La Grande Conversation

*La revue de **Terra Nova***

Aux miroirs de l'Intelligence Artificielle

Société

Par Anne Alombert

Henri Verdier

Philippe Lemoine

Entrepreneur et essayiste, président du Forum d'Action Modernités

Publié le 2 juin 2026

« La technologique n'est ni bonne ni mauvaise, mais pas neutre non plus », disait le sociologue Melvin Kranzberg. L'IA, bifurcation technologique majeure, n'est pas seulement affaire de risques et d'opportunité, elle véhicule son lot de transformations psychologiques, sociales ou anthropologiques, sur lesquelles travaille, depuis des années, la philosophe Anne Alombert, ancienne membre du Conseil national du numérique. Echange à trois voix sur un monde d'interrogations qui s'ouvre…

Sommaire

HENRI VERDIER

Vos livres participent de la grande tradition française de la philosophie des sciences et des techniques : *De la Bêtise artificielle*, *Schizophrénie numérique*, *Penser avecBernard Stiegler* et votrelivre sur l'économie politique de l'IAavec Gaël Giraud. Mais a-t-on encore besoin de philosophie des sciences et des techniques ? En quoi est-ce important ?

La philosophie permet d'interroger les technologies non seulement du point de vue de leurs performances ou de leurs fonctionnements internes, mais aussi du point de vue de ce qu'elles font aux humains, au niveau individuel et au niveau collectif, en ouvrant sur des enjeux anthropologiques plus larges, notamment des enjeux psychiques, sociaux et politiques. Par exemple, en posant la question de savoir ce que les transformations technologiques et industrielles actuelles font à nos esprits individuels et collectifs.

La question de savoir ce que la technique fait à nos esprits est au fondement de la philosophie : la philosophie commence même avec cette question. Quand la philosophie voit le jour, l'écriture alphabétique est en train de se diffuser

dans la société grecque. Et la question que Socrate et Platon soulèvent, notamment dans le dialogue intitulé le *Phèdre*, est de savoir comment ce nouveau support technique affecte les capacités psychiques, en l'occurrence la mémoire. En rendant possible l'extériorisation des savoirs, l'écriture va transformer notre manière de nous souvenir, notre manière de transmettre les connaissances, notre manière de les pratiquer, etc. La philosophie s'ouvre donc avec cette question des enjeux de la technique de l'écriture pour les esprits, au sens des capacités psychiques et de la culture collective — cette idée, je la dois vraiment au philosophe Bernard Stiegler, qui a beaucoup insisté là-dessus.

Ce sont ces questions que j'ai tenté de réactiver dans *De la bêtise artificielle :* aujourd'hui en effet, la technique de l'écriture alphabétique n'est plus le support d'enregistrement dominant, les supports d'enregistrement dominants sont les technologies numériques. Dans ce nouveau milieu technique, qui est devenu un milieu technologique et industriel, les activités de pensée sont profondément bouleversées : les pratiques de lecture et d'écriture qui sont au fondement de la pratique même de la philosophie ne sont plus un acquis aujourd'hui, elles sont sans cesse remises en question par les innovations technologiques, en particulier par les nouvelles machines d'écriture que constituent les IA génératives. En ce sens, je ne tente pas seulement de produire un discours philosophique sur les transformations technologiques en cours, mais je me demande aussi comment les transformations technologiques actuelles affectent la possibilité même de la philosophie — mais il en va de même de la science, de l'art ou de la politique !

HENRI VERDIER

Les biologistes nous disent que la lecture nous demande de connecter des aires cérébrales. C'est même une transformation biologique puisqu'on connecte l'aire de la vision et l'aire de l'audition. Mais on voit bien, dans vos écrits, que c'est une définition de ce qu'est l'humain qui se joue dans ces innovations ?

Il est tout à fait juste de dire que ces transformations de nos esprits ne se font pas seulement au niveau des capacités psychiques ou cognitives mais au niveau des organisations cérébrales elles-mêmes. C'était aussi un point très important dans la pensée de Bernard Stiegler, qui avait un mot pour cela : il parlait de mémoire « organologique », pour souligner le fait que la mémoire humaine n'est pas seulement organique, c'est-à-dire qu'elle ne repose pas seulement sur un organe biologique, le cerveau, mais qu'elle est organologique, au sens où elle implique aussi des organes artificiels ou instrumentaux, elle repose sur les relations entre nos cerveaux et tous les artefacts techniques ou médiatiques qui les entourent. Les travaux de Maryanne Wolf, chercheuse en neuroscience et en littérature, montrent très bien cela : le passage des supports imprimés aux supports numériques transforme la synaptogenèse, c'est-à-dire la manière dont les neurones se connectent dans le cerveau. Quand j'apprends à écrire à la main et à lire sur du papier imprimé, ou bien quand je tape sur un clavier et que je lis sur écran en étant sollicitée par toutes sortes de notifications, ce ne sont plus les mêmes connexions neuronales qui s'opèrent, donc certaines zones cérébrales vont être plus ou moins abandonnées, et il deviendra difficile de les remobiliser. C'est d'ailleurs ce qu'une récente étude prépubliée par des chercheurs du MIT a montré au sujet de ChatGPT, en évoquant la « dette cognitive » qui tend à s'installer suite à l'usage de l'outil : à force de déléguer nos activités d'expression,

on devient dépendant de la technologie, on ne peut plus s'en passer !

C'est très important de prendre conscience de cela, notamment pour questionner l'anthropologie sous-jacente aux discours sur l'IA. En effet, quand on parle d'intelligence artificielle, on affirme implicitement qu'il existe quelque chose comme l'intelligence humaine d'une part et l'intelligence machinique d'autre part, et l'on tente d'ailleurs souvent de les comparer, comme s'il s'agissait de deux entités autonomes et séparées, qui seraient dans une sorte de compétition. Or, dans la perspective organologique que je viens d'évoquer, il n'y a pas d'intelligence purement humaine : l'intelligence a toujours déjà été artificielle, depuis les silex taillés jusqu'à ChatGPT. Nos activités intellectuelles sont toujours conditionnées par des prothèses techniques, et tout l'enjeu est de comprendre comment ces prothèses affectent nos esprits, afin de limiter les risques et de développer les potentialités. Ce qui devrait nous intéresser, c'est la relation entre nos esprits individuels et collectifs et nos milieux numériques, et non la comparaison entre de supposées intelligences humaines et de supposées intelligences artificielles.

Je me méfie beaucoup de ces comparaisons : vouloir produire des systèmes algorithmiques aussi « intelligents » que les humains peut aussi être une manière de rendre les humains aussi obéissants et contrôlables que des ordinateurs. Ces comparaisons entre humains et machines reposent la plupart du temps sur une vision mécaniste et comportementaliste de l'humain, compris sur le modèle de l'ordinateur, comme un système entrées/sorties effectuant des calculs sur des données. Or cette vision computationnelle de l'humain comporte des enjeux politiques : si nous ne sommes rien de plus que des machines à calculer, alors pourquoi ne pourrions-nous pas être remplacés ? Se cache derrière ces comparaisons l'idée selon laquelle nous devrions être aussi programmables et performants que les machines algorithmiques, alors même que

nous ne faisons pas la même chose qu'elles. Bien sûr, nous intériorisons toutes sortes d'automatismes corporels ou psychiques, mais la spécificité de l'humain, comme le soutenaient les philosophes Gilbert Simondon et Bernard Stiegler, c'est de pouvoir désautomatiser, créer du nouveau, de l'improbable, du singulier – notamment à travers des activités d'interprétation ou de décision qui ne sont pas automatisables, et ce sont ces activités qu'il faudrait valoriser, plutôt que de mécaniser ou d'automatiser les comportements humains à travers des services standardisés. Enfin, le fait d'anthropomorphiser les machines peut aussi être dangereux : l'anthropomorphisation conduira à vouloir attribuer des droits aux robots (alors qu'il faudrait plutôt penser à protéger les humains qui n'auront plus d'emplois) ou à vouloir laisser l'IA apprendre à partir de données numérisées (alors qu'il faudrait plutôt penser à rétribuer les humains qui ont produit ces données).

Bref, je pense qu'il faut éviter de mécaniser l'humain et d'anthropomorphiser la machine, non pas pour les opposer, mais pour interroger les relations entre humains et machines, justement. L'humain n'est pas l'opposé de la technique : il s'est toujours déjà extériorisé techniquement dans toutes sortes de machines, mais la question qui doit être posée est celle de savoir si cette extériorisation technologique se fait pour son bien ou à son détriment, et quelles sont les dynamiques politiques et les enjeux de pouvoirs à l'œuvre dans ce processus d'extériorisation. Les comparaisons entre humain et IA ont souvent pour effet de dépolitiser les questions.

L'IA, un pharmakon comme un autre ?

