LAURENT
BIBARD
NICOLAS
SABOURET

L'intelligence artificielle n'est pas une question technologique

Échanges entre le philosophe et l'informaticien

l'aube

LAURENT
BIBARD
NICOLAS
SABOURET

L'intelligence artificielle n'est pas une question technologique

Échanges entre le philosophe et l'informaticien

l'aube

L'intelligence artificielle n'est pas une question technologique

La collection *Monde en cours*
est dirigée par Jean Viard

www.editionsdelaube.com

ISBN 978-2-8159-5422-8

Laurent Bibard
Nicolas Sabouret

L’intelligence artificielle n’est pas une question technologique

Échanges entre le philosophe et l’informaticien

éditions de l’aube

Du même auteur, chez le même éditeur

Laurent Bibard, *Terrorisme et féminisme. Le masculin en question,* 2016

1
Intelligence artificielle : De quoi parle-t-on ?

Qu'est-ce que l'intelligence artificielle (IA) ? Peu de gens comprennent réellement ce qui se cache derrière ce terme à la mode.

L'intelligence artificielle est une discipline de l'informatique apparue au milieu du XX^e^ siècle, à une époque où les premiers ordinateurs commençaient à être construits. Le brillant mathématicien Alan Turing a alors imagé qu'un jour peut-être, avec ces machines, nous arriverions à faire des choses comme jouer aux échecs ou conduire une voiture, choses que nous, humains, accomplissons à l'aide de notre intelligence. Si cela arrivait, propose Turing, nous pourrions parler de *machine intelligence*, qu'on peut traduire par « intelligence mécanique », et que John McCarthy a rebaptisé en 1956 « intelligence artificielle ».

Un génie visionnaire

Il faut bien comprendre la difficulté qu'il y a derrière cet énoncé. Le mythe de la machine à l'image de l'homme n'est pas nouveau : bien avant Turing, on peut citer le Frankenstein de Mary Shelley au XIX^e^ siècle et, encore avant, le mythe de Prométhée, capable de créer un être vivant. Mais au début du XX^e^ siècle, nous avons commencé à construire des machines de calcul pour pouvoir faire ce qu'on appelle du traitement automatique de l'information, ce qui a donné en français le terme « informatique ». En anglais, on parle de calcul, *to compute*, ce qui a donné le mot *computer*, ou « ordinateur » en français. Nous construisons donc des machines qui effectuent des calculs et qui, à l'aide de ces calculs, sont capables de traiter de l'information en suivant un programme. C'est ce qui caractérise un ordinateur tel qu'il a été pensé au XIX^e^ siècle, réalisé au milieu du XX^e^ siècle, et tel qu'il existe aujourd'hui dans ce terme un peu général « numérique » (en anglais *digital* qui a donné lieu en français à l'adjectif « digital »).

Il faut donc s'imaginer les machines de l'époque de Turing, au milieu du XX^e^ siècle, qui n'avaient rien à voir avec nos smartphones, nos ordinateurs personnels ou nos montres connectées. C'était de gigantesques constructions, de grandes armoires comprenant des pièces mécaniques – des roues qui cliquettent, des bobines qui tournent, des picots qui attrapaient des fiches perforées – et beaucoup moins d'électronique ! Se projeter, comme l'a fait Turing, dans l'idée qu'un jour, avec de telles machines capables de

faire uniquement des calculs, nous arriverions à accomplir des choses que nous, les humains, nous faisons avec notre intelligence, c'était extrêmement visionnaire, voire révolutionnaire.

C'est de là que vient le terme « intelligence artificielle » : l'idée n'est certainement pas de dire que nous faisons une machine intelligente, ou que nous avons fabriqué de l'intelligence. C'est que nous utilisons un artifice, le calcul, pour accomplir à l'aide d'une machine des choses que nous, humains, faisons avec notre intelligence.

Logos et arithmos

Le terme « intelligence » est malheureux, parce qu'il n'y a évidemment rien d'intelligent dans ce que fait la machine : il s'agit uniquement de calculs, mais il est intéressant de remarquer que, dans notre vocabulaire, ce terme d'« intelligence » renvoie explicitement à la notion de rationalité. Or le mot « rationalité » vient de *ratio* en latin, et le *ratio* est un rapport qui fait l'objet de calculs. Et donc les notions de calcul, de rationalité et d'intelligence sont très proches.

Mais l'intelligence ne s'épuise pas dans le calcul : elle renvoie aussi au langage. L'origine du mot, qui est *logos* en grec, renvoie à une capacité par le langage à révéler le réel, donc de donner du sens à la réalité et pas seulement de calculer des données. Ce n'est donc pas la même chose et le terme d'« intelligence » renvoie potentiellement aux deux : langage et rationalité, *logos* et *arithmos*.

L'ambivalence que nous sommes en train d'évoquer vis-à-vis de la notion d'intelligence artificielle est donc fondée ; elle a du sens, mais elle demande la plus grande vigilance.

L'intelligence artificielle n'existe pas

Les effets de cette ambivalence sont potentiellement problématiques et c'est une des choses que nous allons traiter dans ce livre, car lorsqu'on parle d'intelligence humaine, on sait aujourd'hui qu'il y a plusieurs dimensions : la capacité d'apprendre à lire, par exemple, mais aussi la capacité à comprendre et à réagir aux émotions des autres, ce qu'on appelle l'« intelligence sociale », et encore plein d'autres aspects. C'est ce qui fait qu'il est très difficile de définir précisément l'intelligence, car il y a plein d'autres compétences que juste la capacité de faire du calcul ou la capacité de comprendre du langage.

Le point important à retenir lorsqu'on parle d'intelligence artificielle est cette difficulté fondamentale de résoudre, à l'aide d'une méthode de calcul, des problèmes de traitement d'information que nous, humains, savons

résoudre avec notre intelligence. Nous savons en effet démontrer mathématiquement certaines propriétés de ces problèmes et en particulier que certains de ces problèmes, pour lesquels nous utilisons notre intelligence, sont en fait extrêmement compliqués à calculer. C'est pourquoi il faut beaucoup d'intelligence humaine chez les programmeurs pour arriver à résoudre par le calcul ces problèmes de l'intelligence artificielle.

Il n'y a donc pas vraiment d'intelligence artificielle dans ce que fait une machine, aussi malin que cela puisse nous sembler lorsque nous regardons le résultat.

Où allons-nous ?

Une fois qu'on est bien conscient du fait qu'il n'y a pas d'intelligence complètement artificielle, mais bien des programmes de calcul pour résoudre des problèmes bien posés sur le plan scientifique, le fond du problème n'est pas ce que fait l'intelligence artificielle (qui est souvent magnifique), mais plutôt le rapport que nous, humains, entretenons avec ces machines que nous qualifions d'intelligentes et qui en fait ne le sont pas. C'est le thème central que nous voulons aborder dans le livre.

Derrière les questions que nous allons aborder, c'est non seulement la compréhension de ce qu'on appelle l'IA qui est en jeu – acronyme désormais convenu pour désigner l'intelligence artificielle comme technique, telle que définie dans ce chapitre –, mais c'est aussi la compréhension que nous avons, nous, humains, de notre propre humanité. À ce titre, la question de l'usage que nous faisons de toute technologie est centrale.

Il y a malheureusement de nombreux exemples d'enfermement dans la technologie et cet enfermement est largement, comme nous allons le discuter, du fait des humains. Une anecdote illustre bien notre propos. Il s'agit d'un échange entre l'un des auteurs et un ingénieur lors d'une formation permanente. Celui-ci explique qu'un jour, grâce à l'IA, il n'y aura plus besoin de pilotes (humains) dans les avions. Le formateur répond alors immédiatement : « Ah mais s'il n'y a pas de pilote (humain) dans l'avion, je ne monte pas ! » Et la réponse, fort éloquente, a été : « Mais vous ne le saurez pas ! »

Ce genre d'exemples montre qu'en investissant toute l'énergie possible dans la conception des systèmes, les ingénieurs pensent qu'ils vont aider les gens. Mais ils pensent à leur place, sans savoir ce dont les gens ont besoin de leur propre point de vue ni s'y intéresser. C'est en fait une bienveillance dangereuse, comme on peut en avoir avec le handicap ou la pauvreté. Loin d'apaiser le handicap ou la pauvreté, elle l'approfondit, car quand on traite

quelqu'un comme un enfant, comme c'est le cas dans le dialogue précédent, on maintient en enfance les enfants en question. Une telle posture peut malheureusement être comparée à ce que l'on trouve dans les régimes dictatoriaux : le gouvernant sait pour les autres ce qui est bien pour eux et les autres n'ont qu'à se taire.

Dans ce livre, nous allons aborder un certain nombre de questions autour de l'intelligence artificielle et de la place qu'on veut bien lui donner : pourquoi et comment la machine devient un obstacle à la réalisation de nos activités au lieu de nous aider, pourquoi nous en arrivons à croire que les machines peuvent faire des choses toutes seules, sans erreur, et comment nous leur attribuons une responsabilité au lieu de nous poser la question : À quoi servent les technologies et que veut-on en faire ?

2
Pourquoi les machines nous embêtent-elles autant ?

« De l'inconvénient d'être né. »
EMIL CIORAN[1]

L'une des toutes premières questions que nous avons rencontrées en commençant cette réflexion sur la relation de l'humain à l'intelligence artificielle est de comprendre d'où provient cette impression que les machines sont « contre » nous.

Les services en ligne

Chacun et chacune d'entre nous, en tant qu'utilisateur de la technologie, a un jour ou l'autre eu le sentiment que « la machine fait un peu ce qu'elle veut » ou, en tout état de cause, que nous n'arrivons pas à obtenir ce que nous voulons de la technologie. Les déboires de l'application de réservation de billets de la SNCF début février 2022 en sont un très bon exemple : les utilisateurs se sont plaints de « ne pas pouvoir consulter les tarifs des billets », ni « obtenir un train direct », alors qu'ils savent qu'il existe, puisqu'ils le prennent tous les jours.

Nous nous sentons souvent prisonniers des fonctionnalités de ces outils et c'est en général assez agaçant. Nous avons parfois le sentiment que l'ordinateur ne « veut pas » faire ce que nous voudrions qu'il fasse. Or, dès que nous rencontrons ce type de problème, que nous n'arrivons plus à obtenir un tarif ou que nous souhaitons obtenir un trajet particulier, nous avons besoin d'un interlocuteur humain. Et là, ce n'est pas possible… Il faut passer par une « foire aux questions » préétablies, qui nous promène dans un ensemble de pages web dans lesquelles nous ne trouvons pas la réponse à notre question et nous revenons au point de départ. Nous nous promenons de page en page sans jamais trouver ce que nous cherchons.

Nous pouvons citer l'exemple similaire du site web de l'Assurance maladie, Ameli, qui a récemment remplacé son formulaire de contact par un chatbot, une interface de communication prétendument capable de répondre à toutes nos questions. Le formulaire de contact permettait d'écrire à des employés de l'Assurance maladie *via* la messagerie interne. Le chatbot qui la remplace n'a souvent pas la réponse à notre question. On tourne en rond en lui répétant « voilà ma question », « j'aimerais savoir ce qu'il se passe

sur mon dossier ». Mais le chatbot nous promène de question en question, sans jamais nous donner de véritable réponse. Finalement, on se bat avec l'outil au lieu d'obtenir une réponse, et il n'y a plus aucun moyen de contacter un humain.

Ces exemples illustrent parfaitement le sentiment d'emprisonnement ou d'enfermement dans la technologie informatique que ressentent les utilisateurs. Cela contribue à cette croyance que la machine décide ce qu'on peut faire et ce qu'on ne peut pas faire, ou des réponses qu'on est en droit d'attendre ou pas du service – ici la SNCF ou l'Assurance maladie. Mais en fait, ce ne sont pas des machines qui ont décidé cela, ce sont les programmeurs de ces outils – le chatbot ou le site Internet. Ces outils, comme tout outil, ont été conçus et programmés par des humains qui ont fait des choix d'utilisation possible et qui ont cru qu'ils allaient, à travers ces choix, répondre à toutes nos questions. Ils ont alors débranché la possibilité de dire « votre système ne répond pas à la question, il me faut un autre interlocuteur ».

Ce que montrent ces exemples, c'est que les concepteurs ont choisi de ne pas permettre à l'utilisateur de s'adresser à un être humain, et ils ont choisi les réponses et les questions que la machine était capable de traiter. En dehors de ces questions, il n'y a plus aucune autre solution. Et quand il y a une demande de passer autrement que par le système qui est prévu, et donc conçu par les concepteurs, la réponse est « non, ça n'est pas possible, la machine ne le fait pas ». Comme si c'était la machine qui avait décidé de ce qui est possible.

C'est là une erreur de conception, qui provient d'une mauvaise compréhension de la relation entre les utilisateurs et les systèmes, du fait des fournisseurs des services concernés. Ce n'est pas du tout un problème d'intelligence artificielle mais de conception des systèmes !

Les aides de l'État à l'isolation

Un exemple supplémentaire dans lequel la technologie est à tort présentée comme responsable est celui des aides de l'État pour l'isolation énergétique et thermique des maisons. L'insistance sur la pertinence et l'importance de ces aides dans le cadre de la transition énergétique est considérable. Mais les systèmes qui sont destinés à préparer les dossiers pour l'obtention de ces aides sont conçus de telle sorte qu'il n'y a aucun contact personnel direct ou par des voies simples technologiquement, par exemple, envoyer des documents par email. Il faut passer exclusivement par le site. Or, ces sites sont la plupart du temps extrêmement difficiles à comprendre, voire incompréhensibles, avec, qui plus est, des interfaces difficiles à lire, par

exemple, pour des personnes malvoyantes.

Et quand on demande de l'aide à des opérateurs humains – on peut quand même joindre des opérateurs humains –, ces derniers ont en fait une position complètement intenable : ils ne peuvent que répondre : « Il faut aller sur le site. » Ainsi, quand bien même des opérateurs humains sont maintenus en emploi pour être les interlocuteurs des usagers, la seule chose, quasiment, qui leur reste à dire, c'est « je ne vous sers à rien », « je ne suis là que pour vous dire que je ne peux pas vous aider et qu'il faut tout faire *via* le site pour déposer votre dossier et en suivre le traitement ».

La difficulté que nous constatons n'a donc rien à voir avec les technologies prises comme telles. Elle dépend strictement des choix humains qui ont été faits lors de la conception des systèmes.

Il faut ajouter que quand une organisation décide d'imposer un passage obligé par un système pour la constitution de dossiers, par exemple, et maintient malgré tout des opérateurs humains qui n'ont plus qu'à dire « je ne sers à rien », c'est un problème – si ce n'est un scandale ! – strictement managérial.

La machine ne « comprend » pas

Face à ces systèmes qui nous gênent dans nos activités quotidiennes, nous adoptons une attitude qui consiste à essayer de « faire comprendre » notre besoin à la machine. Nous sommes ici dans des dynamiques d'« anthropomorphisation » de la technologie, c'est-à-dire qu'on entretient spontanément toutes et tous des relations avec les machines comme si elles étaient des humains, ou comme si elles étaient dotées de facultés, de compétences, de qualités humaines, comme l'est la capacité de « compréhension » de quelque chose.

Cette attitude est spontanée : quand nous interagissons avec un système, nous sommes en effet dans l'intention de faire « comprendre » notre besoin à l'interlocuteur virtuel qu'est le système, le logiciel ou la machine. Mais lorsque le besoin en question n'est pas prévu, nous nous retrouvons face à des réponses générales et cela provoque le genre de difficultés que nous avons évoquées. L'anthropomorphisation inévitable des systèmes n'aide pas à se sortir du problème, au contraire ! La raison est simplement que nous nous trompons d'interlocuteur : les machines ne sont pas nos interlocutrices, ce sont celles et ceux qui les conçoivent et les imaginent qui le sont.

On peut comprendre que les concepteurs préparent *a priori* des réponses générales à des questions générales, imaginées à l'avance. Cela est évidemment à la fois pertinent et efficace. Mais la difficulté vient de ce que l'on croit qu'en préparant ces réponses *a priori* générales, on épuise toutes

les questions. Or, cela n'est jamais vrai et ne pourra jamais l'être. Il y aura toujours des situations particulières non prévisibles à l'avance. L'erreur est bien de croire qu'on épuise *a priori* tout le champ des possibles. Et comme il n'y a pas de recours alternatifs pour passer à un humain et à des questions plus fines, la conception du système bloque l'opération qu'il est censé servir.

L'élimination des humains des services et ses enjeux

Imaginons un système très simple qui ne permettrait que d'acheter des billets pour une destination donnée, bien précise, connue à l'avance, avec des tarifs parfaitement clairs qui ne poseraient aucune question, sans mécanisme complexe avec des cartes de réduction ou des tarifs variables. Bref, rien qui puisse poser question à l'utilisateur. C'est déjà très difficile à concevoir, car il faut être sûr qu'il n'y a aucune incertitude possible sur les données fournies à la machine. Mais imaginons une telle situation. Dans ce cas, il s'agit, pour l'utilisateur, de dérouler une procédure en saisissant un ensemble de paramètres bien précis, bien définis à l'avance dans une machine qui va ensuite faire le traitement. Là il n'y a pas de problème : c'est une mission qu'un ingénieur est capable de remplir.

La difficulté, c'est que, dans la vraie vie, on n'est jamais dans cette situation-là. Il y a toujours des demandes d'utilisateurs qui sortent du cadre de ce que le concepteur du système a prévu. Évidemment, il s'agit d'une minorité d'utilisateurs, c'est-à-dire que la majorité des gens vont rentrer dans le cadre. Les concepteurs du système sont contents parce qu'ils ont répondu à 95 % des demandes. Mais ce qu'il faut regarder, ce ne sont pas ces 95 %, ce sont les 5 % restants, ces imprévus pour lesquels il faut admettre qu'on ne va pas pouvoir répondre à travers un système entièrement automatique, parce qu'un tel système n'a pas la capacité d'adaptation nécessaire pour faire face aux situations imprévues que créent les utilisateurs. Il faut impérativement, si l'on veut bien faire les choses, prévoir la place et le rôle d'un humain dans le dispositif. Il faut permettre aux agents de dire « oui, votre demande n'est pas traitable par la machine, mais moi je vais être capable de la traiter ».

La problématique à laquelle font face les ingénieurs et les concepteurs de systèmes est qu'on leur demande, voire on leur fait la commande, de réaliser un système qui fonctionnera sans humain. Ce qui est une impossibilité sur le plan technique : un système qui opère sans humain ne peut fonctionner correctement que lorsque le cadre est complètement prévisible.

Un problème politique qui n'a rien de nouveau mais qui risque de s'aggraver

La demande de la part des organisations qui sont gérées par des humains, demande humaine donc, est de manière dominante actuellement qu'il n'y ait plus besoin d'humains. Ceci n'est pas impossible en tant que tel. Mais il faut circonscrire de manière très précise les cas où cela peut fonctionner et être utile : ce sont les cas où la demande particulière colle parfaitement avec la demande générale. Tout est alors prévisible, mais cela est à la fois extrêmement rare et toujours local, provisoire, fragmentaire, comme nous l'avons vu sur l'exemple du billet de train. De manière générale, il y a sans cesse des cas particuliers. Une fonctionnalité générale rencontre donc la plupart du temps tôt ou tard des exceptions.

L'équivalent en philosophie politique consiste à dire qu'une loi est toujours potentiellement « tyrannique », au sens où la loi vise par principe le général. Et évidemment, il y a toujours des exceptions à la loi et, donc, si l'on s'en tient strictement à l'application de la loi sans considérer les cas particuliers, il y a injustice.

Le problème des technologies est comparable à cela : la technologie ne peut répondre qu'à des situations précises, mais accueillant le général dans le contexte précis qui est visé. C'est aussi le cas dans le domaine juridique : une loi est faite pour un objet précis et elle accueille le tout-venant dans le cas précis. La société repose alors sur les juges (humains) pour gérer la particularité de chaque cas dans ce cadre général de la loi. Là où, pour les technologies, cela peut devenir vraiment problématique (si ce n'est dramatique), c'est que les situations non prévues conduisent à exclure des utilisateurs. Par exemple, lorsque les sites empêchent des personnes malvoyantes de surfer sur Internet. Nombreux sont les exclus du net pour des raisons économiques, sociales, culturelles…

L'exemple du Boeing 737 MAX

Il arrive malheureusement que ces raisons conduisent à exclure un usage adéquat d'un système censé servir l'intérêt des usagers. Cela peut parfois se révéler encore plus problématique.

Prenons le cas du tristement célèbre Boeing 737 MAX, très révélateur de ce problème. Le Boeing 737 MAX est équipé d'un système appelé MCAS (pour Maneuvering Characteristics Augmentation System), qui est un système automatique installé sur l'avion destiné à éviter le décrochage en pilotage manuel. Quand un avion se cabre, ses ailes le portent moins et il risque de commencer à chuter : c'est le décrochage. C'est entre autres

choses ce qu'il s'est passé en 2009 sur le vol Rio-Paris.

Il y a donc une bonne intention des concepteurs de systèmes qui consiste à dire : lorsque l'avion se cabre et risque de décrocher, et si les pilotes ne s'en rendent pas compte et ne réagissent pas assez tôt, nous allons faire en sorte que, automatiquement, le système fasse plonger l'avion vers le bas pour lui permettre de reprendre de la vitesse et de récupérer son « assiette », c'est-à-dire la position de vol normale. Mais ce qui n'est pas prévu, c'est que le système puisse dysfonctionner. Or, les deux catastrophes arrivées en 2018 et en 2019 sur deux Boeing 737 MAX, à trois mois d'écart, sont des catastrophes qui ont eu lieu lors du décollage. Dans chacun des cas, l'avion était évidemment cabré, puisqu'il était en phase d'ascension. Les deux fois, le système MCAS a « calculé » que l'avion était indûment cabré et a fait plonger l'avion en vue de récupérer de la vitesse, de manière erronée.

Dans le premier cas, les pilotes n'étaient pas même informés de l'installation du système dans l'avion. Ils n'ont *a fortiori* rien pu faire et rien compris à ce qui leur arrivait. Dans le deuxième cas, les pilotes étaient informés que le système existait, mais ils n'avaient pas été formés au désenclenchement du système en cas de défaillance. Et donc ils ont assisté à la mise en route automatique de cette plongée totalement intempestive sans rien pouvoir faire non plus.

Hormis le problème directement éthique qui s'est posé dans le management chez Boeing à ce moment-là, nous voyons là qu'il y a eu un problème relatif à la culture sous-jacente à la mise en place d'un système automatique embarqué, et pas du tout un problème strictement technique. L'intention de fabriquer un système qui favorise la sécurité aérienne est excellente ! Mais nous sommes dans la présupposition que les pilotes n'ont même pas à savoir que le système est installé parce qu'il n'y aura jamais de défaillance, ce qui est une erreur totale, finalement assez similaire à ce que nous avons décrit pour sites web d'achat de billet, par exemple. Dans le cas des pilotes, on ne forme pas les usagers à débrancher le système et à reprendre la situation en main. Dans le cas des sites web, on ne propose pas d'alternative humaine pour obtenir le service souhaité. Les opérateurs humains ne sont là que pour dire « je ne peux rien faire pour vous ».

L'enfer est pavé de bonnes intentions

Ce que nous montrent ces exemples, c'est qu'il est important que les concepteurs ne décident pas seuls de ce que l'utilisateur va faire du système.

Ceci se produit soit par fantasme technologique, soit par souci de faire des économies par tous les moyens, ce qui est un véritable problème. Il est impératif que les décideurs des organisations, qui sont les commanditaires,

ne demandent pas aux ingénieurs de fabriquer n'importe comment dans n'importe quel contexte des systèmes dont on imaginerait (à tort) que les humains pourraient être complètement absents.

Comme nous le verrons par la suite, ces erreurs de conception conduisent à construire deux points de vue antagonistes dans nos sociétés. Il y a, d'un côté, le point de vue « technophile » du manager, qui s'imagine que la technologie va pouvoir tout faire et tout prévoir, et résoudre toute seule les problèmes. De l'autre, il y a une position « technophobe », qui est de plus en plus adoptée par les utilisateurs ayant l'impression que la technologie va à l'encontre de leur propre volonté, de leur demande, qu'elle est là pour les asservir, pour les empêcher d'être et de vivre bien… Ces deux points de vue se retrouvent depuis longtemps dans la littérature, avec des romans dystopiques ou utopiques, montrant, d'un côté, des situations où la technologie a permis à l'humanité de s'élever à un niveau incroyable et, de l'autre, au contraire, des situations où la technologie aurait emprisonné et réduit l'humanité entière en esclavage.

Il est essentiel de sortir de ce débat ! Tant que nous sommes dans cette opposition entre « la technologie va bien faire » ou « la technologie va mal faire », nous oublions l'élément important : ce n'est pas la technologie qui fait, ce sont les concepteurs de la technologie qui lui font faire et, derrière les concepteurs, les décideurs strictement humains des entreprises, des organisations, des services.

Tant qu'on maintiendra l'idée qu'il y a un débat entre progressistes technophiles et réactionnaires ou conservateurs technophobes, on paralysera la discussion qu'il faut absolument dépasser en direction de la bonne question, qui est de savoir comment on use de manière ajustée des technologies. Le fond du problème, c'est notre rapport à la technologie, ce n'est pas la technologie elle-même.

1. Emil Cioran, *De l'inconvénient d'être né*, Paris, Gallimard, 1973.

3
La machine peut-elle faire toute seule ?

« Se demander si une machine peut penser est aussi absurde que se demander si un sous-marin peut nager. »
EDSGER DIJKSTRA[2]

Le mythe de la créature intelligente, autonome, capable de « faire toute seule », revient avec force, porté par les progrès de l'IA. Il est accompagné de l'idée que la machine « apprend », à la manière d'un être humain, devenant chaque fois plus intelligente. On parle d'ailleurs d'« apprentissage automatique », voire d'« apprentissage profond » (*deep learning* en anglais) et même de « réseaux de neurones ». Tous ces éléments suggèrent que la machine se comporte comme un cerveau humain. Il n'en est rien.

Pour comprendre ce qu'un programme d'intelligence artificielle peut ou ne peut pas faire « tout seul », il est essentiel de revenir sur cette notion d'apprentissage.

Qu'est-ce que l'« apprentissage automatique » ?

L'informatique est la science du traitement de l'information : il s'agit d'étudier comment l'information peut être manipulée de manière automatique et de réaliser des machines qui effectuent ces traitements.

Pour comprendre ce que les machines peuvent ou ne peuvent pas faire, il faut d'abord se pencher sur la manière dont ces ordinateurs rassemblent, comparent et traitent l'information. Les modèles, pensés dès le XIX^e^ siècle avec les métiers Jacquard, puis la « machine à différence » de Charles Babbage, première machine à calculer programmable, représentent l'information à l'aide de nombres binaires, c'est-à-dire une suite de 0 et de 1. Le texte, les images ou les sons que nous manipulons aujourd'hui sur nos ordinateurs suivent tous ce modèle : une information est une suite de nombres et les ordinateurs sont des machines qui calculent. Autrement dit, un ordinateur est une machine qui, à partir d'une information (une suite de nombre) et d'un programme (une autre suite de nombre), produit une nouvelle information (toujours une suite de nombre), le tout en faisant des additions et des multiplications de nombres.

Tout ce que fait un ordinateur tient strictement à du calcul. Ce qu'on appelle l'« apprentissage automatique », le *deep learning*, ou encore les

« réseaux de neurones » ou l'« intelligence artificielle » ne font pas exception. De la même manière, nous pourrions dire que tout ce qui se passe dans notre cerveau, de la réception des signaux à travers nos cinq sens jusqu'à la réflexion mathématique, est le résultat de processus chimiques.

Ce qu'on appelle l'« apprentissage automatique » est une technique particulière (ou un ensemble de techniques particulières) pour effectuer un traitement de l'information. Il est caractérisé par le fait que le programme ne décrit pas directement le calcul de la réponse, mais plutôt la manière dont ce calcul peut être construit (calculé) à partir d'un ensemble d'informations, appelées « corpus » ou « base d'apprentissage », fournis à l'ordinateur. Pour simplifier, nous pourrions dire qu'un programme « classique » calcule un résultat à partir d'une valeur, alors qu'un programme d'« apprentissage automatique » calcule des paramètres à partir d'un corpus, lesquels paramètres permettent ensuite de contrôler un programme pour effectuer le traitement de l'information souhaité. Tout se passe comme si la machine écrivait elle-même un programme (ou du moins certains éléments d'un programme) à partir des données (le corpus d'apprentissage). C'est pour cela que les informaticiens ont baptisé cette technique « apprentissage », car c'est comme si la machine trouvait toute seule comment faire pour traiter l'information.

Mais la machine ne fait rien toute seule ! De même que le terme « *machine intelligence* » employé par Alan Turing en 1950 ne doit pas être confondu avec l'intelligence au sens où nous l'entendons pour les humains, le terme « *machine learning* », que nous avons traduit en français par « apprentissage automatique », ne doit pas laisser croire que la machine apprend comme le ferait un humain. Pour commencer, la machine ne produit pas n'importe quel programme. Elle produit les paramètres du programme de traitement qui correspond à ce que l'ingénieur concepteur a programmé. Elle écrit donc un ensemble de choses bien spécifiées, dans un cadre contraint. Même dans le cas des réseaux de neurones artificiels pour lesquels le nombre de paramètres est gigantesque et laisse une très grande « liberté » à la machine, les informaticiens doivent encore contrôler la structure du réseau, ce qu'ils appellent « méta-paramètres », de manière à ce que le traitement fasse bien ce qui est attendu.

Il ne faut pas oublier non plus que ces paramètres sont calculés à partir des données fournies par les humains (éventuellement collectées à leur insu) et formatées par les programmeurs pour qu'elles correspondent à ce dont le programme a besoin pour régler correctement ces paramètres. Non seulement la machine n'écrit pas n'importe quel programme, mais en plus elle ne le fait pas avec n'importe quelles données !

L’apprentissage de la machine

Prenons un exemple : la reconnaissance faciale à partir d’images pour identifier les individus. C’est un sujet qui fait beaucoup parler d’IA en ce moment en raison de son utilisation dans le contexte du « crédit social » en Chine.

L’objectif est d’écrire un programme qui reçoit en entrée une image et donne en sortie le nom de la personne sur cette image. Une image en informatique, c’est encore et toujours une suite de nombres ; chaque nombre ou groupe de nombres représentant les pixels de l’image, c’est-à-dire les points de couleur. À partir de cette suite de points, donc de cette suite de nombres, il s’agit de déterminer s’il s’agit plutôt du visage de Monsieur X ou de celui de Madame Y. L’idée de l’apprentissage automatique est la suivante : plutôt que d’écrire péniblement des règles informatiques disant que Madame X a les cheveux bruns et que Monsieur Y a les yeux marrons et des lignes de codes compliquées pour retrouver ce qui correspond aux yeux ou aux cheveux dans l’image, nous partons d’une base de données avec l’ensemble des photos de visages de Monsieur X, Madame Y (et de toute la population) à partir duquel nous écrivons un programme de reconnaissance avec des paramètres. Un programme d’apprentissage va alors régler automatiquement les paramètres, de manière à ce que le programme de reconnaissance faciale réponde bien « Monsieur X » quand on lui donne une image de Monsieur X et « Madame Y » quand on lui donne une image de Madame Y. Bien sûr, cette reconnaissance n’est jamais parfaite : nous reviendrons sur cette limite dans un prochain chapitre. Mais le programme d’apprentissage essaye de trouver les paramètres qui donnent les meilleurs résultats.

L’apprentissage automatique n’est donc pas un apprentissage à proprement parler : il s’agit en réalité d’un mécanisme de réglage de paramètres pour obtenir un programme de traitement de l’information conforme à ce qui a été souhaité par le programmeur et aux données du corpus.

Réseaux de neurones artificiels ?

Il n’y a donc pas de réflexion au sens strict dans l’apprentissage automatique, même dans le cas d’apprentissage dit « profond » à l’aide de réseaux de neurones artificiels. Ce terme est, à nouveau, assez trompeur : « réseaux de neurones ». Cela fait forcément penser au cerveau, alors qu’il s’agit juste d’une technique de calcul à base d’additions et de multiplications. Le nom provient de ce que dans les années 1960, on

essayait de simuler le fonctionnement d'un neurone humain (ou du moins la compréhension que nous en avions à l'époque) à l'aide d'un ordinateur, sans représenter l'ensemble des processus chimiques, mais en reproduisant le lien entre les entrées et les sorties. C'est d'ailleurs cette simulation qui a donné lieu ensuite aux très beaux résultats en termes de performance que nous connaissons aujourd'hui. Mais il ne s'agit en rien d'un processus qui s'apparenterait à de l'apprentissage humain.

Il en va de même pour l'« apprentissage profond », qui désigne simplement une technique particulière de réseaux de neurones artificiels dans laquelle on utilise plusieurs couches successives de « neurones », donc de calculs d'additions et de multiplications du même type. Le fait de les mettre en couche les unes derrière les autres va permettre d'obtenir de meilleurs résultats dans l'algorithme d'apprentissage. La profondeur est le nombre de couches, en aucun cas une « profondeur de réflexion », comme nous l'entendons à propos de la pensée humaine.

Réflexivité et apprentissage

Nous pouvons remarquer qu'un programme d'apprentissage automatique présente une certaine forme de réflexivité : c'est un programme qui ordonne à la machine de fabriquer son propre programme à partir d'un ensemble de données. Pour autant, la machine ne « réfléchit » pas à proprement parler sur elle-même : il n'y a pas de réflexion au sens humain du terme. Pour expliciter cela, il nous faut interroger le sens de la notion de réflexion propre à l'humain.

La réflexivité proprement humaine s'exprime comme possibilité de mise en question d'évidences, de préjugés au sens technique du terme.

De l'imperfection humaine

Il est indispensable pour bien comprendre ce qu'il en est de souligner le point suivant : une spécificité de l'humain est la pauvreté de notre mémoire biologique. Nous sommes, comme il est de mieux en mieux reconnu et compris, « prématurés ». C'est-à-dire que nos corps ne sont pas spontanément adaptés comme le sont ceux des animaux (nous n'avons, par exemple, pas de carapace, de griffes suffisamment acérées, de crocs, de pelage, etc.) et nos esprits, nos « âmes » non plus ; nous ne savons pas spontanément comment vivre, nous devons l'apprendre.

Nous l'apprenons de notre entourage : les parents, les frères et sœurs, le monde social où nous sommes éduqués, etc. Depuis notre naissance, nous passons toutes et tous notre temps à apprendre. Apprendre à s'habiller, apprendre l'hygiène, apprendre à se nourrir, apprendre à entrer en relation

avec les autres, apprendre à parler, à compter, etc. Or, apprendre revient à incorporer les normes qui nous permettent de vivre, de faire notre travail, etc. Cela donne, dans un sens qui déborde largement le monde de l'entreprise et du travail, toutes nos compétences.

Une fois « incorporé », ce que nous apprenons devient réflexe : nous n'y pensons pas pour appliquer, pour mettre à l'œuvre ce que nous avons appris. Nous le faisons « automatiquement ». Nous pouvons ainsi nous « saturer » d'évidences, ce qui nous permet de ne pas penser à une quantité considérable de choses pour évoluer dans le monde. La vie sinon ne serait tout simplement pas possible !

Sur cette base, la capacité de réflexion s'exprime comme une capacité spécifiquement humaine de prise de distance, d'interrogation, de doute par rapport aux évidences préalablement apprises. La mise en doute ou la problématisation du réel correspond à la capacité humaine à interroger le sens de ce que nous faisons de bien ou de mal, nos manières de vivre, et à réfléchir à d'autres manières souhaitées comme meilleures. On peut appeler cette capacité de mise en doute du réel et de nos évidences comme une capacité de désapprentissage de nos évidences.

Cette dynamique spécifiquement humaine de désapprentissage rend possible l'élaboration de meilleures règles, de meilleures manières de vivre en fonction de notre souhait d'une vie bonne. En retour, l'élaboration de ces meilleures manières de vivre revient à nous orienter vers un nouveau processus d'apprentissage, donc en direction d'un réapprentissage. La dynamique de réflexion spécifiquement humaine peut donc être approchée comme une dynamique faite d'un apprentissage, potentiellement suivi d'un désapprentissage, en direction d'un nouvel apprentissage donc d'un réapprentissage.

Un exemple contemporain de cette possibilité fondamentalement humaine est la prise de recul mondial à laquelle oblige la crise du climat. Nous devons interroger l'évidence de notre déploiement industriel, de nos manières de vivre en direction d'une vie qui permette la durabilité de notre monde. La radicalité de cet exemple souligne l'effort, la ténacité, le courage, la capacité d'écoute et de dialogue que demande la réflexion au sens proprement humain du terme. On peut ajouter sur cette base que les machines n'ont pas ce genre d'états d'âme, une machine ne se pose pas de question de sens. Elle ne se pose pas de question tout court et elle n'évalue pas par elle-même la qualité de ce qu'on lui demande de faire, elle fait ce qu'on lui dit de faire.

Si nous, humains, nous contentions de nos réflexes, nous ressemblerions à des robots. C'est donc sur le plan des réflexes que robots et humains sont

proches. Pas sur celui de la réflexion.

2. Extrait d'une conférence donnée en 1984 à l'université d'Austin (Texas) dans le cadre de la conférence ACM South Central Regional Conference.

4
Les machines font-elles des erreurs ?

« Il arrive que l'erreur se trompe. »
GEORGES DUHAMEL[3]

Nous attendons de l'intelligence artificielle, comme de toute construction humaine, une certaine forme d'infaillibilité. La machine ne devrait pas dysfonctionner, l'IA ne devrait pas faire d'erreur. À mesure que la science progresse, peut-on espérer atteindre une telle perfection ?

Le dangereux rêve de perfection

Le rêve de perfection, qu'il concerne les machines que nous fabriquons ou la société en tant que telle, est plus que dangereux, comme l'a montré Albert Camus dans *L'Homme révolté*[4]. Dès 1951, en plein régime soviétique, il met sur un même plan le nazisme, le fascisme et le communisme. C'est un scandale à l'époque pour tous ceux qui croyaient à l'humanisme que recèle le communisme. Et il est vrai que Marx est tout sauf un anti-humaniste. Mais la thèse de Camus est solide et pose une problématique encore d'actualité.

Il montre que les révolutions de son temps ont trois points communs majeurs :

1. Un rêve de solution définitive ou « finale » à ce qu'on pourrait appeler le « problème de l'homme » (le problème de la finitude, de la souffrance, du manque, de l'injustice, etc.) ;
2. Le rêve d'une réponse mondiale à ce problème de l'homme : la solution doit être mondiale ou elle n'aura pas lieu ;
3. Enfin, le rêve d'une solution qui permettrait à l'humanité de faire table rase du passé.

Si l'on observe ce qui se passe au travers du désir contemporain de perfection, nous y retrouvons les mêmes caractéristiques : le rêve d'une réponse définitive qu'apporteraient les technologies au problème de l'homme, le rêve que cette réponse soit mondiale et qu'elle nous permette de faire définitivement table rase du passé (c'est ce que proposent les transhumanistes).

Cela est alors très inquiétant, car Camus observe que ces grandes révolutions conduisent toutes à ce qu'il appelle un « terrorisme d'État » !

Au vu de la situation mondiale actuelle, nous subirions sans aucun doute non pas un terrorisme d'État, mais un terrorisme d'États au pluriel, ou bien un terrorisme socio-économique mondial, un terrorisme d'entreprises qui déborde largement la puissance des États.

L'impossibilité mathématique de l'IA parfaite

Ainsi, déjà au XX^e^ siècle, avant même l'intelligence artificielle, se posait cette question de la révolution qui conduirait à la perfection mondiale. Le mythe de la machine parfaite (aujourd'hui de l'IA parfaite) n'est donc pas neuf. Il est même assez ancien.

Ce mythe est alimenté depuis la Seconde Guerre mondiale par la croyance forte que la technologie allait pouvoir tout résoudre et qu'on allait pouvoir atteindre la perfection grâce à la machine. Dans le cas particulier de l'intelligence artificielle, il y a cependant une impossibilité mathématique à cette perfection. Nous l'avons brièvement évoquée au premier chapitre.

Comprenons-nous bien : cette impossibilité n'est pas juste le fait des ingénieurs ou des chercheurs qui feraient mal leur travail. Au contraire, dès le développement de l'informatique et l'invention de l'intelligence artificielle au milieu du XX^e^ siècle, Alan Turing a posé les bases théoriques du calcul : c'est ce que les scientifiques appellent la « théorie de la calculabilité et de la complexité ». Il s'agit de pouvoir caractériser ce qui peut se calculer ou non, et de mesurer la difficulté de l'obtention d'une solution exacte, donc « parfaite », à un problème de calcul donné.

Aussi surprenant que cela puisse paraître, certaines choses ne sont pas calculables. Par exemple, il n'est pas possible de construire un programme qui détermine si un programme s'arrête. En tout cas, il n'est pas possible de répondre à cette question à l'aide d'un calcul, comme le font les ordinateurs et les intelligences artificielles.

Cependant, aussi surprenant que cela puisse paraître, même pour les choses qui sont calculables, donc pour lesquelles nous pouvons écrire un programme d'intelligence artificielle, il existe certaines difficultés. Les scientifiques ont montré que tous les problèmes dits « calculables » se heurtent à une certaine complexité pour être résolus de manière exacte. Pour simplifier, nous pouvons dire que chaque problème nécessite un certain temps de calcul qui dépend avant tout de la nature du problème et de la taille des données. Or, la plupart des problèmes auxquels l'intelligence artificielle s'attaque font partie d'une catégorie de problèmes dont la complexité est trop élevée : reconnaître un piéton sur une image ou détecter une tumeur cancéreuse, gagner à coup sûr aux échecs ou au jeu de go, interpréter une commande vocale ou traduire un texte dans une autre

langue… Pour tous ces problèmes, obtenir une réponse parfaite, c'est-à-dire complète ou qui épuise le réel, nécessiterait des temps de calcul astronomiques.

Pour résoudre ces problèmes, les chercheurs et les ingénieurs en intelligence artificielle développent donc des algorithmes qui calculent une solution « pas trop mauvaise », à défaut d'être parfaite. Dans la plupart des cas, le résultat est d'ailleurs assez impressionnant, que ce soit pour gagner au jeu de go, pour reconnaître des images ou pour traduire des phrases. Mais aussi impressionnante que soit la performance de la machine, le programme d'IA ne peut jamais promettre qu'il ne se trompera pas. Pire : il a été conçu pour pouvoir se tromper ! C'est ce qui lui permet d'obtenir un résultat dans un temps de calcul acceptable.

Jouer aux échecs

Tout cela vous semble probablement un peu abstrait. Pour illustrer cette difficulté, prenons un exemple concret : le jeu d'échecs.

Il existe un algorithme inventé au début du xx^{e} siècle, donc avant l'avènement des ordinateurs modernes, permettant de gagner à coup sûr aux échecs. Cet algorithme a été proposé par le mathématicien hongrois John von Neumann et il consiste à comparer, coup après coup, toutes les parties possibles. Or, si l'on en croit les travaux du mathématicien Claude Shannon, le nombre de parties différentes que l'on peut faire aux échecs est de l'ordre de 10 puissance 120, c'est-à-dire un 10 avec 120 zéros derrière. Pour se représenter ce nombre, il faut imaginer le nombre d'atomes qu'il y a dans l'univers visible[5], autrement dit beaucoup ! Le nombre de parties d'échecs qu'il faut examiner dans l'algorithme de von Neumann est encore des milliards de milliards de milliards de fois plus grand que le nombre d'atomes dans l'univers visible.

Il semble donc qu'écrire un programme qui garantisse le coup parfait aux échecs n'est certainement pas une bonne idée : même avec des ordinateurs plusieurs milliards de fois plus rapides que ceux que nous connaissons aujourd'hui, nous serions encore loin du compte. C'est pourquoi les chercheurs ont inventé des programmes qui jouent aux échecs autrement, en ne cherchant pas à gagner à tous les coups. Pour cela, ils font un calcul avec ce qu'on appelle des « heuristiques », terme technique qui vient du grec *heuriskô* (« je trouve »). Il s'agit de méthodes de calcul consistant à dire « je vais jouer un coup dont je ne sais pas si c'est le meilleur, ce n'est d'ailleurs probablement pas le meilleur, mais en tout cas, c'est un coup pas trop mauvais ». Et le résultat est là ! Ces programmes arrivent désormais à battre n'importe quel joueur humain.

Les heuristiques ne sont pas l'apanage des ordinateurs. Dans la vie de tous les jours, nous appliquons en permanence des raisonnements heuristiques. Pour aller d'un point A à un point B dans la rue, nous suivons grosso modo le chemin le plus court, mais pas au millimètre près. Si je traverse sur ce passage piéton plutôt que sur le suivant, peut-être que c'est un peu plus long, mais peut-être pas. Globalement, cela semble être une solution acceptable et je ne vais pas passer ma journée à mesurer les deux distances pour choisir la plus courte.

Pragmatisme ou approximation ?

L'erreur serait de croire que la programmation heuristique est le résultat d'une forme de pragmatisme, voir de paresse, de la part des programmeurs. C'est tout le contraire : il s'agit d'un travail long et difficile qui consiste à identifier la meilleure manière de renoncer à la solution exacte pour écrire un algorithme capable de calculer un résultat rapidement sans que ce résultat soit trop mauvais. C'est grâce à ce long travail que nous avons maintenant des logiciels qui reconnaissent assez bien les visages des gens, des assistants de conduite qui restent plutôt bien sur la route et des programmes qui jouent plutôt bien aux échecs.

Il s'agit donc de faire des approximations : pour obtenir un résultat correct la plupart du temps, il faut accepter aussi que la machine va de temps en temps donner un résultat faux. Car si nous écrivons un programme qui ne fait jamais d'erreur, les mathématiques nous disent qu'il faudra un temps de calcul quasi infini.

Les chercheurs en IA ont donc renoncé au fantasme de la perfection que dénonce Albert Camus et ils l'ont fait en toute conscience.

« Le silicium ne se trompe jamais »

Pourtant, on entend souvent dire que « le silicium ne se trompe jamais ». Qu'est-ce que cela veut dire ?

Ce qui réalise les calculs dans un ordinateur, donc ce qui exécute le programme, est appelé « microprocesseur » (on parle aussi de « puce » informatique). Il s'agit de circuits électroniques gravés sur une couche de silicium. Pour un calcul donné, par exemple 2 712 + 3 418, le résultat que va nous donner le microprocesseur sera toujours correct. Sauf intervention extérieure[6], il n'y a aucune raison que le résultat de ce calcul donné par les circuits électroniques soit faux : chaque calcul fait par la machine dans un algorithme est correct. Autrement dit, le silicium ne se trompe jamais.

Mais pour savoir quel est le meilleur coup à jouer aux échecs ou pour savoir comment conduire une voiture de manière parfaite, le problème n'est

pas que chacun de ces petits calculs soit correct, c'est qu'il faudrait en faire plus que le nombre d'atomes qu'il y a dans l'univers visible ! Donc ce n'est pas possible, même avec des machines qui feraient des milliards de milliards de calculs par seconde.

Pour faire un parallèle avec une question que nous avons abordée dans un chapitre précédent, l'achat de billet sur Internet, il est possible de complètement tenir sous contrôle une opération précise d'achat d'un billet pour une destination donnée avec un tarif donné et un horaire donné. Mais on ne peut pas garantir qu'on pourra répondre à toutes les demandes des utilisateurs. Dans un cas comme dans l'autre, on fait faire un calcul donné, précis, à une machine et elle le fera de manière impeccable. Mais c'est la quantité de calcul, sur le fond d'une attente de notre part, nous les humains, de réaliser une tâche complexe, faite de plusieurs tâches différentes et indépendantes les unes des autres, qui ne pourra pas aboutir à un résultat exact.

Un problème de prise de conscience qui nous incombe

Quel est le degré de lucidité dans notre société par rapport à ce que nous venons de dire ? Nous pouvons supposer que parmi les scientifiques qui font de la recherche fondamentale en informatique, il y a une conscience très claire de ce problème. Mais dans les entreprises, dans les endroits où l'on construit des logiciels à vocation pratique et donc dans le milieu des techniciens qui élaborent des logiciels, des ordinateurs et des systèmes automatiques embarqués, est-ce que la conscience de cette limite du calcul est aussi présente ?

Il faut espérer que les ingénieurs qui écrivent des programmes en utilisant des bibliothèques d'intelligence artificielle sont le plus souvent bien conscients du fait que la machine ne donnera pas un résultat exact à tous les coups, en particulier lorsqu'ils utilisent, comme c'est le cas aujourd'hui, ces programmes d'apprentissage automatique dont nous avons discuté dans le chapitre précédent. Ces ingénieurs diront « mon programme donne un résultat correct dans 98 % des cas ». Leur métier consiste justement à mesurer le risque d'erreur et à en tenir compte lors de la construction de systèmes techniques.

L'erreur de raisonnement, c'est-à-dire la croyance que la machine serait infaillible, pourrait, si l'on s'en tient à cette hypothèse, sembler être surtout le fait du grand public, en raison des attentes qu'il exprime vis-à-vis des systèmes. Mais croire cela serait aller un peu vite en besogne ! Nous avons vu dans le chapitre précédent que c'est également une déduction

managériale : il vaut mieux tenter de tout automatiser et tant pis pour les 2 % restants.

Ainsi, bien qu'il y ait à tous les étages une certaine conscience de la limite par rapport à l'exactitude, un pari est en train d'être fait et accepté communément par l'ensemble des acteurs. Ce pari consiste à dire qu'il vaut mieux de toute façon, pour sécuriser une quantité de résultat maximal (c'est-à-dire pour atteindre 98 % de bonnes réponses), écarter les cas problématiques du système.

Mais en éliminant tous les cas particuliers impossibles à coder à l'avance dans lesdits systèmes, nous nous orientons vers l'élimination des humains de ces systèmes ! C'est là que réside l'erreur fondamentale de notre relation à la technologie : la penser sans y inclure de manière délibérée et systématique les utilisateurs, y compris ceux qui rendent le traitement automatique compliqué, voire impossible.

---

3. Georges Duhamel, *Défense des lettres. Biologie de mon métier*, Poitiers, M. Texier, 1937.

4. Albert Camus, *L'Homme révolté*, Paris, Gallimard, 1951.

5. C'est-à-dire la partie de l'univers située à moins de 13 milliards d'années-lumière de nous, que l'on peut observer avec un téléscope.

6. Il est toujours possible de détraquer une machine : un bon coup de marteau sur un microprocesseur donne généralement des résultats surprenants…

5
Peut-on « épuiser » le réel grâce aux machines ?

« Pour réussir, il ne suffit pas de prévoir. Il faut aussi savoir improviser. »
Isaac Asimov[7]

De même qu'il est tentant de croire que la machine ne peut pas commettre d'erreur, il est tentant de penser que l'infaillibilité est nécessairement et « automatiquement » une bonne chose. Or, dans ce chapitre, nous montrons en quoi ce que l'on pourrait appeler la « faillibilité » humaine, loin de n'être qu'un problème, est une ressource.

Attention aux effets d'annonce

Comme nous l'avons vu dans le chapitre précédent, il y a forcément des erreurs, c'est une conséquence mathématique de ce qu'est un problème d'IA. Face à cette inévitable faillibilité des systèmes, nous observons à tous les niveaux de la société, y compris chez les scientifiques dont les financements dépendent de plus en plus de leur capacité à mettre en valeur leurs travaux de recherche, une tentation de « l'effet d'annonce » : mettons en avant les succès en omettant les limites.

Nous pourrions penser que les entreprises ont cette problématique parce qu'il leur faut bien vivre *via* leur commerce avec leurs clients. Une entreprise qui irait voir ses clients en disant « notre outil marche, mais pas tout à fait » ne ferait probablement pas un bon chiffre d'affaires. Les manageurs semblent donc obligés, dans notre système économique, de prétendre que la machine ne se trompera pas, quand bien même les scientifiques disent : « Non, c'est impossible. »

Pourtant, il est possible d'adopter une démarche plus nuancée. On peut très bien imaginer un manageur ou un leader d'organisation plus subtil et qui ferait passer le discours suivant : « Nous savons que nos machines ne sont pas parfaites ou exactes systématiquement, c'est impossible par ailleurs. Mais nous comptons sur quelque chose de fondamental : c'est que ce n'est pas la machine qui va résoudre votre problème. Ce sont les interactions entre la machine que nous vous vendons et les usagers que vous êtes. » Cela veut dire que ce qui va faire l'efficacité de l'ensemble du système, c'est l'interaction entre la machine et les utilisateurs.

La controverse Perrow-Weick

On pourrait dire ainsi qu'une approche intelligente de l'intelligence artificielle consisterait à réapprendre, de manière lucide, que le bon fonctionnement de quelque système technologique que ce soit n'est garanti qu'en interrogeant la relation de la machine avec les humains qui l'utilisent. Certes, le système fera des erreurs, mais c'est dans son utilisation, dans l'interaction que les utilisateurs vont avoir avec ce système qu'un bon fonctionnement va pouvoir s'opérer, à défaut de perfection.

Dans cette perspective, il est très utile d'évoquer une controverse entre deux chercheurs américains, Karl Weick et Charles Perrow, qui date des années 1980. Charles Perrow publie en 1984 un livre intitulé *Les Accidents normaux*[8]. La thèse du livre est la suivante :

1. Nos systèmes techniques deviennent tellement sophistiqués qu'ils sont de plus en plus intradépendants (*highly coupled* en anglais). Perrow anticipait et comprenait bien l'évolution des nouvelles technologies ;
2. On ne peut jamais épuiser le réel à l'avance, c'est-à-dire que, tôt ou tard, un système technique, par exemple un logiciel, va recevoir des données qu'il ne pourra pas traiter parce qu'il ne sera pas préparé pour. Ces données qu'il ne pourra pas traiter vont jouer le rôle de grain de sable dans le système ;
3. Or, les systèmes sont tellement intradépendants que dès qu'il y aura un grain de sable à l'entrée du système, il y aura, par effet domino, une catastrophe du système.

La thèse de Perrow est donc la suivante : nous ferons de plus en plus face à des « accidents normaux » inévitables. L'argument est malheureusement assez convaincant. Il revient à dire : « On va inévitablement vers des accidents qui seront normaux parce qu'on peut prévoir qu'on ne pourra pas tout prévoir et que, par effet domino, il y aura des implosions des systèmes. »

Karl Weick lui répond dans une analyse de l'une des catastrophes industrielles les plus importantes de l'histoire de l'industrie, la catastrophe de Bhopal : « Perrow a parfaitement raison, dit Weick, on va sans doute vers des accidents normaux. Sauf qu'il oublie quelque chose d'essentiel : le facteur humain[9]. »

Le facteur humain

L'argument de Weick consiste à dire que, par sa vulnérabilité même, par sa flexibilité inséparable de sa vulnérabilité, l'humain introduit la flexibilité nécessaire dans les systèmes[10]. C'est-à-dire que la vigilance des utilisateurs

et la vigilance de ceux qui maintiennent les systèmes[11] introduit l'« élasticité » nécessaire au fonctionnement des systèmes en cas d'imprévu ou d'erreur que le système ne peut gérer. Cette vigilance – spécifiquement humaine – introduit la flexibilité nécessaire que l'on fait progressivement disparaître des systèmes tellement ils sont inter et intradépendants.

Cette controverse dit beaucoup quant à notre rapport à l'IA. D'un côté Perrow nous dit : « il y aura forcément un grain de sable » et les informaticiens ne disent rien d'autre lorsqu'ils affirment : « Nous ne pouvons pas tout calculer donc, à un moment, il y aura des choses que nous n'aurons pas calculées et quand cela va arriver le programme va se tromper. » Perrow rajoute : « Comme il y aura un grain de sable quelque part et que tout est de plus en plus technique, forcément, par effet domino, nous allons avoir des catastrophes qui vont se répandre. » Mais Weick nous dit : « L'humain, parce qu'il est faillible et attentif, va voir les erreurs et saura y remédier. Alors tant qu'on garde des humains dans la boucle, les accidents resteront évitables. »

Le robot trader

Malheureusement, la vigilance des humains n'est pas du tout automatique, bien au contraire. C'est à l'occasion d'une analyse de la catastrophe industrielle de Bhopal, justement en 1984, que Karl Weick évoque particulièrement la nécessité de la vigilance humaine qui vient pallier les défaillances des systèmes[12]. Cette difficulté est toujours présente, comme le montre l'exemple à la fois tragique et comique des robots tradeurs sur les marchés financiers qu'il faut régulièrement « débrancher ». Par cupidité strictement humaine, nous fabriquons des robots tradeurs capables d'opérer des interactions financières à une vitesse incomparable par rapport à nous humains. Mais cela crée des catastrophes et donc il faut régulièrement arrêter ces robots traders. Et qui les débranche ? Nous, les humains.

On pourrait imaginer concevoir un logiciel qui observe les interactions et débranche régulièrement les robots traders, mais, là encore, la théorie de la complexité nous dit que c'est impossible. Si un tel logiciel existait, capable de repérer automatiquement et sans se tromper les erreurs des autres logiciels, nous aurions un programme qui calcule une solution exacte au problème du robot trader. Ce qui n'est pas possible à moins de disposer d'un temps infini de calcul ! Le logiciel qui calculerait, à partir d'indicateurs définis par l'humain lui-même, quand débrancher les robots serait lui aussi empreint d'erreurs.

La complexité restreinte

Cette impossibilité de tout calculer peut être abordée de la manière suivante au sujet de la notion de complexité. Edgar Morin a prononcé en 2005 une conférence qui s'intitule « Complexité restreinte, complexité générale ». Son argument consiste à dire ceci : il y a la complexité des systèmes techniques et il y a une complexité beaucoup plus vaste, qui englobe cette complexité des systèmes techniques. C'est celle des systèmes également humains. Or, Edgar Morin inclut dans les « systèmes humains » la notion d'incertitude qui renvoie à quelque chose de capital qui est l'inimaginable. L'inimaginable n'a rien à voir avec quelque chose dont on pourrait calculer la probabilité de réalisation. Quand on parle de probabilité de réalisation, on peut parler d'un risque calculable. Mais si on n'imagine même pas ce dont on parle, on ne peut *a fortiori* pas calculer quelque probabilité de réalisation que ce soit. Cela tient de quelque chose d'« infigurable » à l'avance.

Ainsi, nous pouvons dire que l'humain est fait d'infigurable. Mais attention : l'infigurable ce n'est pas seulement des catastrophes. C'est aussi la possibilité d'un émerveillement, d'un amour qu'on n'avait pas du tout imaginé, d'une amitié qu'on n'avait pas du tout anticipée. L'infigurable, ce n'est pas seulement le chaos et la crise liée au Covid-19, c'est aussi le surgissement d'une amitié possible. Cela ne tient pas du calcul.

Quand nous expliquons qu'il existe une complexité inhérente au fonctionnement des systèmes, qu'on peut circonscrire par la théorie sur le fond d'une approche mathématique, c'est vrai. Mais cela n'épuisera jamais la complexité du réel. Il y a toujours quelque chose qui déborde nos systèmes techniques. C'est un peu comme si la notion de complexité « restreinte » au sens de Morin concernait le fonctionnement des « objets », alors que la notion de complexité « générale » engage tout autant la complexité du monde – l'environnement, la « nature » – et la complexité propre à l'humain.

L'infaillibilité

Ainsi, il est heureux, en un certain sens, d'admettre que la machine fait des erreurs et s'adosse à ces erreurs pour que cela fonctionne *in fine*. Cela tient d'une lucidité des humains à l'égard de ce qu'ils ont fabriqué, que ces humains soient des scientifiques, des ingénieurs, des managers ou des utilisateurs.

L'erreur fondamentale que nous commettons est de rêver, comme nous l'avons évoqué au chapitre précédent, d'une IA parfaite. Ce n'est pas la perfection qu'il faut chercher, c'est un fonctionnement pragmatique, ajusté aux circonstances de relations entre humains et système, avec la part

d’erreur, autant du côté des systèmes que du côté des humains. Mais loin d’être seulement une défaillance, c’est une chance, car cela permet aux systèmes – toujours socio-techniques donc et jamais seulement techniques – de fonctionner.

Nous retrouvons ainsi les deux points fondamentaux de notre livre : 1) il n’y a jamais en pratique de système seulement technique, cela est à la fois théoriquement faux et absurde et 2) les mathématiques, pas plus que le langage humain, n’épuisent et n’épuiseront jamais le réel. Et c’est une chance !

7. Isaac Asimov, *Fondation,* Paris, Gallimard, 1956.
8. Charles Perrow, *Normal accidents. Living with High-Risk Technologies*, Princeton, Princeton University Press, 1999.
9. Pour la référence exacte, on pourra consulter : « Enacted Sensemaking in Crises Situations », *Journal of Management Studies*, vol. 25, n° 4, juillet 1988, p. 305-317 ; « Reflections on Enacted Sensemaking in the Bhopal Disaster », *Journal of Management Studies,* vol. 47, n° 3, mai 2010, p. 537-550.
10. Cela vaut d'ailleurs sur des plans qui dépassent largement le problème des technologies comme telles, car cela concerne ce que l'on peut sans hésitation appeler la force inhérente à notre humaine vulnérabilité. Mais c'est une autre question.
11. Nous parlions dans un chapitre précédent de la vigilance de ceux qui conçoivent ces systèmes, en imaginant par exemple qu'on interroge les pilotes humains des avions pour élaborer les logiciels embarqués.
12. Bhopal est liée à une défaillance de communication strictement humaine, mais Weick exploite son analyse sur Bhopal pour dire que Perrow a tort si on sait maintenir le degré de vigilance humaine nécessaire.

6
Les machines sont-elles irresponsables ?

L'utilisation de systèmes techniques faisant appel à l'intelligence artificielle amène la société à se poser la question de la responsabilité des décisions prises par la machine. Or, les machines ne sont, au sens propre, responsables de rien. De la même façon que la notion d'intelligence implique l'intériorisation d'une réflexion, d'un effort de compréhension sur le fond de la capacité spécifiquement humaine à interroger le réel (comme nous l'avons vu au chapitre 3), la notion de « responsabilité » est spécifiquement humaine.

Les véhicules « autonomes »

Cette question de la responsabilité est actuellement couramment posée au sujet des véhicules dits « autonomes ». Commençons donc par un détour à leur propos. Que serait un véhicule « autonome » ? Au sens strict, ce serait un véhicule qui se donnerait à lui-même ses propres règles : l'autonomie, c'est se donner à soi-même ses propres règles, sa propre loi (l'origine du mot renvoie aux deux mots grecs *autos nomos*, « soi-même » et « règles »). Or, le véhicule n'est que le récipiendaire des systèmes électroniques embarqués destinés à faire qu'il se « débrouille » pour circuler. Il s'agit ici non seulement d'un abus de langage, mais aussi d'un contresens.

Entendons-nous bien : l'intention d'automatiser au mieux l'évolution d'un véhicule sur une route dans le but d'éviter des accidents, tout comme l'intention d'optimiser la fluidité du trafic, sont d'excellentes intentions. Mais le rêve de faire qu'un véhicule soit totalement « autonome » n'a pas de sens.

Le dilemme du tramway

Cela s'exprime très clairement au travers de ce qu'on appelle le « dilemme du tramway », qui renvoie à une réflexion théorique sur le moins mauvais des choix à faire. Imaginons la situation suivante : un aiguilleur voit un tramway arriver et constate une situation dramatique : sur la voie que doit suivre le tramway il y a un groupe de six personnes qui ne l'entendent pas arriver et que le tramway ne pourra pas éviter. Mais l'aiguilleur a encore le temps d'agir pour diriger le véhicule sur une voie secondaire, actuellement en travaux, sur laquelle un employé travaille à l'entretien du réseau. Si l'aiguilleur décide de modifier la circulation, l'employé est certainement

condamné, mais il sauve les six autres personnes. Faut-il que l'aiguilleur oriente le tramway vers la voie où il n'y a qu'une personne pour éviter un accident avec le groupe des six ?

Il s'agit évidemment d'un dilemme qui n'admet pas de « bonne » solution. Laisser faire l'accident avec six personnes n'est pas acceptable moralement, mais diriger volontairement un véhicule vers une autre personne non plus. Le dilemme du tramway reste donc un cas d'école théorique, imaginaire, mais il illustre bien ce que nous avons vu au chapitre précédent : tôt ou tard, il y aura de l'inattendu, des situations qui ressembleront à des impasses comme l'impossibilité de ne pas percuter des passants. Il faut choisir la solution la moins pire au problème inévitable de la percussion des passants.

Un véhicule qui serait véritable autonome, donc capable de se fixer ses propres règles, devrait alors être capable de prendre une décision face à une situation comme le dilemme du tramway. Or, répondre à cette question suppose une préoccupation que l'on peut qualifier d'éthique et que les technologies ne portent pas en propre. En effet, le véhicule charrie en lui les décisions *a priori* que les humains ont inscrites, soit explicitement en fonction de l'imagination *ex ante* d'une situation réelle, soit à l'aide d'un algorithme d'apprentissage automatique (comme vu au chapitre 2) qui a généralisé la décision à partir de données, elles aussi fournies par des humains. Ce qui veut dire, pour être très clair, qu'évidemment ce choix face au dilemme est encodé à l'avance dans les algorithmes.

Nous faisons face ici à une confusion entre la fabrication d'un logiciel par des humains, qui fera que le véhicule opte pour telle ou telle solution dans telle ou telle situation, et le fait qu'on attribue au véhicule une sensibilité morale. Cette attribution d'une sensibilité morale au véhicule est une erreur qui nous abuse par rapport aux bonnes questions à poser, dont celle de savoir ce que nous voulons anticiper et comment nous voulons l'anticiper, nous les humains et non les machines que nous fabriquons.

L'impossibilité d'encoder le réel à l'avance

Dans la dynamique de préfabrication des réponses est supposée une connaissance *a priori* du réel. Or, le réel ne peut jamais être exhaustivement prévu ou anticipé, on ne peut pas l'épuiser à l'avance, comme nous l'avons vu au chapitre précédent.

Donnons un exemple proche du dilemme du tramway. Imaginons qu'un véhicule arrive dans une situation où il ne peut pas éviter de percuter soit un vieillard d'un côté de la chaussée, soit, de l'autre côté, une femme encore jeune avec deux enfants et un landau. On peut supposer que, dans le landau, il y a un bébé. On peut dire alors dans la délibération, évidemment, le

vieillard a fait son temps, les enfants symbolisent et réalisent l'avenir, il faut les sauver. La femme aussi. Il est évident qu'il faut imposer aux véhicules (donc inscrire dans l'algorithme) de percuter le vieillard.

Mais si le vieillard se révèle être un génie du style de Einstein, il a peut-être encore des choses fondamentales à révéler sur l'univers, son origine, son évolution, son avenir. Il faut sauvegarder si possible le vieillard. Dès lors, tant pis pour la femme et ses trois enfants. Mais imaginons que le génial vieillard soit atteint d'une maladie incurable, connue de lui seul, qui le condamne irrémédiablement dans les prochains jours. Il redevient évident qu'il faut faire en sorte que le véhicule « opte » pour sauver la femme et sacrifier le vieillard. Mais attention, parce que, parmi les enfants de la femme, il y a peut-être un futur tyran. Hitler bébé n'avait pas encore l'allure de Hitler devenu Hitler. Donc, au fond, il vaut mieux percuter la femme et les enfants… Et ainsi de suite.

Que voulons-nous signifier avec cette fiction qui tend à l'absurde, en y soulignant l'impossibilité de prévoir le réel à l'avance ? On ne peut jamais savoir à l'avance qui est qui. Tenter d'inscrire dans un logiciel des préférences morales est peut-être parfaitement possible, mais ce sont en réalité des prescriptions qui tiennent de choix politiques, pas même moraux ou « éthiques » comme tels. Il y a des présupposés moraux, par exemple en faveur de la femme et des enfants, ou alors en faveur de l'homme qui se révèle être un grand chercheur. Mais puisqu'on ne peut rien savoir du réel à l'avance, on sait que de toute façon, il y a un parti pris réducteur de la réalité, dont le véhicule va être simplement le réalisateur. Le véhicule animé par les logiciels biaisés, non pas de manière inconsciente, mais biaisés de manière délibérée pour résoudre ou être censé résoudre le dilemme du tramway, constitue un problème de nouveau strictement humain qui n'est pas du tout lié à la machine en tant que telle. C'est lié à ce qu'on veut lui faire faire dans le contexte de l'intention d'autonomiser de plus en plus le véhicule.

Différents aspects de l'autonomie

Du point de vue des sciences de l'information, il y a deux plans à considérer dans ce qu'on appelle l'autonomie : l'autonomie par rapport à un opérateur humain qui aurait ou non la possibilité d'intervenir (c'est le sens attribué par les ingénieurs lorsqu'ils parlent de véhicules « autonomes ») et l'autonomie par rapport à l'environnement dont nous venons de voir qu'elle n'existe pas puisqu'il s'agit toujours, pour la machine, de réaliser ce que les programmeurs ont pensé pour elle.

L'autonomie vis-à-vis de l'opérateur humain reste toutefois

problématique. On peut en effet imaginer que la machine a produit, par le calcul, une analyse de la situation qui la conduit à préférer une action différente de celle choisie par l'humain. Or, plusieurs cas sont envisageables :

– soit la situation est prévue et correspond à une situation que l'algorithme devrait savoir traiter. Mais la machine n'a aucun moyen de déterminer si c'est l'humain qui a tort, car elle ne peut pas savoir si elle a obtenu une solution exacte à son problème, comme nous l'avons vu au chapitre précédent. Il faudrait alors qu'elle fasse fi de ce qu'a décidé l'opérateur humain, ce qui pose tout un tas de questions que nous avons évoquées dans un chapitre 2 ;

– soit la situation est « imprévue », c'est-à-dire que le programme n'aurait pas dû se retrouver dans cette situation, mais la machine a été programmée pour agir en tout état de cause. Elle ne sait pas qu'elle est en train de prendre une décision en dehors de son espace de compétence, à partir de données qui n'ont rien à voir avec ce qu'elle aurait dû et pu traiter. Se pose alors à nouveau la question de savoir comment « rendre la main » à l'opérateur humain. Car si la machine ne fait pas ce que dit l'opérateur humain, c'est effectivement une machine très autonome par rapport à son opérateur. Mais si cette machine se trompe, cela peut avoir des conséquences dramatiques, surtout si l'humain n'a pas la possibilité de reprendre la main sur la décision.

En d'autres termes, il n'y a pas d'« interprétation » du réel par la machine au sens où elle lui donnerait du sens. Il y a simplement des calculs d'options, mais pas d'interprétation. Lorsqu'une machine prend la « décision » de bouger le volant à droite ou à gauche ou d'accélérer ou de ralentir de manière à ne pas percuter les obstacles qu'elle a repérés autour d'elle, elle ne met pas de sens, ni de morale, dans ses options. Elle ne mesure pas qu'elle est en train d'éviter des obstacles. Elle ne mesure pas qu'elle est en train de protéger les humains. Elle fait juste ce pourquoi le calcul est fait.

Calculer n'est pas comprendre !

Il faut aussi souligner que le dilemme du tramway est un dilemme parce qu'il n'y a pas de solution tierce. S'il y avait un choix qui était possible, une fois informé, ce ne serait pas un dilemme. Autrement dit, du point de vue de l'informaticien : si on était capable de quantifier ce que les mathématiciens appellent l'« espérance d'utilité » dans le fait de percuter plutôt le vieillard ou plutôt la femme et ses enfants, il n'y aurait pas de dilemme. On choisirait l'option qui a la meilleure espérance d'utilité, ce qui traduit en quelque sorte

l'idée d'opter pour l'action pour laquelle la probabilité de faire quelque chose de mal est la plus faible.

Mais parce qu'il est impossible de quantifier cette espérance d'utilité, le dilemme n'a pas de bonne réponse *a priori* indépendamment de choix qui sont donc strictement politiques. Jamais : toutes les actions sont forcément mauvaises et incomparables entre elles. C'est là l'une des difficultés auxquelles les scientifiques font face lorsqu'ils essayent de programmer des machines autonomes. Ils inscrivent dans les algorithmes des choix politiques (légaux ou sociétaux) qui conditionnent les choix du véhicule. Ces choix portent sur des décisions ex-ante par rapport à un réel dont on est conscient qu'on le réduit. On sait qu'il y aura sans doute des erreurs, mais si vraiment on veut un véhicule autonome, il faut bien accepter ces réductions. Une machine autonome, c'est une machine qui doit être capable de faire quelque chose (et on ne sait pas quoi) dans n'importe quelle situation.

S'impose alors une deuxième difficulté : comment faire si on a une machine autonome qui doit remplir une tâche dans des situations que personne n'a pu anticiper ? Tôt ou tard, la machine sera confrontée à une situation de ce type. Face à une telle situation où la réponse de la machine n'avait pas été programmée, nous aurons des comportements imprévisibles. Bien sûr, nous serons capables de les interpréter *a posteriori*. Nous pourrons ainsi dire : « La machine a décidé cela de manière erronée parce qu'elle a détecté telle chose ou parce qu'elle a utilisé tel paramètre de calcul, d'où elle a obtenu des résultats. » Il en va de même lorsque nous analysons des erreurs humaines *a posteriori* : qui ne s'est jamais demandé « pourquoi est-ce que j'ai fait cela ? ». Nous sommes ainsi capables de trouver des explications rationnelles à des comportements visiblement erronés ou absurdes. De la même manière, nous pourrons expliquer les décisions, bonnes ou mauvaises, que prendront ces machines autonomes, même lorsque ces décisions seront sorties de l'espace de calcul prévu initialement.

Dans la conception d'une machine autonome, il y a donc ce renoncement, de la part des chercheurs et des ingénieurs, à la possibilité de tout maîtriser. Et donc il y a implicitement l'acceptation que certaines situations ne seront pas prévues, qui conduiront à des comportements imprévisibles.

Les machines irresponsables ?

Au contraire, le rêve d'encoder à l'avance le réel dans la machine et de pouvoir présupposer qu'elle a un sens moral, en parlant de l'« éthique » des véhicules dits autonomes, en supposant qu'il serait possible de leur indiquer, *via* la programmation, quel choix prendre à l'avance, résulte strictement d'une intention humaine d'épuiser le réel avant qu'il existe.

Il est toujours possible de programmer un algorithme qui prendrait de telles décisions. Les algorithmes de reconnaissance d'image, dans un futur probablement pas si lointain, pourront peut-être identifier suffisamment rapidement le vieillard et la dame et peut-être aller chercher dans une base de données des informations sur ces deux personnes pour décider qui percuter. Tout cela est interdit dans nos pays grâce, entre autres, au Règlement général sur la protection des données (la RGPD), à la vigilance de la Commission nationale de l'informatique et des libertés (CNIL) et de tous ces organismes qui sont justement là pour nous protéger. Mais il n'est pas difficile d'imaginer une dictature où il serait autorisé que la machine accède à l'identité des personnes, aille chercher leur « crédit social », comme ce qui se développe en Chine, et décide qui doit être sacrifié.

Ces décisions, implémentées dans les programmes des véhicules dits « autonomes » mis en œuvre par les algorithmes, relèvent exclusivement de choix d'ordre politique et donc humains. De tels choix n'ont strictement rien de technique ou de dépendant des machines ou de ce qu'on veut appeler l'IA.

Nous nous sommes souvent entendu rétorquer par des ingénieurs : « Oui, mais de toute façon, quand les humains font face à des situations similaires, ils ne décident pas mieux ! » C'est très possible, mais cela ne permet pas de légitimer le fait que, *ex ante*, hors situation, d'autres humains s'arrogent l'idée qu'ils peuvent légitimement planifier à l'avance des décisions qui, de toute façon, sont biaisées, voire n'ont pas de sens, parce que le réel qui sera concerné par cette décision n'existe pas encore.

Il existe un cas d'accident célèbre qui illustre bien cette situation : le vol US Airways 1549 du 15 janvier 2009 au cours duquel les pilotes ont décidé, en urgence, de se poser sur l'Hudson River, ce qui leur a été ensuite reproché et a donné lieu à une commission d'enquête retentissante. L'enquête a montré deux choses intéressantes.

Premièrement, les pilotes n'avaient que quelques instants pour prendre leur décision. Un groupe d'experts qui prétendrait pouvoir faire mieux qu'eux, et donc qui reproche aux pilotes leur mauvaise décision, n'est en réalité jamais dans la situation du pilote. Dire « voilà la décision rationnelle que, moi, j'aurais prise » est de toute façon quelque chose de faux puisque si un tel expert avait réellement été à la place du pilote, il aurait aussi été dans une situation émotionnelle différente, avec un temps de réflexion plus court, et donc il n'aurait probablement pas pris la décision rationnelle qu'il prétend être la bonne.

Deuxièmement, les simulations ont montré que les solutions censées être raisonnables proposées par la tour de contrôle étaient impossibles. Si

l'équipage avait obtempéré, c'était malheureusement un crash assuré. C'est sa capacité d'improvisation qui a sauvé l'équipage et avec eux les passagers.

Les machines ont du mal à « improviser »

Karl Weick, chercheur américain que nous avons déjà évoqué, parle d'improvisation en prenant comme exemple les musiciens de jazz[13]. Le principe est que les joueurs de jazz connaissent des séquences musicales par cœur et sont capables en fonction du contexte rythmique et mélodique que leur proposent leurs collègues de repérer dans leur répertoire de compétences ce qu'ils ont comme possibles qu'ils peuvent inscrire dans le contexte fabriqué par leurs collègues.

Dans le cas de l'Hudson, c'est cette improvisation extraordinaire qui a permis aux pilotes de réaliser quelque chose qui, normalement, n'aurait pas dû être possible. Un avion qui arrive sur l'eau avec les poids et vitesses des avions actuels fait que l'eau a une résistance de béton, donc en principe l'avion se disloque. C'est à l'époque une option considérée comme absolument impossible. Pour poser l'avion sur le fleuve, les pilotes sortent des compétences relatives au pilotage d'un Airbus, car un A320 n'est ni un hydravion, ni un planeur. Il faut donc mobiliser des ressources et les adapter pour les appliquer à un cadre qui y ressemble vaguement. Grâce à des compétences anciennes de pilotage de planeur de la part du commandant de bord, au lieu d'interpréter l'Hudson comme seulement une catastrophe (il est plus que probable que l'avion se disloque), il interprète l'Hudson comme une chance possible (celle de se poser). C'est de l'improvisation au sens de Weick.

Les machines ne savent pas faire cela. Les chercheurs et les ingénieurs développent des solutions dites d'intelligence artificielle « faible ». Ces solutions ne sont pas faibles en elles-mêmes, elles sont même impressionnantes et résultent de fortes compétences de la part de leurs concepteurs. On parle d'IA « faible » par opposition au mythe de l'intelligence artificielle « forte » telle que la qualifie le philosophe John Searle, c'est-à-dire d'une machine qui serait capable de résoudre tout et n'importe quoi.

Il faut bien comprendre que l'intelligence, ce n'est pas seulement gagner aux échecs ou conduire une voiture. C'est aussi savoir étendre son linge ou préparer son petit-déjeuner, ce qui paraît tout de suite moins impressionnant… Les programmes d'IA dite « faible » sont ceux qui savent très bien jouer aux échecs ou très bien conduire un véhicule sur route mais, pour autant, ils ne savent pas tout faire – par exemple aussi étendre le linge ou préparer un petit-déjeuner. Quand on programme un système

d'intelligence artificielle faible, on s'attaque à la résolution d'un problème donné dans des conditions données. Nos machines sont capables d'être très performantes dans des conditions données. Mais quand ces conditions ne sont plus satisfaites, non seulement le programme va très probablement donner une mauvaise solution, mais en plus il peut se mettre à faire vraiment n'importe quoi. Les être humains, eux, sont dotés d'une incomparable plasticité qui leur permet de s'adapter, de transposer un ensemble de réflexes et de connaissances acquises dans d'autres contextes à des situations qu'ils n'ont jamais rencontrées.

C'est un problème sur lequel se penchent actuellement les informaticiens : on voudrait bien pouvoir réutiliser le résultat d'un apprentissage dans un autre contexte que celui sur lequel il a été fait (ne serait-ce que pour économiser des ressources de calcul). Les informaticiens parlent de « généralisation » d'un apprentissage, c'est-à-dire du transfert de ce qui est appris à partir d'un ensemble de données pour une situation précise à de nouveaux problèmes dans des situations qui n'ont pas été montrées à la machine. C'est un problème de recherche difficile qui n'est pas résolu et qui occupera très probablement les chercheurs pendant encore de nombreuses années. Mais de toute façon, il ne se fera pas dans un contexte aussi général que celui que nous venons d'évoquer.

La responsabilité du conducteur

Cela nous renvoie à quelque chose qui devrait être toujours possible : aussi « autonome » que soit un véhicule, il est capital que les gens qui sont dedans, et évidemment ceux qui sont censés le piloter, comme dans le cas de l'Hudson, n'estiment jamais qu'ils n'ont rien à faire.

L'aviation, aujourd'hui, c'est à peu près cela : le pilotage des avions est quasiment totalement automatisé, mais il y a des pilotes humains. Et leur rôle est en particulier de maintenir vive la possibilité d'opter pour une solution qu'il n'est *a priori* pas possible d'encoder dans les logiciels. Cela illustre la raison pour laquelle il n'est pas possible de dire que les machines sont « responsables » des « décisions » qu'elles prennent et qui ne sont, en réalité, que le résultat d'un calcul, sans aucune « improvisation ». La problématique n'est pas d'obtenir chez les machines la même flexibilité que chez les humains – quand bien même nous y arriverions, la question s'impose de savoir pourquoi le faire, et nous y reviendrons au prochain chapitre –, mais de savoir comment concevoir des systèmes qui aident au mieux les opérateurs humains.

De la même façon donc que l'on a dit plus haut que la notion d'« intelligence » artificielle n'a pas de sens, parler d'autonomie des

systèmes n’a pas de sens.

13. Karl Weick, « Improvisation as a Mindset for Organizational Analysis », *Organization Science*, vol. 5, n° 9, octobre 1998, article introductif à un numéro spécial consacré à l'improvisation en jazz et aux organisations.

7
À quoi servent les technologies ?

« Il faut imaginer Sisyphe heureux. »
ALBERT CAMUS[14]

Les technologies sont omniprésentes dans nos sociétés et on peut légitimement se demander pourquoi nous en sommes si friands. Quel est le rêve de l'homme qui se cache derrière ?

Pas d'humanité sans technologie

Il faut à notre époque impérativement et sans cesse garder à l'esprit qu'il n'y a jamais eu d'humanité sans technique. Nous naissons infiniment faibles et vulnérables par rapport aux animaux. La nudité de l'humain est comme doublée par rapport à la nudité animale. Nos corps ne sont pas du tout spontanément dotés des outils de défense qui seraient liés à un environnement donné. Les fauves ont des fourrures et des crocs, des griffes. Les insectes ont des carapaces. Les oiseaux, les plumes et un bec… Nous, c'est comme si nous étions complètement nus. Il est donc vital que nous complétions nos corps par des outils, des instruments, des vêtements que nous fabriquons et aussi, malheureusement, des armes. Ainsi, il n'y a pas d'humanité sans technique. Et l'on peut évidemment généraliser le propos : il n'y a pas d'humanité sans technologie.

Dit comme cela, on peut dire que c'est formidable : nous avons fondamentalement besoin des technologies, elles sont vitales et nous font faire des merveilles. Par ailleurs, sur un plan non plus seulement physique, mais sur un plan qui nous oriente vers notre spécificité disons « intellectuelle », on peut dire ceci. Contrairement aux animaux, nous ne sommes pas faits pour un environnement donné. Alors que pour tuer un animal, on peut se contenter de détruire son environnement sans action directement mortelle à son encontre, les humains sont capables de survivre même lorsqu'on détruit leur environnement[15]. L'humanité est considérablement adaptable et cette adaptabilité fait qu'on ne dépend pas d'un environnement donné. Mais cela fait aussi que nous pouvons rêver à l'infini. Nous imaginons donc toujours des choses plus loin, ailleurs, autrement.

En ce moment, au travers du transhumanisme et de l'expression anglo-

saxonne qui parle d'« augmenter » l'humain, on rêve à une espèce de puissance infinie qui nous serait apportée par les technologies. Les transhumanistes radicaux vont jusqu'à dire que nous serions proches d'atteindre un point « singulier » qui se caractériserait par deux choses : nous deviendrions immortels, nous, les humains, en même temps que l'intelligence artificielle dépasserait l'intelligence humaine.

Les technologies pour atteindre l'infini ?

Il y a tendanciellement une absence de limites dans cette dynamique, dans les possibilités qui seraient offertes par les technologies. Le rêve d'infini a bien existé. D'abord, le rêve d'inventer des technologies qui permettraient à l'homme de tout faire. Icare, en Grèce, rêve de voler. Il se fabrique des ailes, il vole, mais ses ailes sont en cire. Or, il veut s'approcher de la puissance du soleil et la cire fond. Il retombe.

Mais le rêve humain d'infini s'est adossé à d'autres choses que les technologies. Il s'est adossé à des notions comme celle de Dieu ou de plusieurs dieux, la notion de Tao en Chine, qui est une notion de la totalité, donc l'idée d'être en contact avec quelque chose qui est tout ou qui est sans limite a toujours existé. L'humanité est inséparable de cette idée-là. Ce qu'il y a de spécifique de nos jours, c'est qu'on croit qu'on va réussir à réaliser l'infini en quelque sorte, grâce aux technologies.

La question qui se pose alors n'est pas tant de savoir à quoi servent les technologies, puisque la réponse est qu'elles peuvent servir à tout. C'est « que voulons-nous faire des technologies ? » C'est encore ici notre rapport aux technologies qui est en question, et non les technologies elles-mêmes. La bonne question n'est pas de savoir à quoi elles servent. C'est « à quoi voulons-nous qu'elles nous servent ? »

La fiction de l'infini

Prenons un exemple simple. Supposons que nous devenions immortels grâce aux technologies, ce qui semble peu probable – on voit, malheureusement, combien la mortalité nous caractérise, même si on parvient à vivre plus âgé. Si l'on pouvait devenir immortel, cela aurait des effets existentiels totalement destructeurs. Imaginons qu'on puisse se dire « je vis infiniment ». Ce que je pourrais faire maintenant pour jouir de la vie, puisque j'ai l'éternité devant moi, je le reporte à demain. La notion souvent utilisée de façon ironique de « procrastination » deviendrait complètement inséparable de l'humain. Même si j'ai pris naissance un jour, si je deviens immortel, ma naissance n'a plus de sens.

Cette fiction théorique sur l'immortalité nous renvoie donc à la question

du sens et au fait que la vie n'a de saveur que parce qu'elle a un début et une fin. En rêvant à tous ces possibles que les machines pourraient faire devenir réalité, finalement, ce qui est en question c'est notre humanité, c'est ce que c'est qu'être un être humain. Le plus souvent, cette question est à la fois sous-jacente et inconsciente. Le titre du livre du philosophe Jean-Michel Besnier le dit très bien *Demain les posthumains*[16]. Le rêve de fabriquer des « posthumains » est souvent d'ailleurs accolé au rêve de se débarrasser de ce qui fait l'ambivalence, l'ambiguïté, la douleur de l'homme, de l'humanité.

Nous sommes complexes, comme aime à le dire Edgar Morin. Il y a en nous le pire et le meilleur. On se laisse par ailleurs aisément fasciner par le pire, en croyant qu'il n'y a que le pire à venir parce que nous les humains ne serions que « mauvais » ou méchants. Cette croyance, cette hypothèse, n'est pas fondée, ne serait-ce que parce que, dès qu'un humain se lamente que les humains sont méchants, il y a le signal que l'on n'est pas que méchant : on aspire au sens, à la bonté, à la beauté, aux saveurs des choses, etc. Et c'est bien le cas : nous ne sommes jamais que « méchants » – ni d'ailleurs que « gentils ». Nous sommes faits des deux, et de méchanceté et de gentillesse[17]. La vraie vie est faite des deux. La vie est tôt ou tard souffrance autant que joie. Et en philosophie, on le sait très bien, c'est-à-dire que le sens de la vraie joie est inséparable du sens de la tragédie. Celui qui l'a dit peut-être le mieux est le philosophe Nietzsche. Il dit très bien cela quand il parle d'ailleurs de « gai savoir ». C'est savoir que la vie est tragique, mais se réjouir tout de même quand il y a de la véritable joie.

Le cas du bouddhisme est très intéressant dans cette perspective. Les bouddhistes ont une approche de l'infini tout à fait radicale. Ils constatent la chose suivante : vivre, c'est tôt ou tard souffrir. La vie humaine est faite de souffrance et de joie, c'est comme ça. Les bouddhistes insistent sur la souffrance en disant « on souffre tout le temps, mais en fait on souffre pourquoi ? Parce qu'on est toujours dans le désir : on veut toujours quelque chose qui soit mieux que ce qu'on a ou qui soit « juste bien », comme on dit. Et donc la vie, c'est du désir. Et le désir, c'est l'élan vers ce qu'on n'a pas. Or, quand on souffre et qu'on tend vers ce qu'on n'a pas, qu'on n'a jamais eu, qu'on n'aura peut-être jamais, on est fondamentalement en manque, en porte à faux et en souffrance. Et les bouddhistes disent que tant qu'on désire et tant qu'on n'a pas vécu une vie sage où on arrive à abolir le désir, c'est-à-dire le manque, on va se réincarner. Et l'idéal pour les bouddhistes, c'est de parvenir à se débarrasser à tel point du désir, donc de tout manque, qu'on ne se réincarne plus. L'idéal, c'est de se satisfaire de ce que l'on est et de ce que l'on a à tel point qu'on meurt complètement sage.

Et mourir complètement sage, cela implique d'atteindre le paradis, qu'ils appellent le Nirvana, pour ne plus jamais se réincarner. Tant qu'on meurt malheureux, c'est-à-dire tant qu'il y a encore du désir en nous, il y a réincarnation en un autre vivant. Alors que si on atteint la sagesse, qu'on meurt comme, peut-être, Emmanuel Kant, en disant « c'est bien » (*Es ist gut*), que l'on peut entendre comme « j'ai fait, ce que je devais, je suis satisfaite ou satisfait, tout va bien », alors en mourant, disent les bouddhistes, on sera tellement sage qu'on va aller au paradis, c'est-à-dire dans le néant. On va enfin tomber dans le néant. On ne retombera plus dans l'être qui est tout autant saturé de désir. On ne se réincarnera pas. Voilà la satisfaction absolue.

De la même façon, mais sur un plan qui n'a plus rien à voir avec une forme de sagesse intérieure, imaginer un monde où les technologies feraient tout, absolument tout à notre place, c'est rêver à un monde dont nous serions exclus. Nous n'aurions plus qu'à mourir, idéalement en ne se réincarnant nulle part, ni jamais. Nous serions enfin débarrassés de l'effort que représente le fait de vivre.

La vie, c'est le chemin

En fait, toute philosophie ou sagesse ou religion a pour but, en quelque sorte, l'élimination ou le dépassement du chemin, c'est-à-dire le paradis dans les religions monothéistes que nous connaissons, du côté de l'Occident. Les sagesses également, dont l'aboutissement est une forme d'« ataraxie », c'est-à-dire un état où l'on n'a plus d'émotion, ou en tout cas où on ne se laisse plus emporter par elles. Un état d'âme totalement apaisé. On est complètement satisfaite ou satisfait et on dépasse alors le plan de la souffrance et du manque. Si on pense cependant qu'on pourrait, par les technologies, parvenir à un état de cet ordre, c'est une erreur. La « neutralité » propre aux machines n'est pas l'ataraxie rêvée par les Anciens. Elle présuppose que nous pourrions nous défalquer de nos émotions avec le secours des machines. On est peut-être en train de le faire, mais c'est alors en ravalant l'humain au rang de nos machines, non pas en élevant au plus haut l'humanité dont nous sommes pourtant toutes et tous capables.

Il est intéressant de noter qu'au tout début de l'intelligence artificielle, les chercheurs se sont appuyés sur le fonctionnement neuronal, ou ce qu'on en savait à l'époque, pour tenter d'imaginer certains algorithmes (qui sont devenus, plus tard, le *deep learning* si populaire aujourd'hui). Il s'agissait de reproduire informatiquement des mécanismes biologiques. De nos jours, la tendance inverse apparaît : nous tendons à étudier le cerveau humain à

partir de ce que nous savons des ordinateurs. Nous sommes en train de laisser se déployer une conception de l'humain qui ravale l'humain à l'état de machine. Nous sommes en train d'oublier que l'humain, ce n'est pas l'absence totale de sagesse, ni non plus l'accomplissement de la sagesse censé nous transformer en Dieu ou ce qu'on voudra. Notre humanité est faite du chemin. Notre humanité, notre vie, c'est d'essayer. La vie est une tentative. Paul Valéry dit dans *Le Cimetière marin* : « Il faut tenter de vivre ! »

Il est intéressant d'inverser le propos. Vivre, c'est tenter. Il faut vivre de tenter. On essaye sans cesse quelque chose et on peut échouer. Mais l'on peut aussi réussir. Nos joies et nos peines sont faites de cette dynamique d'essayer de vivre. Et là, cela a du sens : nos vies prennent sens dans ces cas-là. Ce n'est pas l'aboutissement qui compte, c'est le chemin lui-même.

Et les émotions ?

Outre ce que l'on vient d'en dire, la question des émotions mérite un détour spécifique. Si un aspect essentiel de la vie humaine est d'avoir des émotions, on peut se demander si quelque chose qui n'a pas d'émotion est un être humain. Cela nous amène à une question fréquente au sujet des machines : est-ce que les machines, un jour, pourront aussi avoir des émotions ?

Il s'agit là d'un cas très intéressant de personnification des machines, sous-tendu par la volonté d'une créature à l'image de l'homme, comme le monstre de Frankenstein. Nous rêvons de machines qui seraient comme nous, qui penseraient comme nous, qui agiraient comme nous et qui seraient capables, comme nous, de souffrir, d'avoir des émotions, de désirer. Donc de vivre.

Sur le plan scientifique, un travail considérable est fait aujourd'hui sur la réalisation de programmes capables de détecter (à partir de la voix ou d'images) et de reproduire (à travers des voix synthétiques ou des personnages animés) les émotions humaines. Ceci quand bien même on peut affirmer que les machines ne peuvent pas avoir d'émotion pour une raison bien simple, c'est qu'elles n'ont pas de corps. Or, les émotions passent par le corps, elles s'enracinent dans les sensations, perceptions, sentiments que nous avons à l'occasion des relations que nous entretenons avec le monde, avec nous-mêmes, avec les autres. L'émotion désigne un état physiologique qui est une réaction à notre environnement. L'ordinateur, lui, ne peut pas avoir peur que son programme se trompe, que son électricité n'arrive plus ou que le calcul puisse ne plus fonctionner. L'ordinateur fait le calcul qu'on lui a demandé de faire. Il le fait bien, mais c'est tout ce qu'il fait, et il le fait

sans émotion. Les machines peuvent simuler des émotions mais elles ne les éprouvent pas : elles ne peuvent pas devenir humaines.

Il n'en demeure pas moins que s'impose la question de savoir à quoi ça sert. À quoi sert d'apprendre à des machines de simuler des émotions ? Premièrement, pourquoi est-ce qu'on voudrait des machines qui simulent les émotions ? Au Japon, il y a de remarquables robots, dont l'apparence est celle d'animaux de compagnie, pour aider les personnes âgées à se sentir moins seules. Cela fonctionne bien car la culture japonaise admet la possibilité d'une âme dans les machines. En donnant au robot une apparence animale ou humaine, les machines deviennent attachantes. Même si elles n'ont pas de réelle émotion, elles expriment des émotions.

Ici encore, lorsqu'on attribue des caractéristiques humaines aux machines, lorsque l'on parle précisément ici des « émotions » des machines, nous parlons en réalité de nous-mêmes et de notre besoin d'avoir des contacts affectifs avec le vivant.

À quoi servent les émotions dans les machines ?

Introduire des émotions dans les machines répond non seulement à notre envie de fabriquer un être à notre image, mais cela a aussi des applications concrètes. N'oublions pas que l'être humain, par nature, est un animal politique, donc social, comme l'a dit Aristote. Utiliser au quotidien des machines qui ne sont pas capables d'exprimer des émotions, c'est aussi pour nous une source d'insatisfaction. À mesure que la machine prend de plus en plus de place dans notre société et que nous interagissons avec des artefacts (qui font du calcul : ce sont des technologies), nous avons de plus en plus besoin et envie que ces artefacts soient capables d'exprimer des comportements sociaux et des émotions.

Il y a même des résultats surprenants : une étude menée au CHU de Bordeaux auprès de patients atteints de troubles du sommeil a montré que la présence d'agents conversationnels sociaux, par opposition à une simple interface texte pour le suivi de patients, a un impact réel sur la diminution des troubles du sommeil. Dans ce cas, pouvoir disposer de programmes capables d'exprimer des émotions est une amélioration de la condition humaine !

Il y a donc localement, et dans des circonstances précises, des vertus à ces dispositifs. Mais il y a aussi un danger, celui de ne plus chercher entre nous, humains, l'émotion. Au fur et à mesure que montent en puissance les artefacts que nous sommes capables de fabriquer, nous oublions de nous contacter les uns les autres, c'est-à-dire que nous devenons de plus en plus fermés sur nous-mêmes. Nous prenons de moins en moins le risque de

rencontres véritables. La rencontre véritable avec un humain est toujours risquée parce qu'on peut se tromper, on peut entrer en conflit alors que l'on croyait entamer une amitié, etc. Toute rencontre au sens fort peut aboutir à une merveilleuse relation d'amour, d'amitié, de travail ou à une catastrophe. Et par conséquent, derrière ce détournement en direction des machines, il y a quelque chose comme ce qu'on constate actuellement : « Je ne veux recevoir que du positif et je ne supporte plus la contradiction. »

La question que cela pose est donc de nouveau : à quoi servent les technologies ? C'est finalement de dire « jusqu'où est-ce qu'on peut aller dans cette direction ? » Ne risque-t-on pas d'atteindre finalement un monde dans lequel il n'y aurait que des machines qui expriment, comme nous, des émotions les unes avec les autres en étant positives, et dans lequel finalement il n'y aurait plus de place pour l'humain puisque les êtres humains n'interagiraient plus entre eux[18] ? Si c'était le cas, ce serait évidemment la fin de notre humanité.

Il n'y a pas de machine sans humanité

On peut remarquer sur ce point que présupposer qu'il faut des machines qui simulent des émotions car les émotions entre humains deviennent insupportables aux humains, c'est être d'emblée enfermé dans une pétition de principe. Celle selon laquelle il vaut mieux, décidément, éliminer l'humanité, trop insupportable à la fois « en soi » et pour elle-même. Ceci, au profit – fait éminemment paradoxal ! – des machines que cette même humanité a été capable d'imaginer, de rêver, d'inventer, de fabriquer et de faire fonctionner…

Une telle situation démentirait l'affirmation que nous sommes des animaux sociaux. Quelqu'un qui voudrait n'avoir que de l'assentiment autour de soi, donc n'interagir qu'avec des machines, s'isole, c'est du solipsisme ou c'est de l'autisme, c'est du narcissisme, c'est de l'enfermement et il n'y a plus du tout de contact avec une réalité qui pourrait lui dire non.

Le mythe de Narcisse illustre parfaitement cela : nous risquons de nous isoler des autres par amour pour un artefact qui fait semblant d'imiter nos désirs. Dans le cas de Narcisse, l'artefact est son miroir d'eau. Dans notre cas, ce sont des machines qui ont été programmées pour imiter des comportements sociaux qui ne sont pas réels, mais qui satisfont nos désirs sans nous contredire. Ces machines, et c'est important de bien l'avoir en tête, sont faites par des humains pour des humains : elles restent en tout état de cause à notre service. Mais dans le contexte que nous abordons ici, il est capital de garder à l'esprit qu'« être à notre service », pour quoi que ce soit

(ou qui que ce soit), implique de pouvoir tôt ou tard nous contredire de manière inattendue, parce que la contradiction nous « réveille » et l'on ne peut pas savoir à l'avance à quel propos et dans quel contexte une contradiction sera fertile. Une contradiction attendue ne fera que continuer ce qui est déjà en train de se jouer, mais parce qu'elle nous surprend, la contradiction inattendue nous éveille à une véritable appropriation de nos vies.

Cela suppose de rencontrer tôt ou tard une situation qui ne nous convient pas d'emblée, qui n'est pas le seul reflet de ce que nous adorerions spontanément. Cela, les machines ne peuvent pas le fournir, à moins d'y introduire une forme de « hasard », mais alors les « contradictions » n'auraient aucun sens. Notre miroir ne serait brisé que de façon chaotique, et non sur le fond d'une contradiction sensée.

14. Albert Camus*, Le Mythe de Sisyphe,* Paris, Gallimard, 1942.
15. La crise du climat met sans doute en question de manière limite cette observation qui était jusqu'à aujourd'hui à la fois fondée et évidente…
16. Jean-Michel Besnier, *Demain les posthumains, Le futur a-t-il encore besoin de nous ?*, Paris, Hachette, 2009.
17. On pourra sur ce point consulter la dernière partie d'un ouvrage de l'un des auteurs : Laurent Bibard, *Phénoménologie des sexualités. La modernité et la question du sens*, Paris, L'Harmattan, 2021.
18. C'est comme cela que se termine, par exemple, le film *Her* de Spike Jonze.

8
L'IA est-elle dangereuse ?

« Rien n'est réel sauf le hasard. »
PAUL AUSTER[19]

On se pose souvent la question de savoir si l'intelligence artificielle est dangereuse. Le propos que nous tenons depuis le début de ce livre revient à dire que l'intelligence artificielle n'est pas dangereuse par elle-même. C'est bien l'usage qu'on va faire de l'intelligence artificielle, et plus généralement des technologies, qui peut être un usage mauvais, destructeur ou dangereux. Cela dépend de ce que l'on attend que l'intelligence artificielle fasse.

Du mésusage des technologies

Prenons l'exemple de la vidéosurveillance à des fins politiques, comme cela se pratique par exemple en Chine, où les algorithmes d'IA permettent de reconnaître les individus pour les pénaliser (en réduisant leur crédit social) lorsqu'ils ont des comportements que la société réprouve. Selon le point de vue que nous adoptons, cet usage de l'IA est soit néfaste, soit vertueux. Les dirigeants politiques chinois vous diront qu'avoir une population qui se comporte bien, c'est bien pour la société dans son ensemble et donc c'est une bonne chose. Au contraire, dans notre vision d'Occidentaux attachés à la liberté individuelle, l'idée qu'on surveille tous les individus nous révolte et donc cette surveillance est jugée comme un acte directement malveillant. Dans ce cas, on peut émettre un avis moral sur l'usage qui est fait de l'intelligence artificielle et de ce que cet usage représente comme danger pour la société, en fonction de nos convictions éthiques, mais ce n'est pas l'IA qui est dangereuse par elle-même.

Toutefois, il existe des cas où l'usage de l'intelligence artificielle peut se révéler dangereux, et ce de manière plus insidieuse…

L'enfer est pavé de bonnes intentions

L'un de ces cas réside dans la fabrication des profils quand on circule sur Internet. En effet, lorsqu'on navigue sur les réseaux, il y a une accumulation de données par le biais des *big data* sur les personnes qui surfent. Vont ainsi être fabriqués des profils et, ainsi, des attentes, en tout cas des attentes supposées. Si un écrivain va régulièrement voir quel type de stylo plume il

peut s'acheter pour son travail et qu'il ou elle est très intéressé par les stylos Mont-Blanc, au fur et à mesure, les offres qui seront envoyées par le biais de l'identification de la personne qui surfe vont concerner les stylos plume, éventuellement de luxe.

Autrement dit, la fabrication progressive de l'identité des utilisateurs d'Internet va avoir pour conséquence que les utilisateurs vont recevoir des propositions d'information de plus en plus conformes à ce qu'ils ont déjà fait par le passé. Ainsi, en circulant sur Internet, si eux-mêmes n'introduisent pas de la créativité ou d'autres attentes, lorsqu'ils retournent surfer, ils vont être progressivement réduits, rangés, identifiés à une certaine catégorie de profils.

Dans un premier temps, on peut dire que c'est une bonne chose parce que les personnes vont bien recevoir des informations à propos d'objets ou de contenus qui les intéressent et c'est facilitant. Mais en même temps, c'est enfermant au sens où, potentiellement, si on n'y prend pas garde, on va être progressivement absorbés par notre propre passé, par nos expériences déjà vécues, par ce que nous recevons d'informations réputées pertinentes, intéressantes, utiles, mais seulement à partir de ce que nous avons fait par le passé. Pour des personnes qui ont un regard critique, une capacité de prise de distance, d'interrogation, des personnes déjà mûres, habituées à réfléchir, autonomes, qui pensent « par elles-mêmes », ce n'est pas grave parce que nous réenclenchons sans cesse, dans ces cas-là, l'ouverture des possibles.

Mais pour certaines personnes qui n'ont pas l'éducation nécessaire pour avoir un regard critique sur le réel, le risque est qu'au travers des premières expériences sur Internet, elles soient progressivement enfermées dans une identité donnée contraire à ce qui fait notre humanité. Car il est de la nature humaine de ne pas avoir de « nature », c'est-à-dire qu'il y a une part de l'humain qui est fondamentalement tournée vers le projet, vers l'imprévisible, vers l'inconnu. Et cela fait toute la capacité de création, toute la créativité, la capacité d'improvisation et d'invention qui nous caractérise. En construisant des profils à partir des données du passé, les technologies entravent cette capacité et compromettent l'avenir au sens fort (à-venir).

Et ça, c'est un très gros problème provoqué par les technologies, mais ce n'est pas du fait de l'intelligence artificielle. Cela résulte de la manière dont nous fonctionnons commercialement pour identifier les usages et les attentes et les intérêts des acheteurs potentiels, des consommatrices et consommateurs potentiels et donc pour leur offrir ce dont on est à peu près sûr, puisqu'ils l'ont aimé à quelques reprises, qu'ils vont peut-être l'aimer encore. C'est ce qui se joue, qu'on le veuille ou non, dans ce qu'on appelle « l'expérience client ». Fabriquer une expérience client, c'est enfermer le

client dans ce qu'il a déjà été par le passé et potentiellement compromettre son avenir.

Une question éminemment politique

Ces deux exemples que nous venons de voir, le cas de l'identification faciale en vidéosurveillance et celui de la publicité sur Internet, sont des cas qui tiennent de ce qu'on peut appeler le politique, au sens où la vie politique est faite d'évidences. Ainsi, dans la vie de tous les jours, l'on tient pour acquis que l'on connaît les personnes et leurs profils. L'élément de la vie collective spontanée n'est pas celui du questionnement, du doute, de la prise de recul et de l'écoute. C'est l'élément des évidences à partir de quoi l'on élabore tout projet.

Donnons un exemple simple. Dans un pays, on sait à première vue qui sont les élites et qui sont le contraire des élites, les gens qui n'ont pas les moyens. On fait des statistiques et des calculs pour connaître les différentes classes sociales, les niveaux de revenus, les profils. On sait qui a des positions dominantes, qui a des positions plutôt de dominé, d'inféodation par rapport au pouvoir.

La vie collective est ainsi saturée d'identités tenues pour acquises, identités dont le sociologue Pierre Bourdieu a montré combien elles sont construites[20]. Mais une fois qu'elles sont stabilisées dans des habitudes, ce que Bourdieu appelle l'« habitus », dans les modes de relations sociales, nous nous retrouvons enfermés dans des rôles et dans des manières de faire.

Quand un gouvernement, pour garantir la sécurité et la paix civile, organise délibérément l'identification faciale des personnes pour repérer qui se comporte bien ou qui se comporte mal, cela revient de manière active et répressive à faire quelque chose qui se fait spontanément dans des groupes, dans des situations politiques qui ne sont pas des situations politiques dictatoriales. C'est-à-dire que même dans des régimes politiques censés être libres, le collectif s'adosse tôt ou tard à des identités présupposées, stables et fixes. Et cet adossement est éventuellement déterminant pour la marge de manœuvre des différentes personnes appartenant aux différentes catégories sociales. C'est le même phénomène qui se produit aujourd'hui dans les réseaux sociaux sur Internet.

Un phénomène archaïque

Ce phénomène n'est évidemment pas nouveau : pour s'en tenir à il y a quelques décennies, les gens lisaient *Libération* ou *Le Figaro* selon qu'ils étaient plutôt de droite ou plutôt de gauche. Il est d'autant moins nouveau que même le cas extrême d'un gouvernement autoritaire comme l'est le

gouvernement chinois qui dit à la population voilà ce qui est bien pour la vie collective et, par conséquent, voilà ce qui est bien pour vous, est malheureusement tout à fait classique. Autrement dit, nous sommes en train de revenir à un point que nous avons abordé au tout début du livre : l'affirmation par un petit nombre de dirigeants que « nous savons, nous, gouvernants, ce qui est bien pour vous, les gouvernés ».

Mais il ne s'agit pas ici des seuls gouvernants « politiques ». Même quand ce n'est pas délibéré, mais qu'on utilise les *big data* pour repérer les goûts des personnes dont le profil se fabrique sur Internet, nous avons le même résultat : celui de pouvoir dire qu'on finit par savoir ce qui est bien pour les gens. Puisque la fabrication des profils fait qu'on va progressivement, par la mécanique des algorithmes, leur renvoyer des informations dont on a cru constater, d'après les calculs, que ce sont celles qui les intéressent. Le point commun entre une réduction délibérée, comme en Chine, ou involontaire, mais *de facto*, comme sur Internet, est qu'on fabrique, qu'on le veuille ou non, une identité réductrice qui tend à provoquer deux choses : réduire quelqu'un à une identité ou à un mode de comportement, du fait qu'il a déjà eu lieu, ou le réduire parce qu'une autorité gouvernementale le souhaite.

Dans tous les cas, le risque est celui d'une sédimentation des identités. C'est un problème qui n'a pas d'âge, qui n'a pas de lieu. C'est un problème éternel de l'humanité. La vie collective, d'une manière ou d'une autre, nous réduit à des rôles, à des identités pré-jugeant de ce que l'on doit faire ou de ce que l'on fait. Or, ce manque de possibilités d'avancer dans des mondes autoritaires, cela s'appelle de la dictature et de l'impossibilité d'être libres. Et cette impossibilité d'avancer dans un monde qui n'est pas dictatorial, mais qui provoque le même effet, revient à l'altération de l'imagination créatrice tout court, à petite et à grande échelle.

Un problème d'éducation

Le risque principal de cette « dictature » délibérée ou *de facto* des algorithmes est que ces identités enferment la société. À commencer par sa jeunesse. Lorsqu'on est jeune, on est encore plus vulnérable à cette dynamique de réduction à une identité. Si l'on est réduit à un profil pré-identifié, il va falloir faire un effort délibéré, volontariste, pour se sortir de ce profil en rêvant à autre chose. Et si on n'a pas la maturité de cette capacité critique, on risque de ne pas trouver en soi-même dans ses propres expériences la possibilité d'aller voir des choses inconnues. Or, le rapport à l'inconnu est une deuxième caractéristique spécifiquement humaine, la première étant qu'on est fabriqué par l'éducation que l'on reçoit, et donc par toutes les informations que l'on perçoit.

Ainsi, les informations qu'on reçoit sur Internet font partie, pour le meilleur ou pour le pire, de l'éducation. Nous sommes ainsi progressivement modelés par les contenus que l'on reçoit, d'où que ce soit, par des humains, par Internet ou par d'autres supports. On peut dire, d'une part, que nous, humains, sommes faits de cette capacité fondamentale à nous réduire à quelque chose que nous avons déjà fait par le passé, à ce que nous avons appris. Mais d'un autre côté, nous sommes la possibilité tout aussi fondamentale de nous jeter vers l'inconnu. Être capables de nous jeter vers l'inconnu, donc vers l'avenir au sens fort, fait spécifiquement notre humanité. L'on ne sait pas où l'on va, mais on y va quand même. C'est un aspect essentiel de notre humanité, qui est menacé par la puissance des algorithmes qui ne nous renvoient que ce qu'on a déjà fait.

Le philosophe Leo Strauss dit quelque chose de très intéressant à ce propos : si quelqu'un a comme alternative exclusive un Big Mac chez McDonald's et une aile de poulet de Kentucky Fried Chicken, si c'est cela le choix, et qu'il ou elle ne connaît pas ce que c'est, par exemple, qu'une poêlée de champignons en persillade ou du céleri en salade ou encore un avocat aux crevettes, il ou elle ne pourra pas même rêver à une gastronomie autre que celles de McDonald's ou du Kentucky Fried Chicken : il va être enfermé dans le choix.

Cet exemple souligne la responsabilité de ceux qui donnent accès à l'information. Si l'exemple de Leo Strauss peut paraître secondaire, léger et comique – mais quand on voit les effets catastrophiques de certains produits alimentaires sur la santé, ce n'est pas si comique que ça –, il montre que ceux qui ne font connaître qu'une chose du réel à leur population sont responsables de la réduction du réel, effet de leurs opérations. Or, c'est bien ce que font les réseaux sociaux lorsqu'ils nous emprisonnent dans notre propre passé. Et c'est ce que fait toute dictature.

Sortir de la caverne

Cette analyse nous renvoie au mythe premier de la philosophie : les algorithmes nous font retourner dans la caverne dont Platon s'est efforcé de nous sortir. Si l'on n'y prend pas garde, ils contribuent à fabriquer un monde sans question, sans recul, sans capacité de doute ni d'ouverture à l'inconnu, à l'inimaginable, à la créativité.

Ce n'est donc pas l'intelligence artificielle qui est dangereuse, c'est notre naïveté devant nos attentes par rapport à elle et le fait que l'on soit toutes et tous susceptibles d'être dupes de mécanismes aveuglants. Ce problème est complètement archaïque : c'est un problème éternel, celui de la liberté par rapport à toute vie collective. Cela est évidemment intensifié par la

puissance de l'IA. Mais cela n'est pas modifié qualitativement, sauf à dire que c'est qualitativement plus intense.

Mais alors, pour rester sur l'image de la caverne platonicienne, la difficulté est que c'est aux individus, pour s'en dégager, de prendre conscience de cet enfermement dans lequel ils risquent d'être pris avec l'IA. Comme le souligne Platon, les gens qui essaieraient de faire sortir de la caverne leurs collègues se feraient rejeter par ceux-ci, qui n'ont pas envie qu'on leur dise qu'ils sont dans une caverne, qu'ils observent des ombres. Il faudrait peut-être que les systèmes eux-mêmes disent en quelque sorte « attention, je ne suis qu'un mirage ! ». Mais ce serait là jouer le jeu d'une déresponsabilisation supplémentaire des humains.

Éveiller les consciences

Ce livre a pour but de permettre de prendre de la distance. Cela veut dire que nous jouons d'une manière ou d'une autre, au travers de cet essai, le rôle de ceux qui disent : « Attention, ne soyez pas dupes. » Et peut-être que tout le monde n'appréciera pas une réflexion comme celle que nous sommes en train de mener.

Prenons l'exemple des GAFAM et de la fabrication délibérée, en particulier pour augmenter leurs profits, de systèmes qui nous enferment. Dire aux gens « n'utilisez plus ces technologies » comme nous semblons le faire est forcément source de rejet, car ces technologies nous apportent un confort de vie auquel nous ne voulons pas renoncer, pas plus que les habitants de la caverne ne veulent, pour voir d'où viennent les ombres, s'éloigner du feu.

Cet enjeu est celui de l'éducation. C'est donc un enjeu politique au sens le plus noble possible : un gouvernement ne peut pas ne pas s'interroger sur la manière dont il conçoit l'éducation. La conclusion vers laquelle ce chapitre se dirige est donc assez différente de celle que nous avions dans les chapitres précédents : ce ne sont pas seulement les managers ou les techniciens qui vont devoir changer leurs discours et arrêter de dire que les machines sont omnipotentes, ce sont bien évidemment tout autant les politiques qui doivent former les citoyens. Mais il s'agit aussi des citoyens parents d'enfants jeunes ou moins jeunes. C'est en fait la responsabilité de toutes et tous. Nous sommes tous responsables de nous demander : voulons-nous que nos concitoyens soient aliénés et soient des troupeaux de moutons ? Ou voulons-nous qu'ils soient libres et vivent leur vie ?

La question est valable à tous les niveaux. C'est l'ensemble du système éducatif depuis la plus petite enfance, jusqu'à l'enseignement supérieur, qui doit aujourd'hui s'emparer de cette question. Nous concevons volontiers

notre travail dans ce contexte-là comme un livre d'alerte. Il y en a d'autres, sans aucun doute.

Un dernier exemple pour la route

Un ami d'un des auteurs avec son épouse ont deux enfants. Les enfants ont six ans et quatre ans et les parents se demandent : « Mais comment on va faire avec les écrans et comment on va arriver à ce qu'ils ne soient pas fascinés ? » Le père est très débrouillard, il n'est pas informaticien de formation, mais il l'est devenu par expérience et avec le temps. Il a pris la décision de mettre les enfants devant un téléphone portable et devant un ordinateur. Il leur montre images, vidéos, etc. Les enfants expriment à grands cris leur éblouissement : « Wow, wouah, super ! » Il arrête les deux machines, il les démonte complètement. Une fois démontées, il dit aux enfants : « Vous voyez, ce n'est que ça. » Puis il remonte les appareils et les remet en route. Le couple a laissé les enfants utiliser les machines quand ils voulaient. Jamais les enfants n'ont été victimes des écrans.

Sans doute y a-t-il également dans cette famille une culture particulièrement ajustée. Mais l'effet a été que les enfants utilisent très bien les téléphones portables et les ordinateurs, pas du tout de manière addictive.

---

19. Paul Auster, *Cité de verre*, Arles, Actes Sud, 1987.
20. Voir Pierre Bourdieu, *Esquisse d'une théorie de la pratique*, Genève, Droz, 1972 et Pierre Bourdieu, *Le Sens pratique*, Paris, Minuit, 1980.

9
Est-ce que les technologies détruisent des emplois, polluent et asservissent l'homme ?

« La science n'est jamais qu'une succession de questions conduisant à d'autres questions. »
TERRY PRATCHETT et STEPHEN BAXTER[21]

Il est régulièrement fait aux nouvelles technologies le reproche d'être nuisibles à l'être humain[22]. On entend beaucoup dire des intelligences artificielles, et plus généralement des machines, qu'elles vont détruire des emplois, qu'elles polluent énormément, qu'elles vont asservir les êtres humains. Notre souhait dans ce chapitre est de reprendre ces points et de les discuter au regard de la connaissance scientifique et de la compréhension qu'il faut en avoir pour répondre du mieux possible à ces inquiétudes.

La machine qui remplace l'humain

Commençons par les emplois. Que les machines remplacent les humains dans le travail, ce n'est pas nouveau, c'est même pour cela qu'elles sont construites. Le principe d'une machine, c'est d'utiliser de l'énergie, de la décupler et de permettre de réaliser des choses qu'un humain ne pourrait pas faire ou de le faire plus rapidement et mieux. Pour ne nous en tenir qu'à cette époque, les métiers à tisser de la Renaissance permettaient de faire des vêtements plus vite que les ouvrières au point de croix dans leur foyer. Évidemment, des gens qui faisaient ce métier se retrouvèrent face à une concurrence complètement déloyale : une machine qui le fait mieux qu'eux, du moins suivant les critères économiques.

Avec l'arrivée de l'intelligence artificielle, cette problématique se pose pour des métiers qualifiés qui, jusqu'ici, semblaient préservés, car faisant appel à l'intelligence humaine[23]. Si, demain, les voitures dites « autonomes » peuvent circuler sans contrainte sur nos routes, tous les emplois de chauffeurs VTC et taxis, qui sont importants, peuvent être amenés à disparaître.

L'utilisation de technologies, IA ou autre chose, conduit nécessairement à la disparition de certains emplois. Mais elle conduit aussi à la création d'autres emplois : il faut des gens pour concevoir ces systèmes, outils ou intelligences artificielles, pour les faire fonctionner, pour les entretenir. Et

elles permettent indirectement l'apparition de nouveaux métiers rendus possibles par les produits de ces technologies, en faisant évoluer la société.

Les technologies ont donc toujours éliminé des emplois parce qu'elles sont justement faites pour remplacer l'homme de manière à éviter de la peine et à augmenter l'efficacité de production éventuellement. Il n'en demeure pas moins qu'avec l'intelligence artificielle, la situation est relativement nouvelle, car il y a une espèce de fascination pour les technologies qui a pour effet, de manière réflexe, de « digitaliser » alors qu'économiquement, ce n'est pas toujours rentable ni carrément pertinent. Croire qu'une nouvelle solution numérique pourra résoudre tous les problèmes est une grossière erreur, comme le montre l'échec de la plateforme numérique de General Electric ou encore le déficit chronique de la société Uber qui a perdu 25 milliards de dollars entre 2017 et 2021. La transformation numérique a un coût qu'il ne faut pas ignorer.

Ce qu'en dit la simulation informatique

Il est intéressant de renvoyer ici aux travaux d'un de nos collègues qui dirige des thèses depuis plusieurs années sur la simulation informatique du marché du travail, ce qui permet d'étudier pas mal de choses sur l'emploi[24]. Il s'est intéressé récemment à la simulation de l'impact de la digitalisation des entreprises dans le marché du travail. Le constat est sans appel : il faut évidemment s'attendre à une augmentation du chômage lorsqu'on remplace les humains par des machines. C'est inévitable, en l'absence de création de nouveaux métiers, il y a une disparition d'un certain nombre d'emplois, malgré des gains en efficacité.

L'avantage de la simulation, c'est qu'on peut fabriquer des mondes qui n'existent pas, pour étudier des hypothèses fantaisistes. C'est ce qu'a fait notre collègue en supposant que toutes les technologies numériques pourraient être produites en France. Les résultats de ses expérimentations montrent que dans ce cas – irréaliste – où tout l'appareil de production des technologies numériques serait situé en France, il y aurait probablement une création nette d'emplois parce que la digitalisation nécessiterait le développement de plein d'entreprises de conception et de construction d'outils numériques. L'intelligence artificielle peut permettre de rêver à un monde meilleur, ou au moins de comprendre ce que nous faisons avec notre monde actuel. En affirmant cela, nous sommes bien conscients que les éléments historiques, d'une part, et de fiction théorique, d'autre part, que nous présentons ici ne suffisent pas à établir un état des lieux sur les effets de l'intelligence artificielle sur l'emploi en général. Nous souhaitons simplement donner quelques éléments que nous espérons utiles pour une

réflexion sur la question.

L'intelligence artificielle pollue et contribue à l'épuisement des ressources naturelles

Dans le contexte de la transition vers un monde où on affronterait la crise du climat avec sérieux, il y a un problème de fond : les technologies sont la source d'énormément de pollution. D'une part, en raison de la consommation énergétique des centres de données et des réseaux informatiques. D'autre part, à cause de l'extraction des métaux rares pour la fabrication des systèmes. Nous sommes littéralement en train d'épuiser les ressources de la planète.

Pour aborder cette question aussi sereinement que possible, regardons les chiffres. Le calcul, le stockage et la transmission des données dans les réseaux sont des activités qui consomment de l'énergie. Et consommer de l'énergie, c'est créer de la pollution : au niveau mondial, 80 % de l'énergie finale est carbonée et donc génère d'immenses quantités de CO_2 dans l'atmosphère qui ne sont pas absorbées par la terre. Il n'y a donc aucun doute, l'utilisation de technologies informatiques, et en particulier de technologies d'intelligence artificielle – qui reposent beaucoup sur le calcul et le stockage de données –, est une source de pollution comme d'épuisement des ressources mondiales. Concrètement, le numérique représente 2,5 % de la consommation finale au niveau mondial (entre 3 % et 4 % en France). Cela peut paraître peu, mais cela reste une source de pollution importante : 3,4 % des émissions de gaz à effet de serre proviennent actuellement du numérique et on s'attend à ce que cela augmente rapidement dans les prochaines années – selon les prédictions, 7,6 % en 2025. On ne peut donc pas faire comme s'il n'y avait pas de problème et les technologies d'intelligence artificielle sont un élément important de l'équation.

L'impact du matériel

Mais en réalité, les trois quarts de la pollution engendrée par les ordinateurs proviennent de leur fabrication. En particulier parce que les ordinateurs sont renouvelés très souvent : un smartphone tous les deux ou trois ans en moyenne, car les gens veulent les dernières versions. Les ordinateurs personnels durent un peu plus longtemps mais, globalement, il y a beaucoup de renouvellements et donc comme c'est renouvelé très souvent et que cela coûte cher à fabriquer en termes de métaux rares, en termes de consommation d'énergie pour la fabrication elle-même, la plus grosse part

de la pollution due au numérique provient en réalité du matériel informatique.

Les centres de calcul constituent un exemple frappant. Il faut imaginer des salles de plusieurs centaines de mètres carrés remplies d'armoires avec des calculateurs.

Tous les quatre ans, il est nécessaire de changer l'intégralité des machines et tout est jeté parce que la technologie a évolué : les calculs se font plus vite avec les nouvelles machines, plus efficacement, et les anciennes machines ne sont plus entretenues, il n'y a plus les pièces de rechange. Il n'est même pas possible de redonner ces « vieilles » machines à des universités ou à d'autres centres de calcul dans des pays démunis sur ce plan, parce que de toute façon les composants nécessaires pour réparer le matériel lorsqu'il tombe en panne ne sont plus disponibles sur le marché.

Nous sommes donc en train de parler de choses qui ont des durées de vie très courtes qu'il faut, si l'on veut suivre le rythme des innovations tel qu'il va, renouveler très souvent. Cela pose forcément des questions qui ne sont pas liées aux technologies directement, mais bien au système économique dans lequel nous sommes : pourquoi est-ce qu'il faut avancer sans cesse et vers quoi ? Où est le moteur de ce mouvement, à quoi sert-il ? Nous sommes au cœur d'interrogations qui portent inévitablement sur le système économique lui-même et ses modèles.

De fait, même si nous avons du mal à l'admettre, la tendance est plutôt technophile dans notre société. Malgré certaines résistances qui s'expriment à travers la demande de « moins de technologie » et « plus d'humain », la tendance générale est que la société se « technologise » de plus en plus, sans forcément réfléchir collectivement aux raisons pour lesquelles elle le fait. Et cela nous conduit à ce genre de situation où l'on renouvelle les machines très souvent et où l'on calcule à tout va, donc en générant énormément de pollution.

L'humain esclave de la technologie ?

Cela nous fait une transition malheureusement tout à fait pertinente pour la troisième question : lorsque nous observons des usagers de téléphones portables qui veulent systématiquement se tenir au niveau des appareils les plus récents et qui le font sans du tout savoir pourquoi, si ce n'est que c'est à la mode, nous pouvons estimer que c'est une forme d'asservissement. Mais ce n'est pas un esclavage aux technologies, c'est un esclavage hyperbanal qui a lieu depuis toujours. C'est un asservissement à la mode, qui est, pour plagier le grand philosophe Spinoza, le degré le plus bas de la socialisation. S'asservir aux modes revient à s'asservir à ce que croient, à ce

que font, à ce que désirent les autres, au collectif pour le collectif – coûte que coûte. Nous ne sommes pas du tout sur une question technologique, mais bien sur une question fondamentalement, strictement, humaine.

Il faut bien distinguer, dans ces outils de calcul qui coûtent très cher et qui polluent, d'une part, les calculs qui sont faits dans un but collectif de progrès pour la société – que ce soit par les chercheurs ou par des entreprises qui essayent de travailler pour le bien commun –, d'autre part, les usages individuels qui peuvent parfois relever de cet esclavage moderne. La mode et la diffusion des photos de chatons sur les réseaux sociaux est très coûteuse sur le plan écologique, pour un gain économique qui ne repose que sur les recettes publicitaires. Lors de la pandémie de Covid-19, les centres de calcul ont permis de simuler et de comprendre les modes de diffusion de la maladie, de suivre l'évolution des variants, de manière indéniablement efficace. Dans les deux cas (la publicité et la recherche sur le Covid-19), tout cela est rendu possible par les algorithmes d'intelligence artificielle. Ce sont donc aussi ces algorithmes, dont nous critiquons l'usage sur les réseaux sociaux, qui nous ont permis de sortir de la pandémie en deux ans, avec un nombre de victimes qui, malgré, tout reste limité – dix fois moins, par exemple, que la grippe espagnole du début du XX^{e} siècle. Grâce aux technologies, nous avons su relativement bien maîtriser une situation potentiellement catastrophique. Cela s'est fait avec ces mêmes centres de calcul très polluants que nous critiquions précédemment.

L'individu et le collectif

Il faut donc faire la différence entre la photo de chaton qu'on va mettre sur Internet et des technologies mises à disposition dans l'intérêt collectif du bien commun, pour problématiser la question du rapport aux technologies sur le plan directement de la philosophie morale et politique. Cela montre encore ici que, par elles-mêmes, les technologies ne posent pas de problème. C'est le rapport que nous entretenons avec elles qui pose problème, par rapport à nos attentes, par rapport à la conformation à des groupes, par rapport à la mode, etc. Comme nous l'avons déjà souligné, la notion d'éducation est décidément capitale.

La question de l'asservissement est éminemment une question politique plutôt que directement technologique. La technologie en tant que telle ne recèle aucun bien, ni aucun mal. Il s'agit avant tout d'un point d'équilibre à trouver entre l'individu et le collectif.

Pour nous aider à approcher de façon pertinente la question, nous pouvons dire ceci. L'histoire a très probablement une structure fractale, au sens où certaines problématiques structurantes se présentent à différents niveaux, les

niveaux les plus petits, « locaux », et les niveaux les plus grands ou globaux. Un exemple par excellence de cette dynamique est la problématique de l'individuel et du collectif dans l'exemple de Socrate qui veut penser librement et à qui Athènes dit : « Non, il faut que tu arrêtes. » Et cela pour des raisons de fond, ce n'est pas du tout de l'arbitraire – bien qu'il y ait eu la condamnation à mort de Socrate pour avoir vécu et pensé trop librement. Dès Platon, il y a une compréhension très profonde de l'enjeu : il y a une tension entre ce que représente « Athènes », à la fois comme cité et comme symbole de tout collectif, et ce que représente « Socrate », à la fois comme homme réel et comme symbole de toute vie et pensée libres.

Cette tension se joue à tous les niveaux, au niveau de Socrate à Athènes, d'un côté, et à des niveaux mondiaux, de l'autre. Nous la retrouvons dans notre rapport aux technologies, parce que l'enjeu, c'est que nous restions libres par rapport à nos productions et que nous ne devenions pas prisonniers des systèmes techniques – à la fois du point de vue de la fascination de nos imaginations et de nos vies quotidiennes – et que nous arrivions à trier au cas par cas, à faire un repérage de la pertinence au cas par cas de l'utilisation que nous faisons et voulons faire des technologies. Et que cela ne devienne pas ce que c'est déjà trop souvent : un réflexe compulsif de vouloir la technologie pour la technologie. La technologie n'est pas une fin, elle est un moyen.

La place des sciences modernes

On peut, comme on vient de le voir, problématiser la question sur le plan de la philosophie morale et politique, mais on peut aussi la problématiser sur le plan que l'on peut qualifier d'« épistémologique », c'est-à-dire qui concerne la solidité de nos savoirs. Il y a sur ce sujet une réflexion tout à fait fondamentale d'un philosophe peu connu qui s'appelle Jacob Klein. À titre d'exemple, Jacob Klein observe qu'à partir de la Renaissance, l'élaboration de la physique mathématique sur la base de l'algèbre conduit les sciences à prendre inévitablement pour but leurs propres méthodes. Il écrit à peu près ceci : « Jamais les anciens n'auraient pris la méthode comme but[25]. » Il veut dire par là que l'élaboration des sciences modernes a comme adossement sous-jacent l'idée que la méthode est un but. L'idée que le mode de fonctionnement est le but, en quelque sorte, de la recherche. Et s'il y a du vrai dans ce qu'il dit, il est capital de savoir s'en dégager pour n'utiliser la recherche que comme un moyen d'un but autre que la recherche elle-même, qui n'est pas faite pour se servir elle-même, mais bien pour servir la vie en société, le bien commun.

Cela se transpose dans nos vies concrètes, dans la vie sociale, au travers

de ce qu'on appelle en sociologie le fonctionnalisme. Le fonctionnalisme c'est une manière d'approcher les organisations qui identifient de manière très convaincante qu'une organisation risque toujours de se prendre pour son propre but. Il y en a un exemple remarquable dans le tout début du film *Vol au-dessus d'un nid de coucou* de Stanley Kubrick, avec Jack Nicholson. La scène montrée par Kubrick au début du film est révélatrice de la difficulté que nous abordons ici. Elle a lieu à un moment où des patients dans un hôpital psychiatrique sont censés prendre du repos. Et l'on voit que le personnel de soins, très subtilement, au lieu d'apaiser les patients, les énerve. Ainsi, les soignants sont de nouveaux utiles : on appelle même les brancardiers pour mettre Nicholson dans sa camisole de force. Nous en rions car c'est un film, mais cela est malheureusement tragiquement vrai, et arrive plus que fréquemment tous contextes confondus : au lieu de faciliter le fonctionnement de la structure, de se mettre au service des usagers, les acteurs agissent de manière à rester utiles, c'est-à-dire à ce que leur fonctionnement soit confirmé dans sa pertinence. Autrement dit, on entretient le problème, parce qu'on en est la réponse. