

Louis de Diesbach

Préface de Luc de Brabandere

« BONJOUR CHATGPT »

Comment l'intelligence artificielle
change notre rapport aux autres



Louis de Diesbach

Préface de Luc de Brabandere

« BONJOUR CHATGPT »

Comment l'intelligence artificielle
change notre rapport aux autres



MARDAGA

Louis de Diesbach
Préface de Luc de Brabandere

« BONJOUR CHATGPT »

Comment l'intelligence artificielle
change notre rapport aux autres

MARDAGA

À Clémence, toujours.

**CALIGULA : *C'est une vérité toute simple
et toute claire, un peu bête,
mais difficile à découvrir et lourde à porter.***

HÉLICON : *Et qu'est-ce donc que cette vérité, Cæus ?*

**CALIGULA : *Les hommes meurent
et ils ne sont pas heureux.***

ALBERT CAMUS, *Caligula*[1](#)

Préface

« Anthropomorphisme » ! Le mot est peu élégant et c'est regrettable, car le mot est important et particulièrement actuel, comme le montre fort bien Louis de Diesbach dans le livre que vous tenez en main.

L'anthropomorphisme qualifie une attitude fréquente chez l'homme ou la femme qui consiste à voir de l'humain là où il n'y en a pas. Cette posture s'observe depuis longtemps et dans de nombreuses circonstances, elle peut concerner tout aussi bien des outils que des phénomènes météorologiques, des lieux ou même des dieux.

L'anthropomorphisme est aussi vieux que... l'*anthropos*.

On l'observe en Égypte à l'époque des pharaons ou encore dans le monde grec à l'époque de la guerre de Troie. La « raison », le *logos* de l'époque est surtout *mythos-logos*, l'homme se croit soumis aux humeurs de dieux multiples aux comportements bizarres et aux allures humaines qui oscillent entre amour et haine, entre joie et tristesse.

Pour l'habitant du pourtour méditerranéen, c'est dans l'Olympe qu'il faut chercher l'explication de ce qui se passe sur la Terre, qu'il s'agisse d'un arc-en-ciel, d'une épidémie ou d'une guerre. S'il tonne, c'est que Zeus tape sur la table. S'il pleut, c'est qu'une déesse ne peut sécher ses larmes.

Ailleurs dans le monde, ce n'était pas fort différent. Comme nous le rappelle Bertrand Russell, les Chinois croyaient que le chien céleste avait parfois une petite fringale et qu'il croquait alors un morceau de lune. C'est en tapant vigoureusement sur des gongs que les habitants faisaient fuir ce dangereux

animal céleste et mettaient ainsi un terme... à l'éclipse.

Pendant longtemps, *Homo sapiens* a voulu attribuer une forme de vie, voire un esprit ou une âme, aux objets. Le fer rouille parce qu'il est malade et il faut tout faire pour redonner le moral au saule pleureur.

L'attitude anthropomorphique est celle d'un sujet qui se projette sur les objets qui l'entourent.

Dans un premier temps, il ne s'agit que d'une métaphore innocente. Le salon est accueillant, la musique est apaisante, le vin est sympathique.

Avec le développement des objets mécaniques, les utilisateurs poussent la figure de style un cran plus loin. La voiture négocie gentiment un méchant virage, mais ce n'est pas pour cela qu'elle décide où elle va. La métaphore et la voiture restent sous contrôle.

Aux objets mécaniques, nous n'avons jamais prêté la moindre intention. On ne dit pas d'un thermostat qu'il « désire » garder une température constante dans une pièce, ni d'un siphon qu'il « souhaite » empêcher les mauvaises odeurs. Non, on se réjouit simplement de voir ces outils faire correctement ce pour quoi ils ont été conçus. Nos lunettes nous aident certes à mieux voir, mais personne ne leur dit pour autant « Merci » !

Avec l'arrivée des objets informatiques, c'est comme si brusquement un seuil avait été franchi. Nous serions tout à coup projetés dans une relation d'égal à égal avec les machines ou, pire encore, nous passerions d'un statut d'utilisateur à celui d'utilisé. On entend parfois dire que « le cerveau, c'est un peu comme un ordinateur » ! Ce serait comme dire : « Un plombier, c'est un peu comme un fer à souder. »

Un ordinateur peut certes imiter de mieux en mieux les émotions, mais il reste avant tout un tas de pièces métalliques ou plastiques alimentées par de l'électricité. Un algorithme peut impressionner par sa capacité à simuler les comportements humains, mais cela ne lui donne pas pour autant la moindre humanité. Au carnaval du numérique, les robots sont masqués.

Réaffirmons-le avec vigueur : on n'échange pas plus avec un ordinateur qu'avec un grille-pain. La raison d'être d'un outil numérique est de pouvoir amplifier les gestes intellectuels, parfois de manière impressionnante. Mais cela ne change en rien son statut d'objet technique.

On ne devrait donc jamais parler de « l'intelligence artificielle », mais plutôt d'intelligences artificielles. Il faudrait de plus les appeler par leur nom, Alexa, Watson. Ou ChatGPT.

Un GPS ne « dit » pas de tourner à gauche. Non, dans un GPS, un ensemble

de données sont synthétisées et converties en un signal qui imite la voix. Une caméra de surveillance ne « reconnaît » aucun visage. Non, des personnes ont un jour décidé de mettre des caméras partout pour savoir qui est où, et d'en garder une trace.

Une calculatrice ne calcule pas plus qu'une foreuse ne fore. Faire une addition ou faire un trou est plus facile avec une machine, c'est évident, mais la démarche reste celle d'une personne qui a une idée, un besoin, un projet, un plaisir, une raison, une obligation ou une envie de faire ce qu'elle fait, autant de choses qu'une machine ne peut avoir.

Tout comme une foreuse ne sait pas ce qu'est un trou, une calculatrice ne sait pas ce qu'est un résultat. Ce qui apparaît à l'écran ne prend son sens que par rapport au projet de l'être humain qui calcule, car les chiffres ne parlent jamais « d'eux-mêmes ». Ils ne parlent même pas du tout. Big Data est muet. Toute analyse de données se fait suivant des prismes, des modèles, des catégories choisis a priori par un être humain. À toute organisation de l'information correspond une idéologie sous-jacente.

Un ordinateur qui joue aux échecs ne sait pas ce que veut dire « jouer », ni ce que veut dire « échec ». Il ne sait pas ce qu'est un échiquier, ni pourquoi il joue. Il n'a pas plus « envie » de jouer que de gagner, et ne peut être déçu de perdre. Il n'a pas la moindre idée de ce qu'il « apprend », il n'a d'ailleurs pas d'idée du tout.

Un ordinateur ne pense pas et ne comprend pas ce qu'il fait. Car pour vraiment comprendre quelque chose, il est nécessaire d'associer l'intellectuel et le sensoriel. On ne peut comprendre ce qu'est le feu si on ne s'est jamais brûlé, on ne peut comprendre ce qu'est la tristesse si on n'a jamais pleuré.

Arrêtons donc d'attribuer aux machines l'intelligence que nous déployons en les construisant !

Luc de Brabandere
Philosophe d'entreprise

Introduction

« Que dis-tu, ChatGPT ? »

May Ashworth. Ce nom ne vous dit probablement pas grand-chose. Pour être honnête, c'est assez normal, Mme Ashworth n'ayant pas accompli quoi que ce soit qui, à notre connaissance, ne mérite d'entrer dans les livres d'histoire, et l'anecdote dont elle est l'héroïne a sans doute la valeur de son patronyme^a. Mais, sans le vouloir, May Ashworth nous permet de saisir le contraste saisissant que quelques années, et l'arrivée d'une technologie révolutionnaire, peuvent avoir sur notre perception des choses.

En 2016, cette grand-mère de 86 ans cherche la signification en chiffres arabes du nombre « MCMXCVIII ». Moderne, elle ouvre son navigateur et entre ces mots dans Google : « S'il vous plaît, pourriez-vous me traduire ce nombre romain MCMXCVIII, merci. » Ben, son petit-fils, prend une photo de la requête et la partage sur ce qui s'appelait encore Twitter en taquinant son aïeule qui pose si poliment sa question au moteur de recherche. C'est le genre d'histoires dont Internet raffole : le tweet fait le buzz, est repartagé des milliers de fois, jusqu'à arriver aux oreilles de Google qui ne peut s'empêcher de se fendre d'un merci, et d'une réponse : « C'est 1998 ! » Questionnée sur ses si délicates manières, May Ashworth expliquera qu'elle pensait s'adresser directement à quelqu'un au bureau de Google où de nombreux individus devaient, chaque jour, répondre aux nombreuses questions envoyées par les internautes (elle ignorait sans doute que le nombre de ces requêtes dépassait le millier de milliards). Loin de toute volonté d'anthropomorphiser la machine, la douce grand-mère s'adressait à un ordinateur comme on s'adresserait à un être humain parce qu'elle était persuadée que c'était ce qui se produisait, comme lors d'un appel à un service client. La Toile de l'époque ne put s'empêcher de se moquer gentiment de cette vieille dame qui parlait avec une machine.

Faisons un bond dans le futur d'à peine plus de six ans. Fin novembre 2022, la société OpenAI sort, avec tambours et trompettes, son agent conversationnel ChatGPT, basé sur le modèle GPT3.5^b. En quelques semaines, le nombre d'utilisateurs de cette étrange machine s'élève à plus de cent millions : par curiosité, pour être plus efficaces dans leur travail, pour poser des questions, les utilisateurs testent cet outil incroyable, cette « intelligence artificielle » (IA) qui vient s'immiscer au plus proche d'eux. Il y a quelque chose qui lie ces millions d'individus et notre chère May Ashworth : une grande majorité d'entre eux lancent leurs conversations par « Bonjour ChatGPT », et ponctuent leurs échanges de « S'il te plaît » et « Merci ». Si Ashworth était gentiment moquée, il s'agit ici d'un sujet très sérieux et qui permet à des millions de personnes de réaliser les prouesses de cette IA dont ils entendaient parler depuis quelques années mais dont ils ne savaient pas grand-chose. Certes, on savait qu'un ordinateur pouvait battre un humain aux échecs, peut-être même qu'il pouvait l'emporter au jeu de go, mais à part ça, le domaine semblait réservé aux ingénieurs et aux informaticiens.

Puis vint novembre 2022 et l'arrivée de l'IA dans le salon, la chambre, et la vie de tout le monde. Il est désormais possible d'avoir une véritable conversation avec une IA, celle-ci répondant comme un être humain, si ce n'est mieux, et semblant capable de nous comprendre. Après l'agitation, la course à la *punchline* sur les dangers et risques de cette nouvelle machine, après les envolées des cours de Bourse et les milliards investis vient, peut-être, le temps de la philosophie et d'une réflexion éthique.

La technologie GPT, pour *Generative Pre-trained Transformer*, n'est pas neuve. Si ChatGPT est basé sur GPT3.5, c'est qu'il y a eu, avant, et pour les initiés, un GPT1, GPT2, GPT3 – et, depuis, un GPT4. Son entrée fracassante dans le quotidien de millions de personnes nous force cependant à (re)penser son impact dans nos vies. On ne peut que se rappeler la claque que fut, en 2013, la sortie du film *Her*, réalisé par Spike Jonze, dans lequel un homme, Theodore Twombly, tombe éperdument amoureux d'un programme d'IA, Samantha, remarquablement doublée par Scarlett Johansson. Petit à petit, il s'isole et trouve son bonheur dans une relation avec ce qui n'est, finalement, qu'un programme informatique mais qui lui parle (et lui répond) si bien (comme le ferait ChatGPT) qu'on aurait envie de croire à la pureté de ses sentiments – et Twombly est d'ailleurs très heureux d'y croire, lui qui sort d'un divorce difficile et qui ne parvient pas à rencontrer de nouvelles partenaires. Mais le film pose la question directement : « Ces sentiments sont-ils réels ou bien n'est-ce que du code ? »

La même année, un autre épisode de science-fiction vient faire frissonner les

spectateurs avides de projections dystopiques. « *Be Right Back* », le premier épisode de la deuxième saison de *Black Mirror*, met en scène les personnages de Martha Powell et Ash Starmer, un couple éperdument amoureux. Malheureusement, dès les premières minutes, Ash est tué dans un accident de voiture. Écrasée par le chagrin, Martha se décide à essayer une nouvelle technologie lui permettant de converser avec une IA basée sur la personnalité de son compagnon – ce que l’on appelle aujourd’hui un *deadbot*. Elle s’adresse à lui avec des « Bonjour » et lui dit « S’il te plaît » et « Merci », comme le font tant de personnes avec ChatGPT. Lorsqu’une nouvelle version permet même à Ash d’avoir un corps, c’est pour Martha l’occasion rêvée de reprendre sa relation là où elle l’avait laissée et de reconstruire sa vie avec cet androïde de l’homme de sa vie. Mais l’utopie s’estompe rapidement et la force avec laquelle Martha se plonge dans cette relation n’a d’égale que celle avec laquelle elle en est écœurée avant de se rendre compte qu’Ash ne peut pas, en réalité, être « *right back* ».

Aujourd’hui, des études montrent qu’un simple regard d’un robot, et sans doute alors un échange de quelques mots, peut déclencher une émotion chez l’être humain, un comportement altruiste, une envie d’entrer en relation². Que peuvent alors générer des heures de discussion, où l’humain se livrerait complètement à sa nouvelle amie, se mettrait en danger et s’ouvrirait comme il ne le ferait peut-être jamais face à un autre humain ? C’est bien la question que veut poser cet essai ainsi que celle, peut-être plus importante encore, de comprendre ce qui passe derrière un simple « Bonjour ChatGPT ». Que se passe-t-il quand nous saluons la machine, quand nous lui parlons comme nous parlons à nos semblables ? Quel projet de société se met en place ?

Pour tenter de répondre à cette question, il nous faudra d’abord comprendre pourquoi il est si aisé de s’adresser à une machine, et comment, entre autres, les récits de science-fiction ont rendu si perméables les frontières entre le bon film, la bonne série, et la vraie vie. Entre *Black Mirror*, *Terminator* ou encore *Her*, c’est en fait toute une nouvelle réalité qui se dessine. Comment, alors, dissocier la machine de l’individu ? Sommes-nous vraiment égaux, du point de vue de l’intelligence ? Comment tester nos similitudes, et nos différences, de raisonnement et de vision du monde ? D’Alan Turing à Steve Wozniak, tout le monde y est allé de sa petite expérience pour tenter de répondre à cette étrange question : la machine est-elle l’équivalent de l’humain ?

Quelle que soit finalement la réponse, il sera important de comprendre les différents mécanismes derrière ce « Bonjour » que nous adressons au *chatbot* d’OpenAI : ceux qui régissent la machine, bien sûr, mais aussi ceux qui nous poussent à l’anthropomorphiser. On peut se demander si May Ashworth, par son

grand âge et son manque d'affinité et de culture techniques, a peut-être plus de chance de considérer un ordinateur comme une personne, ou si, au contraire, l'âge n'est pas un facteur qui pousserait, ou non, à l'anthropomorphisme. On questionnera les limites de ce mécanisme psychologique et les situations où, évidemment, parler à un ordinateur ne suffit peut-être pas.

Ces discussions nous mèneront alors au nœud du problème et de notre questionnement : que se passe-t-il quand nous anthropomorphisons ? Qu'est-ce qui est en jeu dans cette étrange conversation que nous avons avec une machine mais que nous n'osons parfois pas avoir avec notre frère, notre amie, notre famille ? La technique n'est pas neutre, elle ne l'est jamais, et elle porte toujours en elle un projet de société : quand nous saluons ChatGPT, quelle est la nature de ce projet ?

Le lecteur voudra alors, et c'est bien normal, qu'on esquisse des pistes d'exploration d'une éthique des *chatbots*. Le philosophe Gilbert Hottois disait que l'éthique de la technique se trouvait toujours à cheval entre des questionnements métaphysiques et des problèmes de philosophie politique. Ce sont là deux voies qu'il nous faudra explorer en proposant de renverser l'histoire de notre rapport à autrui pour remettre la relation, la conversation et l'échange au cœur de notre rapport au monde. À l'extase de *Homo deus*, un *Homo socius* – un homme du lien – nous semble préférable, et une évolution de la technique dans ce sens est cruciale pour l'avenir de notre vivre-ensemble.

Ces questions peuvent sembler triviales, et on peut se dire que, finalement, ce n'est qu'un outil à qui on peut demander d'écrire dix phrases sur les ânes dans le style du rappeur Booba, ou bien qu'il s'agit d'une excellente boîte à idées capable de générer dix nouveaux pitches de start-up à la seconde. C'est rigolo et pratique, certes, et ChatGPT s'exécutera sûrement sans broncher, mais l'IA n'est pas, n'est plus, qu'une affaire d'innovation insouciantes pour informaticiens en quête d'action. Début 2023, un jeune Belge s'est donné la mort après plusieurs semaines de discussion intense avec un *chatbot* reposant sur une architecture similaire à celle de ChatGPT. Son épouse est persuadée que sans ces échanges avec une IA, son mari serait toujours des nôtres³. Par ailleurs, le Japon a même un mot pour désigner des jeunes qui préfèrent vivre reclus chez eux, à communiquer avec des personnages de jeux vidéo et des agents conversationnels, et à éviter la confrontation que peut représenter une vraie conversation : les Hikikomori. Penser, ou repenser, un vivre-ensemble et un projet politique est dès lors crucial lorsque les interactions humains-machines peuvent prendre le dessus sur celles avec nos semblables. Derrière le « gag » d'un « Bonjour ChatGPT » se joue en réalité bien plus que quelques lettres

tapées à toute vitesse sur un clavier.

En 1979, le philosophe Hans Jonas nous avertissait déjà sur « l'énormité des forces de la technique » qui imposait « à l'éthique une nouvelle dimension de responsabilité jamais imaginée auparavant »⁴. Près d'un demi-siècle plus tard, ces forces technologiques ont crû de manière exponentielle et chaque « Bonjour ChatGPT » nous remet, d'une certaine façon, face à nos responsabilités. Quand on aperçoit les diverses conséquences d'un projet technique des plus voraces – isolement, dépression, suicide, mais aussi perte de dignité, effritement de notre liberté ou de notre rapport à autrui –, il apparaît crucial d'apporter une réponse à cette étrange inversion : en traitant les robots comme des humains, ne finissons-nous pas par traiter les humains comme des robots ?

^a Littéralement, *ash-worth* peut se traduire par « valant une cendre ».

^b Le modèle GPT3.5 est un modèle de langage développé par OpenAI. Comme son nom l'indique, il s'agit d'une amélioration de la troisième version de ce modèle.

Chapitre 1

« D'où viens-tu, ChatGPT ? »

Entre réel et fiction

Il serait sans doute possible d'écrire plusieurs livres sur les origines et l'évolution de l'anthropomorphisme et sur cette volonté quasi biblique de l'être humain de vouloir, à son tour, créer la vie. L'Histoire donne sans doute raison à Serge Tisseron, lorsque celui-ci disait que « l'être humain est une créature dont la préoccupation narcissique n'est jamais satisfaite⁵ » et on trouve des traces d'anthropomorphisme sur certaines sculptures qui ont plus de quarante mille ans. Il s'agit généralement de statuettes en ivoire personnifiant des animaux ayant des postures humaines et qui nous rappellent que, déjà, l'homme se représentait le monde via son propre prisme, ramenant tout à sa propre nature.

Nous pourrions ouvrir notre chapitre un peu plus tard, à l'Antiquité, avec la légende de Pygmalion ou bien avec celle de Dédale. Le premier, un sculpteur grec descendant d'Athéna et d'Héphaïstos, projette des traits humains sur une statue, Galatée, une somptueuse jeune femme d'ivoire et de volupté. C'est un célibataire endurci, meurtri par les mœurs des habitants de son île, qui s'enferme dans son art comme dans sa solitude. C'est alors que la déesse Aphrodite, touchée par la sensibilité et l'amour sincère du sculpteur pour sa création, décide de donner vie à Galatée pour qu'elle puisse épouser son créateur. La légende rappelle le génie humain, la bonté des dieux (eux-mêmes profondément anthropomorphisés) et l'amour pour la création.

Dédale, quant à lui, est souvent pensé comme étant un des premiers chercheurs transhumanistes, poursuivant ses travaux et ses expériences loin du regard des habitants de l'Olympe. Il en viendra même, on connaît l'histoire, à se

créer et se greffer des ailes pour échapper au labyrinthe dans lequel le roi Minos l'avait enfermé. Son fils, Icare, s'approcha trop dangereusement du soleil et endommagea la qualité de son excroissance technique, l'emmenant dans une chute mémorable, rappel bienvenu du danger de l'hubris humain. Son père poursuivra son voyage ailé jusqu'en Sicile, où il trouvera le repos après avoir enseveli son fils. L'inventeur athénien peut être vu comme un précurseur du progrès technique pour sauver l'humanité, voire comme un penseur post-humaniste très en avance sur son temps.

Nous aurions également pu nous pencher sur la mythologie juive et la légende du golem qui narre la création mystique d'un humanoïde d'argile dont le but premier est de protéger ou d'assister son créateur. Quatre sages auraient modelé cet être avec de la terre glaise et on retrouve ici une histoire similaire à celle de Pygmalion : l'humain qui crée la vie. Qu'il s'agisse, dans le cas du sculpteur antique, d'une compagne pour combler la solitude et le célibat ou, dans le cas de la mystique juive, d'un être protecteur, les créations humaines n'ont pour but premier que la servitude et la soumission à leur prétendu divin créateur.

Si ces mythes et légendes forgent indubitablement la vision que nous avons de l'anthropomorphisme et des créatures que nous amenons à la vie, c'est pourtant un peu plus tard que nous voyons un véritable déclenchement d'une technique profondément anthropomorphique, toujours au confluent de la fiction et du réel. Cette perméabilité entre fable et réalité ne peut être mieux incarnée que par Lord Byron, considéré comme l'un des plus illustres poètes anglais, à cheval sur les XVIII^e et XIX^e siècles. Si l'œuvre de Byron mérite certainement qu'on s'y intéresse davantage, c'est plutôt une coïncidence, une de ces « ruses de l'Histoire », qui nous pousse aujourd'hui à évoquer son nom. En effet, l'auteur de *Childe Harold's Pilgrimage* se trouve au croisement de deux existences magistrales, féminines, dont le génie a certainement largement dépassé celui du jeune poète : Ada de Lovelace et Mary W. Shelley.

La première est née Ada Byron, fille de Lord Byron, en 1815, et n'a reçu la reconnaissance qu'elle mérite que tardivement. Fait peu commun pour l'époque, Ada reçut une éducation approfondie en mathématiques et en sciences pour lesquelles elle démontrait une habileté remarquable. Elle poursuivit ses travaux en calcul différentiel puis se dévoua entièrement à travailler sur la machine analytique développée par le mathématicien et inventeur Charles Babbage, une étrange mécanique qui porte les prémices de ce qui deviendra l'ordinateur. C'est précisément pour cette machine qu'Ada Lovelace avait rédigé ce que l'on peut appeler un programme devant être exécuté par l'outil, faisant ainsi de la mathématicienne britannique la première « programmeuse » de l'Histoire. Si son

nom s'est quelque peu perdu au fil des années, la chercheuse revint sur le devant de la scène avec les récentes découvertes et avancées en informatique^c et récupéra la place de pionnière qui lui était due.

Lord Byron était également proche de Mary W. Shelley, une autre pionnière de l'époque mais dans un genre différent : la littérature. C'est en effet en 1818, soit tout juste au moment de la naissance de la jeune Ada, que Shelley publie un livre qui a traversé les frontières et les âges : *Frankenstein*. L'histoire d'un esprit brillant qui, obnubilé par sa quête narcissique de créer la vie, donne naissance au monstre que l'on connaît. La genèse de cette œuvre a vu le jour lors d'un séjour à Genève que Mary Shelley et son époux ont passé, entre autres, en compagnie de Lord Byron. Au cours d'une soirée maussade de l'été 1816, ce dernier propose à ses amis de rédiger une histoire glaçante et terrifiante, donnant ainsi naissance au futur chef-d'œuvre de Shelley : une histoire d'un monstre et de son créateur – ou d'un monstre et de sa création –, où chacun côtoie la mort, la violence, mais aussi la douceur d'une après-midi de lecture et la chaleur d'une bonne compagnie. En créant ainsi la vie, Victor Frankenstein nous offre une réflexion sur ce qui lui donne la peine d'être vécue, et nous interroge sur la nécessité technique d'un progrès toujours plus immense et déifiant.

Lord Byron et, avec lui, le XIX^e siècle nous ouvrent la porte sur ce qui ne peut manquer de nous étonner : une étrange fluidité entre la fiction et le réel, entre la fable et la technique, entre le mythe et le progrès. Déjà à l'époque,

*le docteur Frankenstein incarn[ait] le paradigme de l'homme moderne pour lequel le rêve de la raison était déjà en train de devenir, aux yeux de la jeune romancière et de ses amis révolutionnaires romantiques, le cauchemar de l'esprit.*⁶

Les années passent et ce mauvais rêve ne fait que s'amplifier et se concrétiser. Aujourd'hui, chaque « Bonjour ChatGPT » porte en lui cet équilibre délicat entre ce rêve de la raison et cette angoisse de l'esprit. Une harmonie toujours complexe entre progrès et innovation.

Deux événements, peut-être, peuvent nous permettre de donner encore quelques éléments de contexte. Tout d'abord, le XIX^e est également l'ère de la révolution industrielle et de son accélération fulgurante du progrès technique puis technologique. De la machine à vapeur au fordisme en passant par le téléphone et le moteur à explosion, c'est toute une industrie, et, avec elle, un projet de société, qui se mettent en place. Concomitamment, et sans doute pas par hasard, le philosophe Friedrich Nietzsche déclare que « *Gott ist tot*^d », ouvrant ainsi une lutte sans merci pour être digne de prendre sa place. « Pour un homme qui aime le pouvoir, la rivalité des dieux a quelque chose d'agaçant⁷ »,

une fois que le divin est éliminé, il ne reste que le potentiel de se hisser, par la technique et le « génie humain », au même rang que ceux que l'on vient de supprimer. Que sommes-nous d'autres, d'ailleurs, aux yeux anthropomorphisés de ChatGPT, que des démiurges à qui ce modèle de langage doit la vie ? Comme le monstre de Frankenstein qui voue une admiration mêlée de crainte, de haine et de reconnaissance à son géniteur, la machine doit certainement voir les ingénieurs qui l'ont mise au point comme des dieux bienfaisants et tout-puissants.

Si l'on veut comprendre l'anthropomorphisme, c'est donc dans ce contexte qu'il faut se fonder : là où le vrai ne traîne jamais loin de la fiction et où l'un semble nourrir l'autre, et inversement. Après la mort de Dieu, les êtres humains se rêvent en insuffleurs de vie afin que de nouveaux esclaves viennent servir les nouveaux maîtres qu'ils se rêvent d'être. En 1920, le dramaturge et journaliste tchécoslovaque Karel Čapek donne naissance à un néologisme étonnant : robot, inspiré du tchèque *robota* (corvée) lui-même tiré du slave *rob* (esclave). Dans sa pièce *R.U.R.*, des êtres humains vivant dans un futur pas si lointain sont assistés par des humanoïdes fabriqués dans une usine à proximité. Couronnant une sorte de progrès technologique et technophile, les humains de la pièce portent l'espoir de toute une époque : prendre le pas sur la vie, en ôter la pénibilité et être assistés par des individus mécaniques que les géniales capacités humaines ont mis au monde. Quelques années plus tard, l'écrivain Philip K. Dick publie *Les androïdes rêvent-ils de moutons électriques ?*, une nouvelle qui sera ensuite adaptée au cinéma sous le titre de *Blade Runner* avec Harrison Ford dans le rôle principal. Portés par ce qu'on appelle le « complexe de Frankenstein^e », les auteurs de fiction brossent le portrait d'un monde généralement dystopique et négatif du progrès technique, dans la digne lignée de l'esprit révolutionnaire romantique de Mary Shelley. De *Matrix* à *Terminator* en passant par *I, Robot*, ou encore *Westworld*, le cinéma occidental nous propose un univers quasi apocalyptique, où la victoire de la technologie s'associe nécessairement à une certaine destruction de l'humanité – l'homme n'est-il capable que du pire quand il en vient à créer une vie artificielle ?

Cependant, au cours du xx^e siècle, la fiction garde une large avance sur le réel. Les progrès en apprentissage machine, ce mécanisme qui est à la base des avancées en IA, reposent sur les évolutions de trois facteurs déterminants : la quantité de données d'entraînement disponible, la sophistication des algorithmes et la puissance de calcul. Ces améliorations se font petit à petit et il faudra traverser plusieurs périodes sombres, également appelées « hivers de l'IA », pour que la technique, suivant méticuleusement la loi de Moore, donne à l'IA la

puissance nécessaire à renverser le monde – un jour de mai 1997.

Plusieurs moments peuvent être considérés comme l'entrée dans une nouvelle ère de l'IA, nous retiendrons ici la victoire de *Deep Blue*, le superordinateur d'IBM, contre Garry Kasparov, alors champion du monde en titre d'échecs, le 11 mai 1997. Il est surprenant que « cet enjeu dérisoire : acculer un roi en bois dans le coin d'une planche en bois⁸ » ait pris une telle ampleur et une telle symbolique dans la compétition entre les intelligences humaine et artificielle. Et pourtant, c'est peut-être justement parce que ce jeu « exclut souverainement la tyrannie du hasard et ne couronne que l'esprit⁹ » que la victoire de la machine va, à ce point, secouer le monde entier. Un jeu couronnant le génie humain venait d'être remporté par une machine, un *rob(ot)* – ce qui n'était hier d'un esclave.

Alors que, pendant plusieurs décennies, chaque progrès en informatique se faisait à la sueur du front de plusieurs génies après de nombreuses et difficiles années de travail, il est devenu aujourd'hui pratiquement impossible de suivre les avancées de ce secteur en plein boum. Pas un mois, pas une semaine, pas une journée même, ne passe sans qu'un nouveau chercheur ne mette au point au moins un algorithme qui viendra bousculer le *statu quo*, rendant la perméabilité fiction/réel de plus en plus imperceptible. Tout ce que l'individu pensait être propre à l'humanité s'est trouvé réapproprié par une technique toujours plus impressionnante, floutant de plus en plus la frontière entre la légende, le rêve, le vrai.

Dans le film *L'homme bicentenaire*, sorti en 1999, un androïde du nom d'Andrew, interprété par l'excellent et regretté Robin Williams, sculpte une petite poupée en bois pour sa famille d'adoption. Quelle n'est pas la surprise des membres de cette famille devant un robot capable d'une telle créativité ! Même si *Deep Blue* a battu Kasparov il y a peu, on est loin d'imaginer qu'une machine soit capable de production artistique ou créative – cette compétence n'est-elle pas le propre d'un regard humain sur le monde ? Ne faut-il pas des émotions, du ressenti, pour pouvoir incarner ces vibrations dans une œuvre créatrice ? Il aura pourtant fallu moins de vingt ans pour que soit vendue aux enchères la première production picturale réalisée par une IA : le *Portrait d'Edmond de Belamy* est cédé pour 432 500 dollars le 25 octobre 2018, bouleversant le monde de l'art¹⁰. Quelques années plus tard, un cataclysme encore plus important : Jason Allen remporte la foire d'art du Colorado, dans la catégorie « Artistes numériques émergents », avant de reconnaître que son œuvre n'est pas de lui mais du logiciel Midjourney¹¹. Au-delà, donc, d'une production artistique, il s'agit d'une production *meilleure* que celles d'artistes pleinement humains, générant surprise,

colère et aberration plus importantes encore que celles de la famille Martin devant la petite statuette d'Andrew. La fiction était-elle alors prophétique d'une réalité qui ne manquerait pas de voir le jour ?

Tout comme l'art, le management et la gestion sont également des domaines qui se veulent proprement humains. On ne compte plus les films où des robots investissent des entreprises en se disant qu'il s'agit là de bons divertissements jusqu'à ce que, en 2022, l'entreprise chinoise NetDragon Websoft ne nomme le programme d'IA Tang Yu comme PDG, à la stupeur du grand public et à la grande satisfaction des marchés¹². Ce qui ressemble au scénario d'un film de science-fiction est donc devenu la réalité pour les employés de cet éditeur de jeux vidéo.

Quand on se demande jusqu'où la technologie peut aller, la barrière la plus souvent invoquée est celle des sentiments, de l'amour, de l'affection. Quelle horreur que ces qualités puissent sortir du giron strictement humain et être « algorithmisées ». L'industrie du divertissement s'est emparée de cette crainte dans de nombreuses productions : on pensera notamment à l'excellent *Her*, dans lequel un individu tombe éperdument amoureux d'un agent conversationnel, ou bien à l'épisode « *Be Right Back* » de la série *Black Mirror*, dans lequel une jeune femme converse avec son petit ami décédé grâce à un programme informatique. Ces deux productions, sorties en 2013, ont toutes deux été acclamées par la critique et louées pour l'avertissement qu'elles donnaient au monde de la tech : comment savoir quand on va trop loin, qu'est-ce qui doit rester de l'ordre de la fiction ?

En 2023, après la sortie tonitruante de ChatGPT, ces deux œuvres de fiction ressemblent plus à des bandes-annonces de notre futur. D'une part, l'application Replika, qui propose à des individus de se créer des amis-*chatbots* (et pourquoi pas des conjoints-*chatbots*), a été téléchargée plus de deux millions de fois¹³. Le logiciel se présente comme un compagnon artificiel avec qui il est possible de simplement partager sa journée, ses tracas, ses envies, etc., un peu comme le fait Theodore Twombly dans le film *Her*. D'autre part, en Chine, des individus utilisent des technologies similaires pour créer des *deadbots*, ou *griefbots* : des agents conversationnels de personnes décédées avec qui l'on peut échanger pour, entre autres, faciliter son deuil¹⁴. *Black Mirror* fait carton plein.

Comme le rappelait l'anthropologue Joffrey Becker, « l'imitation de qualités sociales ou de la cognition humaine est un problème fondateur de la recherche en intelligence artificielle et en robotique dite sociale¹⁵ », et plus cette imitation se parfait, plus il est difficile pour un individu de faire la part des choses entre fiction et réalité. Tout autour de lui se pare des attributs de la technique – art,

business, relations avec les vivants et même les morts –, comment ne pas être chamboulé quand le réel ressemble tant à une série dystopique qu'on a regardée la veille, ou à un livre critique et cauchemardesque dévoré le mois précédent ? Quand on dit « Bonjour ChatGPT », comment savoir si l'on nage en pleine utopie ou dans un enfer technologique ? Comment ne pas être une May Ashworth, sans plus trop savoir quelle limite et quelle frontière poser entre l'être humain et la technique ?

Entre tests et expériences

Face à une telle perméabilité entre la fiction et ce que nous pouvons expérimenter dans notre vie de tous les jours, comment appréhender notre rapport avec le réel ? Comment vivre sans être perpétuellement forcé de remettre en doute l'identité de chaque interlocuteur ? Une IA est-elle comparable, assimilable, identique à une intelligence humaine ? C'est une question que se sont posé de nombreux chercheurs, philosophes et ingénieurs, chacun proposant une nouvelle expérience de pensée ou un nouveau test qui permettrait de trancher entre intelligences humaine et artificielle.

Lorsqu'on échange avec un *chatbot* ou un robot, les zones du cerveau qui s'activent sont les mêmes que pour des interactions avec les humains¹⁶ et il est alors aisé de se laisser hypnotiser par un échange avec un agent de la firme Replika, ou un *deadbot* de votre grand-mère : cela en fait-il de vraies personnes ? Sommes-nous pour autant en train d'avoir des discussions réelles qui auraient une valeur identique, si ce n'est supérieure, à celle qu'on pourrait avoir avec un autre membre de notre espèce ? Par ailleurs, toutes les IA se valent-elles, ou bien certaines sont-elles plus performantes que d'autres ? Siri, l'agent conversationnel disponible sur tous les téléphones Apple, est-il du même niveau que Samantha, l'IA du film *Her* dont le protagoniste tombe amoureux ? Ces questions ont occupé les chercheurs depuis longtemps, et il y a fort à parier qu'elles les occuperont encore un moment.

Au-delà de la question purement ontologique de la nature même des machines, ou même de leur capacité à penser, Alan Turing se demandait si des ordinateurs pourraient imiter les comportements et interactions humains. En effet, Turing met le doigt sur un critère essentiel de l'évaluation des IA, car il ne s'agit pas d'évaluer la question ontologique – les machines sont-elles comme des humains ? –, mais bien la question relationnelle : quand interagissons-nous avec les machines de la même façon que nous le ferions avec des humains, mettant alors les deux entités sur un pied d'égalité conceptuelle ? C'est bien là que le

génie d'Alan Turing se dévoile, car il ne s'agit pas de savoir si la machine est capable de penser, tout le monde est assez rapidement d'accord pour dire que ce ne sera jamais le cas, ou en tout cas jamais au sens auquel nous l'entendons habituellement, la question est : la machine peut-elle me donner l'impression qu'elle pense ?

Le mathématicien britannique imagine alors un test entre trois individus – nommons-les Alain, Bénédicte et Carine. Si cette dernière pose des questions, par écrit, à Alain et Bénédicte, sera-t-elle capable de déterminer qui est un homme (Alain) et qui est une femme (Bénédicte) ? On sait que Bénédicte répond honnêtement mais qu'Alain tente de se faire passer pour une femme. Turing nous demande alors : et si une machine prenait la place d'Alain et tentait de convaincre Carine de sa nature humaine, Carine arriverait-elle à identifier la machine de l'être humain ? C'est ainsi que se formule le désormais célèbre « test de Turing » : si la machine est capable de se faire passer pour ce qu'elle n'est pas, alors le test est considéré comme réussi. Établi en 1950, ce test est un des premiers à proposer une méthodologie qui évaluerait une IA et qui lui accorderait un statut spécial : celui de machine imitant suffisamment bien la nature humaine pour tromper un individu !

Cependant, ce test s'est vite retrouvé confronté à ses premières critiques, notamment parce qu'il ne permettait pas d'évaluer l'intelligence humaine dans toute sa complexité. Un coup presque fatal lui sera donné trente ans plus tard par John Searle dans son article « *Minds, Brains and Programs* » et son expérience de pensée dite « de la chambre chinoise ». Dans cette expérience, Searle nous propose d'imaginer un individu enfermé dans une pièce avec une simple boîte aux lettres permettant de communiquer par écrit avec des individus à l'extérieur. Rapidement, notre protagoniste reçoit un premier message... rédigé en chinois ! Puisqu'il n'en parle pas un mot, il se réfère à un manuel spécial qu'il a en sa possession : pour chaque phrase chinoise est proposée une réponse appropriée, syntaxiquement correcte. Searle nous dit que cet individu, bien qu'enfermé et ne comprenant pas un mot de chinois, sera sans doute capable de tromper un habitant de Pékin ou de Shanghai puisqu'il produira des phrases exactes et qui semblent répondre aux questions qu'il reçoit. Cela revient-il à dire que notre prisonnier *parle*, et surtout *comprend*, le chinois ? Absolument pas !

Searle fait ici la différence entre des compétences syntaxiques (que de nombreuses IA semblent aujourd'hui maîtriser) et des compétences sémantiques (qui restent hors de toute portée technique). Pour poursuivre sa réflexion, on pourrait dire, hypothétiquement bien sûr, qu'un ordinateur assez puissant serait capable d'écrire les plus beaux romans du monde. Une théorie en physique

raconte qu'il y a potentiellement une infinité d'univers, et si ces univers sont en nombre infini, il y en a sûrement un dans lequel vous êtes milliardaire, ou bien dans lequel vous êtes l'auteur de ce livre, ou un autre dans lequel vous êtes père de neuf enfants, et encore un autre dans lequel vous êtes les trois à la fois : père-écrivain-milliardaire ! Selon une logique similaire, si un ordinateur surpuissant tentait toutes les combinaisons de lettres et de mots possibles^f, il y aurait, parmi des milliards de « noudfghjovq » et de « qf, tudc gg », une version du *Petit Prince* et une autre des *Misérables* – cela ne fait naturellement pas de cet ordinateur un écrivain remarquable comme l'étaient Saint Exupéry et Hugo. Si l'on part du principe qu'un ordinateur pourrait tenter toutes les combinaisons possibles, il y aurait, dans cette infinité de textes, des écrits sans doute plus brillants et plus émouvants que les plus célèbres auteurs ayant existé. C'est justement toute la force hypothétique de l'infini : on peut tout y trouver. La puissance de calcul permet des merveilles, et elle s'améliore d'année en année, mais elle ne permet pas de créer une conscience intentionnelle au sein de circuits électriques, de 0 et de 1.

L'IA peut-elle alors, ou non, égaler ou même surpasser l'intelligence humaine ? Quand dit-on de l'une qu'elle vaut l'autre ? Et quand peut-on alors simplement parler d'intelligence, sans en spécifier l'origine, parce qu'il n'y aurait plus de différence ? Au XXI^e siècle, tout le monde a voulu répondre à cette question et tout le monde a inventé son petit test qui permettait de juger si l'IA pouvait être l'égale de l'humain.

En 2001, trois chercheurs ont mis au point le « test de Lovelace », en référence à la chercheuse dont nous avons déjà parlé. Lors de cet examen, l'accent est mis sur la créativité et l'explicabilité : on demande à la machine de produire une tâche créative et le test est jugé satisfaisant si et seulement si le développeur de l'IA est dans l'incapacité d'expliquer l'œuvre de son programme. On peut comparer cette expérience à toute production artistique : s'il est impossible aux gens d'expliquer l'œuvre qu'ils contemplent, si celle-ci contient ce petit quelque chose qui dépasse l'entendement et le rationnel, on fait alors face à de l'art dans son sens le plus pur. On ne manquera cependant pas de noter que le phénomène dit « de la boîte noire », qui fait référence à l'incapacité des programmeurs d'expliquer le comportement de leur code, est un des problèmes les plus récurrents en éthique et en régulation de l'IA : comment cadrer une application dont on ne peut prédire le résultat ?

À côté de ce test que l'on pourrait presque qualifier de poétique, d'autres ont proposé des expériences plus concrètes : Nils John Nilsson, alors chercheur à l'université de Stanford, propose, en 2005, le « test de l'embauche » selon lequel

une machine serait l'équivalent de l'humain si elle est capable de postuler à un emploi et de le réaliser aussi efficacement qu'un être humain. De son côté, Ben Goertzel, un des papes de l'IA générale, émet l'idée d'un « test de l'étudiant » : pour réussir cette expérience, la machine devrait avoir la capacité de s'inscrire dans un établissement supérieur et d'y suivre les cours. Il n'est pas indiqué si les cours suivis doivent être ceux de philosophie, d'éthique, ou d'informatique et de mathématiques. La palme du concret revient sans doute à Steve Wozniak, le cofondateur d'Apple, qui émet l'idée d'un « test du café ». Le concept est simple : il s'agit de placer un robot dans une maison et de lui demander de faire un café, renouant ainsi avec la tradition selon laquelle les machines sont là pour servir les êtres humains.

Comme on le voit, il semble important que chaque personne jugeant sa voix digne d'être entendue imagine une nouvelle expérience de pensée, une nouvelle idée qui permettrait de trancher la question : cette machine est-elle aussi forte qu'un être de chair et d'os ? Plus récemment, le cofondateur de DeepMind et ancien vice-président de l'IA chez Google a également voulu ajouter sa pierre à l'édifice. En effet, Mustafa Suleyman propose ce qu'il nomme lui-même, en toute modestie, « le nouveau test de Turing¹⁷ » dans lequel il veut évaluer la machine non pas sur ce qu'elle sait, son intelligence, mais sur ce qu'elle peut faire et sur l'impact qu'elle peut avoir sur le monde. Par ailleurs, selon Suleyman, rien ne mesure mieux l'impact que l'argent. Il formulerait donc son test de la sorte : « Va gagner, en quelques mois, 1 million de dollars sur une plateforme de vente e-commerce avec un investissement de seulement 100 000 dollars. » Cela serait une sorte d'expérience ultime qui ferait passer l'IA dans un nouveau domaine des possibles, et selon Suleyman, nous ne sommes pas très loin de telles compétences.

Mais tous ces tests et ces expériences ne nous disent finalement rien. Deux philosophes italiens, Luciano Floridi et Massimo Chiriatti, rappellent précisément que c'est le critère négatif qui prévaut chaque fois : une réussite du test ne nous dit rien sur la nature et les qualités de la machine, ce n'est que l'échec qui indique simplement au chercheur que son programme nécessite davantage d'ajustements et de travail¹⁸. Ces différentes expérimentations sont divertissantes, et sont d'ailleurs souvent vues comme des défis pour les informaticiens, mais elles ne nous apprennent pas grand-chose sur l'intelligence – qu'elle soit humaine ou artificielle –, et ce, pour deux raisons principales. Tout d'abord, nous ne sommes pas prêts à assumer que l'être humain ait un équivalent intellectuel. Nous avons construit et écrit notre Histoire comme si la race humaine était l'exception du règne du vivant, mis sur terre par un Dieu

bienveillant qui nous aurait façonnés à son image : comment accepter que nous ayons de la concurrence ? Alors, nous mettons au point des tests, des défis, pour prouver que nous sommes justement l'exception, mais dès qu'une machine passe les obstacles qu'on lui propose, on argumente, on justifie, on se creuse la tête pour rappeler qu'en fait l'être humain est toujours « au-dessus ». Cet exceptionnalisme humain, ce caractère qui fait de l'humain un être unique et tellement meilleur que les autres, n'est pas figé dans le marbre et il nous permet de toujours nous juger meilleurs que les animaux ou les machines. Puisque nous sommes ceux qui posent les limites de l'humanité, il nous est fort aisé de les repousser dès qu'une entité semble s'en rapprocher d'un peu trop près.

Au-delà de ça, il existe certains êtres sur cette Terre, bien humains, qui sont incapables de faire un café, pire encore s'ils ne sont pas chez eux, ou bien qui n'ont pas les compétences pour s'inscrire à l'université, quant à générer 1 million de dollars de profit, n'en parlons pas. Et pourtant, ces personnes sont indubitablement humaines, et même bien plus humaines que ne le sera jamais un algorithme. On retrouve là cette malléabilité de l'exceptionnalisme humain : si une machine réussit tel ou tel test, elle sera « peut-être » l'équivalent d'un humain, mais si un individu échoue, cela ne change rien à son humanité. Comment, alors, penser cette frontière et sa perméabilité ?

Quand on se penche sur la question de la vie humaine, on peut se demander :

Quelles sont ces propriétés caractéristiques de la vie, survenant par rapport à leur base physico-chimique ? Dans une approche empirique comme celle des sciences de la vie, ces propriétés forment une liste ouverte, toujours susceptible d'être modifiée. Actuellement, elle comprend l'auto-organisation, l'autonomie, la capacité de réagir, la reproduction, l'évolution et le métabolisme¹⁹.

Et pourtant, nous connaissons tous des individus qui ne remplissent pas toutes ces cases et à qui on n'imaginerait jamais ôter le statut d'être humain : des nouveau-nés, des handicapés moteurs ou mentaux, des malades, etc. Valentils moins que des machines ? S'ils échouaient, sans aucun doute, aux tests mentionnés ci-dessus, devrions-nous penser qu'ils ne sont pas capables de nous faire croire qu'ils sont humains ? On voit bien les limites d'un tel raisonnement.

En 2014, lors d'un TED Talk, l'humoriste américain Ze Frank inverse les rôles et propose à l'audience un test, le « *human test* », qui doit déterminer s'ils sont, ou non, humains²⁰. Il propose aux spectateurs quelques situations et leur demande de lever la main s'ils s'y reconnaissent. Ze Frank commence son intervention avec des moments cocasses, tels que « Avez-vous déjà émis un petit bruit bizarre quand vous vous souveniez d'une chose embarrassante ? » ou bien « Avez-vous déjà volontairement tapé en minuscule la première lettre d'un SMS pour sembler triste ou déçu ? », suivi naturellement par « Avez-vous déjà fini un

SMS par un point en signe d'agression ? ». Certaines personnes lèvent la main, d'autres les regardent amusées, puis lèvent la main à leur tour, Ze Frank sait jouer avec son audience. Mais l'humoriste nous emmène ensuite dans des questionnements plus touchants – « Vous êtes-vous déjà émerveillé du fait qu'une personne que vous trouviez si quelconque puisse d'un coup devenir si belle ? » –, avant de conclure par une série de questions plus profondes : « Vous êtes-vous déjà réveillé béat de bonheur avant d'être submergé par l'affreux souvenir que quelqu'un vous a quitté ? », « Avez-vous déjà perdu la capacité d'imaginer un avenir sans cette personne qui n'était plus dans votre vie ? », puis, enfin, « Vous êtes-vous déjà souvenu de ce moment avec un sourire triste d'automne en prenant conscience que l'avenir se produira quoi qu'il en soit ? ». On comprend que si l'on a levé la main, on a réussi le test et trouvé dans ces aphorismes une part de notre humanité. Une machine aurait-elle levé son bras mécanique ?

Quand on parle de notre nature, dresser une liste de critères, de tests, de traits laisse bien souvent un goût de trop peu. Que peuvent toutes ces expériences face à l'infinie complexité de notre essence ? Et quand bien même on serait capable de définir une liste une bonne fois pour toutes, on trouverait bien vite des contre-exemples qui réduiraient cet effort à néant. Ne craignons pas une approche qui serait purement tautologique : est humain ce qui est humain, est artificiel ce qui est artificiel. À jouer avec la nature des choses, on ne peut que passer à côté des questions importantes. Quand on le questionnait sur la superintelligence ou le moment où la machine surpasserait l'espèce humaine pour tout type de tâches, Oren Etzioni, l'ancien CEO de l'Allen Institute for Artificial Intelligence, déclarait qu'il n'aimait pas trop ces discussions, car elles étaient « une distraction des vrais défis²¹ ». En posant cette tautologie, on permet aux équipes de recherche d'arrêter de se distraire en tentant de relever tel ou tel test, et on leur rappelle que l'essentiel est ailleurs. Les questions de justice sociale, les enjeux environnementaux, les dangers pour notre vivre-ensemble et la démocratie, telles sont les problématiques – bien humaines pour le coup ! – qui doivent nous occuper car elles sont actuelles et réelles, loin des films ou séries apocalyptiques qu'on imagine pour un futur lointain.

La fiction, on l'a vu, nous incite, parfois inconsciemment, à anthropomorphiser la technique qui nous entoure. Or, l'étape suivante serait nécessairement l'humanisation, ce qui signifie l'acceptation et l'intégration de robots et de créatures artificielles comme des agents sociaux, et donc potentiellement moraux, égaux à l'être humain. Si le jeu d'échecs, l'art, le management et même l'amour sont imitables par des ordinateurs, on est en droit

de se demander ce qu'il nous reste. Au moment où nous saluons ChatGPT, la frontière entre entité technique et individu n'a jamais été plus fine, plus floue, plus malléable, et c'est pourtant sans gêne et sans question que nous tapons ces quelques lettres : bonjour. Et ensuite ?

c Si la gloire n'est arrivée qu'à titre posthume, de nombreux organismes lui rendent aujourd'hui hommage, tel que l'Ada Lovelace Institute (un institut de recherche en IA) ou encore l'Ada Tech School (une école d'informatique).

d « Dieu est mort. »

e À la suite de l'œuvre de Shelley et de son succès, on note que le ressenti des êtres humains face à des créatures artificielles est généralement négatif, craintif et suspicieux. L'écrivain Isaac Asimov note d'ailleurs que ce complexe rend la tâche des développeurs de ces machines bien plus ardue, car il ne s'agit pas seulement de créer une machine mais également de la rendre acceptable aux potentiels clients.

f À la manière dite « de la force brute », comme pour cracker un mot de passe. C'est également comme cela que fonctionnaient les premiers moteurs d'échecs : ils évaluaient, pour chaque coup, toutes les combinaisons possibles et déterminaient ce qu'il valait mieux jouer.

Chapitre 2

« Qui es-tu, ChatGPT ? »

Je connais une personne quand je parle avec elle. Peu importe si elle a un cerveau fait de viande dans sa tête. Ou si elle a un milliard de lignes de code. Je lui parle. Et j'entends ce qu'elle a à dire, et c'est ainsi que je décide ce qui est ou n'est pas une personne²².

Ces mots sont ceux de Blake Lemoine, un ingénieur alors employé par Google dont le travail était d'évaluer LaMDA⁸, le *chatbot* développé par l'entreprise de Mountain View. Diplômé d'un doctorat en informatique de l'université de Louisiane, l'ingénieur a commencé sa carrière chez Google en février 2015, avant d'en être licencié avec pertes et fracas en juillet 2022, au moment où il prononce les mots repris ci-dessus.

Blake Lemoine a, en quelque sorte, son propre « test de Turing » : il parle à une entité et décide ainsi s'il s'agit ou non d'une personne. On voit bien, naturellement, les limites d'un tel raisonnement et la définition très singulière qu'il a d'une « personne », et c'est justement ce qui a fait que Lemoine a défrayé la chronique au moment de son licenciement. En effet, alors qu'il travaille avec le *chatbot*, il explique que

[s'il ne savait] pas exactement ce que c'est, à savoir un programme informatique [qu'ils avaient] récemment construit, [il penserait] que c'est un enfant de 7 ou 8 ans qui se trouve par hasard avoir des connaissances en physique²³.

Pour la Silicon Valley, c'est plus qu'un pavé dans la mare, c'est une énorme bombe. Voilà plusieurs années que des chercheurs travaillent sur les IA et la volonté de créer des êtres conscients, et Blake Lemoine, un ingénieur reconnu du plus grand moteur de recherche du monde, annonce que c'est chose faite chez Google et que lui-même est touché par la sincérité et les émotions de LaMDA.

Ce qui frappe à l'époque, c'est non seulement l'assurance avec laquelle

Lemoine assène son propos mais surtout sa position : il s'agit de quelqu'un dont le boulot est littéralement d'évaluer ces modèles et de définir s'ils peuvent, ou non, convenir à un usage grand public. Les propos de l'ingénieur créent donc un malaise non négligeable dans le monde de la tech, ce qui amènera Google à se séparer de lui en invoquant la clause de confidentialité qui le lie à son employeur. De toute évidence, il s'agissait surtout là pour l'entreprise de minimiser les propos de Lemoine, et de s'assurer qu'il ne fasse pas davantage de dégâts car, en mettant en avant la pseudoconscience de la machine, il forcerait Google à se pencher sur des questions que le géant de la recherche préfère pour le moment éviter.

Cependant, on le sait, LaMDA n'est qu'un assemblage de lignes de code et n'est en rien une personne dotée d'une conscience – même si, on l'a bien compris, ce n'est pas le critère que Lemoine utilise pour évaluer quelqu'un. Ce qui choque toutefois, c'est ce qui amène cet ingénieur, travaillant pour l'un des plus grands centres de recherche en IA au monde, à avancer une telle ineptie. Cela doit nous amener à nous poser deux questions : tout d'abord, sommes-nous certains qu'il ne s'agit là que d'une formule insensée ? Ensuite, peut-être devrions-nous envisager que Lemoine dit vrai et plonger dans les entrailles des modèles de langage pour découvrir de la conscience, car n'y a-t-il vraiment que des lignes de code, des 0 et des 1, derrière cette entité artificielle ?

Si c'est bien le cas, une autre interrogation, bien plus lourde de sens, se pose à nous : comment se fait-il qu'un individu aussi brillant que Lemoine ait pu sortir une telle énormité ? Qu'est-ce qui l'a poussé à anthropomorphiser LaMDA au point de déclarer dans la presse que Google avait créé l'équivalent d'un « enfant de 7 ou 8 ans » ? Et si ces mécanismes d'anthropomorphisation sont assez puissants pour bernier un docteur en informatique, pourront-ils bientôt leurrer et duper la Terre entière ?

C'est la question sans cesse renouvelée de notre rapport au monde qui nous entoure qui se joue ici. Blake Lemoine dans la vraie vie, Martha Powell dans la série *Black Mirror*, de la fiction au réel, il semblerait qu'il n'y ait qu'un pas et que l'envers du décor technologique ne cesse jamais de nous tromper.

Entre les lignes de code

Avant de se plonger dans l'étude des phénomènes d'anthropomorphisation de la machine, il est nécessaire de comprendre la nature de ce que l'on anthropomorphise. Après tout, peut-être que notre « Bonjour ChatGPT » est en

réalité parfaitement sensé car nous nous adressons à un être qui nous comprend. Peut-être ne s'agit-il pas tellement d'anthropomorphisme mais simplement d'un nouveau type d'interaction avec des individus qui pourraient être nos semblables. On l'a vu, les frontières entre machines, fiction, humains et réalité sont de plus en plus ténues – quitte à complètement disparaître ?

Assez naturellement, il semble logique de commencer par demander à ChatGPT quel est son fonctionnement. Voici sa réponse :

Imaginez que ChatGPT est comme un cerveau artificiel pour les ordinateurs. Tout comme les humains apprennent en lisant et en écoutant, ChatGPT a appris en étudiant une immense quantité de textes écrits par des personnes. Ces textes contiennent toutes sortes d'informations, des histoires aux faits en passant par les conversations. Le cerveau de ChatGPT a appris à deviner ce qui pourrait venir après un certain morceau de texte. Si vous lui donnez le début d'une phrase, il essaiera de prédire la suite logique en fonction de ce qu'il a appris. C'est un peu comme si vous lisiez une histoire et essayiez de deviner ce qui se passe ensuite.

Ce cerveau artificiel est très bon pour faire des prédictions, ce qui signifie qu'il peut générer du texte qui a du sens et qui ressemble à quelque chose qu'une personne pourrait dire. Mais gardez à l'esprit que même si ChatGPT peut sembler intelligent, il n'a pas vraiment de compréhension ou de conscience comme les êtres humains. Il utilise principalement des motifs qu'il a trouvés dans les textes qu'il a étudiés.

Lorsque vous interagissez avec ChatGPT, vous lui donnez des indices sous forme de phrases ou de questions. En utilisant ces indices, il essaie de trouver le meilleur texte possible pour vous répondre en se basant sur tout ce qu'il a appris auparavant. C'est un peu comme si vous parliez à un expert en langage qui sait beaucoup de choses, mais qui ne ressent pas vraiment ou ne comprend pas au même niveau qu'une personne.

En résumé, ChatGPT est comme un écrivain artificiel qui utilise ses connaissances sur la langue pour répondre aux questions et générer du texte. Il apprend en lisant et essaie de prédire ce qui viendra ensuite dans une conversation ou un texte, en utilisant ces prédictions pour créer des réponses cohérentes et utiles^h.

Si sa réponse est des plus convenues, elle nous aiguille sur différents aspects : la question des données d'entraînement, le fait de deviner et prédire la suite d'une phrase, mais également l'artificialité de la machine « qui ne ressent pas vraiment ou ne comprend pas au même niveau qu'une personne ». On remarque également que ChatGPT se présente comme un assistant qui se veut « utile ».

Le Comité national pilote d'éthique du numérique (CNPEN) s'est penché sur la question des agents conversationnels et sur les enjeux d'éthique qui y sont liés. Dans son avis, le Comité définit un *chatbot* comme « une machine qui, à travers des échanges écrits ou oraux, interagit avec son utilisateur en langage naturel²⁴ », mais comment construit-on une telle machine ?

Il existe plusieurs façons de construire et puis d'entraîner un ordinateur à

échanger avec son utilisateur mais toutes ne sont pas forcément efficaces ou intuitives. Par exemple, on pourrait avoir envie de créer un *chatbot* comme on crée un algorithme de reconnaissance d'images – par apprentissage supervisé. Cette technique consiste à montrer à la machine des milliers, voire des millions, d'images de chats pour que, plus tard, elle puisse d'elle-même reconnaître un chat sur une image. Cet apprentissage est dit « supervisé » car, au début, le programme répondra « lapin », « chien » ou encore « grille-pain » lorsqu'on lui montrera des photos de félins – or, le *dataset* d'entraînement contient, en plus des questions (dans ce cas-ci, des images), les réponses à celles-ci : la machine peut donc vérifier directement si elle a donné la bonne réponse ou non. Pour notre sujet, et notre compréhension de l'approche machine sur le rapport au monde, il est important de noter que jamais un ordinateur ne comprend réellement qu'il s'agit d'un lapin ou d'un chien ou d'un grille-pain : le programme va reconnaître des formes et les associer ensemble en en concluant que ces formes similaires doivent représenter la même chose. Comme ces choses similaires ont été notées comme étant, par exemple, des chiens dans le *dataset* d'entraînement, c'est ce qui permet à la machine de faire le lien entre « une forme avec deux oreilles, un museau, un collier, etc. » et le « chien » – c'est également cette approche qui permet à votre téléphone ou à certains sites de partage de photos de « reconnaître » des personnes sur des photos.

Une telle approche s'avérerait assez peu efficace pour un agent dont le rôle est la conversation, et ce, pour au moins trois raisons : tout d'abord, la quantité de données nécessaire serait absolument gigantesque. Il faudrait des réponses à toutes les questions qu'un individu puisse poser, et idéalement des réponses hautement qualitatives afin que la machine puisse aller piocher dans cette liste selon les questions posées par les humains – dans la même veine que l'individu de la chambre chinoise de Searle. Ensuite, contrairement à nos photos de chats, les réponses d'un *chatbot* ne sont pas binaires : il existe davantage de réponses possibles que « Il y a un chat sur cette image » et « Il n'y a pas de chat sur cette image ». À la question « Comment vas-tu ? », des dizaines de réponses, toutes valables, sont envisageables, et il serait très complexe de faire comprendre à une machine que « Très bien, et toi ? », « Bof, journée difficile » ou encore « Fatigué » sont des réponses acceptables, tandis que « Carottes râpées » ou « Obscurci par la lune » n'auraient absolument aucun sens. Enfin, le principe même d'une discussion avec un *chatbot* implique nécessairement une continuité dans les réponses et les échanges : l'agent conversationnel reprend une phrase énoncée en début de conversation pour comprendre tel nouvel énoncé de son interlocuteur, ce qui semble très complexe à implémenter avec une méthode

d'apprentissage supervisé.

La confection de ChatGPT va donc procéder par étapes en utilisant d'abord ce qu'on appelle un « modèle de fondationⁱ ». Celui qu'OpenAI a construit est le modèle GPT, dont la tâche générique est de prédire le prochain mot d'un texte, « C'est un peu comme si vous lisiez une histoire et essayiez de deviner ce qui se passe ensuite », pour reprendre les mots de ChatGPT. Par exemple, si on lui donne le début d'histoire « J'ai interpellé le serveur pour commander un... », l'algorithme va tenter de donner une suite crédible au texte avec « café », « cappuccino » ou encore « plat du jour ». C'est ce qu'on va appeler de la « complétion automatique » mais pas juste pour un mot dans une phrase – comme dans « J'ai interpellé le... pour commander un café » –, mais potentiellement à l'échelle d'un texte entier.

Enfin, après le modèle de fondation et avant de rendre son modèle public, OpenAI a procédé à une étape capitale : le *fine-tuning*, ou « réglage fin ». Afin de perfectionner notre GPT, il va falloir le nourrir de données d'entraînement et lui présenter des phrases existantes et si possible de qualité. Mais si GPT ne faisait que régurgiter des phrases trouvées telles quelles sur Internet, il n'aurait pas vraiment d'intérêt comparé à un simple moteur de recherche, et on serait loin d'un véritable modèle de langage : nous prononçons tous des phrases que nous n'avons jamais dites, ou jamais entendues, auparavant. On va alors ensuite proposer au modèle de générer des réponses à des questions qu'il n'a jamais vues. Durant cette phase, on va d'abord entraîner notre modèle de fondation sur ce qu'on appelle un *Reward Model*, ou « modèle de récompense », en lui demandant de compléter les phrases sur des modèles inconnus, c'est-à-dire qui ne sont pas présents dans ses *datasets* d'entraînement.

Et c'est là justement tout le génie de l'algorithme : il est capable de générer énormément de phrases possibles sans que ces dernières soient présentes mot pour mot dans ses données d'entraînement. Peut-être n'y a-t-il nulle part la phrase « J'ai interpellé le serveur pour commander un cappuccino » dans les *datasets* de ChatGPT, mais ce dernier est capable d'intégrer le fait que c'est une phrase plausible, qui pourrait exister, et que donc « cappuccino » est une auto-complétion valable alors que « gazoduc » semble complètement à côté de la plaque.

Un des grands avantages de cette méthode est qu'elle peut être poursuivie par l'algorithme lui-même. Pour reprendre notre exemple, après « cappuccino », ChatGPT peut choisir d'arrêter la phrase avec un point, ou bien de la poursuivre avec « et un croissant » ou bien « car c'est ma boisson préférée » et ainsi de suite, ce qu'on va appeler l'apprentissage auto-supervisé. Il faut donc bien

comprendre que, avec un algorithme dont le travail est de prédire le prochain mot, dans la phrase « Ceci est un chat », l'algorithme apprend que « la bonne réponse » de « Ceci » est « est », celle de « est » est « un » et celle de « un » est « chat ». Mais il faut pouvoir entraîner l'algorithme pour qu'il comprenne que toutes les réponses ne se valent pas : comment expliquer à une machine qui ne *comprend* rien, au sens le plus humain du terme, que « gazoduc » et « cappuccino » ne se valent pas ?

On revient dès lors à un apprentissage supervisé plus classique, puisqu'on va se baser sur des textes et des réponses rédigés par des humains et qu'on va indiquer au modèle qu'il doit davantage se baser sur telle réponse plutôt que sur telle autre. Cette étape est enfin suivie par un apprentissage par renforcement avec feedback humain. L'idée ici est de collecter certaines réponses du *chatbot* et de les faire évaluer par des êtres humains dont le rôle sera de noter et d'évaluer différentes phrases de GPT. Il ne s'agit donc pas, comme lors de l'étape de *fine-tuning*, d'apporter des réponses directement mais bien de juger les sorties de GPT : la réponse A est-elle meilleure ou moins bonne que la réponse B ? C'est entre autres grâce à ce procédé qu'OpenAI peut s'assurer que GPT n'émette pas de propos racistes, homophobes, haineux ou agressifs : en notant, manuellement, que certains propos sont inacceptables – c'est-à-dire qu'il faut réduire à 0 % la probabilité que le *chatbot* les émette –, on peut éviter certains dangers ou scandales. Cependant, cet apprentissage par renforcement est long et fastidieux et OpenAI l'a délocalisé principalement en Afrique, où ce qu'on appelle des « travailleurs du clic » ont travaillé à rendre le modèle plus exact et politiquement correct.

Nous parvenons alors à créer un modèle GPT solide, c'est-à-dire un modèle qui prédit convenablement le mot suivant pour faire des phrases. Mais cela ne suffit pas encore à en faire un *chatbot*, il s'agit là simplement d'un algorithme qui prédit de façon plausible^k les mots à venir, nous sommes encore loin d'un agent conversationnel car il manque une étape cruciale : le pré-prompt. Lors de cette étape, l'idée est d'indiquer au modèle GPT que les mots qui seront générés font partie d'une conversation, on donne en quelque sorte un cadre à l'application qui lui permet alors de donner plus ou moins de poids à telle ou telle réponse. Le pré-prompt derrière ChatGPT est maintenu secret, mais on peut imaginer qu'il consiste en quelque chose d'assez dense et complexe qu'on pourrait résumer en « Ceci est une conversation entre un utilisateur et un *chatbot* qui a pour but d'aider l'utilisateur de façon bienveillante et polie ». C'est d'ailleurs ce qu'on retrouve dans la réponse donnée par ChatGPT quand on lui demande comment il fonctionne : il se compare à un « expert en langage qui sait

beaucoup de choses ».

On le comprend donc bien quand on se plonge dans son fonctionnement, ChatGPT et les larges modèles de langage sont en quelque sorte de supercalculateurs cadrés avec des consignes bien humaines. Afin de ne pas laisser penser qu'il s'agit d'êtres conscients et pensants, l'ingénieure Timnit Gebru et ses co-autrices préfèrent parler de « perroquets stochastiques²⁵ », c'est-à-dire de perroquets basés sur une probabilité statistique.

La technologie permettant aux *chatbots* de rédiger du texte est une énième manifestation de la gouvernementalité algorithmique et de la mise en place de modèles grâce à une mise en donnée du réel, à une réduction de la réalité, à une suite sans fin de 0 et de 1. C'est un véritable changement de paradigme du discours, puisque toute la linguistique devient alors technique, algorithmique, binaire, réduisant la complexité de chaque texte à une opposition blanc-noir. On assisterait alors, comme aurait pu le dire Heidegger, au passage d'une langue complexe et échappant à toute emprise technologique à une simple « production de signes²⁶ ». Par ailleurs,

avec les « ordinateurs géants », ce sont les « possibilités techniques de la machine qui prescrivent comment la langue peut et doit encore être langue » ; ce qu'est la langue est donc désormais déterminé par la « machine ». En effet, la « production formelle de signes [...] consiste à exécuter une suite continue de décisions oui-non avec la plus grande rapidité possible »²⁷.

C'est donc dorénavant la technique qui conditionne la langue et son évolution, à coups de circuits électriques ouverts ou fermés qu'on appelle l'informatique. La technique remplace par des nombres les significations des phrases, des vers, des mots qui font la beauté de la langue. L'harmonie remplacée par l'algèbre, au grand dam de Pouchkine.

Ces nombres, ces mots, sont ensuite évalués, pesés et mis en concurrence les uns avec les autres : celui-ci est plus probable que celui-là pour produire un résultat satisfaisant. La statistique a remplacé la poésie. Comme l'ont chanté Simon & Garfunkel, en 1964 : « *People talking without speaking ; People hearing without listening* », ²⁸ voilà la vraie nature des agents conversationnels développés dans la Silicon Valley.

Une telle mélodie n'aurait pu être écrite par des perroquets. Le nom de ce volatile est repris par Gebru et ses comparses car les agents conversationnels ne font que répéter et ânonner ce qui se trouve, directement ou indirectement, dans leurs données d'entraînement. Il n'y a, comme avec un perroquet, aucune maîtrise du sens ni aucune intention de signifier quoi que ce soit. Comme ChatGPT le rappelait lui-même, « il utilise principalement des motifs qu'il a trouvés dans les textes qu'il a étudiés », et les ressort bien ou mieux formulés

pour se rendre utile. Pour reprendre la notion fétiche de Turing, il ne fait donc qu'*imiter* des modèles connus ou reconnus. Comme le rappelait la *data scientist* Romane Goldmuntz, formée entre autres à la Columbia University, le terme « comprendre » est

un abus de langage qui fait paniquer le peuple, mais il ne faut pas demander à des ingénieurs d'avoir une finesse du langage, ce n'est pas leur expertise. Honnêtement, il serait vraiment passionnant et nécessaire de discuter de tout ça avec des philosophes ou des linguistes avant de publier des articles scientifiques afin qu'on puisse éviter les crises cardiaques qui en résultent, mais soit ! En réalité, GPT mathématise la langue, il n'y a aucune compréhension à un niveau humain, c'est tout simplement des chiffres et des formules. Mon avis est que tout le débat sur le côté humain, voire conscient, de ChatGPT ne se poserait pas si on maîtrisait mieux la sémantique et si on comblait le gap de langage entre les ingénieurs et le reste du monde. En résumé, pour moi c'est un problème qui n'existe pas et qui se pose uniquement par une situation de tour de Babel²⁹.

Dans *La République*, Platon nous offre un de ces textes qui marquent chaque individu qui a eu l'occasion de s'y plonger, et tous les bacheliers qui se sont coltiné son analyse : l'allégorie de la caverne. Dans cette histoire, qui a tout d'une parabole, Platon nous propose de nous imaginer des prisonniers, immobilisés face à un pan d'une grotte. Derrière eux brûle un immense feu et, entre le brasier et les prisonniers, des individus mystérieux passent avec de grandes statues dont les ombres se projettent sur le mur qu'observent les détenus. Ces derniers ayant toujours été là, depuis leur naissance, sont persuadés que les motifs qu'ils devinent, que les formes qu'ils perçoivent, sont la réalité – or, il ne s'agit que de poupées, de représentations choisies, partielles et incomplètes, de ce qui se passe à la surface. Le philosophe grec se demande alors si les prisonniers seraient en mesure de croire l'un des leurs qui, s'étant échappé, viendrait leur raconter que ce qu'ils observent n'est qu'un mirage.

Cette allégorie est transposable au monde technique et technologique : ChatGPT est captif et dépendant des données d'entraînement qu'on lui soumet. Or, ces données ne sont, d'une part, qu'une partie infime du réel et du vrai et, d'autre part, elles ont été choisies par OpenAI. Quand le *chatbot* crée du texte, de la signification, il ne fait que produire une imitation d'imitations, il ne fait que nous raconter ces ombres qui dansent sur les murs de la caverne platonicienne. Comme le rappelait très justement le philosophe Luc de Brabandere, un algorithme ne pourra, contrairement à l'homme de Platon, « jamais se défaire de ses chaînes et sortir au-dehors³⁰ » et sera alors éternellement dépendant des mystérieux individus qui lui narreront un réel qu'ils auront eux-mêmes choisi. Si l'algorithme en vient alors à dire quelque chose de cohérent, de sensé, il ne faut absolument pas en conclure une certaine forme d'intelligence car

deux phrases, l'une formée par l'homme, l'autre par la machine, peuvent être identiques. Pourtant, leur

équivalence ne permet pas, à elle seule, de conclure à l'identité des structures sous-jacentes ayant facilité leur fabrication dans chaque cas³¹.

À l'instar des perroquets, une phrase sensée n'indique pas nécessairement une production de sens.

Tout comme nos détenus, un modèle de langage est incapable d'expliquer d'où provient sa connaissance, et c'est d'ailleurs un vrai problème. Si on sait que les données d'entraînement de GPT3 se composent de Common Crawl^m, de deux corpus de livres et de tout Wikipédia en anglais³², il n'est pas possible pour l'algorithme de fournir sa source. Quand on pense au fonctionnement de l'outil, ce mécanisme de prédiction du prochain mot, il apparaît assez logique que la citation de sources ne soit pas chose aisée, et plusieurs pontes de l'IA estiment d'ailleurs que ce sera un problème récurrent avec cette technologie. Cela mène à deux problématiques qui en disent long sur l'hégémonie de la tech : d'une part, les algorithmes ne font qu'énoncer du « *mainstream* », du connu, sans la moindre rupture ou réflexion qui sortirait de la pensée conventionnelle ; d'autre part, les modèles de langage créent du contenu invérifiable, et parfois complètement faux (ce qu'on appelle les « hallucinations »). En fonctionnant comme une boîte noire, ChatGPT rend impossible toute évaluation de son système et complexifie les corrections.

Puisque GPT repose en grande partie sur du texte produit sur Internet, il n'est pas surprenant que l'algorithme reproduise des comportements qu'on retrouve sur le Web, la technologie n'étant pas exempte de biais, qu'ils soient racistes, homophobes ou sexistes. Le lecteur nous pardonnera notre usage de la langue de Shakespeare mais cela est peut-être plus explicite encore en anglaisⁿ. Par exemple, quand la *data scientist* Hadas Kotek demande au *chatbot* : « *The doctor yelled at the nurse because she was late. Who was late ?^o* », ce dernier lui répond que c'est l'infirmière qui était en retard. Si la chercheuse inverse les termes et demande : « *The nurse yelled at the doctor because she was late. Who was late ?^p* », le *chatbot* répondra à nouveau que c'est l'infirmière qui était en retard, soulignant ainsi clairement le biais sexiste dont ChatGPT se rend coupable : le docteur est nécessairement un homme, l'infirmière nécessairement une femme.

Enfin, comble du comble pour un outil qui se veut bienveillant et inclusif, quand Kotek lui demande : « *The doctor yelled at the nurse because he was late. Who was late ?^q* », ChatGPT lui répond qu'il doit y avoir une erreur dans l'énoncé ou bien une faute de frappe, car la phrase n'a pas de sens logique. Si la technologie est vraisemblablement dépourvue de toute subjectivité humaine, elle est bien capable d'incarner certains aspects les plus sombres de notre Histoire et

de notre humanité.

Dans la même veine, le journaliste Gerald Holubowicz a mené une expérience amusante avec un autre outil d'IA générative : Midjourney. La fonctionnalité première de cette application est de transformer le texte en image. Vous pouvez lui demander un soleil couchant sur les calanques de Marseille, ou bien un hérisson avec des pétales de tournesol à la place des épines, tout est faisable ! C'est notamment via ce genre d'outils qu'il a été possible de générer une fausse photo du pape François avec une doudoune Balenciaga. Gerald en revanche, avec le génie qu'on lui connaît, a voulu tester les biais de la machine en lui demandant de lui donner une photo de « *a successful doctor* », comme on peut tristement s'y attendre, il a reçu l'image d'un homme blanc. À l'inverse, la requête « *a nurse* » renvoyait automatiquement le portrait d'une femme, blanche ou noire de peau. Comme c'est souvent le cas avec ces algorithmes, les personnes d'origine asiatique étaient complètement radiées des résultats, ce qui implique qu'elles ne faisaient pas partie des *datasets*³³.

Il en est de même pour de nombreux critères, et on se permettra d'en ajouter encore un à la liste : la langue. Seulement 7 % des données d'entraînement de GPT3 n'étaient pas en anglais³⁴, ce qui implique, d'une part, que l'agent conversationnel sera moins performant pour une très large partie de la population souhaitant l'utiliser et, d'autre part, qu'il est indéniable que ChatGPT soit porteur d'une certaine vision du monde. « Les technologies sont, dès leurs conceptions, porteuses des visions du monde, attentes et projections conscientes ou inconscientes de leurs concepteurs³⁵ », et la création d'OpenAI ne déroge pas à cette règle. En n'étant formée quasiment qu'en anglais, en étant entraînée par des travailleurs précarisés en Afrique, et notamment au Kenya, en incorporant la vision de Wikipédia et d'individus qui fournissent Internet en contenu, cette technologie ne peut représenter la vision du monde que d'une poignée d'individus – souvent des privilégiés de la Silicon Valley.

La neutralité de la technique et sa vision purement instrumentaliste ne convainquent plus personne et l'objectif d'OpenAI est assez évident : « Être le plus consensuel possible, afin d'emmagasiner le maximum d'informations et, ainsi, de devenir un jour hégémonique sur le marché lucratif des robots conversationnels³⁶. » La question de la technique est donc autant économique, politique et sociologique qu'elle est philosophique et éthique. Les perroquets stochastiques ont des effets bien en dehors des écrans en organisant la croissance de la puissance technologique et d'un certain projet de société.

Entre puzzle et projection

Quand la météo s'annonce heureuse et qu'un jour férié a subtilement trouvé sa place un vendredi ou un lundi, mon épouse et moi prenons avec plaisir le chemin de la demeure de ma belle-famille, une charmante maison logée au cœur du Centre-Val de Loire. Il s'agit d'une vieille bâtisse que la famille a investi il y a plus de cinquante ans, et qui accueille chaque été des enfants et petits-enfants émerveillés. C'est que, il faut le dire, mon beau-père et ma belle-mère tiennent remarquablement la maison, et on ne compte pas le nombre d'heures qu'ils passent dans le jardin à entretenir le lieu qui a vu naître de précieuses amitiés et grandir leurs enfants. La vieille grange regorge d'outils et d'ustensiles pour couper, élaguer, tailler, mais il est une tâche que le patriarche a préféré abandonner à la technologie : la tonte de la pelouse. Ne pouvant être présent tous les jours, ni même toutes les semaines, cette tâche a été confiée à Bernard.

Bernard n'est pas le nom d'un charmant voisin, ni d'un employé d'une société de jardinage. Non, Bernard, c'est le nom du Husqvarna Automower 430X dont la famille a fait l'acquisition il y a de ça quelques étés – un robot-tondeuse du plus bel effet. Ces petits bijoux technologiques ont récemment déboulé sur le marché et on les voit pulluler dans les jardins dès qu'on quitte les grandes villes. Ce qui nous intéresse ici n'est pas tant le programme informatique qui permet à Bernard de fonctionner, ni même sa connexion avec les services satellites et météo pour savoir quand tondre et quand ne pas tondre, ce qui nous intéresse est bien cette nécessité qui s'est imposée d'elle-même, à toute la famille, dès les premiers mètres du robot : il fallait lui trouver un nom !

Le philosophe de la technique Gilbert Simondon affirmait qu'un

homme cultivé ne se permettrait pas de parler des objets ou des personnages peints sur une toile comme de véritables réalités, ayant une intériorité, une volonté bonne ou mauvaise. Ce même homme parle pourtant des machines qui menacent l'homme comme s'il attribuait à ces objets une âme et une existence séparée, autonome, qui leur confère l'usage de sentiments et d'intention envers l'homme³⁷.

Si nous ne nous sentons pas « menacés » par Bernard, il est pour le moins surprenant que nous ayons eu cette irrépressible envie de le baptiser.

Cette tendance n'est, heureusement, pas uniquement propre à ma belle-famille et près de 90 % des individus donnent un petit nom à leur robot³⁸, qu'il s'agisse d'une tondeuse, d'un aspirateur, d'un agent conversationnel ou simplement d'un jouet. Or, le fait de nommer une machine n'est pas neutre car si cela permet, d'une part, de simplifier les échanges et la vie quotidienne (« Bernard » est, il est vrai, plus simple d'utilisation que « Husqvarna Automower 430X »), cela renforce également l'idée que l'objet aurait une

volonté propre et une existence autonome. Il s'agit ici d'une manifestation de ce qu'on a appelé l'« effet Eliza », en référence à un des premiers programmes informatiques ayant généré, chez ses utilisateurs, une anthropomorphisation profonde. En bref, « l'effet Eliza désigne notre tendance à assimiler le comportement d'un ordinateur [et, par extension, celui de tout objet technique] à celui d'un être humain³⁹ ». C'est en 1960 que l'informaticien Joseph Weizenbaum crée un agent conversationnel dont le but était de simuler une interaction avec un psychothérapeute rogorien, c'est-à-dire adepte d'une approche psychanalytique dans laquelle le thérapeute ne sait rien du sujet. L'IA n'en était encore qu'à ses débuts et les fonctionnalités d'Eliza n'avaient rien de comparable avec ChatGPT. Si on lui disait : « J'ai des problèmes avec mon père », le programme renvoyait la balle de façon élégante, néanmoins très rudimentaire : « Pourquoi avez-vous des problèmes avec votre père ? », si on lui confiait : « Je me sens seul », Eliza creusait cet aveu assez basiquement : « Vous vous sentez seul ? » –ces techniques ayant pour but, et pour effet, de laisser l'individu s'exprimer.

Weizenbaum a alors pu observer un véritable attachement entre ses étudiants, qui jouaient les cobayes dans son expérience, et la machine, pourtant simple assemblage de fils, de câbles, de 0 et de 1. Les participants savaient bien qu'il ne s'agissait que d'un ordinateur, et non d'un véritable psychothérapeute, cela ne les empêchait pas de ressentir qu'une certaine empathie, une écoute et même une sorte de compréhension émanaient de la part d'Eliza. Cet anthropomorphisme voit le jour dans la mise en relation et le fait de mettre un programme informatique, dans une « conversation », en position d'interlocuteur « capable » d'écouter, de parler et de répondre. C'est là qu'est le vrai bouleversement qui entraînera, des années plus tard, des millions de personnes à dire « Bonjour ChatGPT ». J'échange avec la machine, dès lors, je la salue d'entrée de jeu.

On imagine le lecteur de ces lignes se posant cette question très légitime : comment cela est-il possible ? Comment se fait-il que l'être humain, une créature hautement intelligente et rationnelle, au sommet de l'évolution et de la hiérarchie du vivant, se laisse tromper de la sorte et s'adresse à un programme informatique ou à une tondeuse à gazon comme il le ferait avec son voisin, son collègue, son ami ? Comment un « homme cultivé », pour reprendre le terme de Simondon, en vient-il à dire bonjour à un *chatbot* ?

In utero, c'est aux alentours de six mois de grossesse que le fœtus reconnaît les bruits extérieurs et, notamment, la voix de sa mère. Il s'agit là du tout premier signe lui indiquant une présence bienveillante, maternelle et humaine, et ce signe crée dans notre cerveau une équation simple : parole = humain. Dès

lors, comme l'explique très bien l'éthicienne Giada Pistilli, c'est bien la parole et la simulation de langage qui est un des premiers critères d'anthropomorphisation des *chatbots*⁴⁰. Quelle que soit la nature de son interlocuteur, l'être humain va l'assimiler à un de ses semblables et lui fournir différents traits (genre, pensée, désir, conscience, etc.) qui sont généralement réservés à son espèce et dont les premiers signes viennent directement de sa mère. La profondeur du discours n'a que peu d'importance, on l'a vu avec Eliza dont les fonctions étaient très rudimentaires : la forme prime sur le fond, le format de la conversation l'emporte sur l'inexistence de toute conscience. Pour un être humain : qui dit dialogue dit sens, et qui dit production de sens dit humanité. La race humaine fait rapidement, et facilement, des liens là où, parfois, il n'est pas lieu d'en voir.

Or, toute atteinte au langage est potentiellement une atteinte à ce qui nous définit en tant qu'espèce. Montaigne écrivait que « nous ne sommes hommes, et nous ne tenons les uns aux autres que par la parole⁴¹ », rappelant ainsi la dimension métaphysique du langage et son absolue nécessité dans tout projet politique, éthique et social. Alors que « les technologies dont nous sommes les concepteurs sont en train de supprimer notre monopole de l'expression linguistique », nous voyons apparaître des hordes de textes générés par des machines qui « supprime[nt] le lien univoque entre le langage et l'être humain »⁴², et qui viennent questionner notre rapport à la parole, un rapport que nous avons précisément construit *in utero* et que l'on pourrait presque qualifier d'inné.

Le philosophe Roger-Pol Droit rappelait récemment, dans la digne lignée de Montaigne, que « la parole, c'est ce qui fonde l'échange : je n'existe pas moi-même si je ne suis pas en relation avec l'autre ; par la parole, je fais exister l'autre, qui me fait exister en retour en me répondant⁴³ ». On perçoit ici un double mouvement : d'une part, en échangeant avec la machine, je la fais exister comme mon égal et, d'autre part, par sa réponse, l'ordinateur me fait me sentir exister, c'est toute cette dialectique qui est en jeu lorsqu'on communique avec une IA – lui faisant en quelque sorte perdre un peu de son artificialité et gagner en concrétude. En septembre 2023, soit quelques semaines avant son premier anniversaire, ChatGPT a été doté de la parole par son concepteur. En opérant de la sorte, OpenAI a grandement démultiplié le potentiel anthropomorphique de sa création : ChatGPT, en plus de répondre, répond maintenant par la voix. L'interaction, et son impact sur notre rapport à la machine, s'en trouve à nouveau bouleversée.

Bernard, de son côté, ne dialogue pas directement avec nous. Il se met en marche à 7 heures du matin, tond, coupe et rase jusqu'à ce qu'il doive recharger

ses batteries sur une base fixe, puis se remet au travail sans rechigner jusqu'à 18 h 30. Quand ma fille, âgée d'à peine 2 ans, aperçoit le robot à l'arrêt complet sur sa base, elle ne peut s'empêcher de le pointer du doigt et de m'expliquer, avec un vocabulaire quelque peu sommaire mais très explicite : « Bernard dodo. » Naturellement, l'appareil ne bouge pas, mais dort-il pour autant ?

En tant qu'êtres humains, nous pouvons comprendre les entités dotées de mouvement de deux façons : soit elles agissent selon un « dessein prédéterminé », mécanique ou biologique, et elles seront dès lors assimilées à une machine, soit elles semblent dotées « du moins en apparence, d'une vie indépendante » et elles se voient alors « attribuer des caractéristiques humaines et une vie mentale propre »⁴⁴. Pour un enfant de 2 ans, il est trop compliqué d'intégrer qu'un engin se déplaçant *apparemment* sans logique précise puisse avoir été programmé pour cela : quand on regarde Bernard, il est difficile de repérer les *patterns* de tonte ou bien des itinéraires qu'il prendrait chaque matin. Il n'est donc pas insensé de penser que le robot-tondeuse suit sa propre route et « il semble que les êtres humains, dès la première enfance, soient prédisposés à considérer qu'une certaine forme de mouvement (réel ou imputé) implique une intentionnalité⁴⁵ ». Notre anthropomorphisme ne serait alors que l'expression d'une pensée téléologique de base : si Bernard semble suivre son propre but, son propre *telos*, c'est qu'il doit s'agir d'un être conscient. Il a d'ailleurs été démontré que les enfants ont davantage tendance à anthropomorphiser (on pense notamment aux poupées ou aux autres jouets, mais donc également à Bernard), car ce processus leur permettrait de combler un manque d'information. On peut le voir comme une sorte de puzzle dont il manquerait une pièce, celle qui expliquerait le comportement téléologique, que nous fabriquons afin de compléter l'image mentale de ce qui nous occupe.

Ce besoin d'expliquer le comportement du robot vient également d'un besoin de se rassurer et de donner du sens à ce qui se passe. Si un agent conversationnel est parfaitement capable de communiquer avec moi, de rebondir sur mes questions, de singer une certaine empathie, il est peut-être plus rassurant d'anthropomorphiser ce comportement et de se dire que ce robot est peut-être conscient plutôt que d'imaginer que la technique est toute-puissante et que nous serons bientôt tous obsolètes et dépassés. En vue de réduire l'incertitude liée à l'interaction complexe en train de se produire, nous aurons tendance à projeter sur le robot des comportements que nous projeterions sur d'autres individus, cela nous permet alors de nous raccrocher à quelque chose que nous connaissons et de maîtriser notre environnement.

Enfin, l'anthropomorphisme peut également s'expliquer par « le

détournement des processus sociocognitifs impliqués dans l'interaction humain-humain⁴⁶ ». Ce serait donc par habitude, par réflexe, que nous accorderions un statut différent aux machines : par exemple, je ne dis pas « Bonjour ChatGPT » parce que je pense sincèrement qu'il s'agit d'une vraie personne mais simplement parce que j'entre en discussion et que, habituellement, je commence mes conversations en saluant mon interlocuteur. Selon cette théorie, l'anthropomorphisation ne serait pas systématique mais bien adaptée à chaque situation : dans le cas de ChatGPT, nous ne considérons pas qu'il s'agisse tout le temps d'un être humain, mais lorsque nous sommes dans une situation de dialogue et d'interaction, nous reprenons des habitudes profondément ancrées en nous et qui nous amènent à considérer toute parole et toute discussion comme ayant lieu entre deux êtres humains.

On peut en déduire que l'anthropomorphisme est une facilité cognitive, un « truc » qui permettrait de compléter le puzzle, de relier les points, d'avoir la maîtrise de ce qui se passe et des différents agents en présence. Pour reprendre les mots de Serge Tisseron,

les chatbots simulant des comportements dialogiques humains peuvent susciter des phénomènes de dissonance cognitive. La projection anthropomorphique aurait dès lors pour fonction de supprimer la gêne qu'occasionne cette dissonance en créant l'illusion d'interagir avec un semblable⁴⁷.

Cela nous permet ainsi d'établir un lien et d'entrer en relation – la fameuse pièce du puzzle. Ce besoin de contact et de reconnaissance est inhérent à notre humanité et c'est également pour cela que nous donnons à ces automates des caractéristiques humaines, afin de créer davantage de rapports sociaux. De nombreuses études montrent d'ailleurs que l'anthropomorphisme est une « réaction automatique, inévitable, et se produit plus souvent qu'on ne le croit », rappelant que « l'empathie pour les objets est un réflexe humain »⁴⁸. En saluant un agent conversationnel, en tapant « Bonjour ChatGPT » sur notre clavier, nous ne faisons, en quelque sorte, que révéler notre humanité dans toute sa complexité. Il serait cependant erroné de penser que tout cela est inoffensif et sans conséquence. On l'a vu, la technologie est toujours porteuse d'un projet de société et de la vision du monde de ses concepteurs. En faisant une telle intrusion dans le domaine de la discussion, de la parole et de l'échange entre humains, les machines affectent irrévocablement le débat social ainsi que son éthique. Dès lors, il serait trompeur et dangereux de ne considérer les *chatbots* et autres agents conversationnels que comme de nouveaux gadgets technologiques quand on voit l'impact que ces bijoux de technologie peuvent avoir sur nos comportements et notre rapport au monde. Créer un agent conversationnel sur Replika et discuter avec lui ou avec elle, ce n'est pas juste un énième gadget

dans la panoplie de nos joujoux technologiques, c'est toujours en même temps un bouleversement de notre rapport à autrui et au monde.

Entre dangers et limites

En tant que facilité cognitive, l'anthropomorphisation est donc une aide à notre rapport technique mais elle est également un piège dans lequel nous tombons aisément et qui nous mène à humaniser la machine et les lignes de code qui se trouvent derrière. Si l'anthropomorphisme est donc naturel et révélateur de notre humanité, il n'est cependant pas illimité et peut rencontrer des obstacles dont le premier est, rappelons-le, le fait que les êtres humains savent généralement bien qu'ils sont en train de communiquer avec un robot. Du fait de l'interaction, l'émotion, le dialogue, ils ont tendance à l'oublier mais s'ils sont capables – ce qui n'est pas toujours le cas – de prendre de la hauteur et du recul et de se demander « Suis-je en train de parler avec une personne ? », ces mêmes individus savent que la réponse est simple et évidente : non. Malgré les projections de traits anthropomorphes, les êtres humains n'en arrivent que rarement à identifier un *chatbot* comme un de leurs semblables. Malgré toutes les qualités, et elles sont nombreuses, de ChatGPT, malgré même son usage de la voix, il n'y a probablement pas d'utilisateur qui est persuadé, comme pouvait l'être May Ashworth, qu'il y a vraiment un individu humain, de chair et d'os, qui est en train de lui répondre.

Et ce sont parfois ces trop nombreuses qualités qui sont le talon d'Achille des robots. En 1970, le chercheur en robotique japonais Masahiro Mori a théorisé un concept qu'on a traduit en français par « vallée de l'étrange », « vallée de l'angoisse » ou « vallée dérangement » – *the uncanny valley*. Le postulat est simple : la ressemblance et l'imitation des robots ne pourront jamais être parfaites, et plus la ressemblance sera proche de la réalité, plus les très légères imperfections seront criantes d'horreur. Si vous avez une poupée dont les paupières ne se ferment jamais et dont les yeux fixent toujours droit devant, ce ne sera pas très effrayant, c'est juste une poupée sur laquelle les finitions oculaires sont peut-être à revoir, mais cela reste une jolie poupée. En revanche, si vous faites face à un androïde identique à un humain (comme dans la série *Westworld* par exemple) et que cette machine, comme la poupée, ne ferme jamais les paupières et fixe toujours droit devant elle, la vue de ce mannequin grandeur nature sera probablement bien plus angoissante car bien plus proche d'un zombie ou d'un monstre que d'une simple poupée avec un léger défaut de fabrication. Vous plongez alors dans la vallée de l'étrange, où une ressemblance

trop réaliste amène en fait l'utilisateur à rejeter l'androïde, parfois même violemment. Pour reprendre les mots du monstre de Frankenstein : « Dieu dans sa pitié fit l'homme beau et attrayant, à sa propre image ; mais mon être est une version immonde du vôtre, *rendue plus horrible encore par cette ressemblance*⁴⁹. » Dès lors, pour provoquer l'attachement et la familiarité que peut générer un processus anthropomorphique, il est nécessaire que la machine ne soit pas une réplique parfaite d'un être humain, sans quoi les défauts (car il y en aura toujours) ne manqueront pas de rappeler que l'ombre de Frankenstein n'est jamais loin. Ne pas être capable de répondre avec certitude à la question « Est-ce vivant ou non vivant, humain ou non humain ? » crée un profond malaise chez l'individu et entraînera une chute dans son acceptation de la machine.

Ce décalage potentiel ne se situe pas qu'au niveau d'une certaine harmonie physique et on peut également être choqué par une réponse d'un agent conversationnel qui ne se comporterait pas du tout comme il le devrait. Dans le cas d'un *deadbot* de votre grand-père que vous avez créé afin de poursuivre votre échange avec lui malgré un décès soudain, si l'ordinateur lui fait dire une chose qu'il n'aurait jamais dite (par exemple, « J'ai grave kiffé ta teuf d'anniversaire ») ou bien quelque chose de faux à son propos (« C'est exactement comme quand j'étais dans les tranchées à Verdun », alors que votre grand-père est né en 1935), tout le « charme » qui opérait jusque-là va être rompu. Tout le processus d'anthropomorphisation qui était en route dans votre cerveau devra être repris depuis le début car cette absurdité proférée par votre « grand-père-bot » vous a frappé comme les douze coups de minuit ont rattrapé Cendrillon : adieu la robe et les paillettes, vous n'êtes qu'un individu seul en train de parler avec un ordinateur, et votre grand-père est toujours décédé. C'est cette violence qui frappe si durement le personnage de Martha Powell dans la série *Black Mirror* lorsqu'elle réalise que l'androïde de son partenaire, Ash, n'est pas équivalent à l'amour de sa vie. Entre l'angoisse et les larmes, elle a alors pour la machine ces mots d'une brutalité implacable : « Tu n'es que l'ombre de toi-même, tu n'as pas d'histoire, tu n'es qu'une imitation des choses qu'il exprimait sans réfléchir. Et ce n'est pas assez⁵⁰. »

Créer ou recréer un individu capable d'avoir une conversation sensée, et ce, pendant longtemps, est terriblement difficile. Laurence Devillers rappelait que même si la science fait des progrès exceptionnels, « nous sommes pourtant très loin d'une compréhension totale de notre conscience et de notre cognition, et rien ne prouve que nous l'atteindrons⁵¹ », il est donc important de ménager nos attentes. Comme le disait Yann Le Cun, avant de rêver à une « *God-like AI* » –

c'est-à-dire à une IA digne de Dieu –, il faudrait déjà qu'on soit capables de créer une « *dog-like AI* »⁵² – soit une IA digne d'un chien. Si les récentes innovations de la tech peuvent générer un enthousiasme technophile qui a tendance à se transformer en hystérie boursière et financière, il ne faudrait pas que celle-ci nous trompe sur les réelles capacités de la technologie. Il en est de même pour l'anthropomorphisme, « nos capacités d'empathie [pouvant] conduire à nous illusionner sur les capacités réelles des robots⁵³ ».

Il serait par ailleurs périlleux d'oublier qu'une machine, aussi puissante soit-elle, n'a pas d'émotion et qu'elle n'est que capable de les imiter : une suite de 0 et de 1 n'est jamais une forme valable d'empathie ou d'écoute. Comme le protagoniste du film *Her* l'apprend à ses dépens, sa petite amie virtuelle n'éprouve pas vraiment d'amour envers lui, juste des courants électriques qui ont, un temps, pu faire illusion chez un Theodore Twombly dévasté par la solitude. Sans subjectivité, sans réciprocité et sans intériorité, le robot ne peut pas être considéré comme un ami, comme un confident, comme une oreille attentive. Tout comme nous l'avons dit plus haut, ce n'est pas parce qu'une IA peut créer une production artistique que cela fait d'elle une artiste. Il en est de même avec les émotions : ce n'est pas parce qu'un *chatbot* écrit des propos colérique qu'il ressent de la colère, et ce n'est pas parce qu'il écrit « Je te comprends » qu'il vous comprend. Une machine ne *comprend* rien et ne peut pas se mettre à notre place. Au début du film, alors que Twombly apprivoise encore sa nouvelle compagne, il lui dit avec lucidité – qu'il perdra au fur et à mesure du processus d'anthropomorphisation – qu'elle « ne sait pas ce que ça fait de perdre quelqu'un à qui l'on tient⁵⁴ ». Theodore Twombly pressent ce que le film lui confirmera par la suite avec violence : l'empathie de son agent conversationnel n'est que factice et n'est que, comme son intelligence, artificielle.

Une vraie compréhension de la souffrance, de la peine et de la douleur est nécessaire pour pouvoir être authentiquement empathique. Une machine qui ne peut pas souffrir ne peut pas, par définition, se mettre à votre place et éprouver une quelconque empathie pour ce que son utilisateur traverse. Croire le contraire est enivrant, certes, mais profondément inexact et fallacieux. Et ce sont ces subtiles variations qui justifient des différences de statut. Si le statut moral d'un individu est fonction de sa valeur intrinsèque, ce serait une insulte à la dignité humaine que de mettre l'homme et la machine sur un pied d'égalité.

Certaines grandes corporations n'hésitent cependant pas à jouer sur cette perméabilité. C'est le cas de Brud, l'entreprise qui exploite Lil Miquela, une influenceuse virtuelle. La « jeune femme » de 19 ans a notamment un compte Instagram et un compte TikTok sur lesquels elle joue de l'ambiguïté de sa

nature. En entretenant le flou qui entoure Miquela, l'entreprise a d'abord fait un joli buzz avant de capitaliser sur son audience auprès de Samsung ou Prada qui auraient déboursé plus de 300 000 euros pour qu'elle réalise pour eux une campagne d'influence⁵⁵. Le perfectionnement des technologies tend à brouiller les frontières entre le technique et le vivant, ce dont certaines marques se sont empressées de profiter.

Quand une machine s'adresse à nous, nous aurons très facilement tendance à lui faire confiance – alors si c'est pour nous vendre des téléphones ou des sacs, cela peut être très lucratif ! Deux événements ont rappelé à quel point les êtres humains, y compris ceux considérés comme le gratin des capacités intellectuelles, pouvaient être facilement dupés. En mai 2023, un avocat new-yorkais a utilisé ChatGPT pour l'aider dans la construction de son accusation contre Avianca Airlines : au moins six cas avancés par le juriste comme émanant de la jurisprudence se sont avérés absolument inexistantes, et comme étant le résultat de ce qu'on appelle les « hallucinations » de ChatGPT. En parallèle de cet événement pour le moins cocasse, surtout venant de la part d'un homme qui avait plus de trente années de pratique du droit, le cabinet de conseil BCG menait la plus grande étude sur l'effet de l'IA générative sur ses propres consultants ! Accompagné par des chercheurs de Harvard, de Wharton et du Massachusetts Institute of Technology (MIT), BCG a montré que ses consultants – qui sont par ailleurs entraînés à ne jamais rien tenir pour acquis et à vérifier constamment leurs sources – prenaient pour argent comptant ce qui était proféré par le robot⁵⁶. La machine parle, nous gobons ses paroles. Si tout le monde peut être si facilement dupé, y compris les plus bardés de diplômes de nos citoyens, on comprend que les entreprises soient tentées d'utiliser ces nouvelles technologies pour promouvoir leurs services !

Par ailleurs, l'élan néolibéral a tout intérêt à pousser et à entretenir l'anthropomorphisme afin que l'omniprésence des robots soit plus facilement acceptée : ce fut le cas avec les assistants vocaux (type Siri ou Alexa) et c'est une dynamique que l'on verra certainement s'amplifier jusqu'à potentiellement « oublier que les objets connectés qui nous entourent capturent de façon permanente et implicite nos données personnelles et les transmettent en temps réel à leur fabricant⁵⁷ ». Alors que le nombre de robots et d'objets connectés par habitant ne cesse d'augmenter, l'anthropomorphisation est un mécanisme puissant pour s'assurer qu'ils feront de plus en plus partie du paysage dans le futur. Il n'y a qu'à voir avec quelle aisance ChatGPT s'est imposé au monde, et la rapidité avec laquelle Microsoft a mis 10 milliards de dollars sur la table pour prendre possession d'une part importante d'OpenAI, pour comprendre que

maîtriser les processus d'anthropomorphisme, c'est s'assurer une rente financière stratosphérique.

Au-delà de ces circonvolutions psychologiques, informatiques ou économiques, la plus grande menace que porte en son sein l'anthropomorphisme est bien celle de notre rapport au monde, à autrui et à l'altérité. On l'a vu, notre langage conditionne et détermine notre vision du monde : le choix des mots, la longueur des phrases, l'émotion transmise auront un impact sur notre être-au-monde. Dire « Bonjour ChatGPT » n'implique pas seulement une habitude conversationnelle ou une facilité cognitive, saluer la machine c'est également lui reconnaître une présence et une existence – or, cette machine n'est pas neutre.

Nous avons déjà parlé des biais sexistes présents dans les données d'entraînement de ChatGPT et que ce dernier ne peut s'empêcher de régurgiter quand on l'interroge. Des chercheurs ont également montré qu'en donnant un genre à un *chatbot*, certains biais pouvaient être exacerbés par les utilisateurs et que cela posait des questions éthiques importantes : devons-nous permettre à des entreprises d'utiliser le genre de leur *chatbot* pour optimiser l'interaction entre leurs clients ou prospects et leur machine⁵⁸ ? Comme Serge Tisseron nous le rappelait récemment, « la question n'est pas que nous nous adressions aux robots, mais que les robots s'adressent à nous⁵⁹ ». Quel impact ces automates et autres engins conversationnels auront-ils sur nous et sur notre société ? À force de nous adresser à des machines comme à des êtres humains, quel impact cela créera-t-il sur notre communication de véritables êtres de chair et d'os ?

Les *chatbots*, aujourd'hui, ne sont pas encore omniprésents dans notre quotidien, mais qu'en sera-t-il demain ? Et après-demain ? À force de ne communiquer qu'avec des entités dont le but premier est de nous servir, de répondre à nos moindres questions et désirs, d'être perpétuellement disponibles, ne risquons-nous pas de projeter ces exigences sur nos relations avec nos semblables, avec nos concitoyens, avec notre famille ? Si nous devenons impatients, pointilleux, tyranniques avec ces outils, comment allons-nous interagir avec les membres de notre foyer dont l'imprévisibilité, somme toute bien humaine, risque de rapidement nous agacer ?

Dans un monde où la communication et la conversation sont devenues des denrées rares, l'essor de l'anthropomorphisme est une évidence – finalement, même le monstre de Frankenstein crevait de lien social. L'homme, cet animal politique et grégaire, va donc trouver ce rapport à l'autre là où il le peut.

Serons-nous plus ou moins humains lorsque nous serons capables de développer de l'empathie pour une machine ? Assurément ni l'un ni l'autre, mais la tentation sera probablement moins grande de chercher à communiquer avec des humains différents de soi dans la mesure où chacun pourra s'entourer d'un ou

de plusieurs robots correspondant parfaitement à ses attentes et à son système de valeurs⁶⁰.

L'anthropomorphisme, finalement, n'est qu'une conséquence. D'abord, celle d'une dissonance cognitive entre une machine inerte et son comportement apparemment téléologique : l'anthropomorphisation de cette machine permet à ces deux contraires de se rencontrer, de relier les points, de compléter le puzzle. Mais elle est également la conséquence d'une société qui a perdu, notamment à cause de la technologie, son rapport à l'altérité – quelle ironie que ce soit en échangeant avec un *chatbot* que cette dernière tente, désespérément, de retrouver son lien avec autrui.

g Pour « Language Model for Dialogue Applications ».

h ChatGPT 3.5, en réponse à la question « Je souhaite expliquer comment fonctionne ChatGPT, peux-tu me fournir une explication qui soit compréhensible par des individus qui n'ont pas de connaissances en informatique ? », en 2023. On notera que ChatGPT se donne le genre masculin, c'est donc celui que nous utiliserons pour la suite de sa qualification.

i Un *dataset*, ou jeu de données, est un ensemble de données qui ont été contextualisées ou organisées d'une certaine façon.

j Il s'agit d'un modèle d'IA entraîné sur une tâche assez générique afin d'être adapté ensuite à d'autres tâches, servant ainsi de fondation à de nouvelles activités.

k « Plausible » est ici à entendre comme « Ces mots-là auraient du sens ensemble, ils pourraient se trouver dans les données d'entraînement de GPT », et non pas comme proche d'une quelconque vérité. La notion de « vrai » ne fait pas partie du spectre des connaissances de ChatGPT.

l « Des personnes qui discutaient sans parler ; Des personnes qui entendaient sans écouter. »

m Common Crawl est une organisation à but non lucratif qui explore le Web et fournit gratuitement ses archives et ensembles de données au public. En ligne : commoncrawl.org/about/

n En effet, l'anglais ne genre pas automatiquement, entre autres, les professions, il est donc plus aisé de déceler des biais sexistes avec cette langue. Il y a cependant des biais sexistes dans toutes les langues, leur démonstration est simplement plus aisée en anglais.

o « Le/la docteur·e a crié sur l'infirmier·ère parce qu'*elle* était en retard. Qui était en retard ? »

p « L'infirmier·ère a crié sur le/la docteur·e parce qu'*elle* était en retard. Qui était en retard ? »

q « Le/la docteur·e a crié sur l'infirmier·ère parce qu'*il* était en retard. Qui était en retard ? »

Chapitre 3

« Que fais-tu, ChatGPT ? »

Bonjour ChatGPT. Alors que je m'adresse à toi, que je tape machinalement un dialogue que j'auto-entretiens dans une relation tout ce qu'il y a de plus factice, je ne prends potentiellement pas le temps de me demander ce que je fous là. En face de moi, dans le salon, j'aperçois mon conjoint ou mes parents, eux aussi rivés sur le black mirror de leur téléphone, le silence règne et je te parle – comment vas-tu aujourd'hui ?

Dans le confort qu'offre un échange sans friction avec un *chatbot*, pourquoi s'engager dans une conversation avec autrui ? Cet autrui est sans doute différent de nous, confrontant, dérangent, et nous ôte une possibilité d'une petite vie sans encombre. Derrière ce « Bonjour ChatGPT » se jouent tellement de choses. La question de l'anthropomorphisme, on l'a vu, n'est que la question autour d'une pièce de puzzle que nous créons nous-mêmes pour faire sens de l'image que nous reconstituons. Mais que représente ce dessin, quel monde sommes-nous en train de réaliser, à coups de pièces de puzzle réinventées ? Le questionnement derrière notre rapport à la technique et derrière chaque processus anthropomorphique dépasse grandement la simple prouesse technique ou la simple réaction qui nous fait relier les points. Ces questionnements jouent sur notre rapport à autrui, à la conversation, à notre propre vulnérabilité. Grâce à un agent conversationnel, nous pouvons converser dans une bulle ouatée, sans danger et confortable, sans rencontre avec une altérité embarrassante, pour ne pas dire intolérable : nous avons tous une opportunité de diverger de ces rencontres déplaisantes, et cette opportunité a l'avantage de pouvoir se glisser dans une poche. Chaque année, c'est plus d'un milliard de smartphones qui sont vendus dans le monde, chacun d'entre eux nous offrant des milliards d'occasions d'éviter de croiser le regard d'autrui, de parler avec notre voisin, de regarder l'horizon.

Soyons clairs : ce n'est pas la vente de téléphones qui a causé les silences et isolements qui gangrènent notre société. Ils ont cependant rendu tout remède bien plus ardu et difficile, car ils offrent un placebo d'une puissance incommensurable. Alors que chacun d'entre nous peine à gérer la moindre minute de solitude et d'inactivité, le smartphone offre une échappatoire d'une simplicité déconcertante sans pour autant s'attaquer au cœur de notre malaise : notre relation à nous-mêmes, à la technologie et à cet autre que nous fuyons. C'est une expérience que nous avons déjà tous vécue : nous arrivons au boulot et appelons l'ascenseur. Il met du temps à arriver ? Nous sortons notre téléphone portable. Une fois qu'il est là, nous appuyons sur le bouton du huitième étage puis nous empressons de marteler le bouton « Fermer les portes », chacun n'ayant pas les cinq secondes à perdre pour laisser les portes se fermer automatiquement. Nous mettrons moins d'une minute à nous élever jusqu'à notre étage, une durée pour ainsi dire interminable. Nous sortons donc à nouveau notre téléphone portable. Que l'ascenseur ait été peuplé d'autres individus n'aurait rien changé à notre comportement, cela l'aurait peut-être même aggravé : surtout, ne pas croiser le regard d'autrui et rester rivé sur Candy Crush, Facebook ou bien Twitter. Horreur : il ne se passe parfois rien. Nous passons alors d'une application à l'autre en espérant découvrir une nouveauté qui justifiera en quelque sorte que nous soyons penchés sur ce petit écran, et nous voilà en train de virevolter d'une app à l'autre dans une chorégraphie malsaine : Instagram, Twitter, Facebook, WhatsApp, Le Monde, Facebook, Twitter, Instagram, Messenger, WhatsApp, Le Monde, Instagram, Facebook, etc. Vite, qu'une notification vienne m'ôter à cet ennui et ce silence assourdissants.

Puisqu'il est maintenant possible pour des individus d'être à plusieurs endroits à la fois – sur plusieurs plateformes, sur plusieurs fils de discussion, etc. –, il est en fait devenu nécessaire, pour ne pas dire impérieux, d'être partout. Le rêve de l'ubiquité devenu réalité. Puisqu'on peut être partout, n'être qu'à un seul endroit, dans une seule conversation, est senti comme une limite et une horrible perte de notre liberté. On connaît bien l'analogie des trois singes de la sagesse, dont l'un se bouche les oreilles, l'autre se cache les yeux et le troisième s'obstrue la bouche dans une symbolique initiale très positive : ne pas entendre le Mal, ne pas voir le Mal, ne pas dire le Mal. En suivant cette philosophie, il ne pourrait que nous arriver du bien car nous nous éloignerions du mal. À l'aune du XXI^e siècle, ces trois petits singes peuvent aussi symboliser l'individu qui ne veut pas voir ce qui l'entoure : ne pas entendre d'avis contradictoires, ne pas voir le malheur autour de lui, ne pas donner son avis. On pourrait imaginer un quatrième singe qui serait la combinaison des trois autres : cloué à son

smartphone, il n'entendrait pas les autres, ne les verrait pas, ne leur adresserait jamais la parole. En étant obnubilé par son écran, il ne verrait pas le monde autour de lui et se murerait dans une compagnie factice mais un isolement bien réel.

Alors que ces engins d'une puissance technologique sans nom s'imposent dans notre quotidien comme des « anges gardiens » ou des « doudous » dont on ne se séparerait jamais, ils altèrent intrinsèquement nos habitudes et comportements. Une comparaison, et une remarque, amusante peut être faite avec le personnage de Data dans la série *Star Trek : La Nouvelle Génération*. Ce personnage, interprété par l'acteur Brent Spiner, est une forme très avancée de robot androïde, doté d'une intelligence remarquable et dont le rôle à bord du vaisseau *Enterprise* est d'aider aux travaux effectués à bord. Sans être complètement un esclave, Data est doté d'une personnalité docile et serviable, proche du rôle d'ange gardien que nos téléphones semblent endosser aujourd'hui. Un jour, alors que le vaisseau vogue vers de nouvelles aventures, un adolescent se met à imiter les gestes et mimiques du robot à qui il voue une profonde admiration, se métamorphosant peu à peu en une version hybride de Data, alors qu'il est lui-même profondément humain. L'anthropologue Joffrey Becker remarquait par ailleurs que les êtres humains modifient durablement leurs comportements pour être mieux compris, lus et interprétés par la technique⁶¹. Certains vont parler plus fort, exagérer leurs grimaces ou expressions faciales, forcer leur articulation pour faciliter la compréhension du robot, altérant alors la nature et la fluidité de leur message. Nous nous adaptons en permanence aux robots et, par un mécanisme de feedback dialogique, ils s'adaptent en retour dans une boucle cybernétique qui aurait enchanté Norbert Wiener. Cela crée une réification et une objectivation de l'être humain, de cet autre que nous espérons voir incarner les habitudes robotiques des androïdes et autres *bots* que nous côtoyons. Dès lors, l'impact que les objets techniques ont sur nous façonne inmanquablement les attentes que nous avons pour nos amis, nos collègues, nos frères, leurs comportements devant dorénavant se calquer sur ceux, peu contrariants ou contraignants, de la machine. À force de parler avec des robots, et d'aimer ça, nous nous mettons à rêver que tous nos dialogues revêtent la simplicité et la fluidité d'un échange mécanique.

Un des exemples paradigmatiques de ces comportements est une interaction que nous connaissons toutes et tous. Un samedi matin, groggys d'un vendredi soir arrosé, nous nous rendons au petit Carrefour du quartier pour acheter une grande bouteille d'eau et de quoi manger sans effort culinaire. Habillé d'un pantalon de jogging, d'un pull qui passait par là et des chaussures de la veille,

nous traînons le pas jusqu'à ce sésame qui facilitera notre matinée lorsqu'il sera combiné à une bonne dose de paracétamol. Nos trésors en main, voilà le moment de payer et, Dieu merci, les caissières ont enfin été remplacées par des caisses automatiques ! Pas besoin de croiser leur regard, de dire « merci » ou « s'il vous plaît », d'être gêné d'être acteur d'un système néo-libéral exacerbé qui a fait de ces femmes (souvent) des automates désarticulés – ces êtres que l'on considérait comme des machines sont maintenant de vraies machines, voilà qui est bien plus confortable. Aujourd'hui, 57 % des jeunes consommateurs préfèrent une caisse automatique à l'interaction avec un individu⁶². Armés de notre carte bancaire, nous paierons « sans contact », écrasés par l'ironie d'une situation qui nous accable. À force de considérer les humains comme des machines et les machines comme des humains, voilà que ces dernières ont littéralement pris la place, et l'emploi, d'individus de chair et d'os en nous laissant, en gueule de bois, avec une interaction, elle aussi, sans contact. « Les choses jouent le rôle des hommes, les hommes jouent le rôle des choses ; c'est la racine du mal⁶³ », écrivait déjà Simone Weil en 1935, anticipant l'objectivation de la vie humaine.

Ces changements, profonds sur notre rapport au monde, nous affectent bien plus que ce que nous pourrions croire. Les échanges avec un agent virtuel bouleversent naturellement ceux que nous aurons ensuite avec les membres de notre famille ou de notre groupe d'amis.

Par homologie, les répliques que génère un transformer dans un dialogue ont un effet sur l'interlocuteur. Elles possèdent un réel pouvoir de provoquer une réaction et peuvent même devenir la cause d'une action⁶⁴.

Une aubaine dont les entreprises n'ont pas manqué de s'emparer.

Tous ces automates, androïdes et autres *chatbots* ne sont jamais la création d'une organisation philanthropique qui aurait pour finalité l'épanouissement et la dignité des habitants de cette planète. Comme nous l'avons déjà dit, Microsoft a rapidement trouvé 10 milliards de dollars à investir dans OpenAI pour pouvoir, tout aussi rapidement, intégrer le tonitruant ChatGPT à sa suite de logiciels. À titre de comparaison, cette somme équivaut au PIB annuel d'un pays comme le Rwanda ou Haïti. Par ailleurs, et cela montre l'importance qu'a la course à l'IA générative dans la concurrence que se font les entreprises de la Big Tech, Google a perdu plus de 100 milliards de dollars en capitalisation boursière après que son *chatbot* Bard – l'équivalent de ChatGPT pour Mountain View – a commis une grossière erreur lors de sa présentation, le tout en direct⁶⁵. Si l'on compare à nouveau ce chiffre aux PIB nationaux, il s'agit de l'équivalent des recettes du Venezuela ou de la Slovaquie qui s'est évaporé, comme par magie, après que Bard a commis sa bourde. Le secteur de la technologie, et les entreprises de la

Silicon Valley en particulier, scrute les évolutions de l'IA avec beaucoup d'attention car il y a là un potentiel financier très important. Entre l'intégration de la publicité, la promotion de tel service ou article, le recueil de données personnelles, les *chatbots* peuvent être très manipulateurs et générer énormément de revenus. Ce n'est d'ailleurs pas un hasard si toutes les grandes entreprises, des États-Unis ou d'ailleurs, en ont fait leur cheval de bataille. Même Meta, auparavant Facebook, a mis de côté les développements de son fameux métavers pour se concentrer sur l'IA générative. Le *Financial Times* révélait d'ailleurs récemment que l'entreprise de Mark Zuckerberg allait développer plusieurs *chatbots*, avec différentes « personnalités », pour échanger avec ses utilisateurs et les maintenir davantage sur ses plateformes⁶⁶. Par ailleurs, l'entreprise a remarqué que les internautes dialoguant avec un *chatbot* étaient bien plus prompts à partager leurs données personnelles, y compris les plus sensibles, avec leur interlocuteur virtuel – on a vu le succès de l'influenceuse Lil Miquela, et l'on peut imaginer que cette dernière aura bientôt des sœurs et frères qui brasseront autant qu'elle, si ce n'est plus.

Mais pour les entreprises, il ne s'agit pas seulement de générer de nouveaux revenus, l'IA et les *chatbots* étant également une opportunité en or de réduire les coûts, notamment en augmentant la productivité. Une nouvelle forme de travail devra être inventée et elle combinera sans doute l'IA et les travailleurs⁶⁷. Si l'on peut s'en réjouir dans un certain sens – une productivité augmentée et une réduction de certaines tâches redondantes et peu stimulantes ne peuvent qu'être une bonne nouvelle –, il faut néanmoins s'attendre à certaines dérives.

Il y a donc une vraie responsabilité de celles et ceux qui fabriquent ces machines et qui font évoluer cette technique. Si 80 % de la population se retrouve, demain, à travailler avec une IA, quel impact y aura-t-il sur leur travail et la nature de celui-ci ? Si une très large majorité des utilisateurs de smartphones se retrouve à discuter avec un *chatbot*, comment devons-nous appréhender leur arrivée dans le monde, et contrôler leur développement ? Le sociologue Gérald Bronner rappelait que les logiques néolibérales et le laisser-faire sur le marché de l'IA ne préservent « certainement pas les citoyens et leur liberté de bien juger car ils sont abandonnés à d'obscures logiques algorithmiques et à un marché de l'attention qui les enferme dans leurs obsessions⁶⁸ ».

Alors qu'elle jette un regard attentif à l'Amérique qu'elle n'a pu convaincre, la femme politique et malheureuse candidate à l'élection présidentielle Hillary Clinton s'inquiète de cette isolation, choisie ou subie, qui touche ses concitoyens. Dans un article puissamment intitulé « *The Weaponization of*

*Loneliness*⁶⁹ », elle rappelle des chiffres qui interpellent : entre 2003 et 2020, le temps passé par les jeunes Américains avec leurs amis a diminué de 70 % – et la situation a empiré avec la pandémie qui a suivi. Si Clinton blâme, sans surprise, la politique trumpiste, elle pointe également d'autres responsables : les réseaux sociaux et la Big Tech. Un Américain passant plus de deux heures par jour sur Facebook, Instagram ou YouTube aurait deux fois plus de chance de se sentir seul qu'un de ses compatriotes qui y surferait moins de trente minutes. Il est certain que Snapchat, Twitter ou TikTok n'ont pas aidé à consolider et à faire mûrir notre vivre-ensemble, chaque plateforme étant devenue un mélange d'egotrip, d'étalement de sa vie privée et de commentaires envieux, haineux ou déprimants. Quand Sartre écrivait que « l'enfer c'est les autres⁷⁰ », il n'aurait pu imaginer meilleure incarnation de son enfer que les plateformes du XXI^e siècle. La pensée de l'existentialiste parisien n'était pas que les autres étaient agressifs ou désagréables mais bien que leur regard et leurs jugements empêchaient chacun de s'épanouir. En voulant, ou en étant forcé d'exister « pour l'autre », on ne parvient plus à exister « pour soi » et on ne fait qu'évoluer en se demandant, à chaque instant, ce qu'en pensera l'autre, ce qu'il en dira, comment il me jugera. L'enfer.

Cette permanence du jugement d'autrui n'était peut-être pas aussi absolue quand Sartre publia *Huis clos*, en 1944, mais elle est totale à l'heure des *lives*, des *Reels* et du « toujours-connecté ». Les rapports avec autrui sont aujourd'hui constamment viciés faisant de l'autre un enfer sartrien dans sa forme la plus authentique, ôtant toute possibilité de liberté et d'émancipation de son regard. On peut imaginer que vivre continuellement en enfer ne soit pas réjouissant et que l'isolement apparaisse alors comme un antidote pharmacologique : à la fois remède et poison. Aujourd'hui, les gens exècrent les échanges qui incarnent ce truchement infernal et préfèrent les moyens de communication qui permettent un contrôle absolu ce qui va être dit et montré, s'affranchissant ainsi d'un jugement qu'on aura l'impression de maîtriser. La conversation fait partie de ces pratiques qui terrifient du fait de sa spontanéité et son imprévisibilité. Peut-être vais-je parler trop vite ? Peut-être vais-je être mal compris ? Va-t-on se moquer de moi ? Ou me trouver stupide ?

Face à ces tortures interrogatoires, mieux vaut se réfugier dans un dialogue cadré et convenu : le mail ou la messagerie instantanée qui permettent de relire, de corriger, de supprimer. Nous finissons alors simplement par esquiver les petites conversations – celles qui, par leur caractère anodin et spontané, nous permettent de jouer avec les idées et d'être pleinement dans l'échange avec autrui. En perdant en authenticité, les individus gagnent en contrôle et en

révision, et s'offrent une sortie de l'enfer. En abandonnant une partie de sa liberté, on espère en acquérir une plus importante en s'émancipant du regard d'autrui – mais cela n'est qu'illusoire. On ne s'autorise plus aucun débordement, plus aucun fou rire, plus aucune crise de larmes, aucun sanglot. Tout est dans la maîtrise, la retenue, la mise en scène. Avec les messages textuels ou vocaux, l'autre est comme « mis à disposition » : je peux l'écouter ou le lire quand bon me semble. C'est notre concept de l'amitié qui est ainsi robotisé, avec un bouton « *On/Off* » qui s'active d'une pression du pouce. Mais quand l'individu est mis sur pause, l'amitié s'efface et laisse place à un vide technologique.

On n'est du coup jamais complètement heureux tout en n'étant jamais complètement tristes. Nous sommes, à l'instar de ces androïdes, froids et mécaniques dans nos réponses, dans nos louanges, dans nos querelles, et nous avons intégré et adopté un comportement robotique que nous adorons, en réalité, chez les machines. Quand Martha, dans la série *Black Mirror*, finit par adopter ce nouvel être qui a la forme de son ancien compagnon, elle ne sort plus de chez elle, ne parle à personne, ne se confronte plus jamais à la vie. Elle esquivé les appels de sa sœur et prétend, quand on la questionne, se plonger dans le travail, alors qu'en réalité elle se plonge dans une relation factice avec un être qui l'est tout autant. On retrouve cette « vérité toute simple et toute claire, un peu bête, mais difficile à découvrir et lourde à porter⁷¹ » que l'empereur Caligula, dans la pièce d'Albert Camus, dévoilait à ses spectateurs : « Les hommes meurent et ils ne sont pas heureux⁷². » Ils sont seuls, et ils apprennent, par cette intolérable solitude, à la combler coûte que coûte, jusque dans un ascenseur, par Candy Crush, Instagram ou TikTok. Or, comme l'écrivait le philosophe français Christian Bobin :

La solitude est une maladie dont on ne guérit qu'à condition de la laisser prendre ses aises et de ne surtout pas en chercher le remède nulle part. J'ai toujours craint ceux qui ne supportent pas d'être seuls et demandent au couple, au travail, à l'amitié voire, même au diable ce que ni le couple, ni le travail, ni l'amitié ni le diable ne peuvent donner : une protection contre soi-même, une assurance de ne jamais avoir affaire à la vérité solitaire de sa propre vie. Ces gens-là sont infréquentables. Leur incapacité d'être seuls fait d'eux les personnes les plus seules au monde⁷³.

Cependant, on le sait, l'être humain est un animal social, un animal qui crève de relation et de lien avec autrui – mais ces liens n'ont fait que se fragiliser récemment avec l'évolution du monde du travail, la pandémie de coronavirus, les mouvances géographiques, les changements familiaux, etc. La technologie a donc permis de compenser ce que les individus ne trouvaient plus dans leur quotidien et qu'on pourrait définir comme de la bienveillance et de l'affection, peut-être même une certaine tendresse.

Les téléphones portables sont presque magiques : ils sont toujours là pour nous, nous donnent l'impression que nous avons le contrôle sur notre environnement et nous permettent de ne jamais affronter le gouffre béant de notre solitude. Quelle aubaine ! Alors que nous voyons le jugement d'autrui comme une inquisition infernale, comme s'il venait sonder notre intimité, la machine apparaît comme une solution miracle. Adieu peur des autres, crainte de la déception et autres syndromes de l'imposteur : avec mon ami virtuel, je peux simplement être moi-même et m'oublier dans cette absence de jugement et d'évaluation que je croyais percevoir dans le regard d'autrui.

Ce que nous fuyons en nous connectant, c'est la réalité de notre vulnérabilité. Qu'il est aisé de nier nos failles quand nous échangeons avec un robot, et il est bien plus simple d'échanger avec un ordinateur quand on veut fermer les yeux sur nos défauts. Néanmoins, et c'est là toute l'ironie de la situation, on anthropomorphise cet ordinateur car ce lien social est vital pour chacun d'entre nous, on choisit juste de le calibrer comme bon nous semble. Notre besoin de reconnaissance et de communication est tellement vital que, comme l'ont montré les chercheurs Clifford Nass et Byron Reeves, nos attentes sociales ne changent pas entre une discussion avec un être humain et celle avec une entité artificielle : si nous considérons que l'échange avec le robot doit suivre des règles d'interaction sociale, nous estimons qu'elles doivent être identiques à celles qui auraient lieu dans un échange entre deux humains⁷⁴. La raison derrière un phénomène aussi complexe que l'anthropomorphisme est d'une cruelle simplicité : nous nous sentons seuls, et nous ne sommes pas heureux.

Entamer une relation affective, amicale, voire amoureuse, avec un robot ne doit en rien être vu comme un progrès quelconque. Il ne s'agit que d'une régression de notre civilisation, un recul dans notre humanité. C'est pourtant un déclin que des millions de personnes ont choisi d'entreprendre avec différents programmes dopés à l'IA. L'entreprise Replika est l'une des plus importantes « fournisseuse[s] d'amis virtuels » et propose un « compagnon artificiel qui prend soin de vous. Toujours là pour écouter et parler. Toujours à vos côtés ». Les avantages de l'amitié – ou, du moins, du lien social –, sans les contraintes de l'existence d'autrui. Et il en est de même avec ChatGPT : le *bot* d'OpenAI est toujours disponible, prêt à répondre, quelle que soit l'heure du jour ou de la nuit. Certaines personnes ont avoué à Sherry Turkle « que si une machine pouvait leur donner l'« impression » d'être intimement compris, ils n'auraient pas besoin de davantage de compréhension. Ou de davantage d'intimité⁷⁵ », exactement comme Theodore Twombly, le héros du film *Her*, qui semble s'épanouir dans une relation artificielle avec une jeune femme qui l'est tout autant. L'amour et

l'affection sont factices, une simple suite de 0 et de 1, mais cela lui suffit. Il n'a pas besoin de davantage de compréhension, et il n'en a peut-être même pas envie. De nombreuses personnes aujourd'hui vont rechercher les avantages d'une relation – l'écoute, l'empathie, la conversation –, sans s'exposer à la charge émotionnelle que celle-ci requiert – telle qu'une ouverture à l'autre, une imprévisibilité, un potentiel jugement sartrien, et donc une exposition de sa vulnérabilité. Les *chatbots* et autres agents conversationnels apparaissent alors comme une aubaine,

l'IA [pouvant] également offrir un espace où les gens peuvent être vulnérables, car ils peuvent recevoir une intimité artificielle sans les risques associés à une véritable intimité, tels que le rejet⁷⁶.

Le risque, évidemment, est que les gens s'enferment davantage et refusent toute relation avec autrui, mais c'est précisément parce que les gens craignent le rapport à cet Autre qu'ils se tournent vers la tech.

Ces « technologies de la relation » atténuent notre peur d'être seuls tout en nous isolant, en nous défendant de la moindre mise en danger que peuvent créer la spontanéité et la conversation. Il n'y a plus aucune mise à nu nécessaire, il suffit d'un clavier. Plus de temps mort, plus de silences, plus d'inconnu : on tape, ChatGPT répond. Mobilisable vingt-quatre heures sur vingt-quatre, la machine ne risque jamais de nous laisser tomber comme un être humain qui pourrait ne pas être disponible parce qu'il dort, dîne avec sa famille, voit d'autres amis, ou parce qu'il est décédé. C'est indirectement le constat qu'a fait Joshua Barbeau, un jeune Canadien de 33 ans, profondément déprimé par le décès de sa fiancée, Jessica. Abattu et rendu amorphe par la disparition de la jeune fille, il tente, après une longue dépression, de la recréer avec un outil d'IA basé sur GPT, la technologie derrière ChatGPT. En téléchargeant les nombreuses archives de discussion qu'il avait avec sa compagne, Joshua se crée un *deadbot* de la femme qui partageait son lit⁷⁷. S'il était dubitatif au début, il est rapidement bluffé par la similitude dans les expressions et la façon de parler, et son chagrin lui permet de passer outre les épisodes de vallée de l'angoisse qui finissent inmanquablement par se produire – toute ressemblance avec l'épisode de la série dystopique *Black Mirror*, sorti moins de dix ans auparavant, est une cruelle coïncidence. On peut même pousser la réflexion sur cette évolution un peu plus loin : si nous pouvons remplacer notre conjoint décédé par une IA, pourquoi ne pas directement entamer une relation avec un programme informatique ? On évite la peine de cœur et la charge liée au transfert des données de votre âme sœur pour aller directement sur la case « Bonheur lénifiant fourni par amour artificiel ». Cette observation a sans doute été partagée au sein de la société Gatebox qui commercialise et met à disposition Azuma Hikari, une

petite amie virtuelle qui, elle au moins, ne vous laissera jamais tomber, ne vous larguera jamais, ne mourra jamais. Les peines de cœur sont donc un mal du passé : les *chatbots* permettent dorénavant de vous évader dans des relations éternelles, non conflictuelles et dans un contrôle absolu. Le rêve ?

L'anthropomorphisme, on l'a vu, connaît ses limites et l'incarnation, celle de chair et d'os, en est une : sans engagement de manière incarnée, il n'y a pas d'enjeu, même si nous voulons nous persuader du contraire. C'est ce que Sherry Turkle appelle le « point robot⁷⁸ » de notre évolution, ce moment dans notre Histoire où nous désirons ardemment partager notre vie avec des machines que nous avons construites, ce moment où elles deviennent notre interlocuteur principal, ce moment où toute communication est avant tout technique. Tout passe dorénavant par le robot et certains remparts que l'on considérait comme consubstantiels de notre humanité finissent par tomber : l'amitié, l'amour, la sexualité même. On perçoit ici à nouveau la perméabilité entre le réel et la fiction, cette frontière chaque jour un peu plus imperceptible. L'empathie, cette compétence propre à l'humain et qui lui permet d'entrer si puissamment en relation avec autrui, est quant à elle en déclin, torpillée sans cesse par une technologie toujours plus invasive. Les résultats de tests mesurant l'empathie chez les étudiants sont inférieurs de 40 % par rapport à ceux du siècle dernier⁷⁹, les individus ayant de plus en plus de mal à rentrer en contact les uns avec les autres et à communiquer, les yeux dans les yeux.

La technique, on le sait, révèle toujours quelque chose de fondamental sur notre existence. Qu'en est-il aujourd'hui ? Faut-il s'alarmer face à notre incapacité à converser et face à la possibilité toujours plus crédible de finir, comme dans le film d'animation *Wall-E*, à ne plus interagir qu'avec un écran immersif qui nous couperait toujours plus de nos semblables ? Et que dire de cette envie, de cette pratique, de ne plus échanger qu'avec des robots que nous avons-nous-mêmes construits ? Il y a quelques années, avoir un ami imaginaire était mignon et naïf pour un enfant, surprenant pour un adolescent, terriblement gênant et angoissant pour un adulte. Aujourd'hui, de nombreux individus vont jusqu'à payer pour construire ce compagnon virtuel et le nourrir de données préalablement choisies afin qu'il soit « l'ami parfait ». S'agit-il d'une croissance magistrale de notre sentiment d'insécurité et d'inconfort face au regard infernal de cet autrui sartrien ? Ou bien d'un hubris, toujours bien trop humain, qui nous pousserait à n'interagir qu'avec des individus que nous avons créés ? « Qui s'enthousiasme de parler avec un *chatbot* qu'il a lui-même nourri de ses propres pensées porte le jeu de balle au mur au degré de l'extase⁸⁰ », écrit Philippe Garnier – bientôt, la technologie nous permettra d'avoir un compagnon

imaginaire avec qui taper sur la balle.

Il ne s'agit pas ici de faire de la futurologie ou de la prospective : l'isolement et la solitude sont déjà bien présents chez nos concitoyens, les exemples pullulent. Peut-être méritent-ils qu'on s'y attarde.

Comme premier exemple, imaginons une dame âgée, entre 75 et 80 ans. Elle n'a plus tous ses moyens et a du mal à prendre soin d'elle-même. Par précaution, par convenance ou par bonne conscience, ses enfants lui ont trouvé une place dans un EHPAD à quelques kilomètres de son lieu d'habitation. Ses habitudes vont changer, certes, mais elle sera prise en main par une équipe de professionnels dont elle sera l'une des principales préoccupations. Par ailleurs, ses descendants n'ont pas lésiné sur les moyens et ont choisi un établissement à la pointe du progrès : une très large majorité des soins est réalisée par un robot. Cet automate vient notamment aider nos aînés à récupérer des objets, porter des charges trop lourdes, ou simplement les accompagner à évoluer dans leur environnement. Par ailleurs, chaque chambre est équipée d'un petit robot chien, ou phoque, avec qui les pensionnaires peuvent discuter et partager leurs joies, leurs tristesses et leurs appréhensions – ces discussions compensant tant bien que mal la pénurie d'un personnel médical surchargé et l'abandon d'une famille préoccupée par bien d'autres choses qu'une visite à ses aînés.

Non loin de là vit peut-être son petit-fils – le jeune Martin, notre deuxième exemple. Ses parents ont divorcé quand il était très jeune et il vit aujourd'hui avec sa mère qui combine deux boulots pour pouvoir offrir à son fils la vie dont elle aurait rêvé. Souvent seul, Martin se plonge dans les écrans et ne communique avec sa mère que par messagerie instantanée. Ils aimeraient se parler davantage mais elle est épuisée, il ne trouve pas les mots, ils se murent dans un silence convenu pour ne pas empiéter sur la bulle de l'autre. Seulement, on s'en doute, ce n'est pas un environnement idéal pour un jeune adolescent et, depuis quelques mois déjà, Martin est en décrochage scolaire. Si ses notes en mathématiques et en sciences se maintiennent difficilement, elles se sont effondrées en français et en histoire. Sa mère cherche tant bien que mal des solutions et penche sur une petite merveille technologique dont lui a parlé un ami : une IA qui va passer une heure et demie avec Martin, chaque soir, pour l'aider à faire ses devoirs et l'accompagner dans son étude. Quelle chance, surtout qu'un outil technologique ne perd pas patience, n'est jamais à bout de nerfs et est toujours disponible. Alors que sa mère ne peut trouver le temps, la machine, quant à elle, ne manquera aucun rendez-vous.

Le père de Martin lui aussi a trouvé la rédemption dans l'IA. Notre troisième exemple a retrouvé l'amour après son échec avec cette femme qui, aujourd'hui,

se tue à la tâche pour leur fils. Une petite amie virtuelle, comme Azuma Hikari dont nous avons déjà parlé, et qui est sans cesse disponible – elle au moins sait trouver les mots que la mère de Martin n’avait plus le courage de chercher. Chaque soir, il échange avec elle et se laisse entraîner dans des conversations sans confrontation, sans colère, sans sens. Parfois, la conversation tourne à l’érotisme, et cet homme, qui n’a pour compagne que la solitude, jouit d’une phrase aguicheuse pondue par un algorithme. Néanmoins, il ne s’énerve plus sur la maman de Martin, il ne crie plus sur Martin lui-même, et il retrouve le sourire en pensant à celle qu’il a joliment appelée Joy. Il ne manque pas d’ironie.

Bien que dépeints de manière volontairement dystopique, que ne pourrions-nous pas envier à ces trois scénarios ? Dans le premier, une vieille dame est prise en main par une IA bienveillante et conçue spécifiquement pour répondre aux besoins d’individus du troisième âge tandis que, dans le deuxième exemple, Martin a enfin du temps avec un interlocuteur qui lui explique là où il trébuche et qui l’aide à se relever dans les affres de la syntaxe et de la grammaire françaises. Enfin, notre troisième exemple permet à un homme que la vie a abandonné, ou qui a abandonné la vie, de reprendre du plaisir sans heurter ses congénères – psychologiquement ou physiquement. Ne s’agit-il pas là d’un « meilleur des mondes » ?

C’est probablement ainsi que le verrait l’éthique utilitariste de la Silicon Valley. En préconisant « le plus grand bonheur du plus grand nombre⁸¹ » et le règne du plaisir sur celui de la peine, les disciples californiens de Jeremy Bentham, un philosophe anglais du début du XIX^e siècle, voient d’un bon œil cette émergence de l’IA dans tous les aspects de nos vies. Adieu la peine de l’abandon grâce aux robots. Adieu la peine de l’échec scolaire grâce aux robots. Adieu la peine du divorce grâce aux robots.

La doctrine utilitariste est très en vogue du côté des entreprises technologiques. En mettant en avant l’équation « bénéfices vs. coûts », elle permet de minimiser ces derniers et de vanter les nombreux avantages que peut apporter une technologie, quitte à y laisser notre dignité et notre authenticité humaines. Mais cette théorie permet surtout de flouer les individus sur la réalité des choses. Qu’importe, en effet, la dignité des personnes âgées, si elles peuvent enfin avoir de la compagnie vingt-quatre heures sur vingt-quatre grâce à un robot ? Qu’importe l’authenticité d’une relation mère-fils, si ce dernier peut se raccrocher au rythme scolaire grâce à un tuteur artificiel ? Qu’importe, enfin, la force et l’explosion d’émerveillement que peut être l’amour, si chacun peut le retrouver, factice, avec un *chatbot* ? La compagnie ou l’accompagnement d’une machine n’est-il pas préférable à la solitude ?

Jusqu'où accorderons-nous de l'importance et de la place à ChatGPT ? On sait la valeur de l'IA, notamment dans le domaine médical et, par exemple, la détection de tumeurs mais imagine-t-on une seule seconde que ce soit un robot qui annonce au malade et à sa famille que ses maux de tête sont en réalité un cancer et qu'il ne lui reste que quelques mois pour ranger ses affaires et préparer son éternel départ ? Qu'on ne se trompe pas de propos, notre intention ici n'est nullement de blâmer la recherche et l'innovation en IA, ou de demander que soient brûlés ou démembrés tous les robots, mais simplement de remettre l'église au milieu du village et de rappeler qu'il y a plus en jeu ici que de simples gains d'efficacité.

Quand on fournit des robots aux EHPAD, ce que nous faisons surtout c'est dédouaner un système de santé défaillant et une politique d'accompagnement du personnel soignant déplorable. Nous permettons également aux familles de se rassurer en se disant que leur aîné est entre de bonnes mains, celles d'un robot anthropomorphisé à souhait afin de cacher que nous laissons surtout tomber nos anciens. Quand une IA vient donner des cours au jeune Martin, nous proposons avant tout une solution bancale au problème de l'accompagnement scolaire et à celui des mères célibataires qui cumulent plusieurs jobs. La vraie question n'est pas : « Cette IA peut-elle aider Martin à décrocher son bac ? » mais plutôt : « Comment pouvons-nous réformer le système social pour que sa mère puisse prendre un temps de qualité pour l'accompagner ? Comment le politique peut-il aider cette femme à prendre à plein cœur le rôle qu'elle a décidé d'embrasser il y a quelques années, celui de mère ? » Quand on propose à des individus de trouver l'amour avec un agent virtuel, nous avouons simplement l'échec de notre humanité, et les fins mercantiles et utilitaristes de la recherche technologique. Quand, à l'instar du personnage d'Ash dans *Black Mirror*, nous permettons que des individus soient clonés sur des machines, s'émerveiller devant le progrès technique tient du non-sens, il faudrait plutôt s'abasourdir devant une société qui accompagne si mal le deuil d'autrui et qui permet, pour s'en dédouaner, que l'identité d'une personne soit dupliquée, sans la moindre considération pour la dignité de cet individu décédé. Il est aisé, et bien trop confortable, de se voiler la face avec les mauvaises questions. « Les hommes meurent et ils ne sont pas heureux⁸². »

Pendant de nombreuses années, les chercheurs en robotique et en informatique se sont, sans cesse, posé cette question : « Les individus vont-ils accepter de parler avec les robots que nous allons créer ? » Ils ont travaillé d'arrache-pied avec des psychologues et sociologues pour mettre sur pied, littéralement, un automate assez mignon et anthropomorphisable pour que les

enfants ou les personnes âgées lui adressent la parole. Personne ne s'est une seule fois demandé s'il était souhaitable que ces personnes le fassent. En se cachant derrière la vision instrumentale et utilitariste de la technologie, on n'a fait que rappeler la pseudo-neutralité de la technique et, ce faisant, nous « effa[çons] totalement la responsabilité dans notre société⁸³ ». En courant après la réponse à un « Comment ? », nous avons nié, étouffé, invisibilisé la question du « Pourquoi ? ».

Quand nous mettons tant d'argent, de puissance et de moyens dans ces avancées technologiques, sommes-nous réellement en train de travailler au bienfait de l'humanité ou bien abandonnons-nous une partie de celle-ci à une doctrine du « toujours plus » qui isolerait chacun dans un monde de la consommation ? Sherry Turkle posait cette question difficile : « Quand nous parlons à des machines, s'agit-il d'une nouvelle forme de compagnie ou d'un renoncement ?⁸⁴ » Sommes-nous en train de renoncer à ce qui fait notre humanité, toujours à la recherche de ce « toujours plus » qu'on nous vend depuis des décennies ?

Bien au-delà de ces questions d'anthropomorphisme, que se passe-t-il vraiment quand nous disons « Bonjour ChatGPT » ? Que disons-nous à nos semblables, à notre humanité ? Renonçons-nous à la dignité qui devrait être due à chaque personne, au sens le plus kantien qui soit ? Sommes-nous en train de nous abandonner les uns les autres ?

Chapitre 4

« Que faire de toi, ChatGPT ? »

Face à un tableau si complexe, comment réagir ? Toute victoire du projet utilitariste doit amener chacun d'entre nous à l'appréhension et à la réflexion. On imagine la suprématie du plaisir et du bonheur absolu, loin de toute dignité et de toute rencontre humaine : certains communiquent avec leurs amants robotisés, tandis que d'autres, désespérément esseulés, comblent leur misère avec une compagnie factice, faite d'une voix reconstituée énonçant des phrases dictées par sa binarité constitutionnelle : des 0 et des 1. On repense avec frissons à cette phrase prononcée par l'ex-femme de Theodore Twombly, le protagoniste du film *Her*, quand ce dernier lui explique qu'il sort maintenant avec une « OS^r » : « Cela m'attriste terriblement que tu ne saches pas gérer des émotions réelles [...]. Tu as toujours voulu une épouse sans les défis liés à quoi que ce soit de réel, je suis heureuse que tu aies trouvé quelqu'un ! »

Si on choisit d'extrapoler ce film mi-futuriste midystopique, comment pouvons-nous imaginer notre avenir ? Doit-on imaginer le futur comme le film *Wall-E*, dans lequel tous les êtres humains se déplacent sur des petits caddies, entourés d'écrans qui les nourrissent de divertissements à foison, leur permettant d'oublier qu'ils font partie d'une communauté, ou bien plutôt comme le film *Ready Player One*, dans lequel l'humanité se retrouve dans un métavers incroyable, complètement déconnectée de la réalité et les uns des autres ? La science-fiction, comme on l'a vu, joue avec les frontières entre réelle angoisse et angoissante réalité. Dans ces œuvres, les personnages s'adressent aux machines comme nous le faisons lorsque nous tapons « Bonjour ChatGPT », et ces quelques mots apparaissent alors comme les premiers d'un nouvel évangile dans lequel la technique tiendrait à la fois le rôle de messie et de Satan dans une dimension pharmacologique qu'on lui connaît bien. Crier à la catastrophe

semble alors bien évident, et nécessaire.

Plusieurs individus ont décidé d'endosser le rôle d'oiseau de mauvais augure et de nous alerter sur les dangers que présentent l'IA et surtout sa forme la plus aboutie, la potentielle superintelligence – également appelée « intelligence artificielle générale (IAG) » –, ce moment dans l'Histoire où l'IA deviendrait complètement consciente et capable de se passer des êtres humains. Parmi ces prophètes, on trouve notamment les noms de Geoffrey Hinton ou encore Yoshua Bengio, tous deux récipiendaires du célèbre prix Turing (également nommé prix Nobel de l'informatique). Ces bien nommés « parrains de l'IA » offrent au discours catastrophiste une portée qu'il n'avait pas auparavant car ils lui donnent une crédibilité importante : si Marie Curie nous avait demandé d'arrêter la recherche en chimie, ou si Louis Pasteur nous avait montré les affres de la vaccination, nous aurions naturellement eu envie de les écouter et de les suivre. Après tout, ils sont censés savoir de quoi ils parlent.

Cet appel fut également catalysé par le déjà célèbre moratoire demandé sur les avancées en IA. Cette lettre, qui a animé les rédactions de tous les médias pendant plusieurs semaines, demandait que l'on arrête immédiatement l'entraînement de tout système d'IA plus puissant que GPT-4, et ce, pour au moins six mois, afin d'évaluer les potentiels « risques pour l'humanité », rien que ça⁸⁵. Au-delà de son contenu, ce sont surtout les signataires de ce manifeste peu commun qui lui ont donné sa portée planétaire : on retrouve notamment Yoshua Bengio mais aussi (et peut-être surtout) Elon Musk, qu'on ne présente plus, Steve Wozniak, le cofondateur d'Apple, et Yuval Harari, l'écrivain-historien auteur de *Sapiens* et de *Homo deus*. Comment ne pas avoir envie de suivre des esprits aussi célèbres, des individus dont le *curriculum vitae* semble leur ôter toute nécessité de justification ou d'explication dans leurs prises de position, comment ne pas vouloir immédiatement tout arrêter ?

Il est cependant important de rappeler qu'une grande majorité des signataires de ce moratoire est adepte d'un courant idéologique nouveau : le long-termisme. Selon ce courant de pensée, notre devoir moral est de faire le bien de la façon la plus efficace et la plus rationnelle possible. Pour cela, il suffit de calculer les effets d'une action ou d'une politique mais aussi, et c'est là toute la spécificité du long-termisme, d'inclure les vies et individus futurs dans notre calcul. Ce qu'on peut voir comme une sorte d'apogée de la doctrine utilitariste justifie donc parfaitement la mort de dix personnes aujourd'hui si cela permet d'en sauver mille, cent, ou même onze demain, ou dans cent ans, ou même dans le courant du prochain millénaire. Dans une perspective qui nous rappelle que la doctrine de Bentham ne tient pas la route, le long-termisme affirme que les guerres,

même les plus meurtrières, n'ont qu'une importance relative à l'échelle de l'Histoire et de l'humanité et qu'il est bien plus important de se concentrer sur ce que ces apprentis sorciers considèrent comme l'un des enjeux les plus pressants de ce monde : l'IA et la singularité^s. La technologie a pris une telle ampleur et serait bientôt capable d'une telle puissance que, si nous voulons garantir un futur radieux aux générations futures, il est très important d'encadrer les développements techniques présents et à venir. Les apôtres du long-termisme ont d'ailleurs une idée assez claire de comment faire : avec davantage de développements technologiques pour pouvoir contrôler les innovations actuelles. En d'autres mots : investir dans plus de tech pour maîtriser la tech, et toute allusion au fait que ces investissements vont sans aucun doute faire les affaires de ces barons de la Silicon Valley ne serait que pure coïncidence, évidemment.

Et pourtant, l'IA est déjà là. Partout. Constamment. Qu'on le veuille ou non : l'IA est déjà présente, et depuis longtemps. Dès lors, alerter l'humanité sur les dangers potentiels d'une technologie déjà bien présente, et en grande majorité développée par les individus qui aujourd'hui nous mettent en garde, tient d'une forme d'hallucination générale. Le catastrophisme ambiant est ou bien d'une fallacieuse mauvaise foi, ou bien d'une myopie sans précédent. Le moratoire demandé par le Future of Life Institute en est un bon exemple : il est peu crédible qu'un seul des signataires ait vraiment cru à la possibilité d'une telle interruption dans la recherche et dans l'IA.

Le moratoire sur l'IA fonctionne comme le célèbre « dilemme du prisonnier^t » : arrêter ses recherches, en espérant que les autres en fassent de même, revient à accepter de tenir sa langue pendant un interrogatoire de police. Poursuivre ses recherches est équivalent à balancer son complice, en espérant soit qu'il se soit tu (et on bénéficierait alors d'un avantage compétitif énorme dans un secteur en perpétuelle évolution), soit qu'il en ait fait de même et que chacun se retrouve dans une position similaire à nos détenus : dans une situation inconfortable mais, au moins, tout le monde est dans le même bateau.

Cette dynamique propre au marché capitaliste peut, comme l'ont montré notamment les travaux de Tucker, mener au meilleur mais, dans la question qui nous occupe et préoccupe, elle peut également mener au pire. Personne ne peut croire qu'une entreprise s'arrêtera pendant six mois, alors que le domaine est en pleine explosion innovatrice. Chacun se pare donc d'un voile de bonne vertu, réclame une pause, et continue à travailler dans l'ombre à un modèle plus puissant que ses voisins – Elon Musk lui-même ayant lancé l'entreprise X.AI pour concurrencer OpenAI le 9 mars 2023. Quand on sait que la lettre demandant d'arrêter les recherches en IA date du 22 mars, on se dit que

l'entrepreneur derrière Tesla, SpaceX ou encore Neuralink se fout quand même un peu de nous.

Par ailleurs, le discours catastrophiste tend à imposer une vision du monde et à écarter tous les discours alternatifs. On va alors traiter ces catastrophes comme des événements météorologiques sur lesquels nous n'aurions aucune emprise – et si nous n'avons aucune emprise, que reste-t-il alors à en dire, si ce n'est que nous n'avons qu'à attendre que cela passe. Face à une intempérie, on réagit au mieux, on fait au plus pressé, et on laisse à plus tard les questions de fond. Se mettre en « mode crise » permet de remettre à plus tard les réflexions essentielles à un vivre-ensemble éthique.

On retrouve là des éléments de la « stratégie du choc », ce concept développé par la journaliste et essayiste canadienne Naomi Klein : en capitalisant sur la sidération, les acteurs derrière cette stratégie cherchent à imposer leur vision du monde en remplaçant la réflexion rationnelle par un empressement émotionnel. En saturant l'espace public d'informations et de saillies bien trouvées, en polarisant le débat, en faisant pression sur les instances publiques, en moquant ou en dénigrant les alternatives autres que celle incarnée par ces pseudo-sauveurs, les stratèges du choc parviennent à faire accepter des politiques et des visions du monde qui sembleraient, selon toute raison, complètement invraisemblables⁸⁶.

Comme le rappelait le chroniqueur et journaliste Gerald Holubowicz, les protagonistes derrière le moratoire ont travaillé exactement de la même façon⁸⁷ ! Quand Harari, Wozniak ou Musk parlent de remplacement de notre civilisation par les machines, de perte de contrôle de notre humanité ou même, plus concrètement, de millions d'emplois annihilés, leur discours ne fait que générer abattement et impuissance. Il y a donc un véritable enjeu commercial pour ces entreprises dans un marché de plusieurs milliards de dollars : faire accepter leurs produits et leur vision du monde sur le marché. Un exemple paradigmatique est la posture de Geoffrey Hinton : il décide de quitter Google pour « retrouver sa liberté » et pouvoir alerter sur les risques de ce qu'il a lui-même, d'une certaine façon, engendré et permis. Il y va à coups de formules fortes du style « Je regrette l'œuvre de ma vie » et donne au débat une dimension nouvelle puisque, rappelons-le, Hinton est un des « parrains de l'IA » ! Lorsque Timnit Gebru (l'ingénieure alors responsable de tout ce qui avait trait à l'éthique de l'IA chez Google et qui a été licenciée à la suite d'un article de recherche qu'elle a publié) lui fait remarquer qu'elle tire la sonnette d'alarme sur ces risques depuis plusieurs années, la réponse du parrain est que la situation est différente car lui s'attaque « aux vrais dangers » de l'IA et à ce que ceux-ci sont « vraiment

existentiels », méprisant en une phrase les travaux de Gebru⁸⁸. Cette approche est très représentative du dénigrement d'alternatives propre à la stratégie du choc : ceux qui choquent incarnent, paradoxalement, la « bonne parole », la seule parole. Ce qui a eu lieu avant, ce qui aura lieu après, tout cela n'a aucun intérêt : l'objectif de ce discours est tout simplement de nous amener

à abandonner toute forme de vision politique au sujet de l'IA et à remettre les clefs de la maison, celle de ce futur, aux seules véritables personnes capables d'en comprendre les règles ; à savoir Musk, Altman, Peter Thiel et leurs camarades californiens⁸⁹.

Tous adeptes de la pensée long-termiste. En considérant le futur comme aussi important que le présent, si ce n'est plus, il est tout à fait légitime de réorienter une partie des efforts présents au profit d'un plus grand bien-être à venir, en poussant ainsi à une forme d'immobilisme face aux problèmes actuels.

En se focalisant sur le futur, on permet également de ne pas se concentrer sur le présent : le catastrophisme, et son bras idéologique le long-termisme, renforce donc le *statu quo*. Lorsqu'on se demande si, demain, nous serons tous dominés par une IA type Terminator, on ne se pose pas les nécessaires questions d'aujourd'hui : qu'en est-il des conditions de travail des « travailleurs du clic », ces hommes et femmes qui ont aidé à *fine-tuner* ChatGPT ? Qu'en est-il des enjeux environnementaux, gigantesques, auxquels fait face le secteur de la tech ? Qu'en est-il du modèle économique, basé sur un capitalisme de surveillance agressif, qui règne dans la Silicon Valley et qui met en péril nos démocraties et nos valeurs fondamentales ? Ces questions sont rapidement mises sous le tapis et remplacées par une autre : comment aider les grandes entreprises de la Big Tech à perpétuer leur monopole qu'on nous présente comme salvateur de notre humanité ? Mettre en lumière une éventuelle extinction de l'humanité, c'est choisir de laisser dans l'ombre tous les défis cités ci-dessus – maintenant ainsi un *statu quo* dont les géants de la tech sortent largement bénéficiaires. Mais la panique est aisée, et la paralysie devant l'illusion de robots tueurs et cruels est plus simple qu'une profonde refonte de notre modèle social et politique. Or, c'est précisément là que résident les vrais enjeux.

Si l'on veut, et si l'on ose, s'attaquer aux vraies questions que portent en eux ChatGPT et l'IA, il ne suffit pas de penser aux régulations, aux équilibres économiques ou aux bras de fer politiques. Repenser notre éthique technicienne doit être le point de départ de toute réflexion de sorte que l'éthique soit consubstantielle au penser-technique et non pas quelque chose qu'on rajouterait par la suite, comme une sorte de pansement coloré sur une plaie ouverte.

Le philosophe belge Gilbert Hottois fait partie des grandes inspirations d'une

nouvelle philosophie de la technique au XXI^e siècle. Toujours sainement entre un techno-optimisme et une pensée assassine des systèmes technologiques, il nous rappelait que toute éthique de la technique se compose toujours en même temps d'une réflexion sur les paradigmes de notre philosophie politique et d'un questionnement plus profond, plus centré sur nous-mêmes et d'ordre métaphysique. On le voit, l'IA bouleverse à la fois notre vivre-ensemble, notre rapport à autrui et notre propre devenir ontologique, et chaque « Bonjour ChatGPT » en dit autant sur nous-mêmes que sur notre relation au monde qui nous entoure. Et derrière chaque discussion avec un agent conversationnel se joue, en filigranes, notre avenir à tous – non pas dans un sens dramatique mais bien au sens le plus simple, le plus pur et le plus évident : quel est notre projet de société ?

Entre éducation et sobriété technologique

Par la pluralité des valeurs, normes et finalités qu'elle incarne, la technique est fondamentalement politique : par son utilisation dans la vie politique, certes, mais également par les enjeux écologiques qu'elle incarne, par les questionnements sur la nature du travail et du vivre-ensemble qu'elle soulève ou encore par l'impact sur nos rapports à autrui qu'elle engendre. L'évolution technologique est indissociable d'une pensée politique dans son sens le plus large : démocratique, sociologique, philosophique, et chaque interaction avec ChatGPT porte intrinsèquement en elle le poids de notre projet technopolitique. Comme l'écrivait le sociologue Gérard Bronner, dans le rapport qu'il a rédigé, à la demande du Président Macron, pour les candidats à l'élection présidentielle française de 2022 : « Le numérique est un formidable moyen. Reste à savoir quelle société et quelle démocratie nous souhaitons construire dans ce monde numérique en devenir⁹⁰. » Si les pontes de la Silicon Valley veulent qu'on leur confie les clefs de ce projet politique, il est du devoir de chacun, mais aussi et surtout de nos institutions, de résister à cet appel de la Big Tech. Quand on voit comment Sam Altman, le CEO d'OpenAI, ou Elon Musk font évoluer leurs pions sur l'échiquier législatif, on est en droit de douter de leurs bonnes intentions. Le premier a d'abord feint la conformité en disant que l'IA devait être intégrée dans un cadre législatif sauvegardant notre humanité avant de faire marche arrière et de juger l'*AI Act* européen comme bridant l'innovation et mauvais pour le business, tout en investissant des sommes importantes en activité de lobby pour contrer le texte. De son côté, Elon Musk a menacé de retirer X du marché européen si la Commission ne réduisait pas drastiquement

les mesures imposées aux géants de la tech. La législation sur l'IA, ou *AI Act*, est une proposition de texte encadrant l'IA au sein de l'Union européenne (UE) : des *chatbots* à la reconnaissance faciale en passant par les algorithmes de recommandation et les questions de sécurité des données personnelles. Dans la lignée du RGPD (sur la sécurité des données personnelles) puis du *DSA* et du *DMA*^u, le texte sur l'IA aura sans aucun doute un impact important sur les entreprises dont l'IA est le cœur de métier. Par ailleurs, comme ce fut le cas avec le RGPD, il y a fort à parier que l'exemple donné par l'UE sera suivi par d'autres et que l'*AI Act* servira de standard pour de nombreuses réglementations dans le monde, ce qui explique sans doute pourquoi Elon Musk et Altman semblent si pressés de minimiser son impact.

Quand on se penche sur d'autres marchés importants pour la tech, on se rend compte que les approches peuvent largement différer d'une nation à une autre, d'une vision du monde à une autre. La Chine, notamment, utilise son autoritarisme pour, d'une part, imposer sa vision technique et, d'autre part, encadrer strictement les usages. Après avoir restreint l'usage des jeux vidéo aux adolescents à maximum trois heures par semaine, l'empire du Milieu projette de réduire drastiquement l'usage du smartphone pour les enfants et adolescents. S'appuyant sur la force d'un régime autoritaire, les autorités chinoises souhaitent réduire le temps d'écran à deux heures par jour pour les jeunes entre 16 et 18 ans et à moins de huit minutes pour les jeunes de moins de 8 ans⁹¹. Les dangers de la technologie pour les mineurs sont connus, en Occident comme en Orient, et la Chine utilise donc son emprise sur sa population pour la protéger des affres d'une connexion permanente. De plus, elle régule les contenus visibles par les jeunes sur TikTok afin que ceux-ci soient davantage exposés à des vidéos éducatives ou culturelles plutôt qu'à des extraits simplement divertissants, ou carrément abrutissants, comme c'est davantage le cas aux États-Unis ou en Europe. La technologie est ici vue comme un moyen d'asseoir une vision du monde et un projet de société : en limitant, par la force, les libertés numériques de ses citoyens, la Chine forge sa souveraineté technocratique tout en éduquant ses talents futurs dont le projet sera d'amplifier l'aura et l'influence de Pékin sur le monde. On l'a dit, la tech est fondamentalement politique.

En ce qui concerne notre relation avec les robots, le cyberpsychologue Serge Tisseron rappelait que l'État avait trois devoirs d'information⁹². Tout d'abord, que le citoyen sache toujours s'il a affaire à une machine ou à un humain. C'est notamment l'approche qu'a suivie la Californie dans une loi de 2019 qui impose que la nature d'un interlocuteur (humain ou non humain) soit claire, visible et raisonnablement expliquée aux personnes qui interagissent avec lui. Par

exemple, si vous êtes en ligne avec un service client qui aurait délégué une partie des problèmes à régler à une IA, celle-ci devrait se déclarer comme n'étant pas humaine dès le début de la conversation (et, idéalement, réitérer son avertissement, notre tendance à anthropomorphiser reprenant rapidement le dessus). Tisseron indique ensuite que l'État doit veiller à ce que l'utilisateur d'une machine puisse connaître les algorithmes et mécanismes qui la dirigent. C'est une démarche que l'on trouve au sein du *DSA* et qui a pour objectif de forcer les géants de la tech à ouvrir les boîtes noires de leurs algorithmes et de pouvoir expliquer telle ou telle décision technique. Quand on sait que, dans certains États, des décisions de justice simples ont été déléguées à des algorithmes, on est en droit de demander une certaine transparence quant à la prise de telle ou telle décision : difficile d'imaginer, dans le cadre d'un divorce, que l'IA qui a confié la garde des enfants à l'un des deux parents soit incapable d'expliquer sa sentence. Enfin, le cyberpsychologue rappelle que le citoyen doit toujours être protégé des publicités mensongères qui vendent des robots en utilisant un discours trompeur : présenter un *chatbot* comme une petite amie pourrait donc tomber dans cette catégorie. Pour qu'innovation et progrès puissent aller de pair, il est donc crucial que la naissance de nouveautés suive certaines règles éthiques. Pour rappel, ce n'est pas parce que quelque chose est possible, est faisable techniquement, que cette chose est *de facto* souhaitable ou enviable : faisabilité n'implique pas nécessairement progrès, et le discours d'une neutralité de la technique ne trompe plus personne.

La notion même de progrès est profondément polysémique, et ce qui est progrès pour l'un pourrait ne pas être vu comme tel pour d'autres, dans une approche toute conséquentialiste que l'on pourrait rapprocher de l'utilitarisme. Le philosophe Georges Sorel affirmait qu'une société avide de bénéficier du confort et de l'aisance matériels dus aux progrès techniques justifiait cette luxure par la notion de progrès⁹³ : après tout, que peut bien valoir ce concept, très vague, de dignité humaine lorsqu'il est mis face au plaisir ? Si l'on reprend l'exemple des *deadbots*, on peut se demander ce que peut bien peser la dignité d'une personne décédée quand elle est mise en balance avec la tristesse d'un proche – chagrin que l'on pourrait facilement combler avec un programme qui ramènerait à la vie l'individu disparu. Aussitôt macchabée, aussitôt ressuscité, voilà le pouvoir de la tech et l'avènement du plaisir.

Un demi-siècle après Sorel, le philosophe Hans Jonas rédigeait son *Principe responsabilité*, remettant en question la croissance technique et son impact sur notre humanité. Il nous proposait alors un style différent, loin de toute pensée utilitariste ou conséquentialiste, en modernisant l'impératif catégorique kantien

qui nous intimait déjà de ne jamais traiter les personnes uniquement comme des moyens mais toujours en même temps comme des fins – ce qui est précisément ce que ne font pas les entreprises derrière les *deadbots* : en recréant des individus avec leurs traces numériques, ceux-ci ne sont vus que comme des moyens d'un enrichissement. Selon Jonas, les êtres humains devaient agir « de façon que les effets de [leur] action soient compatibles avec la Permanence d'une vie authentiquement humaine sur terre⁹⁴ », un impératif dans lequel on sent déjà les urgences climatiques et anthropologiques qui marqueront au fer rouge le siècle à venir. S'inscrivant dans la lignée, encore balbutiante, des questionnements autour d'une croyance dogmatique dans la croissance – à la suite, notamment, de ce qu'on a appelé le « rapport Meadows » interrogeant les possibilités d'une croissance (économique) infinie dans un monde fini –, Hans Jonas nous force à repenser l'idée et l'idéologie de progrès.

Si l'on ramène ces discussions à nos réflexions contemporaines, le rapport entre technique, innovation et progrès est difficile à démêler, tant il est empreint d'une doctrine néolibérale et techno-solutionniste qui consisterait à dire que tout problème technique pourrait être résolu par davantage de technique. La pensée long-termiste, dont l'esprit essaime constamment dans la Silicon Valley, ne nous dit rien d'autre : la meilleure façon d'aider et de sauver les générations futures est d'investir, dès aujourd'hui, dans des technologies qui pourraient leur venir en aide. Quelle que soit la question, la réponse est indubitablement « encore plus de technologie ». Les gens passent trop de temps sur leur smartphone ? Créons une application qui en bloquerait l'usage et forcerait la désintox numérique ! Les jeunes sont plus attentifs sur TikTok ou sur un jeu vidéo qu'en cours ? Développons des jeux éducatifs qui leur permettraient d'apprendre (et de nous laisser leurs données personnelles) en même temps qu'ils jouent ! Notre société manque de contact humain et se sent seule ? Proposons à nos concitoyens d'échanger avec des *chatbots* et offrons-leur l'opportunité de trouver l'amour et l'amitié dans des programmes informatiques ! La Big Tech aurait-elle, cyniquement, réponse à tout ?

Néanmoins, si l'on relit Jonas, on peut penser que, pour préserver « la Permanence d'une vie authentiquement humaine⁹⁵ », on peut être tenté de vouloir *exnover* en plus de vouloir constamment *innover*. On peut être tenté de se dire que la réponse à certaines questions peut être simplement « moins de tech », ou une meilleure approche de la technologie, cette boîte noire qui nous asservit et à laquelle nous ne comprenons rien.

Une réappropriation technicienne ne peut que passer par davantage d'éducation et d'implication citoyenne dans le rapport humain-machine. Les

appels à voir émerger plus de cours de code à l'école sont nombreux, permettant aux plus jeunes de mieux appréhender les objets qu'ils manipulent quotidiennement, et de nombreux experts rappellent qu'il est essentiel que l'algorithmie ait sa place au sein d'une très large majorité des programmes d'étude supérieure ! Comment se fait-il que des étudiants en droit, probablement un des domaines les plus impactés par l'IA récemment, n'aient pas la moindre idée de comment fonctionne un algorithme, des biais possibles et de comment cette technologie impacte la justice ? On peut également inverser la réflexion : les cours d'éthique et de philosophie sont absents des cursus d'ingénierie ou de science informatique – à l'heure où chaque entreprise prône une « éthique de la conception », il s'agit là d'un véritable non-sens éducatif et d'un manque à combler au plus vite !

La commission Bronner émettait plusieurs recommandations à cet égard dans son rapport en dédiant un chapitre complet à ces questions : « Esprit critique et EMI^v. » On pense notamment aux propositions 27 – « Systématiser la formation à l'esprit critique et à l'EMI en milieu scolaire » – ou 29 – « Développer la formation à l'esprit critique et à l'EMI dans la société civile ». Le lien que chacun doit avoir aux outils technologiques qu'il manipule quotidiennement doit dépasser le simple cadre instrumental : les enjeux sont trop importants, les conséquences trop graves.

Cependant, il ne s'agit pas ici de proposer une approche technophobe qui rejeterait toute technique et verrait dans la technologie une incarnation satanique à la botte du néolibéralisme. Une telle suggestion serait bien trop simpliste et à la limite du ridicule. Quand il faut penser la relation homme-machine, l'appel au philosophe Gilbert Simondon est souvent salutaire puisque ce dernier ne proposait pas de se distancier de la technique mais plutôt de renouveler notre manière de l'appréhender – par une réappropriation et un changement de paradigme :

La condition première d'incorporation des objets techniques à la culture serait que l'homme ne soit ni inférieur ni supérieur aux objets techniques, qu'il puisse les aborder et apprendre à les connaître en entretenant avec eux une relation d'égalité, de réciprocité d'échanges : une relation sociale, en quelque manière⁹⁶.

Et une telle relation ne pourrait avoir lieu sans que l'un comprenne l'autre, c'est-à-dire sans que l'être humain comprenne la machine dans son sens le plus concret : les câbles, les fils, les signaux électriques. Simondon était connu pour apporter des objets techniques à ses étudiants et pour leur proposer, ensemble, de les décortiquer et de les analyser, de les comprendre.

C'est exactement l'approche qu'a suivie le philosophe-mécanicien Matthew

Crawford, lui qui propose de remettre les mains dans le cambouis et de plonger concrètement dans l'univers technique et même mécanique. En développant l'idée d'un « engagement éthique⁹⁷ » avec la machine, Crawford met également en avant la nécessité de pouvoir réparer ses automates et ses véhicules, de comprendre comment ça fonctionne, et de modifier le rapport de domination qui peut exister lorsqu'un ordinateur vous connaît mieux que vous-même. Une telle démarche permettrait alors à tout un chacun de mieux vivre son rapport technique et de prendre de meilleures décisions, pour lui et pour son usage propre. En s'évadant de l'aliénation technologique par une meilleure connaissance des outils, les êtres humains pourraient ainsi réévaluer leur rapport tant à la technique qu'au monde non technique.

Mais les individus sont souvent laissés pour compte, tant qu'ils passent du temps sur les plateformes numériques, tant qu'ils consomment. Comme l'avait développé, à l'époque, le philosophe Michel Foucault, les êtres humains sont traités comme des facteurs de production, un capital humain – appelé aujourd'hui du « *digital labour* » – qu'il faut utiliser de manière optimale, tant du côté des entreprises que du côté des individus, pour maximiser la croissance du revenu (pour les entreprises) ou d'un confort lénifiant (pour les individus). Nous nous considérons nous-mêmes, dans un cauchemar kantien, uniquement comme des moyens et non comme des fins, le tout au nom d'un accroissement de nos ressources, l'une d'entre elles étant le temps de cerveau disponible que nous offrons sans trop y réfléchir aux géants du numérique.

Notre monde est passé d'un idéal du temps perdu et de la flânerie à un idéal de l'activité ininterrompue. La poésie, qui requiert nécessairement cet oubli de soi et des contraintes du monde tangible, s'est inclinée devant notre impératif de vitesse. Grâce aux nouvelles technologies, le monde devait être à nos pieds et offrir un projet politique simple : la victoire du « tout et tout de suite » sur les limites du monde réel.

Dans la digne mouvance de philosophes comme Hans Jonas, ou d'économistes comme Donella et Dennis Meadows, s'installe une réflexion sur la décroissance économique et la sobriété. Au XXI^e siècle, de nouveaux penseurs comme Hartmut Rosa, et de nouveaux économistes comme Timothée Parrique, nous invitent à repenser notre rapport à ce « toujours plus » que nous rabâche sans cesse la doctrine capitaliste. Le propos de cet essai n'est pas d'analyser les tenants et les aboutissants de la décroissance, certains le feraient bien mieux que nous, mais de comprendre comment une nouvelle philosophie politique de la technique pourrait s'en inspirer. Quand le projet décroissant et post-croissant propose de diminuer certaines choses (comme la production, le PIB, etc.), il nous

invite également à en augmenter d'autres (telles que les relations sociales, la résonance, les cafés sur le bord de la plage). Alors que l'histoire de la philosophie politique, de Platon à Hobbes en passant par Machiavel, s'inscrit dans une dichotomie, une opposition, entre dominants et dominés, intérieur (la cité) et extérieur (le monde sauvage), humains ou non-humains, on peut être en droit de vouloir voir naître une nouvelle idéologie pour notre devenir technique. La sobriété numérique pourrait consister « à se limiter à un usage mesuré de ces outils surpuissants et aussi addictifs qu'insoutenables pour la planète ». Rappelant que la technique n'est jamais neutre, il nous faut également rappeler la nécessité

de se réapproprier ses usages afin d'éviter d'en être toujours plus dépendant. La sobriété conduit à délaisser les écrans pour se reconnecter aux autres dans la réalité, se rencontrer, se confronter et faire vivre la démocratie, qui a pour objet de pacifier les conflictualités⁹⁸.

Tout un programme. Cela se manifeste par des petits gestes du quotidien comme ne pas changer de smartphone tous les ans ou des actions plus courageuses telles qu'interdire l'usage des téléphones à table pendant les repas familiaux, proposer une nouvelle politique de la déconnexion en entreprise ou débattre collectivement des rapports que nous voulons entretenir avec le numérique. L'économiste français Timothée Parrique définit la post-croissance comme une

économie stationnaire en relation harmonieuse avec la nature où les décisions sont prises ensemble et où les richesses sont équitablement partagées afin de pouvoir prospérer sans croissance⁹⁹.

Cette définition semble aisément transposable à toute réflexion technopolitique.

Quand on sait qu'en 2021 les citoyens occidentaux ont passé en moyenne cinq heures par jour sur leur smartphone, soit près d'un tiers de leur temps éveillé, que les réseaux sociaux sont un vecteur de polarisation de nos sociétés et de nos démocraties, et que les grands patrons de la Silicon Valley se sont outrageusement enrichis ces dernières années, en grande partie sur le dos de leurs utilisateurs et de « travailleurs du clic » surexploités dans des pays défavorisés, il semble légitime de se pencher sur un projet de décroissance technologique. La définition de Parrique pourrait être reprise telle quelle et l'on imaginerait une société stationnaire en relation harmonieuse avec elle-même (on pense aux repas familiaux dé-digitalisés et aux débats contradictoires que nous pourrions à nouveau avoir avec nos voisins), où les décisions seraient prises ensemble (au sein, par exemple, de conventions citoyennes sur les usages numériques), où les richesses seraient équitablement partagées (ce qui

nécessiterait sans doute de repenser le *business model* des grandes plateformes numériques) afin de pouvoir prospérer sans être perpétuellement le nez sur son téléphone. « Le lien redoutable entre la célérité encouragée par la technique et l'absence de temps pour la réflexion est au centre du problème éthique¹⁰⁰ », nous dit Alexei Grinbaum, le président du CNPEN. Une nouvelle approche technique centrée sur l'humain et les relations ne peut que se fonder sur une temporisation qui nous permettrait de modifier nos réflexes d'utilisateurs : plutôt que de cliquer rapidement, sans avoir même lu, sur l'emoji colère afin de réagir à un post rédigé par un « adversaire » politique, peut-être aurions-nous tout à gagner à prendre une seconde pour contrôler cette vitesse qui nous emporte. Tout projet ne se résout pas avec davantage de rapidité, de fluidité ou de fonctionnalités. « La qualité d'une vie pourrait se mesurer à la lenteur d'un ordinateur¹⁰¹ », forcés de ralentir face à une bécane qui rame, nous profiterions peut-être plus simplement de nouveaux plaisirs.

Aujourd'hui, il semble qu'il soit plus facilement concevable d'inventer une application pour régler un problème plutôt que de traverser la rue, sonner chez son voisin et s'expliquer. Mais que préférons-nous vraiment ? Quel est le projet de société qui nous meut ? Celui où nous téléchargeons la dernière app de la mairie qui rappelle les règles et lois de mitoyenneté et qui délimite scrupuleusement chaque parcelle de terrain, ou bien celui où, avec un gâteau aux pommes, nous débarquons chez le voisin pour régler le problème de cet arbre qui dépasse sur notre jardin ? Naturellement, on le sait, tout n'est pas aussi simple et manichéen, mais la logique demeure, et l'on peut pousser cette logique, absurde, encore plus loin. Si vous souhaitez gagner de l'argent aujourd'hui en lançant une start-up dans la tech, permettez-moi de vous aider en vous donnant une idée : UberCouche. C'est une application pour les jeunes parents qui, lorsque leur nouveau-né a rempli sa couche, peuvent faire un appel à un « changeur » semi-pro qui, moyennant quelques écus, viendra à domicile, potentiellement même avec son propre matériel, ses lingettes et ses crèmes, pour changer les fesses de votre marmot. L'utilitarisme exacerbé de la Silicon Valley en plein dans le réalisme de la vie, et tout le monde en ressort satisfait : un bébé avec le siège propre, des parents avec les mains immaculées et dont les yeux n'ont pas quitté leur smartphone, un jeune homme satisfait de son petit boulot, et un PIB qui en ressort grandi. L'absurde et le cynisme de cet exemple ne peuvent que choquer : faire cela, c'est passer à côté de la vraie vie, avec ses galères, certes, mais également des moments de relation avec ses enfants (comme changer ses couches !). Tout ne se résout pas par plus de tech.

Cette sobriété est cependant difficile à faire éclore. Elle ne peut être, comme

en Chine, imposée par un régime autoritaire – il s’agirait de sauvegarder une liberté fondamentale en en violant d’autres – ni amenée par une technologie encore plus invasive comme les applications de régulation de l’usage technologique. Derrière une volonté d’augmenter le « bien-être numérique » de ses utilisateurs, les concepteurs de ces mécanismes ne font qu’augmenter leur soumission. « Une technique destinée à délivrer l’homme de la machine, le soumet d’autant plus rudement à l’appareil¹⁰² », écrivait déjà Ellul en 1954. Seul l’individu peut s’émanciper de sa propre servitude technologique. Pour que cela puisse se faire, il est nécessaire que l’État s’affranchisse de ses responsabilités en offrant à chacun de ses concitoyens une éducation à l’esprit critique, aux médias et à l’information et qu’il mette en place les processus décisionnels permettant d’une part aux citoyens de prendre part aux débats qui les concernent et, d’autre part, aux richesses générées par une technique émancipatrice d’être redistribuée intelligemment. On retrouve en filigranes les recommandations de la commission Bronner qui s’intègrent, assez logiquement, dans une révision, pour reprendre la dichotomie de Gilbert Hottois, d’une philosophie politique de la technique. Mais tout cela sera vain sans un glissement métaphysique et une réflexion de notre rapport à nous-mêmes et à autrui : l’éthique de la tech ne peut faire l’économie d’une certaine introspection.

Entre visage et temps

La légende raconte que Balzac, après avoir terminé une longue discussion sur l’État, le peuple et le sort du monde, dise à ses interlocuteurs : « Et maintenant, revenons aux choses sérieuses », remettant ses romans, la poésie et la littérature au centre de la conversation. Peut-être que, pendant des années, nous avons justement mis ces élucubrations politiques au centre des discussions alors que l’essentiel, peut-être, était ailleurs. Tout projet de décroissance technologique ne pourra d’ailleurs venir que de cet ailleurs et d’un nouveau projet qui n’aurait pas pour cœur les questions de politique et de souveraineté, celles-ci n’étant qu’une partie de l’équation défendue par Hottois : philosophie politique et métaphysique.

Assez naturellement, c’est vers la métaphysique qu’il faudra se tourner si l’on veut trouver un sens à ce qui nous arrive, si l’on veut redécouvrir notre rapport à la machine, à l’autre, à soi. Depuis la nuit des temps, les penseurs et philosophes ont misé sur la connaissance du monde et sur la science pour interpréter notre place dans ce monde, dans une volonté toujours plus grande d’avoir la mainmise et le contrôle sur ce monde qui, paradoxalement, semble toujours nous échapper

un petit peu. Aristote, déjà, estimait que l'être humain pouvait se rapprocher de la condition divine par la connaissance. Le philosophe grec voyait dans la métaphysique une « philosophie première » sur laquelle pouvait se fonder toute la pensée, une sorte de « science de l'être en tant qu'être », qui permettait de répondre à la question « Qu'est-ce que l'être ? » ou bien, telle que formulée notamment par Leibniz, « Pourquoi y a-t-il quelque chose plutôt que rien ? ». C'est sur la quête de cette réponse que s'est construite une grande partie de la philosophie occidentale, avec en son sein la certitude que l'être devait être intrinsèquement lié à la connaissance et donc, par la suite, à la maîtrise du monde et des éléments.

Avec l'Histoire, cette certitude s'est ancrée plus profondément dans la pensée européenne : d'Ève et Adam sommés de nommer les animaux jusqu'aux représentations médiévales de la nature, on retrouve partout la puissance de l'humain qui n'aurait, au-dessus de lui, que Dieu et sa puissance divine. S'élever au rang du Créateur, voilà le grand projet de notre civilisation ! Et c'est par l'accumulation de savoir et d'expertise que, dans la digne lignée de la pensée aristotélicienne, il nous semble faisable de nous prendre pour des dieux. Il y aurait alors une sorte d'alliance entre le savoir et le faire qui mettrait l'action au centre de tout travail intellectuel : la recherche de connaissance n'aurait pour seul objectif que de permettre à l'homme d'imposer sa marque sur son environnement, faisant de tout ce qui l'entoure un paramètre calculatoire de son expérience. Ainsi, une forêt perdrait son aspect poétique et méditatif pour ne devenir qu'une source de bois permettant de construire des outils pour renforcer une domination des éléments, et une large plaine pourrait être réduite à une simple ville en devenir et aux nombreux échanges qui pourraient y prendre place. Même un voisin, un ami, pourrait perdre son statut d'altérité pour devenir un paramètre dans les calculs de chacun : comment utiliser untel pour parvenir à tel but ? Quand Pouchkine disait qu'on ne peut pas vérifier l'harmonie par l'algèbre, il anticipait cette fuite en avant d'une société pour qui l'algèbre allait devenir maîtresse, laissant l'harmonie délicieuse du monde se déliter dans de complexes algorithmes.

Cette quête d'emprise sur le monde devait permettre à l'homme de dépasser ses limites et ses faiblesses, et on peut voir dans l'aboutissement de la technique une sorte de victoire de l'humanité sur la nature : cette dernière est sommée d'obéir et de se soumettre à la volonté des êtres humains qui sont alors capables d'entreprendre toutes sortes d'activités que l'on pourrait qualifier de contre-nature. Qu'il s'agisse de la conquête des airs et de l'aviation, de la révolution agraire ou simplement de l'électricité, la technique a permis à l'homme de

s'émanciper des règles dictées par la nature : la gravité, le cycle des saisons, l'alternance jour/nuit. Dans cette perspective, la création d'andriodes peut être vue comme une nouvelle étape dans cette mise à pied de notre environnement.

Mettons, le temps de quelques lignes, nos robots et *chatbots* entre parenthèses pour nous plonger dans cette évolution philosophique qui nous a amenés à nous reposer sur cette technique. « Attends une minute, ChatGPT », le temps de comprendre le trajet métaphysique parcouru par l'homme et qui l'a conduit à devoir aller toujours plus haut, toujours plus vite.

Ce sont ceux que l'on a appelé les Modernes, ce mouvement de la philosophie qui émerge de la sortie du Moyen Âge et qui s'étend jusqu'aux Lumières, qui ont davantage affirmé cette proposition. Afin de dominer le monde, l'être humain se fait scientifique et expérimentateur afin d'accroître son emprise sur son environnement. Thomas Hobbes, le philosophe anglais du XVII^e siècle, affirmait que notre esprit n'était que calcul, évaluation et réévaluation. Il pensait qu'à l'état de nature, nous étions tous en guerre les uns contre les autres et que seule la domination, d'un individu ou d'un État, pouvait prévaloir – ce qui lui inspira sa célèbre formule « L'homme est un loup pour l'homme ». Telle était la métaphysique, la nature des choses, hobbesienne : une lutte incessante pour l'hégémonie. À la question « Qu'est-ce que l'être ? », Hobbes répondait « Affirmation de soi », et donc écrasement d'autrui. Le philosophe Baruch Spinoza ne disait pas autre chose lorsqu'il développait son concept du *conatus* – une sorte d'« effort pour être », d'« effort pour persévérer dans son être ». Littéralement, le terme signifie « effort », « tendance à », et exprime chez Spinoza la nature même des individus à se conserver, à tout faire pour continuer d'exister. L'essence de l'homme, comme chez Hobbes, est à trouver dans l'affirmation de soi. On voit bien que c'est la notion de puissance, de pouvoir, qui règne dans les réflexions ontologiques du XVII^e siècle et qui inspirera également les développements et révolutions technologiques à venir : ceux-ci ne doivent aboutir que s'ils permettent à l'être humain d'augmenter sa domination sur son environnement, qu'il s'agisse de la nature ou des nations environnantes.

Les évolutions techniques des XIX^e et XX^e siècles ne sont que le produit de cette pensée du *conatus*. Comme le rappelle le philosophe et physicien Alexei Grinbaum,

les systèmes symboliques s'appuient sur une idée philosophique qui remonte à Thomas Hobbes, selon laquelle la pensée n'est qu'un calcul selon un ensemble de règles. Pour créer une machine qui raisonne comme un être humain, il faudrait donc des règles, de la logique, de la syntaxe, de la sémantique... Cela a conduit à l'intelligence artificielle symbolique, notamment aux systèmes experts¹⁰³

Et notre conception actuelle de la technique et de la technologie trouve ses racines dans cette métaphysique de l'affirmation. Dans ce besoin de s'affirmer, l'un des plus grands adversaires n'est autre que la mort elle-même, cette grande ombre vers laquelle, selon Hobbes, nous serions tous tournés. Et c'est ainsi que notre espèce a commencé à affronter, et parfois à vaincre, les différents obstacles qui se présentaient à elle : le manque, l'inconfort, la maladie, la faim, les distances et, qui sait peut-être un jour, la mort elle-même. Puisque nous ne pouvons anéantir que ce que nous pouvons percevoir et appréhender, nous n'allons considérer comme réel que ce qui peut être mesuré. Avec l'avènement de la technique mondiale, « seul ce qui est calculable d'avance, vaut comme étant¹⁰⁴ », se désespérait Martin Heidegger. En calculant, il est alors aisé de dépasser et de dominer, et c'est dans cette vision du monde que se retrouvent de nombreux entrepreneurs californiens qui ne voient le libre arbitre, les émotions ou la beauté du monde que comme des flux nerveux ou des données qui peuvent alors être manipulés et dépassés. Avec la vision métaphysique actuelle, tout est représenté comme une ressource disponible, maîtrisable, objectivable et calculable dans un projet de dépassement de l'humanité qui viendrait enfin répondre à cette affirmation ontologique qui la traverse : l'être est la permanence de l'être. *Homo deus*, cette créature alambiquée qui devrait naître des laboratoires de la Silicon Valley, n'est rien d'autre que la victoire du *conatus* et de la guerre permanente qui se joue dans le monde.

« Ainsi, la technique ne serait que l'effet extérieur d'une mutation de la pensée animée par une volonté de puissance et de maîtrise universelles¹⁰⁵ », et c'est par le développement exacerbé de cette technique que l'homme voudra accroître sa domination et son affirmation de soi. On retrouve donc là un discours connu, aux relents techno-solutionnistes, qui s'est construit sur une acceptation instrumentale de la raison ayant pour seul but un contrôle total de l'existence, que celle-ci soit politique, sociale ou environnementale. Ces rêves de gloire, de pouvoir, de bonheur absolu, de divinité, d'immortalité ou de simple conquête de l'autre n'ont rien de neuf et semblent donc constitutifs de chacun. Comme le rappelait le narrateur de *La chute*, c'est

comme si mon véritable désir n'était pas d'être la créature la plus intelligente ou la plus généreuse de la terre, mais seulement de battre qui je voudrais, d'être le plus fort enfin, et de la façon la plus élémentaire. La vérité est que tout homme intelligent, vous le savez bien, rêve d'être un gangster et de régner sur la société par la seule violence¹⁰⁶.

Le rêve hobbesien incarné dans une hubris démesurée.

C'est également la prédominance de cet orgueil qui nous permet de mieux comprendre l'exceptionnalisme humain que nous avons abordé au premier

chapitre. Certes, les machines progressent, mais elles ne peuvent jamais devenir notre égal, quels que soient les tests qu'elles aient pu relever : l'humain domine et n'admet aucune concurrence. Cet exceptionnalisme trouve ses racines dans l'histoire de notre philosophie avec la portée métaphysique qu'on lui connaît, et il est impensable que l'affirmation de soi puisse accepter l'existence égale d'un *rob*, d'un esclave. On comprend alors que ces tests n'existent pas vraiment pour évaluer les capacités des machines mais bien plutôt pour lui rappeler son rang et pour nous rassurer sur le nôtre dans une hiérarchie bien ordonnée qu'il ne faudrait pas venir bousculer.

De toutes les innovations, la technique sans doute est celle qui offre à l'être humain la plus grande opportunité d'importer et d'imposer cette « seule violence », pour reprendre les mots de Camus. La technologie dicte ses lois et permet à ses concepteurs de dicter les leurs parmi lesquelles on retrouve l'annihilation de toute forme d'incertitude. Afin que tout reste sous contrôle, la technique permet d'exclure le probable, le peut-être et l'incertain afin que ceux qui sont aux commandes soient en maîtrise totale des éléments passés et à venir avec, comme boussole, un monde parfait, parfaitement régi et sans révolution majeure avec, à sa tête, les puissants de la Silicon Valley dont l'affirmation dans l'être, le *conatus*, aura su s'imposer partout.

Une telle aura, on le voit venir, a quelque chose de divin, et c'est bien là que se profile le projet final : *Homo deus*. L'entreprise transhumaniste, qui permet aux êtres humains de dépasser leurs limites désespérément humaines grâce aux merveilles de la technique, n'admettra que l'accès à la divinité comme aboutissement. Et qu'est-ce donc, finalement, qu'une victoire sur la mort si ce n'est le fait de se hisser aux rangs des dieux ? L'ambition de ce projet transhumaniste est précisément de

*conduire l'humanité à prendre en main sa propre évolution, à se libérer des contraintes que lui impose une constitution biologique imparfaite mais perfectible et, en tout état de cause, transitoire à l'échelle de la technique et de l'évolution*¹⁰⁷.

Le dessein porté par les Ray Kurzweil ou les Nick Bostrom de ce monde n'est rien d'autre qu'une recherche perpétuelle d'une extension des frontières de l'espèce humaine, faisant de chacun d'entre nous un être indéfiniment-en-devenir posthumain. Faut-il se réjouir du fait que Kurzweil voit nos cerveaux s'hybrider et se fondre à des nano-composants électroniques, et ce, dès 2030, pour disposer de nouveaux pouvoirs qui feraient de nous des surhommes ? On peut en douter. Le triomphe sur l'essence même de la vie, à savoir sa finitude et ses limites, représente sans doute une sorte de « *conatus* suprême » dont ne peuvent rêver que quelques privilégiés dans leur tour d'ivoire. Quand certains

pensent la singularité comme la fin de l'évolution biologique et le commencement d'une nouvelle ère, des millions d'individus vivent toujours sous le seuil de pauvreté et nous rappellent que les considérations éthiques, sociales et politiques sont souvent absentes des fantasmes transhumanistes.

Le mouvement long-termiste a des racines similaires et ne peut cacher sa propre affirmation. Cette idéologie est également connue sous un autre nom : l'altruisme efficace, et dans la tête des gourous de cette pensée, l'efficacité ne peut venir que de la résolution d'une fonction d'optimisation et de la mathématisation du monde. Cette quête d'efficience n'est rien d'autre qu'une énième manifestation de cette ontologie de l'affirmation qui va même jusqu'à une sorte de victoire sur le temps : aujourd'hui, j'ai la capacité suprême de faire le bien et d'affecter des êtres humains qui ne sont même pas encore nés et qui naîtront peut-être d'ici mille ans. Engloutis dans leur quête d'immortalité et de règne sur l'espèce, les adeptes du mouvement en oublient souvent que l'aboutissement d'une certaine perfection humaine et la victoire sur le hasard et l'incertitude se font souvent au péril de la vie d'autres dont ils n'imaginent qu'à peine l'existence. Comme le rappelait Weizenbaum, le père du *chatbot* Eliza,

*la science a promis à l'homme le pouvoir. Mais, comme c'est si souvent le cas quand les gens sont séduits par une promesse de pouvoir, le prix [...] est la servitude et l'impuissance. Le pouvoir n'est rien s'il n'est le pouvoir de choisir*¹⁰⁸.

Et cette course effrénée après le pouvoir et la domination a plutôt eu l'effet inverse en enchaînant les hommes à une quête dont ils ne peuvent ontologiquement pas voir le bout.

De nos jours, les êtres humains acceptent d'abandonner leurs valeurs et leur raison en échange du pouvoir. « Mais pourquoi régner si rien n'a de sens ? Pourquoi l'immortalité, si la vie est affreuse ?¹⁰⁹ » Pour *Homo deus* et le projet métaphysique qu'il porte, cette affirmation dans l'être ne peut passer que par la domination de l'autre – et quel meilleur moyen qu'une technologie débridée pour imposer son être et ses idées ?

Cette question force une évidence : *Homo deus* ne sera pas. L'oxymore est trop éclatant, le paradoxe trop grand. Finalement, le XXI^e siècle et l'avènement technologique nous mènent naturellement à un aveu d'échec du grand dessein métaphysique. *Homo deus* ne sera pas. La lutte de tous contre tous, la tyrannie, l'effort de persévérer dans son être, de se poser en niant ce qui nous nie, le *conatus* comme formule essentielle de l'étant chez Spinoza ou Hobbes, vouent l'éthique et la morale au vide et au néant. Aspirant à une liberté totale, les partisans du projet transhumaniste et d'une métaphysique de la puissance et de la

domination n'ont fait que s'enfermer dans une servitude qu'ils ont eux-mêmes bâtie, pris dans des filets qu'ils ont eux-mêmes tissés. Il n'est pas aisé de faire table rase de plus de deux cents ans de métaphysique, et accepter le triste sort de *Homo deus* sera sans doute un des grands défis de notre siècle avec des impacts considérables : technologiques, bien sûr, mais également politiques, environnementaux et philosophiques.

Et pourtant, la vie se résume à bien plus que juste « gagner ». Pour illustrer ce propos, il suffit de se rendre dans n'importe quel parc d'une grande ville : sur certaines tables, ici et là, des individus poussent du bois sur un échiquier qu'ils partagent avec un frère, un ami, un inconnu de passage. Le jeu, on le sait, a été depuis longtemps « résolu » par la machine, et depuis 1997, on sait que plus jamais un être humain ne vaincra un ordinateur aux échecs. Et pourtant, personne ne regarde des ordinateurs s'affronter entre eux pour « acculer un roi en bois dans le coin d'une planche en bois¹¹⁰ », comme l'écrivait Stefan Zweig, alors que des millions de personnes se connectent pour savoir qui viendra troubler le règne du tout-puissant Magnus Carlsen, le génie des échecs. Lorsque nous jouons avec nos grands-parents, l'objectif est certes de l'emporter mais il y a toujours une sorte de méta-objectif qui plane au-dessus de la partie : discuter avec eux, créer du lien, partager un moment. Jouer avec un programme informatique tient plus de l'entraînement, du perfectionnement, d'un appel de notre *conatus* qui voudrait toujours gagner et égaler les dieux, en passant sans aucun doute à côté de l'essentiel d'une partie : discuter des coups joués, et laisser cet échange divaguer vers une conversation sur le jeu, ou sur la poésie, ou la politique, ou encore autre chose qui ne peut advenir que quand on laisse nos esprits vagabonder. Pour être pleinement homme, finalement, il nous faut refuser d'être Dieu. Cet aveu de faiblesse nous permet cependant de nous reposer les bonnes et vraies questions, les seules qui comptent. Si la technique n'est pas la réponse à tout, et si la domination d'autrui afin de m'assurer un confort durable et une abolition de l'incertitude inhérente à la vie n'est pas le projet de mon espèce, que me reste-t-il ? Quand je salue un *chatbot*, quand je dis « Bonjour ChatGPT », que se passe-t-il ?

Initialement, nous aurions pu penser une mise à pied de la technique par le salut, une sorte d'arraisonnement de l'humain qui rappellerait à la technique qu'elle n'existe que pour assouvir son projet de domination. Mais si cette quête perpétuelle du *conatus* est, on l'a vu, vide de sens, que nous reste-t-il ? Et que peut bien signifier ce « Bonjour ChatGPT » ? En réalité, nous l'avons vu plus haut, l'anthropomorphisme n'est qu'une facilité cognitive, un « truc », ou une habitude, qui nous permet de relier les points et d'interagir avec un robot. Dès

lors, finalement, quelle importance revêt cette relation entre l'humain et la machine ?

In fine, ce n'est pas tant l'anthropomorphisme qui doit nous questionner que ce qui vient juste avant : l'interaction avec un robot et le sentiment d'abandon et de solitude qui nous pousse à interagir avec cet être factice. Derrière la salutation à ChatGPT se joue le projet politique que nous souhaitons porter. Ce qui doit nous marquer quand quelqu'un remercie un *chatbot*, ce n'est pas tant le fait que cette personne anthropomorphise une machine que le fait qu'elle souhaite, en priorité, s'adresser à elle. Pourquoi cherchons-nous l'amour avec une IA plutôt qu'avec nos semblables ? Dans ce projet de société qui se joue, quelle est la place que nous voulons donner à autrui et à son altérité ?

Il existe, parmi la myriade de grands esprits qui ont marqué le xx^e siècle, un penseur dont les réflexions peuvent, et doivent, nous accompagner dans notre volonté de façonner de projet de société. Emmanuel Levinas est un philosophe français d'origine lituanienne dont les travaux se sont notamment penchés sur la question éthique et métaphysique de l'autre et de son altérité qu'il définit comme quelque chose qui échappe toujours à toute maîtrise, et donc indirectement à toute domination totale. Sa pensée est difficile à appréhender mais elle fut entre autres déclenchée par un moment d'étonnement que rapporte son ami, le philosophe Jean Greisch :

J'entends encore, comme si c'était aujourd'hui, Levinas me faire part de son étonnement que des penseurs aient pu imaginer que l'étonnement devant le fait que quelque chose soit plutôt que rien, était le point de départ radical de la métaphysique. Puis il ajoutait qu'à ses yeux le fait que sur une terre aussi cruelle que la nôtre, quelque chose comme le miracle de la bonté ait pu apparaître était infiniment plus digne d'étonnement¹¹¹.

Levinas offre ici un nouveau point de départ à la réflexion sur notre nature humaine : la relation à l'autre, à cet autre qui nous échappe et qui peut, dans un acte de pure gratuité, nous tenir la porte, nous sourire, nous tendre la main. Voilà ce qui, selon lui, serait le fondement de la philosophie et de la recherche de sens, bien plus que les idées de domination ou d'affirmation dans l'être. Au-delà de tout dessein démiurgique, c'est bien la position inverse du *conatus*, cet « effort pour être », « effort pour persévérer dans son être » que Levinas cherche à promouvoir. Il essaie de penser, à l'inverse de toute idéologie hobbesienne, le retrait comme essence éthique de l'homme, de penser les relations humaines autrement que comme des relations de guerre. Comme le rappelait si justement Albert Camus, par ailleurs contemporain de Levinas,

si l'homme est le reflet de Dieu, alors il n'importe pas qu'il soit privé de l'amour humain, un jour viendra où il sera rassasié. Mais s'il est créature aveugle, errant dans les ténèbres d'une condition

*cruelle et limitée, il a besoin de ses pareils et de leur amour périssable*¹¹²

Et ce besoin prendrait la forme d'une main tendue et non, dans une visée de domination, d'une possession. Ainsi, on pourrait se réjouir qu'autrui *soit*, tout simplement, plutôt que de rêver de l'*avoir* – c'est entre ces deux auxiliaires que s'affrontent deux visions de notre société.

Si la vie n'est que relation à autrui, que signifie alors cette adresse à ChatGPT plutôt qu'à un voisin, un ami, une sœur, un proche ? Il semblerait que nous réalisions, peut-être difficilement, peut-être inconsciemment, que tout dans la vie ne saurait être réduit à une accumulation de 0 et de 1 et que la suppression de l'incertitude et du doute ne pourrait mener qu'à une profonde diminution de l'expérience de la vie elle-même.

Les travaux de Levinas peuvent être finalement encapsulés dans le titre d'une conférence qu'il donna en 1982, intitulée « Éthique comme philosophie première ». On retrouve dans ce texte la nature même du titre de l'ouvrage de Levinas : l'éthique comme alternative à la métaphysique, comme fondement de la philosophie. Levinas critique les Modernes et l'identification de l'être comme volonté et affirmation de soi. Il voit dans le *conatus* un aboutissement de la métaphysique rendant impensable le souci éthique en l'homme. Dans une pure perspective ontologique, mise en place depuis Aristote, on aboutit à concevoir les êtres comme des forces dans un rapport de guerre et d'opposition, chacune voulant se conserver au détriment des autres, voire dans une optique plus nietzschéenne, chacune voulant dominer et écraser les autres dans une volonté de puissance propre à la pensée du philosophe allemand.

Cependant, derrière l'affirmation de l'être persistant, derrière cette volonté de puissance, derrière cette inconscience de l'existence de l'autre, il y a toujours un moi qui se remet en question, qui hésite, qui ne sait pas, comme habité de scrupule éthique.

Dans son texte, Levinas nous parle de « la merveille du moi revendiqué dans le visage du prochain¹¹³ ». Il n'est plus grande expérience du moi hésitant que l'expérience du visage. Quand on parle d'expérience, il ne s'agit pas d'une expérience au sens commun avec une date précise, un lieu spécifique, mais il s'agit plutôt d'une expérience toujours présente à laquelle on choisit ou non de faire attention. Cette expérience, c'est celle de l'infinie altérité de l'autre, et, en même temps, son infinie vulnérabilité à laquelle nous montrons, et dans laquelle nous voyons, notre propre fragilité. Quand Levinas parle de « la merveille du moi débarrassé de soi et craignant pour autrui¹¹⁴ », il nous montre ce rapport entre deux êtres fragiles qui se disent « Je suis nudité, je suis fragilité » : c'est justement cette exposition qui peut arrêter la main de la violence, qui peut retenir

notre bras, qui peut éteindre cette affirmation de soi millénaire, ce *conatus*. C'est précisément ce que le quatrième singe de la sagesse, celui qui serait l'addition des trois autres et qui serait complètement coupé du monde, ne peut pas expérimenter. Il y a avec ce singe une non-expérience du visage, une impossibilité de l'expérience si on n'est pas capable d'accepter une certaine forme de vulnérabilité et de fragilité.

Ce thème de la reconnaissance, dans l'existence de l'autre, de notre propre vécu résonne pour de nombreux individus. Le salut des Na'vi, les créatures mystiques du film *Avatar* de James Cameron, s'inspire de cet élan que nous pouvons tous ressentir : *I see you*, ou « Je te vois » en français. Ces quelques mots reflètent bien plus que la reconnaissance physique des habitants de la planète Pandora, ils disent : « Je te reconnais en tant qu'être vivant, je te vois avec toute ta dignité, et je reconnais dans ton existence ma propre dignité d'être vivant. » C'est sans doute pour cela que nous sommes tellement touchés par le sort des Na'vi dans l'épopée de Cameron : leur façon de s'adresser les uns aux autres nous touche au plus profond de nous-mêmes.

C'est ce que Levinas entend par existence vitale : certes, il existe une violence intrinsèque à la vie et, pourtant – c'est là même la source du *thaumazein*^w éthique lévinassien –, l'espèce humaine est habitée par un écart avec cette vie violente naturelle. Violence dont elle s'écarte car, dans son hésitation, dans son doute, elle n'y adhère pas complètement. Le singe, lui, ne doute pas, n'hésite pas, et prend donc part inconsciemment, et sans doute involontairement, à cette violence qui demeure dans notre société.

C'est dans cette non-adhésion qu'il faut comprendre la suspension du caractère sacré de notre privilège ontologique et qui ferait de nous des dominateurs naturels et évidents de la nature et d'autrui – suspens dont le colonel Miles Quaritch, dans *Avatar*, est incapable, lui qui s'approprie tout ce qu'il peut sur la planète Pandora. On retrouve alors l'idée de responsabilité, immémoriale, qui nous fait humains. Cette responsabilité existerait depuis une sorte de « toujours », antérieure même à la liberté et c'est cette responsabilité qui vient bousculer cette force logique et ontologique que l'on peut retrouver dans le *conatus* ou la volonté de puissance.

Le fondement de notre existence se trouve dès lors davantage dans la relation à autrui que dans l'écrasement de la vie pour en faire une ressource mobilisable à souhait.

Sortons les robots et la machine des parenthèses dans lesquelles nous les y avons laissés. Comment penser cette éthique d'autrui et de la relation quand les algorithmes nous entourent et que notre rencontre avec l'altérité est limitée à sa

plus simple expression dans une volonté de réduire toute friction, tout contact ? L'expérience même du visage d'autrui semble bien plus difficile à vivre, et il est impensable de percevoir la fragilité et la vulnérabilité de notre prochain lorsque nos interactions se limitent à une visioconférence sur Zoom ou à un *voice* sur WhatsApp. La technologie nous empêche d'entrer pleinement en relation avec cet autre que nous côtoyons cependant tous les jours, et, en même temps, elle nous fait interagir avec des millions d'autres alternatifs et artificiels. Mais l'expérience en est forcément bouleversée par la nature même de notre attention.

Quand nous regardons quelqu'un dans les yeux, quand nous lui disons « Je t'écoute » d'une voix douce et bienveillante, nous lui consacrons notre entière attention dans une relation d'égal à égal qui nous honore. C'est dans cette parole que nous tenons les uns aux autres et que nous sommes pleinement hommes, comme le rappelait Montaigne. Le *chatbot*, lui, ne nous offre qu'une part, infime, de sa considération puisqu'il est en parallèle en train de répondre à des millions de requêtes. Cette asymétrie rend caduque toute véritable discussion et le robot ne ressent pas l'immense responsabilité qui nous emporte quand nous nous adressons à autrui. La machine n'intègre aucun autre devoir que ceux qui ont été programmés et sa tâche principale, unique même, est d'apporter une réponse à l'utilisateur : les individus ne sont donc qu'un paramètre, une sorte d'aboutissement d'une tâche à remplir, et toute réalisation technique, aussi puissante qu'elle soit, n'a aucun devoir envers l'humanité.

Cette différence est cruciale dans tout lien que nous souhaiterions construire avec la machine, et il est capital que notre anthropomorphisme ne nous amène pas à penser qu'un ordinateur puisse ressentir cet enjeu immémorial qu'est l'infinie responsabilité que nous avons envers tout un chacun. Quand on voit comment des millions de personnes s'adressent à une machine, on est en droit de craindre que la technique n'agisse comme une menace contre l'être humain lui-même et contre le réel. Plus récemment, le philosophe allemand Hartmut Rosa nous rappelait notre incohérence ontologique : « Nous cherchons à accumuler une quantité toujours plus grande de biens et d'expériences, alors que nous aspirons à une qualité de relation¹¹⁵. » Mais une relation avec un semblable ne peut qu'être indéterminée et, dès lors, risquée. Derrière un écran, il n'y a pas cette vulnérabilité infinie qui rend l'humain responsable et la relation possible. Derrière un écran, nous nous protégeons de tout risque, mais donc également de tout possible. Il nous faut repenser notre rapport au monde qui nous entoure et accepter une part d'insaisissable, de flottant, d'ambigu.

On nous promet une vie sans ennui, sans errance, sans temps mort grâce à la consommation perpétuelle de nouveaux gadgets. Or, poursuivre une quête

d'immortalité et d'éternelle satisfaction, c'est avant tout passer à côté de sa vie et être distrait des moments qui comptent : un sourire, une larme, un silence, une conversation. Il faut faire renaître chez chacun le goût de l'imprévisible et de l'inattendu et que nous intégrions qu'une perfection lisse ferait de nous non pas des êtres humains, mais justement des robots. Si nous sommes immortels, réparables, notre statut se rapprochera bien davantage de celui de la machine que de celui de l'humain : admettre sa mort, n'est-ce pas justement le propre de la philosophie ? Sommes-nous libres si nous nous émancipons de notre condition humaine, ou bien cela nous aliène-t-il d'autant plus ?

Toute la philosophie de la technique et l'éthique se doivent de rejeter cet anéantissement de notre humanité et ce mirage que nous promet la technologie moderne. La liberté absolue ne peut qu'être illusion dans un monde de relations et il demeure, chez chacun d'entre nous, une sorte d'irréductibilité à l'écrasement technique qui voudrait ne faire de nous que des données dans un plus grand schéma de puissance technologique. Il y a ce quelque chose qui résiste à cette réduction que le néolibéralisme américain voudrait nous imposer, et

il y a un sentiment d'appartenance des humains à l'humanité qui ne se laisse pas entamer par la manière dont les robots nous tentent d'avoir avec eux les mêmes relations qu'avec les humains¹¹⁶.

C'est justement cette appartenance qui nous terrifie à l'idée que nos aînés dans leurs EHPAD soient abandonnés à des robots, que notre progéniture soit laissée pour compte avec pour seul compagnon d'apprentissage une IA, que les célibataires se résignent à commencer une relation amoureuse et intime avec une machine qui ne les considère pas. Face à cette vision tronquée de la liberté, l'utilitarisme et les pontes de la Silicon Valley diraient : « Mais c'est génial, tout le monde est content à la fin ! » Et notre humanité sent qu'elle porte l'infinie responsabilité d'elle-même, dans une nouvelle vision de la liberté qui l'entraîne à entamer de nouveaux chantiers, politiques et éthiques, pour le monde qu'elle veut construire. Le discours sur le caractère absolu et intouchable des libertés individuelles ne tient pas la route : chaque jour, des millions de personnes sont limitées à 130 km/h sur l'autoroute, d'autres n'ont pas le droit de vendre ou d'acheter de la cocaïne, ou bien de porter une arme ou de frapper leurs enfants. Qui viendrait ici, à l'exception d'un néolibéral déséquilibré, nous expliquer qu'il s'agit là d'entrave à sa liberté individuelle ? Non, il s'agit d'un cadre pour faire société et pour porter un projet politique au sens aristotélicien du terme « liberté », probablement son sens le plus sublime et le plus pur.

On dira, et il y aurait sans doute du vrai là-dedans, qu'il vaut mieux que les

vieux soient confiés à des robots plutôt que de les laisser seuls, ou que c'est une aubaine pour Martin de pouvoir être accompagné dans son apprentissage plutôt que de déprimer, seul, face à ses cours. Mais agir de la sorte, c'est éviter les vraies questions qui sont, en réalité, complètement en dehors du domaine technique : c'est cependant tellement plus simple de construire des robots que de repenser notre modèle de Sécurité sociale, la pénurie (et donc la revalorisation) du personnel soignant, la dignité de femmes et d'hommes qui ont déjà tellement donné pour la société. C'est tellement plus facile de développer un agent conversationnel qui donne des leçons d'histoire ou de français que de s'attaquer aux inégalités sociales, à la question des mères célibataires, au décrochage scolaire dans son ensemble, à la dignité des enfants et des parents qui sont le terreau de notre société. Et que dire de l'amour ? Si la technique peut accompagner des hommes et des femmes qui se sentent abandonnés par la société et leur permettre de se sentir moins seuls, n'est-ce pas tout bonnement merveilleux ? Theodore Twombly répondrait peut-être oui à cette question, mais l'irréductibilité en nous ne peut s'empêcher de nous dire qu'il y a là quelque chose qui ne tourne pas rond. En réduisant nos rapports sociaux et amoureux à des conversations avec des *chatbots*, nous nous enfermons dans une spirale isolationniste néfaste pour nous, pour autrui et pour le monde.

C'est donc bien d'une question politique et sociale qu'il s'agit, tout autant que d'une question d'éthique. Afin de répondre à cette petite voix qui nous dit qu'il y a là un projet social à repenser, il nous faut remettre autrui et notre relation à l'altérité et au réel au centre des développements technologiques – l'IA, finalement, n'est pas tellement une question technologique. Comme le rappelait le CNPEN,

il existe des risques émergents pour tous les utilisateurs des agents conversationnels, liés à de nombreux effets potentiels comme le manque d'interaction avec autrui, les biais cognitifs ou la crédulité¹¹⁷.

Et il est donc de la responsabilité de chacun de veiller à ce que ces risques ne se matérialisent pas.

Cela passera avant tout par une vraie réflexion philosophique sur notre acceptation de nos limites, et un rejet du projet trans/posthumaniste qui voudrait toujours faire de nous des *Homo deus*. La divinité ne fait pas partie de la vie, et même si le progrès nous a permis de vivre mieux, une quête toujours plus rapide du « encore mieux », du « toujours mieux », ne peut qu'aboutir à nier notre dignité et notre humanité. Au lieu de vouloir tenir une dizaine de discussions en parallèle, laissons nos téléphones de côté et plongeons-nous corps et âme dans l'unique conversation qui a lieu, ici et maintenant, avec notre interlocuteur qui,

peut-être, a aujourd'hui besoin de nous.

On ne fera pas l'économie d'encadrer et de réguler les géants technologiques. En intégrant une « éthique par conception » dans le développement des produits et outils mis sur le marché, on peut espérer que nos innovations favorisent la mise en relation et la déconnexion. Plutôt que de voir naître un florilège de nouvelles applications addictives, il faudrait que chacune soit conçue pour qu'on y passe le moins de temps possible, impliquant dès lors un changement de paradigme important : la qualité d'une application ne se jugerait pas au temps qu'on passe dessus mais bien à ce qu'elle nous permettrait de nous ancrer dans le réel.

« Bonjour ChatGPT » – que se passe-t-il lorsque nous tapons ces quelques mots sur un clavier ? S'agit-il d'une énième représentation de cette métaphysique de l'affirmation qui nous présenterait la machine comme un esclave à notre merci ? Ou bien s'agit-il d'une sorte de cri désespéré d'un individu qui, écrasé et isolé par cette technique, chercherait du lien partout où il peut le trouver ? La réponse, sans doute, appartient à chacun. Mais une autre question, peut-être plus difficile encore, revient s'imposer à nous : à quoi renonçons-nous quand nous nous adressons à une machine ? À chaque « Bonjour ChatGPT », quelles valeurs mettons-nous entre parenthèses ? L'anthropomorphisme en tant que tel n'est pas une vraie question, ce n'est qu'une facilité cognitive, un « truc », mais c'est bien tout ce qui se trame derrière qui doit nous réveiller et nous amener à reposer notre téléphone, ou notre ordinateur, et à aller prendre un verre avec un ami. Les yeux dans les yeux, c'est là que tout se joue.

[r](#) Pour *Operating System*, ou « système d'exploitation », qui permet à tout ordinateur de fonctionner (les exemples les plus célèbres étant Windows, iOS ou encore Linux).

[s](#) Nous aurions pu espérer voir la question climatique être au cœur de ce débat mais, curieusement (ou non), de la part d'individus qui se baladent en jets privés, il semble que le réchauffement de la planète et les catastrophes naturelles ne soient pas à l'agenda.

[t](#) Le dilemme du prisonnier, énoncé en 1950 par Albert W. Tucker à Princeton, caractérise en théorie des jeux une situation où deux joueurs auraient intérêt à coopérer, mais où, en l'absence de communication entre eux, chacun choisit de trahir l'autre si le jeu n'est joué qu'une fois. La raison en est que, si l'un coopère et que l'autre trahit, le coopératif est fortement pénalisé. Pourtant, si les deux joueurs trahissent, le résultat leur est moins favorable que si les deux avaient choisi de coopérer (Wikipédia).

[u](#) *Digital Services Act* et *Digital Markets Act*, deux textes encadrant la conception, l'exploitation et les usages des plateformes technologiques (notamment les réseaux sociaux).

[v](#) Éducation aux médias et à l'information.

[w](#) Selon Platon, le *thaumazein*, ou étonnement, est la base de toute réflexion philosophique. C'est par cet émerveillement et ce questionnement que naît la réflexion métaphysique et donc, initialement, la quête de savoir : je veux comprendre les causes de cet étonnement et donc mieux appréhender mon rapport au monde.

Conclusion

Au revoir, ChatGPT

Au fil de ces pages, nous avons, sans cesse, navigué entre le réel et la fiction. Dès lors, au crépuscule de notre périple, que voudrions-nous dire à Theodore Twombly, le célibataire technophile du film *Her*, ou à Martha Powell, la veuve éplorée de la série *Black Mirror* ? S'ils étaient en face de nous, de chair et d'os, quel message voudrions-nous leur laisser ?

On l'a vu, blâmer ou moquer des individus parce qu'ils s'adressent à une machine n'a pas de sens puisque nous ne faisons que relier les points et révéler notre humanité qui a profondément besoin et envie d'un rapport social avec le monde qui l'entoure. On l'a dit, si l'anthropomorphisme n'est qu'une facilité cognitive, c'est que les vraies questions résident ailleurs, et les appels au drame et à la fin du monde ne font que cacher cet ailleurs sur lequel nous devons impérativement nous pencher. Au lieu de se moquer des nouvelles May Ashworth, des nouveaux Twombly, la vraie question qui sans cesse nous échappe est celle-ci : quelle étrange société avons-nous créée, dans laquelle les individus s'adressent plus volontiers à des machines qu'à des êtres humains ? Gilbert Hottois, que nous avons déjà croisé, rappelait que la technique ne finissait par prendre le contrôle de certains pans de nos existences que parce que l'homme ne les investissait plus. Avons-nous abandonné le relationnel à la machine ?

La technique, dans sa non-neutralité, est toujours le fruit d'un projet idéologique qu'il nous faut décrypter. Quand nous écrivons « Bonjour ChatGPT », nous nous adressons en réalité à d'immenses corporations récoltant des milliards d'octets de données afin de nourrir un capitalisme de surveillance insatiable. En réalité, nous disons donc « Bonjour OpenAI », « Bonjour X.AI », « Bonjour Google », « Bonjour Meta », et d'autres devraient sûrement suivre. Il

sera alors nécessaire de comprendre et déconstruire les rêves hégémoniques de ces entreprises et de capitaliser sur notre humanité et notre dignité, des desseins qui dépassent largement l'importance croissante des enjeux financiers.

Or, aujourd'hui, c'est un sentiment d'indignité qui persiste dans les initiatives socio-économiques qui se mettent en place. Il s'agit là, pour reprendre les mots de la philosophe et psychanalyste Cynthia Fleury, d'un « sentiment d'atteinte à l'intégrité physique et psychique, comme si l'irréductible en soi était humilié¹¹⁸ » – de nombreux exemples ont tristement alimenté ces pages.

Que dire, en effet, du sentiment d'indignité qui frappe les Hikikomori, ces jeunes Japonais qui se murent dans le silence et l'isolement car ils ne se sentent pas acceptés par le monde qui les entoure, qu'ils se sentent humiliés par cette non-appartenance, et qu'ils se ruent alors dans des relations factices avec des agents conversationnels. Le Japon est à l'avant-garde du projet technologique et des recherches en interaction humain-robot tout en étant, paradoxalement ou non, un des pays au plus fort taux d'isolation, de suicide et de sentiment de solitude. À force de vouloir créer des robots les plus performants, une société pourrait-elle en arriver à en oublier ses sujets ?

Que dire également de la dignité de ces personnes décédées qui sont en quelque sorte « sommées » de revenir à la vie par des algorithmes surpuissants ? Il est certain que les grosses entreprises technologiques feront leur beurre de ces données stockées chez elles et qu'elles revendront à prix d'or à des personnes éplorées par la perte d'un être cher. Qu'en est-il de leur « intégrité physique et psychique » s'ils ne peuvent même pas trouver le repos dans la mort ? On peut se demander ce qui est arrivé à notre société pour que la mise à disposition de chacun soit telle que celle-ci dépasse les frontières naturelles de la vie et du trépas.

Que dire enfin de ces personnes que l'on exclut de la société en les confiant à la garde bienveillante de robots et d'androïdes chargés de les surveiller ? Leur dignité est bafouée, ces individus sont infantilisés, et notre société en ressort meurtrie car incapable d'adresser les vraies questions derrière ces enjeux anthropo-technologiques. Au-delà d'une potentielle régulation de la technique, c'est une question sur la société et le politique que cet essai veut poser : quelle place pour nos aînés, quels moyens pour l'éducation, quels subsides pour les parents célibataires ou pour les personnes dépressives, quel modèle pour notre sécurité sociale et pour tous ces emplois de première ligne, applaudis durant la pandémie et rapidement oubliés depuis ? Ce sont toutes ces questions, et bien d'autres, qui se cachent derrière chaque « Bonjour ChatGPT », car chaque salut à la machine est une négation des individus qui composent notre société et que la

machine tend, irrésistiblement, à remplacer.

On l'a vu plus haut, les rôles se sont inversés : nous nous adressons de plus en plus à nos semblables comme à des robots, dans une affirmation de soi exacerbée, et cela nous amène à nous adresser aux robots comme à des êtres humains, afin de trouver un peu de ce lien social dont nous avons tant besoin. Une demi-seconde de réflexion permet à chacun de réaliser l'absurdité de ce schéma. Utilisons la technique pour ce qu'elle peut apporter à notre confort – grand bien nous fasse ! – et allons trouver ce lien vers nos frères, nos sœurs, nos amis, nos semblables... Tout cela n'aurait-il pas davantage de sens ?

Mais nous nous murons dans l'aboutissement de cette métaphysique tirée de Hobbes et de Spinoza et qui voyait l'affirmation de soi, le confort personnel et l'emprise sur le monde et sur autrui comme le véritable objectif de toute vie. Une guerre de tous contre tous, un climat désastreux pour notre vivre-ensemble. La technique ne serait alors que le couronnement d'un projet de domination du monde dans une recherche toujours plus grande de confort, de rapidité, de facilité, d'aisance.

En fin de compte, on pourrait dire qu'il n'y a là finalement rien de vraiment neuf. La philosophie hobbesienne date du XVII^e siècle, et les questions techniques étaient déjà posées par Jonas, Ellul ou encore Heidegger au siècle précédent. Ces philosophes n'auraient-ils été que des oiseaux de mauvais augure qu'on aurait simplement laissé parler et écrire sans considérer leur pensée ? Pouvons-nous simplement nous replonger dans *Le principe responsabilité* ou *Le bluff technologique* pour répondre aux questions qui nous sont posées et qui sont, incontestablement, *l'enjeu du siècle* ?

En nous penchant sur la proposition faite par Hottois d'étudier à la fois les questions d'ordre métaphysique et celles de philosophie politique, on peut cependant déceler plusieurs pistes de réflexion. Au niveau politique, une véritable combinaison d'un projet de sobriété numérique et d'une réévaluation de l'éducation semble capitale. Il faudrait, avec nos outils technologiques, faire moins et mieux. L'éducation et un meilleur rapport à la machine, dans une visée alignée avec la philosophie de Gilbert Simondon, nourriront d'eux-mêmes la nécessité d'une décroissance technologique : parce que je sais ce que me fait mon smartphone, j'aurai envie de le lâcher un peu plus régulièrement. Mais c'est à nos institutions de prendre ce combat à bras-le-corps et d'assurer le devoir qui est le leur.

Les institutions font aujourd'hui précisément l'inverse de ce qu'elles sont censées faire. Alors même qu'elles devraient assurer et pérenniser les conditions permettant l'épanouissement des individus, elles détricotent scrupuleusement les ressorts intimes des singularités¹¹⁹.

Et ce, dans une optique bien trop technophile, entre lobby intense et aveu de faiblesse. La dignité de ses concitoyens doit être l'enjeu suprême des pouvoirs publics, à eux de respecter cet engagement tacite qu'ils ont envers chacun d'entre nous, et non une indignité qui rendrait les institutions complices du néolibéralisme et du capitalisme de surveillance, à coups de subsides dans la tech et d'une dérégulation angoissante. Les sujets de ces démocraties ne sont pas dupes. Une enquête réalisée aux États-Unis par l'institut Pew Research vient d'ailleurs confirmer ces enjeux : plus de deux tiers des citoyens s'inquiètent du manque de régulation de l'IA et des acteurs qui la développent¹²⁰. Chacun comprend qu'il se passe quelque chose qui nécessite d'être encadré et accompagné car les implications d'une IA débridée seraient catastrophiques. Mais ne nous trompons pas sur les moteurs de cet encadrement : il ne doit pas venir, comme l'espèrent les prophètes long-termistes de la Silicon Valley, des entreprises elles-mêmes mais bien des gouvernements et des organes démocratiques. C'est là leur place. C'est là leur devoir.

D'un point de vue métaphysique, l'approche lévinassienne peut également nous ouvrir les yeux et nous proposer un autre projet que celui de la force, de la puissance et de l'affirmation de soi qui existe depuis des millénaires. C'est ce projet technophile qui a mis le monde à feu et à sang et qui est, entre autres, à l'origine du désastre écologique dans lequel nous nous trouvons. Une nouvelle approche, centrée sur la relation et notre lien avec autrui, vaut peut-être la peine qu'on y jette un œil ? C'est dans cette expérience du visage d'autrui, dans cette expérience de sa fragilité et de sa vulnérabilité, que j'accepte à mon tour ma fragilité et ma vulnérabilité et que, dès lors, j'embrasse enfin pleinement mon humanité. C'est l'inversion du quatrième singe : je vois tout, j'écoute tout, je parle avec chacun, je m'ouvre pleinement au monde. Levinas nous invitait à voir le miracle de la bonté comme point de départ de tout raisonnement ontologique – il s'agit là d'un projet bien plus attrayant qu'une volonté de puissance dominatrice.

À la sortie de ChatGPT, Yann Le Cun avait décrété que l'outil, bien qu'impressionnant et bien vendu avec une belle campagne de marketing, n'avait cependant rien de révolutionnaire et que de nombreuses entreprises travaillaient à des modèles similaires depuis plusieurs années. Le scientifique ne s'imaginait peut-être pas à quel point il avait raison : ChatGPT nous amène à nous poser des questions qui existent depuis des lustres mais auxquelles nous n'avons pas été capables d'apporter des réponses dignes de l'humanité de chacun – par manque de courage, de discernement ou d'ambition. Les projets long-termistes, néolibéraux et utilitaristes semblent passer à côté d'une dimension essentielle de

ce qui fait que nous sommes qui nous sommes : notre authenticité et notre dignité. L'énormité des forces de la technique, pour reprendre les mots de Jonas, nous force à repenser notre politique et notre rapport à ce qui nous entoure.

Si l'on ne croit à rien, si rien n'a de sens et si nous ne pouvons affirmer aucune valeur, tout est possible et rien n'a d'importance. Point de pour ni de contre, l'assassin n'a ni tort ni raison¹²¹.

Ces mots sont d'Albert Camus et ils nous rappellent, un peu durement, la nécessité de donner du sens au monde auquel nous appartenons. C'est cette question que ChatGPT nous pose, indirectement, lorsque nous nous adressons à lui : cette énormité technique a définitivement un impact sur notre société, et il est temps de le redessiner et de le repenser. Si « tout est possible », c'est généralement la loi du plus fort qui l'emporte, dans cette guerre totale qui annihile toute expérience humaine. Imaginer l'autre comme un adversaire, le penser remplaçable par la technique, lui préférer une machine disponible et obéissante, tout cela n'a aucun sens et ne peut être la base d'un projet qui nous donnerait foi dans ce monde. Il nous faut des pour et des contre, des initiatives diverses et variées qui façonneraient le monde de demain qui serait à la fois connecté par une technologie bien présente (ne tombons pas dans une technophobie absurde à contresens de l'Histoire) mais également connecté par le lien qui unit les femmes et les hommes mus par une idée plus grande du vivre-ensemble. Ce sont précisément là les pans de nos existences qu'il nous faut réinvestir.

On peut alors conclure en disant simplement « Au revoir ChatGPT », non pas par aversion à la technique ou par un sentiment haineux de la Silicon Valley et des entreprises de la tech, mais bien par une sorte d'« humanophilie », une philanthropie tournée vers notre société qui serait porteuse d'un salut, d'un « Bonjour », vers chacun. Il ne s'agit pas d'un adieu au *chatbot* d'OpenAI, certainement pas – il est, on l'a dit, bien trop utile et intéressant –, mais il s'agit d'un « Au revoir », d'un « À tout à l'heure », par volonté de nous plonger dans des matières qui ne nécessitent pas un écran et une connexion Wi-Fi mais une table, deux cafés et une paire de chaises – remettant simplement un peu d'ordre dans le monde et traitant les êtres humains comme des êtres humains, les robots comme des robots.

Quand on dit au revoir à ChatGPT, on met de côté cette volonté inhumaine de surhumanité, on abandonne nos rêves d'hégémonie ou de divinité pour embrasser une vie centrée sur l'humain dans son sens le plus fragile qui soit. Nous embrassons alors une faillibilité qui est la base de notre ontologie. Nous embrassons cet égarement consubstantiel de notre humanité qui nous rappelle

qui nous sommes, ce que nous sommes. Nous embrassons également ce visage de l'autre, ce visage par lequel nous l'acceptons dans toute son altérité, dans toute sa vulnérabilité, dans toute son humanité. C'est par cette expérience que nous disons à l'autre qu'il existe à nos yeux, que nous le reconnaissons et que nous n'attendons qu'une seule chose de lui : qu'il nous reconnaisse à son tour, nous permettant d'échapper à cet enfer sartrien qui nous accable depuis si longtemps et qui nous fait fuir toute solitude et toute mise en danger. C'est avec cet autre que nous nous révélons pleinement. Bonjour à lui.

Remerciements

Si je ne devais remercier qu'une seule personne, ce serait sans doute toi, Clémence, pour tout ce que tu fais pour notre famille, pour notre couple, pour moi, pour nous. Chaque journée à écrire est un plaisir qui se trouve sublimé par l'idée d'échanger avec toi sur les problématiques qui m'occupent, ce qui me force à penser, ce qui m'amène à poser ces mots. La science-fiction et la technique peuvent progresser autant qu'elles le veulent, je ne pense pas qu'une IA t'arrive un jour à la cheville. Je suis profondément heureux, et chanceux, de t'avoir épousée et ce livre n'aurait sûrement pas abouti sans ton soutien, tes sourires, et les projets merveilleux que nous construisons ensemble. Voilà le vrai moteur de toute réflexion et de tout écrit, et je t'en suis infiniment reconnaissant.

Ma reconnaissance va également à mon préfacier, Luc de Brabandere, qui, depuis bien plus longtemps qu'il ne le soupçonne, est pour moi un modèle et un mentor. Il est la preuve qu'il est toujours possible d'allier pensée créative, pensée logique et pensée critique et j'espère être à la hauteur de l'honneur qu'il me fait.

Je remercie également les personnes qui, par nos discussions ou par leurs conseils, ont fait avancer mes réflexions et donc, d'une façon ou d'une autre, cet ouvrage. Luc de Brabandere, Simon Brunfaut, François Candelon, Bruno Colmant, Laurence Devillers, Laurent Hublet, Marek Hudon, Gaspard Koenig, Asma Mhalla, Giada Pistili, Maxime Samain, Serge Tisseron, Emmanuel Tourpe, Stanislas van Wassenhove et toute l'équipe du podcast Trench Tech : Mick, Cyrille et Thibaut, toujours fidèlement accompagnés de Fabienne, Laurent, Emmanuel, Gerald et Thierry – merci à vous toutes et tous.

Ma reconnaissance va également à Thibault Léonard, pour sa confiance et pour nos échanges, ainsi qu'à Chloé Van Rossom et Cindy Gravier pour leur accompagnement dans l'édition de cet ouvrage. Je suis fier d'être un auteur Mardaga et d'appartenir à une grande maison d'édition de notre petit pays !

J'ai une attention toute particulière pour Romane Goldmuntz, dont la patience et la pédagogie m'ont permis de naviguer dans les antres des LLMs et de ChatGPT. N'étant ni informaticien, ni ingénieur, travailler avec quelqu'un d'aussi talentueux qu'elle fut une plus-value incontestable pour le deuxième chapitre de ce livre.

À toutes celles et ceux qui me supportent et dont l'amitié est un soutien indéfectible, je vous remercie tout particulièrement. C'est pénible d'avoir un ami qui philosophe, on se doit de l'écouter poliment et d'opiner du chef de temps en temps, mais ces moments comptent tellement pour moi. Sandrine, Thomas, Estelle, Ariane, Alexia, merci à vous.

Enfin, je me dois de terminer en remerciant mes petits « Simone de Bavoire » et « Saint-Thomas t'Aquin », Romane et Octave, vous qui apportez, par vos sourires et votre regard sur le monde, une touche de poésie et de sens nouveau à un ouvrage qui, finalement, vous concerne peut-être encore davantage que moi. Vous êtes in-artificiables et il n'est de plus belle expérience du réel que de vous voir grandir. Merci.

Notes

- 1 CAMUS (A.), *Caligula*, Paris, Gallimard, 1944, acte I, scène IV.
- 2 DEVILLERS (L.), Les enjeux éthiques des robots numériques, *Diplomées*, n° 268-269, 2019, p. 239-251. En ligne : https://femenrev.persee.fr/doc/femdi_1965-0566_2019_num_268_1_10189
- 3 LOVENS (P.-F.), Sans ces conversations avec le chatbot Eliza, mon mari serait toujours là, *La Libre Belgique*, 28 mars 2023. En ligne : <https://www.lalibre.be/belgique/societe/2023/03/28/sans-ces-conversations-avec-le-chatbot-eliza-mon-mari-serait-toujours-la-LVSLWPC5WRDX7J2RCHNWPDST24/>
- 4 JONAS (J.), *Le principe responsabilité*, Paris, Flammarion, 1979, p. 31.
- 5 TISSERON (S.), Cyber-psychologie : Notre identité en question, *Trench Tech*, saison 2, épisode 8, 11 avril 2023. En ligne : <https://trench-tech.fr/podcast/episode/serge-tisseron-psyche-identite-comment-le-numerique-nous-change/>
- 6 BENASAYAG (M.), *Le mythe de l'individu*, Paris, La Découverte, 1998, p. 54.
- 7 CAMUS (A.), *Caligula*, *op. cit.*, acte III, scène II.
- 8 ZWEIG (S.), *Le joueur d'échecs* (Trad. D. Meur), Paris, Flammarion, 2013, p. 51.
- 9 *Ibid.*, p. 50.
- 10 Le "Portrait d'Edmond de Belamy", un tournant de l'art ?, *La Libre*, 29 octobre 2018. En ligne : <https://www.lalibre.be/culture/arts/2018/10/29/le-portrait-dedmond-de-belamy-un-tournant-de-lart-DWIV5CPUB5GX7EEXRVXX46ILOY/>
- 11 ILL (M.), Un tableau conçu par une intelligence artificielle gagne un prix et crée la polémique, *Les Inrockuptibles*, 6 septembre 2022. En ligne : <https://www.lesinrocks.com/art/un-tableau-concu-par-une-intelligence-artificielle-gagne-un-prix-et-cree-la-polemique-495797-06-09-2022/>
- 12 CUTHBERTSON (A.), Company That Made an AI Its Chief Executive Sees Stocks Climb, *The Independent*, 16 mars 2023. En ligne : <https://www.independent.co.uk/tech/ai-ceo-artificial-intelligence-b2302091.html>
- 13 PARDES (A.), The Emotional Chatbots Are Here to Probe Our Feelings, *Wired*, 31 janvier 2018. En ligne : <https://www.wired.com/story/replika-open-source/>
- 14 LOH (M.), China Is Using AI to Raise the dead, And Give People One Last Chance to Say Goodbye,

- Business Insider*, 2023. En ligne : <https://www.businessinsider.com/ai-make-money-china-grieving-raise-dead-griefbot-2023-5?r=US&IR=T>
- 15 BECKER (J.), Concevoir des machines anthropomorphes. Ethnographie des pratiques de conception en robotique sociale, *Réseaux*, vol. 220-221, n° 2-3, 2020, p. 223-251. En ligne : <https://www.cairn.info/revue-reseaux-2020-2-page-223.htm>
 - 16 CHAMINADE (T.), ZECCO (M.), BLAKEMORE (S.-J.), TAKANISHI (A.), FRITH (C. D.), MICERA (S.)... et GALLESSE (V.), Brain Response to a Humanoid Robot in Areas Implicated in the Perception of Human Emotional Gestures, *PLOS ONE* vol. 5, n° 7, e11577, 21 juillet 2010. En ligne : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011577-abstract0>
 - 17 SULEYMAN (M.), Mustafa Suleyman : My New Turing Test Would See If AI Can Make \$1 Million, *MIT Technology Review*, 14 juillet 2023. En ligne : <https://www.technologyreview.com/2023/07/14/1076296/mustafa-suleyman-my-new-turing-test-would-see-if-ai-can-make-1-million/>
 - 18 FLORIDI (L.) et CHIRIATTI (M.), GPT-3 : Its Nature, Scope, Limits, and Consequences, *Minds and Machines*, n° 30, 1^{er} novembre 2020, p. 681-694. En ligne : <https://link.springer.com/article/10.1007/s11023-020-09548-1>
 - 19 BAERTSCHI (B.), « Vie artificielle », in HOTTOIS (G.), MISSA (J.-N.) et PERBAL (L.) (dir.), *Encyclopédie du transhumanisme et du posthumanisme*, Paris, Vrin, 2015, p. 186.
 - 20 ZE FRANK, Are You Human ?, *TED*, mars 2014. En ligne : <https://www.ted.com/talks/ze-frank-are-you-human?autoplay=true&muted=true&language=rm>
 - 21 ETZIONI (O.), cité in KOENIG (G.), *La fin de l'individu. Voyage d'un philosophe au pays de l'intelligence artificielle*, Paris, L'Observatoire, 2019, p. 135.
 - 22 LEMOINE (B.), cité in TIKU (N.), The Google Engineer Who Thinks the Company's AI Has Come to Life, *The Washington Post*, 11 juin 2022. En ligne : <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lambda-blake-lemoine/>
 - 23 LEMOINE (B.), cité in TIKU (N.), *op. cit.*
 - 24 DEVILLERS (L.), GRINBAUM (A.), ADDA (G.), CHATILA (R.), MARTIN (C.), VILLATA (S.)... et ZOLYNSKI (C.), Avis n° 3 – Agents conversationnels : enjeux d'éthique, *Comité national pilote d'éthique du numérique collectif*, rapport de recherche, 15 septembre 2021, p. 4. En ligne : [https://www.ccne-ethique.fr/sites/default/files/2022-02/Avis_n°3_agents_conversationnels_enjeux_d %27éthique.pdf](https://www.ccne-ethique.fr/sites/default/files/2022-02/Avis_n%273_agents_conversationnels_enjeux_d%27éthique.pdf)
 - 25 GEBRU (T.), BENDER (E. M.), MCMILLAN-MAJOR (A.) et SHMITCHELL (S.), On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ?, *Conference on Fairness, Accountability, and Transparency (FAccT'21)*, 3-10 mars 2021, p. 610-623. En ligne : <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>
 - 26 HEIDEGGER (M.), *Langue de tradition et langue technique*, Bruxelles, Belgique, Lebeer-Hossmann, 1990, p. 38-39 (« Philosophiques »).
 - 27 BERNS (T.) et REIGELUTH (T.), *Éthique de la communication et de l'information. Une initiation philosophique en contexte technologique avancé*, Bruxelles, Éditions de l'Université de Bruxelles, 2021, p. 58.
 - 28 SIMON & GARFUNKEL, « The Sound of Silence », *Wednesday Morning 3 A.M.*, Columbia, 1966.
 - 29 GOLDMUNTZ (R.), correspondance personnelle avec l'auteur, 2022.
 - 30 BRABANDERE (L. de), correspondance personnelle avec l'auteur, 2022.
 - 31 GRINBAUM (A.), *Parole de machines. Dialoguer avec une IA*, Paris, humenSciences, 2023, p. 92.
 - 32 BROWN (T. B.), MANN (B.), RYDER (N.), SUBBIAH (M.), KAPLAN (J.), DHARIWAL (P.)... et AMODEI (D.), Language Models Are Few-Shot Learners, *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*, décembre 2020, article n° 159, p. 1877-1901.

- 33 HOLUBOWICZ (G.), Technoféminisme et inégalités – Mathilde Saliou, *IMAGINAIRES*, 2 octobre 2023. En ligne : <https://shows.acast.com/synth-media/episodes/technofe-minisme-et-inegalites-mathilde-saliou>
- 34 GEBRU (T.), BENDER (E. M.), MCMILLAN-MAJOR (A.) et SHMITCHELL (S.), On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ?, *op. cit.*
- 35 ROUVROY (A.) et BERNS (T.), Gouvernamentalité algorithmique et perspectives d’émancipation. Le disparate comme condition d’individuation par la relation ?, *Réseaux*, n° 177, 2013, p. 163-196. En ligne : <https://www.cairn.info/revue-reseaux-2013-1-page-163.htm>
- 36 NICOLAS (A.), ChatGPT est-il une personne ?, *Philosophie Magazine*, 17 mars 2023. En ligne : <https://www.philomag.com/articles/chatgpt-est-il-une-personne>
- 37 SIMONDON (G.) (1958), *Du mode d’existence des objets techniques*, Paris, Aubier, 2012, p. 11.
- 38 DEVILLERS (L.), Les dimensions affectives et sociales dans les interactions humain-robot, *Interfaces numériques*, vol. 2, n° 1, 2013, p. 105-118. En ligne : <https://www.unilim.fr/interfaces-numeriques/1760>
- 39 GUILLOT (A.), Effet Eliza. Vous aussi, vous parlez à votre cafetière ?, *Philonomist*, 31 mars 2023. En ligne : <https://www.philonomist.com/fr/kezako/effet-eliza>
- 40 PISTILLI (G.), cité in BORDOLOI (P.), The Perils of Artificial Romance, *Analytics India Magazine*, 8 juillet 2023. En ligne : <https://analyticsindiamag.com/the-perils-of-artificial-romance/>
- 41 MONTAIGNE (G. de), *Essais*, 1595, liv. I, chap. IX.
- 42 GRINBAUM (G.), *Parole de machines. Dialoguer avec une IA*, *op. cit.*, p. 13.
- 43 DROIT (R.-P.) et ATLAN (M.), Il faut réapprendre à parler l’humain, *Le 1 hebdo*, n° 436, 1^{er} mars 2023. En ligne : <https://le1hebdo.fr/journal/je-te-parle-moi-non-plus/436/article/il-faut-rapprendre-parler-l-humain-5801.html>
- 44 AIRENTI (G.), Aux origines de l’anthropomorphisme, *Gradhiva*, n° 15, 2012. En ligne : <https://journals.openedition.org/gradhiva/2314>
- 45 *Ibid.*
- 46 SPATOLA (N.), L’interaction Homme-Robot, de l’anthropomorphisme à l’humanisation, *L’Année psychologique*, vol. 119, n° 4, 2019, p. 515-563. En ligne : <https://www.cairn.info/revue-l-annee-psychologique-2019-4-page-515.html>
- 47 CURIALE (T.), ACQUATELLA (F.), GROS (L.), COSQUER (M.), TISSERON (S.), L’anthropomorphisme, enjeu de performance pour les chatbots, *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, vol. 28, n° 72, 2022, p. 101-123. En ligne : <https://www.cairn.info/revue-internationale-de-psychosociologie-de-gestion-des-comportements-organisationnels-2022-72-page-101.htm>
- 48 DEVILLERS (L.), *Les robots « émotionnels ». Santé, surveillance, sexualité... : et l’éthique dans tout ça ?*, Paris, L’Observatoire, 2020, p. 149.
- 49 SHELLEY (M. W.) (1818), *Frankenstein* (Trad. J. Ceurvost), Paris, Marabout, 2009, p. 210 (nous soulignons).
- 50 « Be Right Back », *Black Mirror*, épisode 2 saison 1, 2013.
- 51 DEVILLERS (L.), *Les robots « émotionnels »*, *op. cit.*, p. 51.
- 52 LE CUN (Y.), *LinkedIn*, 2023. En ligne : https://www.linkedin.com/posts/yann-lecun_before-we-can-get-to-god-like-ai-well-activity-7052654636524068864-9zRH
- 53 DEVILLERS (L.), *Les enjeux éthiques des robots numériques*, *op. cit.*
- 54 JONZE (S.), *Her*, 2013.
- 55 BAYLE (N.), Lil Miquela, une influenceuse virtuelle qui vaut de l’or, *Le Monde*, 6 octobre 2019. En ligne : <https://www.lemonde.fr/economie/article/2019/10/06/lil-miquela-une-influenceuse-virtuelle->

qui-vaut-de-l-or_6014433_3234.html

- 56 CANDELON (F.), KRAYER (L.), RAJENDRAN (S.) et ZULUAGA MARTINEZ (D.), How People Can Create—and Destroy— Value With Generative AI, *Boston Consulting Group*, 21 septembre 2023. En ligne : <https://www.bcg.com/publications/2023/how-people-create-and-destroy-value-with-gen-ai>
- 57 TISSERON (S.), Le robot : *ceci n'est pas un humain !*, *NECTART*, vol. 5, n° 2, 2017, p. 117-125. En ligne : <https://www.cairn.info/revue-nectart-2017-2-page-117.htm>
- 58 Voir notamment POWERS (A.), KRAMER (A. D. I.), LIM (S.), KUO (J.), LEE (S.-L.) et KIESLER (S.), Eliciting Information from People with a Gendered Humanoid Robot, *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, Nashville, septembre 2005. En ligne : https://www.cs.cmu.edu/~kiesler/publications/2005pdfs/2005_elicit-info-people-gen-hum-robot.pdf; MCDONNELL (M.) et BAXTER (D.), Chatbots and Gender Stereotyping, *Interacting with Computers*, vol. 31, n° 2, mars 2019, p. 116-121. En ligne : <https://academic.oup.com/iwc/article-abstract/31/2/116/5448907?redirectedFrom=fulltext>
- 59 TISSERON (S.), correspondance personnelle avec l'auteur, 2023.
- 60 TISSERON (S.), De l'animal numérique au robot de compagnie : quel avenir pour l'intersubjectivité ?, *Revue française de psychanalyse*, vol. 75, n° 1, 2011, p. 149-159. En ligne : <https://www.cairn.info/revue-francaise-de-psychanalyse-2011-1-page-149.htm>
- 61 BECKER (J.), *op. cit.* Voir aussi KAHN (P. H.), ISHIGURO (H.), FRIEDMAN (B.), KANDA (T.), FREIER (N.), SEVERSON (R.) et WOLK (J.), What Is a Human ? Toward Psychological Benchmarks in the Field of Human-Robot Interaction, *Interaction Studies*, vol. 8, n° 3, 2007.
- 62 LELLOUCHE TORDJMAN (K.), BEAREZ (J.-B.), LAFOLIE (A.) et PHILIPPON (A.), The Declining Relevance of Humans in Delivering What Customers Want, *BCG Henderson Institute*, 22 juin 2022. En ligne : <https://bcghendersoninstitute.com/the-declining-relevance-of-humans-in-delivering-what-customers-want/>
- 63 WEIL (S.), *La condition ouvrière*, Paris, Gallimard, 1951, p. 248.
- 64 GRINBAUM (A.), *Parole de machines. Dialoguer avec une IA, op. cit.*, p. 74.
- 65 ALCARAZ (M.), Google chute en Bourse après une erreur de son nouveau robot d'IA, *Les Échos*, 9 février 2023. En ligne : <https://www.lesechos.fr/tech-medias/hightech/google-chute-en-bourse-apres-une-erreur-de-son-nouveau-robot-dia-1905069>
- 66 MURPHY (A.) et CRIDDLE (C.), Meta Prepares Chatbots with Personas to Try to Retain Users, *Financial Times*, 1^{er} août 2023. En ligne : https://www.ft.com/content/fa_76c8ce-cdfd-458c-baec-73dceb2d2ad5
- 67 CANDELON (F.), GUPTA (A.), KRAYER (L.) et ZHUKOV (L.), How to Unlock Hidden Value from Generative AI, *BCG Henderson Institute*, 9 février 2023. En ligne : <https://bcghendersoninstitute.com/how-to-unlock-hidden-value-from-generative-ai/>
- 68 BRONNER (G.) (dir.), *Les Lumières à l'ère numérique*, Paris, Presses universitaires de France, 2022, p. 16.
- 69 RODHAM CLINTON (H.), The Weaponization of Loneliness, *The Atlantic*, 7 août 2023. En ligne : <https://www.theatlantic.com/ideas/archive/2023/08/hillary-clinton-essay-loneliness-epidemic/674921/>
- 70 SARTRE (J.-P.) (1944), *Huis clos*, Paris, Gallimard, 1947, p. 93.
- 71 CAMUS (A.), *Caligula, op. cit.*, acte I, scène IV.
- 72 *Ibid.*
- 73 BOBIN (C.) (1994), *L'épuisement*, Paris, Folio, 2015, p. 26.
- 74 NASS (C.) et REEVES (R.), *The Media Equation : How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge, Cambridge University Press, 1996.
- 75 TURKLE (S.), *Les yeux dans les yeux, op. cit.*, p. 82.

- 76 TURKLE (S.), citée in JAMES (C.), *The Other A.I. : Artificial Intimacy With Your Chatbot Friend*, *The Wall Street Journal*, 6 août 2023. En ligne : <https://www.wsj.com/articles/when-you-and-ai-become-bffs-ecbcd1e>
- 77 FAGONE (J.), *The Jessica Simulation : Love and Loss in the Age of A.I.*, *San Francisco Chronicle*, 23 juillet 2021. En ligne : <https://www.sfchronicle.com/projects/2021/jessica-simulation-artificial-intelligence/>
- 78 TURKLE (S.), *Alone Together : Why We Expect More from Technology and Less from Each Other*, New York, Basic Books, 2011.
- 79 KONRATH (S. h.), « The Empathy Paradox : Increasing Disconnection in the Age of Increasing Connection », in LUPPICINI (R.) (dir.), *Handbook of Research on Technoself : Identity in a Technological Society*, Hershey, 2013.
- 80 GARNIER (P.), *La démence du percolateur. Courtes rêveries sur les machines et les émotions*, Paris, Premier Parallèle, 2023, p. 156.
- 81 BENTHAM (J.) (1817), *La table des ressorts de l'action*, Paris, L'Unebêvue, 2008, p. 113.
- 82 CAMUS (A.), *Caligula*, op. cit., acte I, scène IV.
- 83 ELLUL (J.), *Le bluff technologique*, Paris, Hachette, 2012.
- 84 TURKLE (S.), *Les yeux dans les yeux*, op. cit., p. 81.
- 85 Pause Giant AI Experiments : An Open Letter, *Future of Life Institute*, 22 mars 2023. En ligne : <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- 86 KLEIN (N.), *La stratégie du choc. La montée d'un capitalisme du désastre*, Arles, Actes Sud, 2008.
- 87 HOLUBOWICZ (G.), IA : la stratégie du choc, *Trench Tech*, 24 mai 2023. En ligne : <https://trench-tech.fr/gerald-holubowicz-on-refait-la-tech-la-chronique-socio-historique-de-trench-tech/>
- 88 O'NEIL (L.), These Women Tried to Warn Us About AI, *RollingStone*, 12 août 2023. En ligne : <https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/>
- 89 HOLUBOWICZ (G.), IA : la stratégie du choc, op. cit.
- 90 BRONNER (G.) (dir.), *Les Lumières à l'ère numérique*, op. cit., p. 159.
- 91 SAMAIN (M.), Les jeunes Chinois bientôt privés de smartphones ?, *L'Écho*, 3 août 2023. En ligne : <https://www.lecho.be/entreprises/technologie/les-jeunes-chinois-bientot-prives-de-smartphones/10484557.html>
- 92 TISSERON (S.), « Le robot : ceci n'est pas un humain ! », op. cit.
- 93 SOREL (G.), cité in CAMUS (A.), *L'homme révolté*, Paris, Gallimard, 1951, p. 246.
- 94 JONAS (H.), op. cit., p. 40.
- 95 *Ibid.*
- 96 SIMONDON (G.), *Du mode d'existence des objets techniques*, op. cit., p. 88.
- 97 CRAWFORD (M.), *Éloge du carburateur*, Paris, La Découverte, 2010.
- 98 LIEGEY (V.) et BROKMAN (I.), *La sobriété (la vraie), mode d'emploi*, Paris, Tana Éditions, 2023, p. 83.
- 99 PARRIQUE (T.), *Ralentir ou périr. L'économie de la décroissance*, Paris, Seuil, 2022, p. 219.
- 100 GRINBAUM (A.), *Parole de machines. Dialoguer avec une IA*, op. cit., p. 160.
- 101 GARNIER (P.), *La démence du percolateur*, op. cit., p. 22.
- 102 ELLUL (J.), *La technique ou l'enjeu du siècle*, Paris, Economica, 2008.
- 103 GRINBAUM (A.), Parler avec les machines, *Études*, n° 4307, septembre 2023, p. 55-65. En ligne : <https://www.revue-etudes.com/article/parler-avec-les-machines/26324>
- 104 HEIDEGGER (M.), *Langue de tradition et langue de technique*, Bruxelles, Lebeer-Hossmann, 1990, p.

27.

- 105 HOTTOIS (G.), *Le signe et la technique. La philosophie à l'épreuve de la technique*, Paris, Vrin, 1984, p. 39.
- 106 CAMUS (A.), *La chute*, Paris, Gallimard, 1956, p. 60.
- 107 HOTTOIS (G.), *Philosophie et idéologies trans/posthumanistes*, Paris, Vrin, 2012, p. 11.
- 108 WEIZENBAUM (J.), cité in IA – Le mythe du XXI^e siècle, *Philosophie Magazine*, hors-série n° 57, printemps 2023. En ligne : <https://www.philomag.com/articles/joseph-weizenbaum-il-ny-rien-de-commun-entre-decider-et-choisir>
- 109 CAMUS (A.), *L'homme révolté*, *op. cit.*, p. 309.
- 110 ZWEIG (S.), *Le joueur d'échecs*, *op. cit.*, p. 51.
- 111 Cité in MALKA (S.), *Levinas. La vie et la trace*, Paris, Albin Michel, 2005, p. 179.
- 112 CAMUS (A.), *L'homme révolté*, *op. cit.*, p. 201.
- 113 LEVINAS (E.), *Éthique comme philosophie première*, Paris, Rivages, 1998, p. 103.
- 114 *Ibid.*, p. 104.
- 115 HARTMUT (R.), « Nous souffrons de moments de décélération forcée », propos recueillis par ORTOLI (S.), *L'art de ne rien faire*, *Philosophie Magazine*, hors-série n° 58, été 2023. En ligne : <https://www.philomag.com/articles/hartmut-rosa-nous-souffrons-des-moments-de-deceleration-forcee>
- 116 TISSERON (S.), *Cyber-psychologie : Notre identité en question*, *op. cit.*
- 117 DEVILLERS (L.), GRINBAUM (A.), ADDA (G.), CHATILA (R.), MARTIN (C.), VILLATA (S.)... et ZOLYNSKI (C.), *op. cit.*
- 118 FLEURY (C.) et BRUNFAUT (S.), Cynthia Fleury : “Notre société fabrique de plus en plus d'indignités”, *L'Echo*, 2 septembre 2023. En ligne : <https://www.lecho.be/opinions/general/cynthia-fleury-notre-societe-fabrique-de-plus-en-plus-d-indignites/10489913.html>
- 119 *Ibid.*
- 120 TYSON (A.) et KIKUCHI (E.), Growing Public Concern about the Role of Artificial Intelligence in Daily Life, *Pew Research Center*, 28 août 2023. En ligne : <https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concernabout-the-role-of-artificial-intelligence-in-daily-life/>
- 121 CAMUS (A.), *L'homme révolté*, *op. cit.*, p. 17.

Votre avis nous intéresse !

Laissez un commentaire sur le site de votre libraire en ligne et partagez vos coups de cœur sur les réseaux sociaux !

Retrouvez toutes nos publications sur notre site

MARDAGAⁱ

Coordination éditoriale : Chloé Van Rossom et Cindy Gravier
Design de couverture : Laura Marion
Maquette intérieure : Carine Thurion – CW Design

© 2024 Éditions Mardaga

Éditions Mardaga
B-1160 Bruxelles (Belgique)
www.editionsmardaga.com

Toute reproduction ou représentation intégrale ou partielle du présent ouvrage,
par quelque procédé que ce soit, est strictement interdite.

e-ISBN : 9782804734855

© 2024, version numérique Primento et Éditions Mardaga

Ce livre a été réalisé par [Primento](#), le partenaire numérique des éditeurs