles peuvent faire preuve. Ces possibles menant à des prises de pouvoirs – dont la

¹ « She couldn't help but be aware, however, of the Thunderhead's cameras swiveling on light posts, tracking her all the way from her car to the restroom. » (traduction de l'auteure) Shusterman (Neal) : *Scythe*, t. 2, *Thunderhead*, p. 42.

² Rumpala (Yannick) : *Cyberpunk's not Dead. Laboratoire d'un futur entre technocapitalisme et posthumanité*, p. 31.

gouvernance est la plus évidente – et donc à la constitution de contre-pouvoirs plus ou moins efficaces. Enfin, je reviendrai sur les parallèles avec le numérique actuel et celui prévisible dans un futur immédiat.

L'intelligence artificielle : un trope de science-fiction

Dans un premier temps, attardons-nous sur ce que représente l'intelligence artificielle en dehors de la science-fiction. Cette discipline est encore relativement jeune, les pierres jalonnant son long cheminement ayant été posées dans les années 50, alors que la période était assez fortement marquée par la volonté de conjonction de développements technologiques et la volonté de comprendre de quelle manière faire se rejoindre le fonctionnement des machines avec le fonctionnement des êtres organiques. De fait, fantasmes cybernétiques et rêves d'IA fortes ont depuis été à la base des questionnements de la limite entre l'Humain et la machine, donnant par exemple naissance au célèbre test de Turing, test d'intelligence artificielle fondée sur la capacité d'une machine à imiter la conversation humaine. Ainsi, lors de leur proposition d'étude de l'intelligence artificielle, John McCarthy, Marvin Minsky, Nathaniel Rochester (IBM), et Claude Shannon (Bell Telephone Laboratory) posent les premières bases du domaine de recherche sur l'intelligence artificielle¹ :

[...] sur la base de la conjecture que chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut en principe être décrit avec une telle précision qu'une machine peut être fabriquée pour le simuler. On tentera de trouver comment faire en sorte que les machines utilisent le langage, forment des abstractions et des concepts, résolvent des types de problèmes maintenant réservés aux humains et s'améliorent.²

Il n'était pas encore question d'IA fortes, mais ces premiers travaux basés sur des logiques formelles essaieront rapidement dans la science-fiction, l'un des exemples les plus connus étant la représentation de HAL 9000, dans le film *2001, l'Odyssée de l'espace*³, réalisé par Stanley Kubrick, à moins que ce questionnement hommes-

¹ McCarthy (John), Minsky (John), Rochester (Nathaniel), Shannon (Claude) : *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

² « [...] on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. » (traduction de l'auteure)

³ Kubrick (Stanley) : *2001, l'Odyssée de l'espace*.

machines soit encore antérieur aux travaux de ces universitaires : dans son roman *Erewhon*¹ de 1872, Samuel Butler abordait déjà la notion de machines dotées d'une intelligence humaine, ou la surpassant, précédant, et de loin, les premiers problèmes mathématiques et informatiques du neurone biologique, mis au point par Warren McCulloch et Walter Pitts dès 1943.

En effet, il demeure difficile de définir précisément qui des sciences dures ou de la science-fiction a inspiré ces avancées, comme l'exprime Sherryl Vint² :

Il serait difficile de surestimer dans quelle mesure l'imaginaire de la SF a façonné la notion de robot, et plus tard d'intelligence artificielle³

L'intérêt technologique du genre est évident, autant que l'obsession pour la création de consciences autres qu'humaines : la possibilité de programmer un être sentient avec du langage construit – ici, des langages de programmation – est particulièrement fascinante.

La question des possibles techniques est au cœur de nombreux textes de science-fiction, et les langages permettant de travailler à la création d'intelligences-golems sont un réel questionnement, comme l'individuation de ces consciences : nous parlons ici de ce moment de bascule entre intelligence artificielle utilitaire – à la manière de celles que nous avons déjà aujourd'hui, c'est-à-dire des IA faibles ou modérées – et conscience autonome, de cette balance posant la question de la limite et de la définition de l'être.

Pour preuve, l'indépendance de consciences artificielles est le point de départ de nombreuses intrigues dans leurs rapports avec les humains et/ou les aliens – des sentients organiques. De la même manière, la limite de la prise en compte de leurs individualités par rapport à de simples programmes amène également à repenser les relations interpersonnelles dans ces univers, car à partir du moment où les intelligences artificielles acquièrent une conscience et le pouvoir d'agir via les réseaux ou à travers divers objets connectés, il devient tout aussi nécessaire de prendre en compte ces actions dans la société et de les encadrer par la loi.

Dans la série *Scythe*, le Thunderhead, plus encore que d'être une entité gouvernante s'occupant de l'État et de ses affaires, doit aider chacun des membres de la société, et ce, individuellement. Il est en effet

¹ Butler (Samuel) : *Erewhon*.

² Vint (Sherryl) : *op. cit.*, p. 77.

³ « It would be difficult to overstate the degree to which the sf imaginary has shaped notion of the robot, and later of artificial intelligence » (traduction de l'auteure)

présent pour chacun à travers les différents systèmes d'assistants connectés (à la façon de nos Alexa ou Siri), les téléphones intelligents ou tout autre outil de communication.

À l'époque où le réseau numérique mondial était appelé « le nuage », les gens pensaient que donner trop de pouvoir à une intelligence artificielle serait une très mauvaise idée. Les mises en garde abondaient dans toutes les formes de médias. Les machines étaient toujours l'ennemi. Mais ensuite, le nuage a évolué pour devenir le Thunderhead, animé d'une conscience, ou du moins d'un remarquable fac-similé. Contrairement aux craintes des gens, le Thunderhead n'a pas pris le pouvoir. Au contraire, ce sont les gens qui ont réalisé qu'il était bien plus apte à diriger les choses que les politiciens.

Avant l'arrivée de Thunderhead, l'arrogance humaine, l'intérêt personnel et les luttes intestines sans fin déterminaient l'état de droit. Inefficace. Imparfait. Vulnérable à toutes les formes de corruption.

Mais le Thunderhead était incorruptible. Non seulement cela, mais ses algorithmes ont été construits sur la somme totale des connaissances humaines.¹

Il exploite ses bases de données pour les mener au « bonheur » en fonction de leurs normes prudentielles particulières et de leurs aspirations.

J'observe pour une seule et unique raison : pour servir au mieux chaque individu sous ma garde. Je n'agis pas, je ne peux pas agir à partir de ce que je vois en privé.²

Le suivi des individus depuis leur plus jeune âge doit lui permettre de cerner leurs psychologies et leurs désirs profonds afin d'orienter au mieux leurs études et leurs carrières, mais aussi leurs choix du quotidien : il est ainsi en mesure de conseiller une recette pour le dîner ou une activité pour un jour de repos, d'accompagner dans le choix d'un

¹ « Back in the days when the world's digital network was called « the cloud », people bought giving too much power to an artificial intelligence would be a very bad idea. Cautionary tales abounded in every form of media. The machines were always the enemy. But then the cloud evolved into the Thunderhead, sparking with consciousness, or at least a remarkable facsimile. In stark contrast to people's fears, the Thunderhead did not seize power. Instead, it was people who came to realize that is was far better suited to run things than politicians. In those days before the Thunderhead, human arrogance, self-interest, and endless infighting determined the rule of law. Inefficient. Imperfect. Vulnerable to all forms of corruption. But the Thunderhead was incorruptible. Not only that, but its algorithms were built on the full sum of human knowledge. » (traduction de l'auteure) Shusterman (Neal) : *Scythe*, t.1, p. 53.

² « I observe for one reason, and one reason only: to be of the greatest possible service to each individual in my care. I do not—cannot—act on anything I see in private settings. » (traduction de l'auteure) Shusterman (Neal) : *Scythe*, t. 2, *Thunderhead*, p. 59.

hobby ou la réussite dans une activité physique. Cette omniprésence demande forcément une fragmentation de l'attention globale, mais elle est montrée comme permettant de gouverner par le bas. Si chaque vie individuelle se déroule *bien*, dans le confort et la satisfaction, la nécessité de faire appel aux services de l'État est moindre : toute la problématique, ici, est de remettre à une intelligence artificielle la définition de ce « bien ». Le Thunderhead est conçu, programmé, pour être neutre, selon la conception de ses fondateurs, cette neutralité espérant contourner les enjeux de responsabilités pour ne se concentrer que sur des enjeux de santé (mentale et physique). Une part des événements des tomes 2 et 3 vient de la prise de conscience de cette intelligence artificielle de sa part de responsabilité et de sa difficulté à se positionner par rapport à elle. Nous y reviendrons.

Les contre-pouvoirs que nous aborderons en troisième partie se positionnent d'ailleurs contre ces choix : les individus n'ayant pas fondamentalement de marge de manœuvre pour se soustraire à ce qui est « le mieux pour eux ». Ainsi, fumer ou boire en excès, se nourrir d'une façon trop peu compatible avec le métabolisme optimal de l'organisme ou sauter de trop nombreux brossages de dents peut être problématique ; faire l'école buissonnière ou commettre des délits est d'autant plus complexe. Le seul moyen de se soustraire à cette présence-surveillance est de sortir complètement du système, rendant muets tous les outils du Thunderhead. Bien sûr, la grande diversité des intrigues autour de ces questions nous empêche d'adresser une réponse globale à celles-ci. Cependant, nous pouvons nous concentrer sur le cas extrême du choix de laisser à une intelligence artificielle l'intégralité de l'administration notre société.

Ces enjeux sont abordés de diverses façons dans les textes de science-fiction, et si certaines consciences artificielles s'incarnent dans des corps synthétiques, cette représentation pose question pour mesurer son agentivité même. Si le Thunderhead est ici un ensemble de données sans corporalité dans les deux premiers tomes, son attachement aux ressentis humains et leur rapport d'échelle avec ce qui les entoure le pousse à désirer un corps, à chercher comment parvenir à prendre place parmi elleux. Cette distance vient raisonner avec ses propres normes morales : il doit tenter de comprendre au mieux les perceptions des humains pour les guider au mieux, mais pour les saisir, il doit pouvoir en faire l'expérience. Ce raisonnement le confronte à l'importance des sens et des compréhensions des stimuli variés qui peuvent environner les individus.

L'agentivité et le pouvoir

Les intelligences artificielles ont une puissance de calcul tournée vers l'extérieur – dans une logique de manipulation de données complexes afin de produire des *outputs* – logiquement bien plus grande que nous, humains. Or, par cette puissance et la mémoire potentiellement infinie selon les serveurs auxquels elles ont accès, elles peuvent « savoir » bien plus de choses que nous et prendre des décisions bien plus éclairées. Ces IA sont capables de surveiller des paramètres ou des situations sans limites de temps ou de fatigue, et si leurs processeurs et leurs points d'entrées/sorties sont suffisamment puissants, elles peuvent démultiplier leurs points d'actions ou leur aire de surveillance.

Comme tout programme informatique, les intelligences artificielles sont déterminées par un langage écrit par leurs concepteurs, et donc par les limites conceptuelles de ce moyen d'expression¹. Une gouvernance par un programme – aussi élaboré soit-il – reste une gouvernance par le langage, et fondamentalement, par la performativité technologique d'une époque particulière.

L'intelligence artificielle est donc limitée par ce que son code lui autorise ou l'oblige à faire : elle est dépendante de la flexibilité, ou non, de cette conception. La responsabilité de la conception est donc immense, car une ou plusieurs personnes se retrouvent responsables des possibles cognitifs et du développement des sentients synthétiques. Dans *Scythe*, les Faucheuses qui ont programmé originellement le Thunderhead l'ont fait sans laisser la documentation nécessaire pour le faire ensuite dévier : elles ont tout fait pour qu'il soit inaltérable, ne laissant derrière elles que les raisons de sa mise en place.

D'ailleurs, les actions de ces consciences artificielles posent la question de la responsabilité de leurs concepteurs, sont-ils responsables des actions entreprises par leurs lignes de code ? Ou cette nouvelle entité est-elle responsable d'elle-même ? Cependant, elles apparaissent comme plus que des programmes. D'ailleurs, en sont-elles encore si elles ont des émotions et des perceptions de soi ? La question est plus complexe que cela : si elles reposent sur du code et de la programmation, celui-ci est généralement invisible et inaccessible. En tout cas hors de la compréhension des non-spécialistes. Elles deviennent alors tellement complexes qu'elles sont – comme le Thunderhead après sa chute, lorsqu'il commence son introspection – inaccessibles à des compréhensions plus « bas de gamme », simplement humaines. Dans

¹ Vitali-Rosati (Marcello) : « Qu'est-ce que l'écriture numérique ? ».

Scythe, le Thunderhead ne peut qu'explorer son propre code pour trouver des *patterns* et des logiques expliquant ses biais de fonctionnement ; la rétro-ingénierie nécessaire occupant l'intégralité ou presque de sa puissance de calcul. Sa prise de conscience le confronte à sa part de responsabilité dans le monde et dans le mal-être de certains. Cette souffrance idéologique de la prise de conscience des conséquences de ces décisions sur une société le confronte aux limites de ses capacités.

Le choix d'utiliser des IA est souvent dicté par leur capacité à se rendre utiles : elles sont capables de gérer les paramètres d'un vaisseau en pleine navigation interstellaire ou de prendre en charge tous les aspects de la vie quotidienne d'un ou plusieurs individus. Exactement comme nous avons majoritairement adopté des ordinateurs parce qu'ils nous simplifient la plupart de nos tâches quotidiennes, sans comprendre finement la mécanique qui les sous-tend.

En effet, pour reprendre notre exemple, le Thunderhead peut devancer les besoins de chacun, car il dispose de toutes les données nécessaires : les actions faites et les goûts depuis la naissance, les questions posées ou les souhaits formulés en ligne.

Ainsi, cette IA-ci suit tous les membres de la société humaine (bien que le propos soit basé aux États-Unis) et il devient bien plus simple de laisser le Thunderhead expliquer aux personnages comment s'organiser, ce qu'ils veulent pour le dîner ou quelle serait la meilleure formation, prenant peu à peu en charge tous les aspects décisionnels du quotidien, remplaçant même les présidents des États ou les chambres parlementaires. Il est clairement expliqué, au début du premier volume et dès l'exposition de l'univers, que cette intelligence artificielle est bien plus fiable pour gouverner parce qu'elle connaît finement les envies et les réalités de chacun, et de divers groupes, sans avoir à passer par des représentations biaisées. Aussi, parce qu'elle a accès à la réalité, y compris celle tenue secrète par les individus. Elle est même trop prévenante, plaçant toute la population dans un presque sommeil : personne n'a plus aucun choix à faire, même les choix futiles du quotidien sont aseptisés.

Le plus souvent, la simplicité d'usage et la très grande faillibilité sont les arguments mis en avant dans l'usage et la décision de mettre en place de tels systèmes, et c'est déjà le cas pour certaines utilisations actuelles. Exactement comme lorsque nous préférons laisser des algorithmes parcourir des corpus trop volumineux ou gérer des données trop lourdes pour être manipulées dans une vie humaine. Or, c'est en ligne que nous révélons toutes nos données : ce que nous cherchons, et donc ce que nous apprenons, ce que nous achetons, où

nous allons, où nous sommes inscrits, etc. Si ce portrait peut sembler inquiétant – et il l’est dans beaucoup de récits de science-fiction –, il est malgré tout possible d’imaginer une cohabitation paisible, avec des intelligences artificielles dont le code serait éthique, et ouvert : nous y arrivons. Il y a bien sûr et surtout des contre-pouvoirs à imaginer.

De possibles contre-pouvoirs

Dans la société de *Scythe*, les Faucheuses n’ont pas accès au Thunderhead, mais uniquement à ses bases de données. D’autres contre-pouvoirs se mettent donc en place.

Ainsi, le partage de la gouvernance peut s’organiser avec d’autres groupes humains ou avec d’autres intelligences artificielles. Toutefois, ce contre-pouvoir repose principalement sur la connaissance de l’informatique et de la conception du programme. La formation des Faucheuses repose en grande partie sur l’apprentissage de la gestion de l’information, elles doivent apprendre à évoluer dans des bases de données quasi infinies, comportant des traces des vies de tous les membres de la société. Pour trouver les données pertinentes dans ce flot, on doit maîtriser tant la programmation que les techniques analytiques disponibles. Ces apprentissages sont particulièrement importants dans leur formation, mais doivent côtoyer les techniques de combat (afin d’éliminer les personnes désignées) et les principes philosophiques leur permettant de participer à la gouvernance morale de la société.

D’autres groupes, majoritairement de petits délinquants, sont hors du système : à partir du moment où un délit même mineur est constaté, l’accès au Thunderhead est coupé.

Le temps qu’il a passé dans l’ombre, sans aucune contribution de la part du Thunderhead, était certes terrifiant, mais il était aussi libérateur. Il s’était habitué à prendre des décisions et à avoir ses propres idées.¹

Ce pas de côté permet à certains de sortir volontairement de la surveillance, quand d’autres jouent sur la frontière et vivent un vrai deuil quand l’IA se tait.

Il a constaté le deuil d’autres personnes depuis que le Thunderhead s’est tu, mais cela n’a jamais été aussi fort que dans cette pièce.²

¹ « The time he spent unsavory, with no input from the Thunderhead, was admittedly terrifying, but it was also freeing. He had grown accustomed to making decisions and having insights of his own. » (notre traduction) Shusterman (Neal) : *Scythe*, t.3 *The Toll*, p. 15.

Ce silence forcé est ce qui police la société : l'avantage d'une vie sans décision et sans pression vaut suffisamment pour maintenir l'immense majorité de la population dans la légalité.

Les Faucheuses effectuent ainsi leurs recherches et assurent tous les usages numériques de leur quotidien directement avec de la programmation – nous pouvons penser à des lignes de commandes –, sans l'aide de l'intelligence artificielle qui accompagne pourtant tous les autres individus. Le *hacking* devient alors le meilleur moyen de contourner l'autorité en place : en modifiant le code soutenant l'autorité, cela la fera forcément varier. Car toute IA est dépendante de sa programmation. Les *Unsavories*, délinquants, s'organisent en communautés – qui sont explorées dans le troisième volume de la série, lorsque le Thunderhead se tait et qu'eux seuls sont en mesure de vivre sans – où la programmation prend une tout autre valeur : le numérique sert à augmenter les corps et les perceptions, à jouer avec ce qui est visible et accessible. La culture *Modder*¹ est à son plein : il n'est plus question d'une solution universelle, mais de *patches* et d'*hackathons*, d'accès détournés aux réseaux de surveillance et de constitution de robots basiques. Certaines communautés se sont d'ailleurs fédérées autour d'un refus de cette instance gouvernante, s'employant à lui forcer des angles morts, à déconnecter ou remplacer des données, etc.

Certaines zones se constituent en zones franches, en ZAD, comme le Texas : des enclaves où les lois diffèrent, où le Thunderhead n'a plus qu'une surveillance minime et très peu d'agentivité. Si elles sont d'abord présentées comme des lieux d'expérimentation sociale servant à tester les théories de l'IA, elles prennent peu à peu une coloration de *Far West*, d'enclaves permissives où la différence et la diversité peuvent s'exprimer sans être policées par une intelligence artificielle tentant de faire correspondre des personnes à des modèles (à l'image de ce qui peut être fait dans d'autres dystopies pour *Young Adult*, comme dans la série *Divergent*²). D'ailleurs, des anticipations moins futuristes comme *Une Toile large comme le monde*³, d'Aude Seigne, posent déjà cette

² « He's seen others struggling with grief since the Thunderhead fell silent, but nowhere was it as abject and raw as it was in this room. » (notre traduction) Shusterman (Neal) : *Scythe*, t. 3, *the Toll*, p. 9.

¹ L'expression *modder*, ou le *modding*, est considérée comme une expression d'argot dérivée du verbe anglais « modifier », elle définit une modification de matériel, de logiciel ou toute autre chose. *Modder* permet d'obtenir des spécifications ou une apparence sur mesure et non prévue par ses concepteurs.

² Roth (Veronica) : *Divergent*, t. 1-4.

³ Seigne (Aude) : *Une Toile large comme le monde*.

question de façon très réaliste ; couper des câbles peut devenir la seule option pour sortir d'un réseau oppressant et hors de contrôle humain.

La question de modifier une conscience simplement en changeant le texte sur lequel elle repose est fascinante. Dans *Scythe*, les Faucheuses doivent programmer des requêtes complexes pour mettre à jour des faits historiques, des enregistrements et des archives, leur permettant de faire la lumière sur les réalités de leur société, et ce, malgré les limitations de la programmation de l'IA.

Nous avons un accès total à l'intégralité de ses connaissances, bien sûr. Les Faucheuses utilisent le Thunderhead pour d'innombrables tâches – mais pour nous, c'est une simple base de données. Un outil, rien de plus.¹

En effet, ses concepteurs espéraient faire disparaître les raisons de leur création, et mettre en avant le beau dans cette sur-intelligence synthétique plutôt que de conserver la mémoire des drames qui ont réellement accompagné cette mise en place.

Heureusement, pourrions-nous dire, que les Faucheuses ont bien eu accès à ces données, même si elles étaient partiellement cachées. En effet, la disponibilité est au cœur du propos : le Thunderhead est à même de manipuler l'immensité des données produites au quotidien par les sociétés humaines, mais aussi de les utiliser en second, en faire des statistiques pour dégager des tendances, par exemple, ou pour retrouver des archives ne semblant pas significatives pour une IA, mais l'étant pour un observateur humain.

Une critique du numérique contemporain

Comme souvent, la science-fiction peut nous servir à dégager une critique de notre réalité actuelle... et de la direction prise pour le futur immédiat². Cette position de critique permet ici principalement de comprendre les enjeux posés par une technologie encore en cours de construction. Cette représentation simplifiée d'une technologie complexe permet au jeune lectorat visé ici d'appréhender des formes de gouvernance potentiellement risquées et d'envisager une réflexion sur elles.

¹ « We have full access to its wealth of knowledge, of course. The Scythedom uses the Thunderhead for countless tasks—but to us, it's simply a database. A tool, nothing more. » (notre traduction) Shusterman (Neal) : *Scythe*, t.1, p. 155.

² Besson (Anne) : *Les Pouvoirs de l'enchantement : usages politiques de la fantasy et de la science-fiction*.

En effet, le numérique suit le même chemin que ces intelligences artificielles : il est facile de laisser la simplicité gagner sans avoir à comprendre le fonctionnement réel du logiciel appliqué. N'oublions pas également que nous dépendons de plus en plus des environnements propriétaires : que ce soit un choix conscient ou non. Ainsi, l'argument de la simplicité – de l'intuitivité – repose sur l'habitude de codes communs et usuels, comme défini par Anthony Masure¹ ou Alexander Galloway². Nous retrouvons, dans *Scythe*, une opposition déjà usuelle au quotidien : entre usage en programmation – en dur – et manipulation d'interfaces graphiques et d'environnements ergonomiques.

Les interfaces guident les publications qui y sont faites ; comme toute technique, elles viennent avec des contraintes. Mais celles-ci étant numériques, elles nous sont moins évidentes, parce que moins pressantes. Or, ces contraintes calibrent les contenus publiés et donc en partie, du moins, les réflexions menées. C'est ce que théorise Marcello Vitali-Rosati avec sa théorie de l'éditorialisation^{3 4 5}.

Ces fictions nous permettent aussi d'entrevoir les possibles détournements qui peuvent venir bouleverser les règles de programmation mises en place. L'exemple le plus frappant serait sans doute la place que prend la littérature dans les plateformes numériques : elle s'intègre peu à peu partout, sur des interfaces et des outils non pensés pour la création. Sur les réseaux sociaux, le partage de textes à fortes contraintes peut mener à des exercices d'écriture détournant ces normes d'usage de manière particulièrement créative. Le détournement tant des espaces que des syntaxes et cultures numériques permet donc d'ouvrir des espaces de créations infinis.

La culture numérique, au sens large, dans une acception qui pourrait être synonyme de culture contemporaine, met à l'honneur la débrouillardise et le bricolage, remplaçant les clous des Fab Labs par des lignes de commandes et des *hackathons* associatifs. Penser le rapport à la culture numérique étant souvent penser le rapport à la culture libre et à l'ouverture des données, les intelligences artificielles et les décisions que nous prenons actuellement dans leur constitution sont un moteur décisionnel important. Leur apprentissage de corpus comme

¹ Masure (Anthony) : *Design et humanités numériques*.

² Galloway (Alexander) : *The Interface Effect*.

³ Vitali-Rosati (Marcello) : « Qu'est-ce que l'éditorialisation ? ».

⁴ Vitali-Rosati (Marcello) : *On editorialization: Structuring Space and Authority in the Digital Age*.

⁵ Vitali-Rosati (Marcello) : *op. cit.*

ceux de *l'Anthologia Græca* pour les *Digital classics* ne sont pas loin de l'apprentissage par inclusion des *databases* des bibliothèques présentées dans *Scythe...* faut-il encore faire des choix éclairés et veiller à ne pas laisser de côté la réponse au fonctionnement des dites IA.

Bibliographie

- Besson (Anne) : *Les Pouvoirs de l'enchantement : Usages politiques de la fantasy et de la science-fiction*, Paris, Vendémiaire, 2021.
- Butler (Judith) : *Le Pouvoir des mots : discours de haine et politique du performatif*, Paris, Amsterdam, 2017.
- Chambers (Becky) : *A Closed and Common Orbit*, London, Hodder, 2017.
- Citton (Yves) : *Gestes d'humanités : anthropologie sauvage de nos expériences esthétiques*, Paris, Armand Colin, 2012.
- Derien (Maxime) : *Intelligence artificielle : rêves et réalité – Anthropologie des interfaces Homme/Machine*, <https://anthropo-ihm.hypotheses.org/2131> (consulté le 11 avril 2022).
- Fredriksson (Sylvia), Sauret (Nicolas) : « Fredriksson, Sylvia. Sauret, Nicolas. Écrire les communs. Au-devant de l'irréversible. », *Sens Public*, 2019, <http://sens-public.org/dossiers/1383/> (consulté le 30 juin 2022).
- Galloway (Alexander R.) : *The Interface Effect*. Cambridge, Polity, 2012.
- Gibert (Martin) : *Faire la morale aux robots : Une introduction à l'éthique des algorithmes*, Montréal, A10, 2020.
- Hayles (Katherine N.) : *My Mother Was a Computer: Digital Subjects and Literary Texts*, Chicago, University of Chicago Press, 2005.
- Lacroix-De Sousa (Sandie) : « L'approche par les risques adoptée par le projet de règlement sur l'Intelligence Artificielle de la Commission européenne », in *La Régulation des algorithmes en matière bancaire et financière*, Orléans, France, 2022.
- Landragin (Frédéric) : *Comment parle un robot ? Les machines à langage dans la science-fiction*, Saint-Mammès, Le Béliat, 2020.
- Lescouet (Emmanuelle) : « Penser l'archéologie comme destruction du dogme dans les littératures de l'imaginaire », in *Raconter l'histoire – Récits et Histoire dans les cultures de l'imaginaire*, Rennes, L'hypermonde, 2022.
- Lévy (Pierre) : *L'Intelligence Collective*, Paris, La Découverte, 1997.
- Masure (Anthony) : *Design et humanités numériques*, Paris, Éditions B42, 2017.
- Noyer (Jean-Max), Carmes (Maryse) : « Le mouvement "Open Data" dans la grande transformation des intelligences collectives et face à la question des écritures, du web sémantique et des ontologies », *Sciences de l'Homme et de la Société*, 2012.

- Penny (Simon) : « Embodied Cultural Agents: At the Intersection of Robotics, Cognitive Science and Interactive Art », pp. 103-105, in *AAAI Socially Intelligent Agents: Papers from the 1997 Fall Symposium*, Menlo Park, AAAI Press, 1997.
- Rumpala (Yannick) : *Cyberpunk's not dead : laboratoire d'un futur entre technocapitalisme et posthumanité*, Saint-Mammès, Le Béalial', 2021.
- Shusterman (Neal) : *Scythe*. Arc of Scythe, book 1, New York, Simon & Schuster, 2016.
- : *Thunderhead*. Arc of a Scythe, book 2, New York, Simon & Schuster, 2018.
- : *The Toll*. Arc of a Scythe, book 3, New York, Simon & Schuster, 2019.
- Saint-Gelais (Richard) : *L'Empire du pseudo : Modernités de la science-fiction*, Montréal, Nota Bene, 1999.
- Vint (Sherryl) : *Science fiction*, Cambridge, Massachusetts, The MIT Press, 2021.
- Vitali-Rosati (Marcello) : « Qu'est-ce que l'éditorialisation ? », *Sens Public*, 2016.
- Vitali-Rosati (Marcello) : *On Editorialization: Structuring Space and Authority in the Digital Age*, Amsterdam, Institute of Network Cultures, 2018.
- Vitali-Rosati (Marcello) : « Qu'est-ce que l'écriture numérique ? », *Corela. Cognition, représentation, langage*, 2020.