PHILIPPE LEMOINE

Pour quelle raison avez-vous choisi l'expression de « bêtise artificielle » comme titre de votre livre le plus récent ? Tout le début du livre paraît bien plus nuancé : dans le prolongement de la pensée de Stiegler, votre interrogation sur l'IA renvoie à la vision qu'avait Socrate de l'écriture et des techniques en général. Il les analysait comme des *pharmakon*, c'est-à-dire des substances qui peuvent, selon la quantité absorbée, être un remède ou un poison. Tout est affaire de proportions. Pour quelle raison faudrait-il, avec l'IA, écarter les nuances et, d'emblée, situer les enjeux en termes de poison ?

Si j'ai choisi ce titre un peu provocateur, c'était avant tout pour questionner les discours dominants et promotionnels, qui envisagent souvent l'IA comme un remède à tous les problèmes et qui se caractérisent aussi par une certaine fascination à l'égard de ces systèmes. Ce que je voulais signifier, c'est que loin d'être « intelligentes » ou « spirituelles », ces industries numériques pourraient bien engendrer beaucoup de problèmes… C'est une manière de questionner l'idéologie techno-solutionniste, souvent portée par les entrepreneurs de la Silicon Valley, qui consiste à promouvoir ces dispositifs technologiques comme des solutions à tous nos problèmes : par exemple, « Il y a une épidémie de solitude dans la société : pas de problème, les agents conversationnels vont nous permettre de la résoudre » ; « Il n'y a pas assez d'enseignants : pas de problème, les enseignants seront remplacés par des *chatbots* interactifs et personnalisés » ; « Il n'y a pas assez d'aide-soignants dans les EPHAD : pas de problème, vous pouvez compter sur les robots compagnons », etc.

Si j'ai insisté sur le côté « poison » du *pharmakon,* c'était pour prendre à rebours ce type de discours techno-solutionnistes. Sachant que le rôle de la philosophie, en ce domaine, est aussi de déconstruire les discours idéologiques, pour ouvrir d'autres

questions que ces discours ne permettent pas de poser. Ici, la question que je voulais poser était avant tout celle des risques psychiques et politiques du déploiement massif de ces systèmes dans les sociétés, d'où le titre un peu provocateur. Car je crois que l'IA générative telle qu'elle est actuellement déployée par les géants du numérique repose sur un modèle industriel et économique intrinsèquement toxique : ce qui est pharmacologique, c'est la technologie algorithmique, mais si cette technologie est appropriée par des acteurs privés qui ne visent que leur profit au détriment des intérêts de la population, alors cette technologie devient un poison, et il me semble que nous sommes actuellement dans cette situation.

Cela dit, il faut aussi que je précise un peu ce que j'entends par « bêtise » : je ne voulais pas dire que les IA sont bêtes. Pour moi, cela n'a aucun sens de se demander si des systèmes algorithmiques sont bêtes ou intelligents : cela reviendrait à se demander si une chaise est assise ou debout, c'est une question qui n'a pas de sens. En revanche, l'idée est que ces dispositifs pourraient engendrer de la « bêtise », au sens d'une destruction de certaines de nos capacités psychiques et, surtout, d'une production de contenus insipides et insignifiants. En exergue du livre, on trouve une citation du philosophe Gilles Deleuze, qui explique que nous sommes envahis de textes et d'images, et que la bêtise n'est jamais muette ni aveugle : ce qui m'inquiétait dès 2022 avec ces nouvelles machines génératives, c'était la possibilité de générer énormément de contenus textuels ou imagés afin de noyer les contenus pertinents ou singuliers. Cette inquiétude s'est confirmée depuis avec ce qu'on appelle l'« *AI slop* » (la bouillie IA) : l'envahissement des plateformes musicales par des contenus automatiquement générés de médiocre qualité ou l'envahissement des revues scientifiques par des contenus automatiquement générés qui mettent en péril les processus de production et de certification par les pairs… Pour moi, c'est cela la bêtise : ce n'est pas une déficience intellectuelle individuelle, c'est une perte collective de sens.

PHILIPPE LEMOINE

Nicolas Carr posait déjà la question en 2011 : *Internet rend-il bête ?* Est-ce ce que vous voyez l'IA comme un prolongement et une amplification de tout ce qui se passe depuis 20 ans avec Internet, les réseaux sociaux, etc. ou l'IA est-elle, comme vous l'écrivez dès la première page du livre, « une bifurcation technologique majeure » ?

Il est toujours difficile de choisir entre la continuité et la rupture, c'est toujours un petit peu les deux. Je me méfie des discours de la rupture, qui soutiennent que l'on serait face à une nouveauté absolue : ces discours ont souvent pour fonction de créer de la stupéfaction et de rendre inintelligibles les innovations. Mais il faut être attentif en même temps à ce qui change, donc j'essaie de naviguer entre ces deux extrêmes.

En particulier, dans ce cas précis, du côté de la continuité, j'insiste beaucoup sur une certaine continuité entre la logique des algorithmes de recommandation sur les réseaux sociaux commerciaux et la logique des algorithmes de génération dans les *chatbots*. Du point de vue technologique, les deux systèmes reposent sur des technologies de *machine learning*,donc de calculs probabilistes sur des quantités massives de données. Il s'agit, dans les deux cas, de prédictions algorithmiques : en fonction d'un certain résultat à obtenir, les dispositifs prédisent automatiquement des tendances probables (par exemple, quel est le contenu probable que vous auriez envie de regarder en fonction des calculs probabilistes effectués sur les masses de contenus précédemment regardés par les autres usagers ou bien quelle est la séquence probable de signes qui correspond à votre demande dans l'interface du*chatbot* en fonction des calculs probabilistes effectués sur les masses de données sur lesquelles le c*hatbot* est entraîné). La technologie n'est donc pas complètement nouvelle. De plus, certains designs et certains effets sont comparables : les IA de recommandation

nous enferment dans des bulles informationnelles et nous suggèrent des contenus adaptés à nos points de vue tout comme les *chatbots* nous enferment dans des dialogues avec nous-mêmes en nous confortant dans nos opinions. Il y a donc des logiques de captation (qui ne sont peut-être pas des logiques technologiques ou des logiques scientifiques, mais qui sont des logiques économiques) qui sont très proches entre ces deux technologies, qui sont d'ailleurs moins des technologies que des services, produits et développés par des entreprises privées. De ce point de vue, on relève une continuité entre l'économie de l'attention qui s'est déployée avec les IA de recommandation sur les réseaux sociaux et l'économie de l'attachement qui est en train de se déployer avec les IA de génération dans les *chatbots.*

Cela dit, je pense qu'il y a quand même de la nouveauté, évidemment. Sans même parler de l'IA agentique dont la généralisation se prépare (et qui va transformer énormément de choses, notamment dans le champ du travail, de la production, de la vie quotidienne), ne serait-ce qu'au niveau de l'IA générative, il y a tout de même de grands changements. Parmi eux, celui qui me frappe le plus, c'est peut-être celui de l'automatisation du symbolique. J'ai l'impression qu'avec ces dispositifs, on passe un cap dans l'automatisation du symbolique. Alors évidemment, l'expression symbolique repose toujours sur des automatismes, l'alphabet en est un, en un sens, la machine à écrire aussi, les logiciels de traitement de texte également, etc. Mais j'ai le sentiment qu'on assiste, avec ces dispositifs génératifs, à un saut, au sens où il devient possible de produire des contenus (au moins apparemment) symboliques sans faire appel à aucun processus d'interprétation, comme si tous les processus d'interprétation pouvaient être court-circuités : à partir d'un prompt généré par IA, on peut passer une commande à une autre IA qui produira elle-même un texte susceptible de servir de prompt pour une autre IA, et ainsi de suite à l'infini. Je n'arrive pas bien à qualifier ce phénomène, il

s’agit de quelque chose comme une automatisation potentielle de nos milieux symboliques eux-mêmes.

Imitations et prédictions algorithmiques

Par rapport à d'autres techniques, vous insistez à juste titre sur le fait qu'on n'est pas dans les mêmes échelles de temps. Si l'on prend l'écriture, l'écart entre son invention (- 3300 avant Jésus-Christ) et le questionnement de Socrate (- 400 avant J-C) n'a absolument rien à voir avec les délais dont on parle à propos d'une interrogation sur l'IA ! De même, vous avez raison d'insister sur la puissance des acteurs économiques qui sont derrière le développement de l'IA.

Mais j'ai trouvé vraiment particulièrement intéressant une différence d'une autre nature mise en avant dans le livre. C'est celle du processus de désymbolisation qui serait à l'œuvre derrière la progression de l'IA. Vous rappelez que Turing avait prédit l'avènement de l'intelligence artificielle en déplaçant la question qui était débattue dans les années 1950 : « est-ce que les machines peuvent penser ? ». Jusques là, on s'interrogeait en tentant de comparer le fonctionnement interne d'un cerveau humain et celui d'une machine. Pour Turing, ce questionnement ne mène nulle part et il faut raisonner sur les effets, sur les « *outputs* ». Est-ce qu'une machine peut produire un effet de langage comparable à celle d'un être humain ? C'est son célèbre jeu de l'imitation. Et, en 1950, il fait le pari que « d'ici 50 ans [légère erreur…] quelle que soit la question posée, on ne pourra pas distinguer une réponse fournie par un ordinateur d'une réponse fournie par un être humain ». Là on est justement dans un incroyable pari sur un grand bouleversement de l'ordre symbolique !

Oui, sur la simulation, on a fait de gros progrès ! Mais il s'agit de simulation, et c'est précisément ce que dit Turing dans cet article sur le jeu de l'imitation, intitulé « *Computing Machinery and Intelligence* » (1950). C'est un article assez surprenant, auquel on a fait dire beaucoup de choses que Turing ne dit pas du tout en réalité : déjà, Turing soutient que pour affirmer qu'une machine pense, il faudrait changer le sens de ce que l'on

entend habituellement par « penser ». Donc il ne dit pas du tout qu'une machine pense au sens courant de ce terme. J'ai beaucoup d'hypothèses au sujet de cet article que je ne pourrai pas développer ici, mais il y a une manière de le lire comme un texte vraiment problématique, dans lequel Turing dit une chose et son contraire, joue avec son lecteur en le soumettant à des thèses contradictoires entre elles, à des paradoxes dignes du paradoxe du menteur (« Un homme disait qu'il était en train de mentir. Ce que l'homme disait est-il vrai ou faux ? »). Tout le texte tourne d'ailleurs autour de la possibilité du mensonge, de la tromperie et de la simulation.

On croit souvent que le jeu de l'imitation consiste, pour une machine, à imiter le comportement d'un humain. En fait, dans le texte, c'est plus complexe : l'ordinateur n'imite pas simplement l'humain, il imite un homme qui se fait lui-même passer pour une femme, donc il imite un homme qui se fait lui-même passer pour ce qu'il n'est pas, il imite un humain qui essaie lui-même de tromper un autre humain. C'est une simulation de simulation, une imitation d'imitation. Donc, selon mon interprétation, Turing n'est pas en train de nous dire que d'ici quelques années, nous allons réussir à produire des machines capables de *penser,* mais des machines capables de *tromper.* Des machines trompeuses, vous allez réussir à en produire, voilà ce que dit Turing. Et nous avons en effet beaucoup progressé dans la capacité de simuler technologiquement des comportements humains, en particulier des comportements langagiers.

Mais alors, la question qui se pose, c'est celle de la simulation. Et là, il faut faire appel à Simondon. Car Simondon, pour sa part, nous avertit contre la dimension illusoire de la simulation : ce n'est pas parce que vous pouvez simuler le langage humain avec une machine algorithmique que la machine et l'humain réalisent les mêmes opérations. Simondon prend l'exemple de la calculatrice : nous ne calculons pas comme des calculatrices, non seulement nous ne sommes pas des machines électroniques mais nous ne mobilisons même pas le même

ANNE ALOMBERT

système de numération. Donc même si nous aboutissons à un résultat identique, l'identité du résultat ne nous dit rien de l'identité des opérations ou des processus. Selon Simondon, il ne faut pas se fier seulement à l'observation extérieure des résultats, aux *outputs* : il faut chercher à comprendre les processus et les opérations internes, sans quoi on reste ignorant face aux fonctionnements technologiques qui sont par ailleurs de plus en plus complexes.

PHILIPPE LEMOINE

En quoi est-ce différent du problème général de la modélisation ? Lorsqu'on modélise l'arrivée d'un orage en météo, on ne prétend pas connaître réellement toutes les lois physiques qui président à cet événement. Simplement, on a un modèle qui dit qu'il va pleuvoir demain à 14h00 et s'il pleut effectivement le lendemain à 14h00, on considèrera que la météorologie est une discipline scientifique. En quoi est-ce différent pour le langage ?

Il est tout à fait possible de modéliser le devenir de tel ou tel langage dans telle ou telle circonstance ou de formaliser des règles syntaxiques ou grammaticales qui permettent de rendre compte de la génération de textes, ou encore d'entraîner des grands modèles de langage sur des grandes masses de données linguistiques ou textuelles pour permettre aux algorithmes de prédire des séquences de signes probables. Mais, pour autant, on ne peut pas dire que ces dispositifs « parlent », c'est un abus de langage. Ces dispositifs exécutent des calculs ou des programmes qui ont été écrits par des êtres parlants, qui savaient parler avant que ces modèles formels existent et qui ne mobilisent pas ces modèles formels ou ces calculs algorithmiques pour parler. De même, les orages n'ont pas

attendu la météo pour exister même si les prédictions météorologiques nous permettent de prédire les orages à venir.

La différence, néanmoins, c'est que les sciences et technologies météorologiques permettent de *prédire* les orages, mais elles ne *produisent* pas les orages. Les sciences et technologies linguistiques, à l'inverse, permettent de *prédire* des comportements linguistiques mais aussi de les*produire* de manière performative, car ces sciences et technologies sont implémentées dans des machines d'écriture automatisées, que les humains utilisent quotidiennement. Par exemple, quand le logiciel d'autocorrection corrige votre orthographe ou quand le logiciel d'autocomplétions complète votre message, vous vous conformez aux prédictions automatiques, il y a un effet performatif de la prédiction.

PHILIPPE LEMOINE

Je repose ma question : où est la différence ? Après de fortes chaleurs et une grosse évaporation, je prévois qu'il y aura du tonnerre. En quoi est-ce différent de prévoir que, quand tu dis « ce n'est pas la même », le mot qui va suivre est « chose » ? Et de compléter avant toi ton énoncé : « ce n'est pas la même chose » ?

D'abord, la dimension performative de la prédiction est importante et il ne faut pas négliger son impact : les technologies de prédictions algorithmiques ont des effets sur le réel, elles conforment performativement le réel à leurs prédictions, donc elles ont des effets d'uniformisation ou de standardisation. La prévision de phénomènes humains ou sociaux est tout à fait possible, mais elle n'est possible qu'à partir du moment où l'on a une masse suffisante de sujets à étudier, c'est d'ailleurs la raison pour laquelle les probabilités fonctionnent très bien sur des masses importantes de données.

ANNE ALOMBERT

Par exemple, je peux prédire que sur 100 personnes, il y en a 99% qui se lèvent le matin pour aller au travail. Mais je ne peux pas prédire que tel matin, exceptionnellement, ce type-là, parce qu'il a fait tel cauchemar ou parce qu'il ne supporte plus son patron, tout à coup, ne va pas aller travailler et va peut-être changer de vie ou je ne sais quoi. Les algorithmes prédisent toujours ce qui est le plus probable et, comme ces prédictions sont performatives, les usagers sont incités à se conformer à ce qui est le plus probable — ce qui engendre des effets de mimétisme comportemental ou d'uniformisation linguistique par exemple. Bien sûr, on peut faire beaucoup de prédictions, surtout quand on a volé beaucoup de données, mais ces prédictions, comme elles sont performatives, risquent d'éliminer les singularités.

Le processus de désymbolisation

HENRI VERDIER

On parle pour l'instant de simulation cognitive. Mais cette technologie nous arrive presque toujours emballée dans une autre simulation qui, elle, est émotionnelle. Les machines énoncent : "tu as tellement raison, je suis bien d'accord avec toi, je suis désolé". Ça induit une projection anthropomorphique constante. Il est très difficile de penser l'IA sans se dire : " elle veut, elle pense, elle répond". Ça me fait penser à la sociologue Zeynep Turkecki qui plaide pour qu'on oblige les robots à parler comme des robots, pour interdire les dispositifs d'IA qui miment des émotions et des sentiments.

Je suis tout à fait d'accord avec ce type de proposition : je soutiens qu'il faudrait éviter que les *chatbots* emploient la

première personne du singulier, qui incite les usagers à les anthropomorphiser, à se confier et à s'attacher à leurs services. Le *chatbot* vous dit : « Je ne suis pas un humain » mais à partir du moment où il dit « je », c'est trop tard. Le « je » est le pronom qui manifeste la capacité de faire référence à soi, donc une forme de réflexivité ou de conscience, donc même si le système affirme qu'il n'est pas un humain, le fait qu'il utilise le « je » suffit à générer une apparence de réflexivité ou de subjectivité. Même si nous savons qu'il s'agit de calculs algorithmiques et non d'une altérité vivante, nous avons tous cette tendance à anthropomorphiser les objets qui nous entourent : les entreprises numériques exploitent cette tendance psychique pour créer de l'attachement à leurs produits et en tirer du profit.

Le problème est que cela engendre des risques : les individus vont commencer à demander toutes sortes de conseils à leurs *chatbots,* concernant leur santé mentale, leur vie amoureuse ou leurs opinions politiques. Plutôt que de s'adresser à d'autres et de rencontrer des points de vues variés parmi lesquels il s'agirait d'arbitrer, de juger, de sélectionner, certains individus risquent de se renfermer sur leur petit assistant personnel disponible 24h/24 et 7j/7, qui ne demande jamais rien, qui n'exige aucune attention, aucune empathie, aucune patience, aucun effort. Alors que les autres personnes ne sont pas toujours d'accord avec moi, elles n'ont pas forcément envie de m'écouter d'ailleurs, elles ne me comprennent pas forcément, etc. C'est pourquoi je dois mettre en œuvre toutes sortes de savoirs sociaux pour me relier aux autres ! Et c'est précisément en raison de ces frictions que la relation à l'autre est transformatrice pour moi, qu'elle m'oblige à évoluer. Mais, avec ces*chatbots,* nous ne faisons que répéter le même, ce sont des miroirs de nous-mêmes qui risquent de court-circuiter la relation aux autres, de nous prolétariser socialement, de nous désocialiser…

ANNE ALOMBERT

Dans le livre, je parle d'objets anti-transitionnels : l'objet transitionnel, pour le psychanalyste Donald Winnicott, c'est le doudou qui permet au bébé de négocier la séparation avec sa maman, et donc de pouvoir entrer en relation avec elle et avec les autres (car tant qu'il n'y a pas de distance donc de séparation, il ne peut pas y avoir de relation, mais seulement de la fusion). Les compagnons virtuels constituent des sortes de doudous numériques qui nous font régresser à un stade fusionnel, dans lequel la résistance du monde et du réel n'est pas reconnue, comme ne sont pas reconnues les limites des pulsions individuelles…

HENRI VERDIER

Peut-être les jeunes vont-ils inventer une nouvelle éthique… J'ai surpris l'autre jour une de mes 2 filles qui a 20 ans et qui disait : « Ce mec-là c'est un vrai salaud, il a écrit sa lettre de rupture avec Chat GPT ! ». J'ai trouvé cela intéressant.

PHILIPPE LEMOINE

J'aimerais qu'on reste un tout petit peu sur le processus de désymbolisation.
On rappelait tout à l'heure que Turing avait insisté sur la simulation. Cela n'a pas empêché que la recherche en IA se soit pendant longtemps inscrite dans une logique symbolique, connexionniste. On essayait de formaliser des algorithmes permettant de passer d'un système à un autre. Par exemple, des algorithmes de traduction automatique. Tout cela a été une impasse ! Il y a même eu des théorèmes comme ceux de Marco Schutzenberger et de Noam Chomsky qui ont théorisé l'impossibilité d'une telle voie.

Aujourd'hui, l'IA s'inscrit dans une logique complètement différente, celle de l'intelligence artificielle générative. Celle-ci repose avant tout sur le calcul vectoriel, c'est-à-dire sur une sorte d'algèbre de la géométrie. Une branche des mathématiques très liée au traitement de l'image.

L'élément décisif a été la mise au point par des équipes de Google, en 2014 je crois, d'un logiciel très important qui s'appelle « Word2Vec ». Il permettait de traiter l'univers du langage comme si c'était une image. On mesurait les distances entre les mots comme s'ils flottaient dans l'espace et on pouvait donc les soumettre au calcul vectoriel. D'un seul coup, l'IA quittait les approches symboliques pour se rapprocher du monde de l'image et de l'imaginaire. Vous vous intéressez à cette rencontre entre l'IA et l'imaginaire et vous vous référez notamment à des auteurs comme Winnicott. Mais j'ai été étonné de ne pas voir cité Jacques Lacan qui théorisait justement la différence entre trois ordres : le réel, l'imaginaire et le symbolique. Pourquoi ?

En effet, je ne me réfère pas à Lacan parce que je ne connais pas assez bien Lacan, tout simplement, mais je pense que certaines des réflexions de Lacan sont importantes pour penser ces questions, notamment ses réflexions autour du stade du miroir. A cet égard, dans le livre, je m'appuie sur un texte de l'historien du droit et psychanalyste Pierre Legendre : Legendre est très influencé par Lacan, évidemment, même si ses réflexions sont moins psychanalytiques à proprement parler, plutôt anthropologiques, historiques, juridiques ou même peut-être civilisationnelles. En tout cas, Legendre reprend un peu les réflexions de Lacan sur le stade du miroir en évoquant la reconnaissance par le sujet de son image comme image. Le fait de reconnaître mon reflet comme étant une image de moi-même me permet de faire une expérience étrange : j'ai en face de moi une image qui à la fois est moi (puisque c'est *mon* image) et qui en même temps n'est pas moi (puisque ce n'est que mon *image*, justement) — donc je suis, dans une certaine mesure, divisé d'avec moi-même, je ne coïncide pas avec moi-même. A partir de là, je peux développer un rapport réflexif à moi-même (la réflexivité dont je parlais tout à l'heure en évoquant le pronom « je »). Je peux parler de moi, faire référence à moi et je comprends aussi que j'ai besoin des autres, que je ne me suffis pas à moi-même, et donc que j'ai besoin du langage, ou plutôt du symbolique plus généralement, pour me relier à moi-même et aux autres. Car les symboles, ce sont précisément ce que je partage avec les autres, ce qui nous permet de nous relier collectivement.

Quand nous parlons à un *chatbot*, nous parlons à une image de nous-même, à un reflet algorithmique de nous-mêmes, mais nous ne le reconnaissons pas comme une image, nous le prenons pour une altérité, pour une autre personne… Un peu comme Narcisse qui prend son reflet pour quelqu'un d'autre, et qui tombe amoureux de son reflet avant de se suicider… Certaines personnes tombent d'ailleurs amoureuses de leur *chatbot,* et il y a des procès actuellement aux États-Unis contre certaines entreprises (OpenAI et CharacterAI, je crois) car les

ANNE ALOMBERT

chatbots ont prodigué des conseils en matière de suicide à certains adolescents, en les incitant aussi à éviter de parler de leurs idées suicidaires à d'autres personnes. Bref, je développe cela de manière un peu plus structurée dans le livre, mais ce que je crains, pour le dire vite, c'est que la généralisation massive et disruptive de ces pseudo-compagnons virtuels ne conduisent à autant d'impasses narcissiques et d'apories relationnelles, court-circuitant la relation de moi à moi comme la relation de moi aux autres…

HENRI VERDIER

Les chercheurs en IA, notamment dans la Silicon Valley, sont pour leur part très préoccupés d'avoir conçu des modèles qui n'ont pas la moindre idée qu'il existe quelque chose qui s'appelle le réel. Les IA peuvent répondre à une question sur la mécanique quantique mais si tu leur demandes où sont tes clés, le modèle n'a aucune conscience qu'il existe un espace et un sujet qui possède une clé… Il fabrique des phrases mais n'a pas la moindre idée du fait qu'il est situé dans le temps et dans l'espace. Le modèle génère du texte, du texte, du texte Mais il ne se situe pas.

Tous les discours sur le risque d'une IA qui déclencherait une guerre nucléaire, qui provoquerait ceci ou cela, partent toujours du fait qu'elle n'a aucune appréhension du réel. Or, j'ai retenu de la psychanalyse que le langage c'est quand même un peu raté, ça ne décrit pas très bien le réel. C'est pourquoi il me paraît dangereux de bâtir une société sur des technologies du langage sans interroger les limites du langage.

Oui, il y a beaucoup d'incompréhension entre les humains. Et là, on a un langage appréhendé à travers le calcul vectoriel, c'est-à-dire comme un espace à n dimensions où il y a des proximités entre les mots. Exactement comme il peut y avoir des

proximités dans l'imagerie médicale entre une tumeur et des vaisseaux. C'est une analyse spatiale.

ANNE ALOMBERT

Et probabiliste, parce que ces calculs sont effectués à partir de toutes les masses de langage dont on dispose, alors que la manière dont je parle, la manière dont Henri parle, la manière dont Philippe parle, ce sont des manières différentes de parler, les unes des autres. Des manières de se taire aussi, car nos silences parlent aussi, d'une certaine manière…

HENRI VERDIER

Mais on ne peut pas les calculer. Lacan disait aussi que les gens n'écoutent que les signifiants. Donc, en fait, la manière dont tu parles n'a aucune importance parce que les gens n'écoutent pas vraiment.

ANNE ALOMBERT

Certes, mais ils sentent les rythmes, ils entendent les accents, les intonations, les lapsus… Ils entendent même les interruptions et les silences, alors que les machines, elles, ne savent pas se taire…

Tout ce développement appelle une réflexion sur le processus de désymbolisation. La logique des reflets, des images, des simulations ou des simulacres n'est pas de même nature que la logique symbolique. Et ce n'est pas non plus la même chose que le réel. D'où l'intérêt d'un détour par Lacan ! Il n'est pas anodin que, derrière la révolution actuelle de l'IA générative, on trouve une entreprise, Nvidia, qui est au départ un fabricant de cartes graphiques pour jeux vidéo. Le moteur de toute cette technologie n'est pas le produit d'une démarche symbolique. Ce n'est pas non plus le fruit d'un rapport au réel. Ce sont des

PHILIPPE LEMOINE

théories et des outils qui renvoient au traitement de l'image et de l'imaginaire.

ANNE ALOMBERT

Oui, on pourrait aussi se référer à Jean Baudrillard et au simulacre… En fait, j'ai l'impression que ce qui distingue le symbolique et l'imaginaire, c'est aussi, en un certain sens, que le symbolique suppose l'existence d'un tiers, la relation est médiatisée par un tiers, ce n'est pas une relation duelle et spéculaire. Le milieu symbolique, par exemple la langue, constitue un tiers qui me permets de me relier à moi-même et aux autres. Mais, dans le cas des services numériques, le tiers est court-circuité, c'est une relation de clients à produits/services ou de consommateurs à entreprises.

Uniformisation et entropisation culturelle

PHILIPPE LEMOINE

Derrière ces questions de passage de l'imaginaire au symbolique, on trouve une coupure très nette, qui est justement celle de l'accès du sujet à lui-même, en tant que sujet. Cela renvoie à la question de la singularité et à celle du sujet. Vous écrivez à ce propos une très belle phrase : « le véritable danger n'est pas la singularité technologique mais l'élimination des singularités idiomatiques et de tout ce qui contribue à l'évolution des sociétés ». Le monde de l'IA est un monde dans lequel il y a du « je » mais il n'y a personne derrière le « je ».

Oui, c'est cela. Il y a des calculs, il y a des algorithmes, il y a des entreprises, mais il n'y a pas d'altérité ou de singularité.

ANNE ALOMBERT

Effectivement, ce que ce que j'essaie de dire, c'est que nous avons besoin de ces singularités pour l'évolution culturelle, justement : les différents champs symboliques, que ce soient les sciences, les arts, la philosophie, le sport, la cuisine, etc., évoluent et se transforment à partir de ces bifurcations singulières, à partir de nouveaux styles de peinture, d'un nouveau théorème de mathématiques, d'un nouveau concept philosophique, d'une nouvelle manière de cuisiner, de jouer au foot, etc. C'est ainsi que tel ou tel champ culturel s'engage dans des voies nouvelles. Mais si on élimine par avance les singularités à force de prédictions automatiques, les comportements humains risquent de se conformer à des moyennes standardisées…

PHILIPPE LEMOINE

On va vers une forme de stérilisation et même d'uniformisation au carré, comme vous l'écrivez. Dans la masse des données qui sont collectées et traitées, une partie de plus en plus importante, majoritaire bientôt, provient de données produites par l'IA elle-même. Je comparais récemment, dans un article intitulé « Le séisme qui vient », la crise qui menace le numérique à la crise écologique. Un sol qui a été trop labouré et trop moissonné n'est plus fertile. De même, l'humus des données ne sera pas éternellement fertile.

Exactement. Je suis complètement d'accord avec cette comparaison. Je pense qu'il faut poser le problème symbolique et culturel en terme écologique : Bernard Stiegler parlait d'une écologie de l'esprit et Félix Guattari, avant lui, soutenait qu'il y avait trois types d'écologies, l'écologie environnementale, mais aussi l'écologie mentale et l'écologie sociale. Dans mon travail, je m'intéresse aux enjeux des technologies numériques (des IA de recommandation et des IA de génération plus spécifiquement) en termes d'écologie de l'esprit, ou d'écologie

mentale et sociale, car l'esprit est toujours individuel et collectif. De ce point de vue, il est très clair que les systèmes d'IA générative dominants sont extractivistes : de même que la construction de centres de données ou d'appareils électroniques épuisent les matières premières et énergies disponibles (métaux rares, eaux, électricité), de même, l'exploitation statistique des données collectées sur Internet épuisent les ressources symboliques ou culturelles. Sans données de qualité, les IA génératives ne peuvent pas générer des résultats pertinents ou signifiants. Or, comme les entreprises dominantes se contentent de piller les ressources culturelles ou symboliques disponibles (par exemple, les contenus de l'encyclopédie collaborative Wikipédia, les articles publiés sur les sites de journaux ou les livres numérisés, etc.) sans envisager des mécanismes de redistribution des richesses permettant de rémunérer les créateurs ou contributeurs et de renouveler ces ressources culturelles ou symboliques, ces ressources seront bientôt épuisées.

Par ailleurs, comme il devient possible de produire des contenus de mauvaise qualité de manière massive, ces contenus noient les autres et deviennent majoritaires sur Internet, et finissent par conséquent par nourrir les systèmes d'IA générative. Des contenus automatisés ou probabilistes sont en train d'entrer dans les bases de données des grands modèles de langage ou des grands modèles de diffusion, qui effectuent donc leurs calculs probabilistes sur des contenus déjà probables, ce qui aggrave encore les effets d'uniformisation. Je cite dans le livre un article publié dans la revue *Nature*, qui montre que la qualité des contenus se dégrade quand les systèmes sont entraînés sur des contenus déjà automatiquement générés : c'est ce qu'on appelle « *model autophagy disorder* ».

Et, bien sûr, cette dégradation des contenus symboliques aura des effets dans tous les domaines et va se combiner au problème déjà très important de l' « *AI slop* » (bouillie IA) que j'évoquais tout à l'heure : dans le domaine de la musique ou de

ANNE ALOMBERT

la vidéo avec les algorithmes des plateformes musicales ou vidéos qui recommanderont massivement des contenus automatisés pour éviter d'avoir à rémunérer les créateurs, ou bien dans le domaine scientifique. Par exemple, un article publié dans*The Atlantic* raconte qu'une grande conférence dans le champ du *machine learning* en 2026 a été perturbée par la réception d'une grande quantité de propositions, dont beaucoup avaient été générées par IA, et comme les organisateurs n'étaient pas assez nombreux pour relire les papiers, ils ont demandé à des IA de le faire, si bien que certains chercheurs ont reçu des évaluations de leur papier avec des citations qui ne figuraient pas dans leur papier. Tout cela me semble quand même très problématique pour le devenir de nos milieux symboliques…

HENRI VERDIER

Il y a aussi l'exemple de l'étudiant qui avait reçu la meilleure note parce que c'est la même IA qui a écrit le papier et qui l'a lu.

PHILIPPE LEMOINE

Cette uniformisation au carré ne peut déboucher que sur une crise au carré de la confiance.

HENRI VERDIER

On doit se préparer à un monde où, dans deux ou trois ans, on ne pourra plus apporter aucune preuve en justice parce que les vidéos ou les enregistrements seront mis en doute par quelqu'un qui dira "mais c'est évidemment truqué". Je ne sais pas si nous sommes prêts à vivre dans ce monde où l'on ne peut plus croire le moindre témoignage.

C'est un énorme problème, en effet. C'est un changement du régime de la preuve en fait. Et c'est un problème également

dans le champ de l'information. Le problème n'est pas seulement de pouvoir produire du faux, mais de ne plus avoir de critère pour décider de ce qui peut ou doit être cru ou non. Même les informations certifiées pourront paraître automatiquement générées, si bien qu'on ne pourra plus rien prouver, il y aura toujours quelqu'un pour dire que c'est de l'IA. Et à l'inverse, il sera possible de faire passer pour vrai tout et n'importe quoi. On parlait de rupture et de continuité tout à l'heure, je crois qu'il y a une rupture ici, au sens où de la désinformation, il y en a toujours eu et, d'une certaine manière, c'est déjà ce que faisaient les sophistes au temps de Platon. Mais des fausses informations produites dans une quantité si massive et à un rythme si rapide, c'est inédit, et cela engendre peut-être un saut qualitatif…

Dans le livre, j'insiste sur une distinction qui me semble importante, entre la désinformation et la défiance. La désinformation, c'est le fait d'être trompé par telle ou telle fausse information ou contenu truqué,*fake news* ou *deep fakes.* Mais pour être trompé, encore faut-il pouvoir croire. Or, le problème, c'est que chacun risque bientôt d'intérioriser le fait qu'il ne peut plus distinguer entre un contenu généré automatiquement et un contenu certifié humainement, donc qu'il ne peut plus rien croire, ce qui risque de provoquer une défiance ou un discrédit généralisé. Or si on ne croit plus en rien, on ne peut plus être trompé, mais on ne peut même plus échanger. Toute société repose sur une forme de confiance : sans confiance, aucun échange social n'est possible sur le long terme. De manière caricaturale, c'est un peu ce que dit Emmanuel Kant au sujet du mensonge : si tout le monde se met à mentir, le mensonge lui-même devient impossible, car plus personne ne croit en rien, plus personne ne fait confiance à personne, donc plus personne ne peut mentir ni tromper les autres. Mais plus personne ne peut échanger non plus. Ici, c'est un peu la même chose : on est face à une telle généralisation de la possibilité de tromper, que la possibilité de croire elle-même est menacée et donc que la possibilité de tromper est menacée

en retour, mais au prix de rendre impossible la socialité elle-même… Bref, la question n'est pas seulement celle de la désinformation, il ne s'agit pas seulement de savoir comment repérer les fausses informations. La question, c'est aussi celle de la défiance, il s'agit aussi de savoir comment recréer de la confiance dans nos milieux symboliques numériques.

Des alternatives numériques ?

PHILIPPE LEMOINE

Au début de cet entretien, vous faisiez part de votre hésitation à parler de rupture. Pourtant toute la fin de votre livre renvoie à cette notion de rupture puisque vous estimez qu'il peut exister des alternatives numériques. Vous écrivez même que « la bêtise artificielle n'est pas une conséquence nécessaire des algorithmes. Ce ne sont pas les technologies numériques en tant que telles qui la provoquent, mais leur appropriation exclusive par des entreprises quasi monopolistiques. » Est-ce une invitation à penser qu'il peut vraiment y avoir un tout autre régime d'existence de l'IA ?

Je pense que oui. Non pas un autre mode d'existence de l'IA générative telle que nous la connaissons actuellement, mais d'autres technologies algorithmiques, qui ne soient pas consuméristes et prolétarisantes, mais contributives et capacitantes. Il ne faut pas réduire l'IA à l'IA générative conversationnelle telle que nous la connaissons aujourd'hui : cette IA-là repose sur des modèles économiques extractivistes à la fois au niveau des ressources naturelles et des ressources attentionnelles, émotionnelles et culturelles — elle se fonde sur

l'exploitation du travail humain et le pillage des données. Je pense que pour envisager les potentialités bénéfiques des technologies numériques, il faut changer complètement de paradigme et, en particulier, passer du paradigme de l'automatisation et du remplacement à celui de la capacitation et de la contribution.

Nous parlions de l'anthropologie sous-jacente aux modèles technologiques au début de l'entretien : les IA génératives qui dominent aujourd'hui visent à remplacer les activités d'interprétations humaines et le travail vivant par des calculs automatisés et du capital fixe. A l'inverse, je crois que si l'on adopte une anthropologie différente, une anthropologie relationnelle par exemple, on peut concevoir des technologies qui ne visent pas à remplacer les individus en se substituant à leurs activités, mais à les mettre en relation et à soutenir leurs capacités. Je donne quelques exemples de plateformes contributives dans le livre, comme Wikipédia, bien sûr, qui permet de coconstruire des savoirs culturels, ou comme Pol.is, qui permet de coconstruire des propositions politiques, ou encore comme Tournesol, qui permet de s'accorder sur les contenus jugés les plus importants collectivement. Je cite aussi des plateformes comme BipPop ou Entourage, qui permettent de renforcer les solidarités locales en mettant en relation des bénévoles et des personnes isolées dans le besoin : ces plateformes ont des modèles économiques tout à fait différents de ceux des géants du numériques, elles ne cherchent pas à mettre les algorithmes au service de la dépendance ou du profit mais au service de l'intelligence collective et du soin vivant. Là, on est loin des projets des grands patrons des Big Tech, qui veulent remplacer les soignants et toute sortes de professionnels par des *chatbots*, qui sont en fait des services numériques standardisés et privatisés : au contraire, des plateformes contributives comme Entourage ou BipPop permettent de renforcer les pratiques de solidarité, de socialité, de soin. Ici, la fonction de la technologie, ce n'est pas de remplacer les individus, c'est d'être un milieu qui nous met en

relation les uns avec les autres : on est beaucoup plus du côté d'un Simondon ou d'un Stiegler que d'un Sam Altman.

Enfin, dans le champ de l'IA générative, il me semble que le développement de l'intelligence artificielle générale n'est pas du tout une voie idéale, bien au contraire, elle n'est pas soutenable sur le long terme – ni économiquement, ni écologiquement, ni socialement. En plus, ce type de dispositif présente beaucoup de risques en termes de cybersécurité, comme l'IA agentique qui est vulnérable aux attaques par injections de prompts. Je pense qu'on aurait tout intérêt à développer plutôt des petits modèles, beaucoup plus spécialisés, sécurisés et fonctionnels, avec des jeux de données contrôlés, et qui puissent servir à automatiser des tâches spécifiques qui peuvent être automatisées parce qu'il a été décidé collectivement qu'elles pouvaient l'être pour telle ou telle raison. J'ai l'impression qu'il y aurait une stratégie industrielle plus intéressante en allant dans ce sens-là, en développant des petits modèles frugaux qui peuvent fonctionner localement. Parce qu'en réalité, le modèle de la Silicon Valley repose sur le développement de services commerciaux fondés sur la publicité comportementale, mais ces services ne sont pas sécurisés ni viables sur le long terme.

Ma conviction est que le paysage actuel de l'IA est encore très dépendant de la rationalité des Google, des Facebook et compagnie. Ces puissances ont voulu prolonger avec l'IA une certaine logique économique, une certaine logique de dominance. Mais ce n'est pas forcément l'avenir qui s'annonce…

Mondialement se développe un débat intéressant sur ce qu'on appelle la distillation, c'est-à-dire le fait d'entraîner des petits modèles pour qu'ils simulent le fonctionnement des gros. Les États-Unis redoutent cela et la Maison Blanche est en train de mener un combat contre la distillation considérée comme une

PHILIPPE LEMOINE

pratique illicite. L'argumentaire américain repose sur le fait que les gros modèles comme Chat GPT reposent sur des brevets que l'on violerait un droit de propriété en faisant de la distillation. Et on laisse entendre que l'IA chinoise se développerait ainsi.

Ce qui est vraiment intéressant, c'est que la Chine répond que la distillation fonctionne en fait dans l'autre sens. Nombre d'entreprises américaines font aujourd'hui appel à des petits modèles pour leur usage interne et ces petits modèles reposent sur la distillation des modèles chinois qui sont eux en open source !

Derrière le débat sur la distillation, on voit apparaître une interrogation sur la logique économique sous-jacente au développement de l'IA. L'univers des GAFAM raisonne dans le prolongement d'une logique hyper centralisée où la valeur reposait largement sur l'exploitation publicitaire des datas. Alors que l'IA transforme bien plus en profondeur le système économique.

HENRI VERDIER

Une semaine avant notre rencontre, Michael Kratsios, le conseiller numérique de la Maison Blanche, a publié une déclaration disant que l'industrie chinoise était fondée sur du vol des propriétés américaines. J'ai trouvé ça amusant parce que c'est quand même un voleur qui vole un voleur, puisque ces modèles sont principalement faits à partir d'une aspiration complète de tout le web, sans respect d'aucune forme de licence.

Ma conviction, c'est que l'IA va favoriser une disruption technologique forte. Un modèle d'IA de 3 à 7 milliards de paramètres peut fonctionner sur un micro-ordinateur. Pour

PHILIPPE LEMOINE

s'exécuter, il n'a pas besoin forcément d'une grande base de données ni forcément besoin d'un système de calcul très centralisé. Si l'on veut réinventer des tas de machines, produire des voitures autonomes, imaginer de nouvelles générations d'automates programmables locaux, il faut se situer dans une autre logique que celle de l'hypercentralisation dans laquelle nous baignons aujourd'hui.

HENRI VERDIER

J'entretiens le même espoir, avec une lecture un peu différente. Au fond, à part Tesla, le cœur de la Silicon Valley, c'est la publicité comportementale, c'est l'économie de l'attention. Sur ce segment, ils ont créé un quasi-monopole. La publicité est une industrie solide, mais c'est seulement 1% du PIB mondial. Pour le reste, pour les usines, pour les fermiers, pour les avocats, pour les profs, on n'a pas besoin de ces grands modèles qui connaissent tout sur tout le monde. Et en revanche, on a besoin d'autres données dont ces acteurs ne disposent pas. C'est pourquoi je pense qu'on peut assister à l'émergence de micro modèles à pertinence locale. Et vu le principe de construction des LLM, ces modèles seront infiniment moins consommateurs en énergie.

Un des paramètres les plus importants de ce qui va se produire est l'inflexion de la stratégie de l'acteur dominant, Nvidia, le fabricant des puces graphiques sur lesquels tourne l'IA. Il a bien compris que les logiques de centralisation risquaient d'enfermer l'IA dans un tout petit marché par rapport à ce que serait le fait de transformer tous les PC du monde en portes d'entrée de l'intelligence artificielle. Réaliser cette ouverture est l'objectif explicite de Nvidia à l'horizon 2030.

Vous développez à la fin de votre livre le rôle que pourrait jouer l'IA au service de l'intelligence collective et de tous les réseaux

de solidarité, de rencontre, de services, de *care*. Dans une optique très stieglérienne, votre perspective est celle de l'invention d'une nouvelle civilisation industrielle.

Mais comment réaliser une telle ambition ? Je reviens à l'idée de « bêtise artificielle ». Elle me rappelait un superbe numéro de *Beaux-Arts Magazine* coordonné par Jean-Yves Jouannais et qui faisait l'éloge de l'idiotie. Après la Révolution française, la France avait en effet été traumatisée par ce qu'elle avait fait et elle était persuadée qu'en ayant supprimé toutes les digues, elle avait libéré un raz-de-marée de bêtise populaire qui allait finir par la submerger.

Tout au long du XIXe siècle, elle n'avait donc cessé d'en appeler au génie et de célébrer les grands hommes dignes d'entrer au Panthéon. Tout cela jusqu'à Flaubert et à son intuition qu'il y avait beaucoup plus efficace que célébrer le génie pour combattre la bêtise : ce serait de valoriser l'idiotie. D'un seul coup, la France entrait dans une autre dimension et s'engageait dans l'incroyable trajectoire des arts incohérents, du dadaïsme, du surréalisme…

Je me demandais si l'on ne pouvait pas imaginer de nouveaux pas de côté. Si l'on craint la bêtise artificielle, ne faut-il pas faire appel à l'ironie, à l'humour, à la subversion, au non-conformisme ?

Oui, sans doute, des subversions, des détournements, ce genre de choses. Il peut être vraiment intéressant de détourner ces technologies d'IA de leurs fonctionnalités commerciales et consuméristes pour en faire des instruments de recherche ou de partage des savoirs. Par exemple, j'ai travaillé avec des collègues en science de l'information et de la communication et avec un développeur logiciel pour développer un petit outil qui permet d'effectuer des recherches par mots-clés dans des

corpus audiovisuels numériques, grâce à la transcription automatique du son sous forme de texte. C'est extrêmement intéressant, car cela ouvre de nouvelles possibilités de navigation dans l'océan des vidéos disponibles en ligne, dans le champ des sciences humaines notamment : l'outil permet de générer la retranscription textuelle des vidéos ou bien de faire des recherches par mots-clés dans les vidéos, qui conduisent directement au moment de la vidéo durant lequel le mot a été prononcé. Ici, on voit bien la dimension pharmacologique de ces technologies : l'IA peut servir à la fois au développement d'un *chatbot* qui répond à toutes mes questions de manière stéréotypée sans me donner ses sources et qui court-circuite mes activités de recherche en me fournissant du prêt-à-penser, ou bien, à l'inverse, au développement d'un outil qui me permet d'explorer de manière plus précise, plus singulière, plus fouillée les contenus audiovisuels en ligne, ce qui enrichit mes capacités de recherche singulières.

J'insiste sur le fait que dans ce cas précis et plus généralement, la dimension pharmacologique de la technologie ne se limite pas à des questions d'usage : c'est une question de fonctionnalité technologique. Si nous voulons que l'IA serve la capacitation et non la prolétarisation, il faut changer les fonctionnalités technologiques, ce qui suppose aussi d'envisager d'autres modèles économiques et d'autres idéologies politiques. Pour reprendre l'exemple précédent, dans le premier cas, le*chatbot* repose sur un modèle extractiviste insoutenable, il est développé dans le cadre d'entreprises privées qui cherchent à créer de la dépendance pour augmenter leur capitalisation boursière, alors que dans le second cas, l'outil de recherche par mots-clés est beaucoup plus sobre et il est développé dans le cadre d'institutions publiques qui cherchent à augmenter la culture et l'intelligence collective.

On peut prendre un autre exemple avec les algorithmes de recommandation : ceux des grandes entreprises californiennes

ou chinoises sont conçus de manière opaque pour capter l'attention des usagers et vendre leur « temps de cerveau » aux entreprises qui les ciblent avec de la publicité, alors que ceux développés par l'association Tournesol se fonde sur les jugements par les pairs qui évaluent les contenus en amont, afin de décider démocratiquement lesquels sont les plus pertinents en terme d'utilité publique. On assiste ici à un véritable détournement pharmacologique du dispositif : le poison devient remède mais, pour cela, il faut changer les fonctionnalités technologiques, les modèles économiques et les idéologie politiques.

Ce qui est intéressant avec ces dispositifs subversifs, c'est que des pratiques deviennent possible qui ne l'étaient pas du temps du livre ou de la télévision. Par exemple, à l'époque de la télévision, je ne pouvais pas chercher un mot-clé dans un émission, j'étais obligée de regarder toute l'émission même si tout ne m'intéressait pas. De même, à l'époque de la télévision, les individus ne pouvaient pas choisir les contenus qui étaient diffusés, alors qu'avec des algorithmes de recommandation collaborative, ils le pourraient. De même, à l'époque du livre, il était impossible d'écrire une encyclopédie de manière collaborative avec des millions de contributeurs, mais Wikipédia a permis cela. Ou encore, à l'époque du livre, il était impossible de délibérer sur des propositions de lois avec un très grand nombre de citoyens et de trouver des consensus transversaux entre des groupes d'opinions antagonistes, mais Pol.is permet cela. En ce sens, ces dispositifs sont vraiment des innovations, au sens fort du terme, car ils permettent à de nouvelles pratiques scientifiques ou politiques de voir le jour, pratiques qui demeuraient impensables dans un autre milieu médiatique : ce sont des innovations à la fois technologiques et sociales. Alors que l'IA générative, en ce sens, n'est pas vraiment une technologie innovante : ChatGPT peut écrire un roman ou une recette de cuisine, d'accord, c'est peut-être stupéfiant sur le moment, mais en fait on pouvait déjà le faire avant, on n'avait pas vraiment besoin de lui

ANNE ALOMBERT

Je dis cela de manière un peu provocatrice, pour suggérer qu'à mon avis, il est nécessaire de repenser ce que l'on entend aujourd'hui par « innovation » : s'il s'agit d'injecter des milliards pour augmenter les quantités de données et les puissances de calcul, je ne sais pas si l'on doit parler d'innovation en réalité. Alors que s'il s'agit de développer des dispositifs qui ouvrent à des pratiques contributives impensables auparavant, soutenables écologiquement et porteuses d'avenir sur le plan psycho-social, alors oui, il y a de la nouveauté et de l'innovation à proprement parler. En fait c'est la notion même de progrès qu'il faudrait repenser.

Régulation et éducation

HENRI VERDIER

Dans le même registre, je suis étonné par le peu de débats de philosophie politique. On voit bien l'apport de la philosophie de la technologie mais, face à d'autres questions, on a l'impression de manquer de concepts à la hauteur des enjeux. Quand on voit par exemple des entreprises fondées sur la maîtrise, peut-être la capture, de tous les savoirs du monde, on se dit par exemple qu'il nous manque un droit de l'humanité, non ?

Oui, c'est sûr. Il nous manque des concepts et des idées. Sur cette question de la protection des œuvres de l'esprit et des cultures communes, dans le livre que j'ai co-écrit avec l'économisteGaël Giraud, nous avions évoqué la possibilité que les grandes entreprises d'IA générative participent à un fond collectif pour un numérique contributif : ce fond pourrait ensuite être mobilisé pour financer la création de contenus de qualités ou la création de plateforme contributives (permettant la création ou la sélection de contenus de qualités) ou la

création d'algorithmes collaboratifs au service de l'utilité publique, etc. Dans la mesure où ces entreprises pillent les données des auteurs mais aussi de tous les contributeurs, il n'y a pas de raison qu'elles ne participent pas au renouvellement des savoirs humains : s'il est compliqué de rémunérer chaque auteur individuellement, on pourrait très bien imaginer un fond commun géré démocratiquement.

Sur le plan économique, Daron Acemoglu, lauréat de l'équivalent du prix Nobel d'économie, a aussi proposé de taxer les revenus publicitaires des grandes plateformes, pour les inciter à changer de modèles, justement, et pour compenser les coûts et les risques que les modèles publicitaires font peser sur les individus et les sociétés. Sur le plan du droit, il y a aussi beaucoup de choses à inventer. Le philosophe Mark Hunyadi a proposé une Déclaration Universelle des Droits de l'Esprit, pour protéger les esprits des emprises numériques. C'est une idée intéressante. Avec le Conseil National du Numérique, nous avions élaboréla proposition du « pluralisme algorithmique », qui obligerait les réseaux sociaux commerciaux à s'ouvrir à des algorithmes de recommandation alternatifs, pour que ce ne soit pas uniquement Meta ou X ou TikTok qui recommandent les contenus : une telle proposition a été reprise par le rapport des États Généraux de l'Information et par le rapport de la commission d'enquête sur TikTok – je pense qu'elle est vraiment importante pour diversifier nos écosystèmes informationnels numériques, de même que la proposition du dégroupage des réseaux sociaux porté par l'association Article 19 ou la proposition de l'interopérabilité verticale notamment défendue par Jean Cattin et le Future of Technology Institute.

Des choses sur lesquelles on s'est battu dans le cas des IA de recommandation sont également pertinentes pour les IA de génération : par exemple, l'interdiction des*dark patterns* (interfaces trompeuses), qui est mentionnée dans le *Digital Services Act* pourrait être transposée dans le cas des*chatbots,* si l'on considère que l'usage de la première personne du

singulier est une fonctionnalité trompeuse. De même pour la transparence des algorithmes ou le droit au paramétrage, qu'il serait pertinent de penser dans le cas des algorithmes de génération : on pourrait avoir accès aux sources par défaut ou avoir le droit de choisir comment le *chatbot* nous parle, par exemple. Bref, il existe des propositions politiques vraiment pertinentes mais encore beaucoup à inventer, en particulier parce que les technologies ne cessent d'évoluer : l'innovation permanente et disruptive, qui est une sorte de stratégie du choc industriel, rend la tâche politique difficile, mais d'autant plus importante et nécessaire.

C'est vrai que y a un manque de concept aujourd'hui, mais il manque aussi des lieux. Je m'efforce de regarder avec des yeux un peu innocents ce que Dario Amodei essaye de faire avec Anthropic. Il développe une ambition qui paraît être celle d'inventer un nouveau type de lieu innovant.

L'histoire est à écrire de tous les acteurs de l'IA qui se sont définis comme cherchant autre chose qu'une logique à la Google. Sam Altman en était parti pour créer Open AI. À son tour, Dario Amodei a quitté Open AI et a imaginé Anthropic comme réponse à l'enjeu anthropologique que représente l'IA. Il n'a pas recruté que des informaticiens et des ingénieurs mais aussi des chercheurs en sciences sociales. Il a banni tous les recrutements de personnes qui rédigeaient leurs CV avec de l'IA. Il a décliné Claude pour expérimenter et évaluer des IA fonctionnant dans des conditions proches d'un contexte réel. Par exemple, Anthropic a créé de toutes pièces une petite entreprise dont le CEO serait une Intelligence Artificielle dénommée Claudius. Et, de façon expérimentale, elle a simulé la croissance de l'entreprise, ses recrutements, ses investissements, sa concurrence, l'adaptation du modèle d'affaires… et Anthropic a évalué les décisions prises par le chef d'entreprise artificiel. Cette approche pragmatique n'était pas

PHILIPPE LEMOINE

sans intérêt puisqu'il est apparu que la jeune entreprise avait dû rapidement se déclarer en faillite !

De nombreux autres cas d'usage ont été expérimentés. On peut bien sûr soupçonner Anthropic d'afficher son exigence éthique, pour soigner son image et sa renommée. Mais on peut également se demander si ce n'est pas une bonne démarche pour initier et donner forme à une maîtrise de ce gigantesque enjeu qu'est l'IA. En tous cas, Anthropic a prouvé qu'elle savait tirer des conclusions de ses expériences, en prenant le risque de refuser que ses produits soient utilisés pour des activités de renseignement ou pour optimiser le « *killing process* » sur le champ de bataille. Comme on le sait, le Pentagone a réagi en annulant tous ses contrats avec Anthropic.

HENRI VERDIER

De temps en temps, Anthropic en effet montre des signes de "plus grande éthique". J'ai ainsi vu des travaux sur une "IA constitutionnelle" : un effort pour aligner leurs modèles sur la Constitution des Etats-Unis.

Oui, pourquoi pas ! Mais attention aux discours sur l'alignement des IA : bien sûr, il est important d'aligner les systèmes d'IA sur des valeurs ou des objectifs moraux, mais encore faudrait-il que nous soyons capables de nous accorder collectivement sur ces valeurs et ces objectifs. Je ne trouve pas cela très moral de mettre des systèmes d'IA qui parlent à la première personne du singulier entre les mains d'adolescents qui ne comprennent pas leur fonctionnement, pour capter leur attention, leurs affects et leurs données ou pour les cibler avec de la publicité. Même si le *chatbot* est bien aligné sur ce que l'entreprise considère comme des valeurs morales, je ne trouve pas que ce soit très moral, et je pense que l'on devrait tout d'abord discuter de savoir si l'on doit vraiment déployer ce type de « service » pour ce type de

public et si oui, à la limite, comment, dans quel cadre, quels sont les risques, quels sont les garde-fous, etc.

Dans le cas des industries pharmaceutiques, on étudie d'abord les effets secondaires avant de déployer massivement un produit. Pourquoi ne le fait-on pas dans le cas des industries numériques ? On a vu le résultat avec les réseaux sociaux commerciaux : le déploiement massif de systèmes fondés sur la publicité comportementale peut être très toxique sur les plans des activités cognitives et de la santé mentale, et tout cela a des répercussions sur les activités économiques et la vie sociale. Le risque des discours sur l'alignement moral des technologies, c'est de passer à côté de la question de leurs enjeux sociaux et politiques. L'urgence, selon moi, ce serait déjà d'aligner le développement industriel avec les impératifs écologiques et démocratiques. Ce qui passe notamment par la techno diversité, c'est-à-dire par le développement de modèles technologiques diversifiés et adaptés aux besoins locaux des populations, et non le déploiement planétaire de services appartenant à des entreprises quasi-monopolistiques, même s 'ils sont bien « alignés »...

PHILIPPE LEMOINE

L'IA est une technologie de rupture dont l'essence première est d'être une technologie de l'imitation. Nous sommes par ailleurs dans une société où la ségrégation sociale se joue tout autant dans l'assimilation des savoir-être que dans la maîtrise des savoirs et des savoir-faire. Au-delà du seul exemple de la rédaction des curriculums vitae, l'IA peut-elle être utilisée par certains pour combler des manques, compenser des insuffisances et mieux s'insérer dans la société ? Dit rapidement, est-ce que l'IA peut être un outil inclusif ?

J'imagine que oui, dans une certaine mesure, pour éviter certaines formes de discriminations peut-être, ou pour les personnes ayant des difficultés à s'exprimer dans telle ou telle langue, tel ou tel style, avec tel ou tel niveau de langage. Mais, tout de même, il vaudrait mieux que les sociétés elles-mêmes deviennent plus inclusives, plutôt que de s'en remettre à des services produits à des fins commerciales par des entreprises privées. Il me semble périlleux de compter sur ces systèmes, qui pour l'instant sont gratuits dans leur version standard, mais qui ne le resteront peut-être pas éternellement, et qui engendrent aussi d'autres inégalités : tout le monde ne peut pas payer l'abonnement à la meilleure version de ChatGPT. De plus, une dépendance cognitive peut vite s'installer : dans le livre, je cite et commente l'étude prépubliée par des chercheurs du MIT, dont je parlais tout à l'heure, sur les effets cérébraux de ChatGPT. L'étude n'est peut-être pas représentative, mais tout de même : ils montrent que certaines zones cérébrales ne sont plus mobilisées lorsque l'on utilise ce service, et qu'il devient beaucoup plus difficile de s'en passer ensuite (les personnes qui se sont habituées à utiliser ChatGPT ont beaucoup plus de difficultés à rédiger par elles-mêmes que celles qui ont continué à exercer leurs capacités intellectuelles). C'est vraiment la question du *pharmakon*, terme grec qui signifie le remède et le poison, que Jacques Derrida, dans son commentaire de Platon, proposait de traduire par « drogue ». La technologie peut devenir une drogue, surtout quand un service est conçu pour nous rendre accro !

L'IA peut être un outil inclusif, mais elle ne sera jamais cela exclusivement, et cette inclusivité peut se payer du prix de nouvelles formes de dépendances cognitives ou émotionnelles auxquelles il faut rester très attentif… Il peut y avoir des avantages à en tirer, bien sûr, mais il ne faudrait pas tomber dans l'écueil du solutionnisme technologique pour autant : si nous voulons des sociétés plus inclusives, sans doute serait-il plus efficace de réduire les inégalités économiques et de renforcer les services publics… Et, surtout, en ce qui concerne

ANNE ALOMBERT

les savoir fondamentaux, ne croyons pas que nous pouvons les abandonner et nous reposer sur des systèmes d'IA : pour être en capacité de pratiquer une technologie de manière cultivée et critique, il faut avoir reparcouru toute l'évolution technologique passée, par exemple, il faut avoir appris à compter sur ses doigts et à écrire ses multiplications pour faire des opérations avec une calculatrice, de même, il faut avoir appris à écrire avec un crayon, à mobiliser des règles de syntaxes et à formuler des arguments avant de rédiger des prompts adaptés...

HENRI VERDIER

Dans la marine française, un officier commence sa formation sur un bateau à voile. Ce n'est pas par nostalgie. C'est cette idée que pour être acteur de son destin, face à la technologie, il faut maîtriser l'histoire de cette technologie.

Dans l'éducation, tout le monde craint que les élèves utilisent L'IA pour tricher. Mais on constate aussi que c'est une pratique extrêmement courante chez les étudiants français de se faire des coachs personnels. Ils chargent le cours dans un modèle d'IA, et lui demandent de les aider à bachoter. Il émerge des usages actifs, pas seulement des usages paresseux.

Exactement. Et il faudrait les encourager en ce sens. On pourrait évaluer la singularité des productions et la dimension collective des travaux, plutôt que des performances individuelles sur des exercices standardisés. Ce qui est singulier et collectif, on ne peut pas l'automatiser, donc si c'est cela qui est évalué, les étudiants auront sans doute moins tendance à se livrer aux systèmes d'IA générative. Je pense que c'est une perspective qui s'ouvre à nous en tant qu'enseignants, notamment : dans chaque discipline, il faudrait se demander : comment valoriser la singularité des interprétations et le caractère collectif des travaux ? Ce n'est certainement pas en

remplaçant les enseignants par des IA qu'on y parviendra : l'apprentissage adaptatif avec des applications d'IA personnalisée n'évalue que des compétences individuelles et standardisées, et non des savoirs (savoir-faire, savoir être, savoir vivre, savoir penser), qui ne peuvent être que collectifs, partagés et singuliers. Et, bien sûr, la question de la culture technique que posait déjà Simondon, et même de la culture « techno-politique » est devenue fondamentale aujourd'hui : comment former de futurs citoyens qui soient capables de comprendre leurs milieux numériques quotidiens, du point de vue technologique mais aussi du point de vue de leurs enjeux anthropologiques, sociaux, politiques ? Les réponses restent à inventer.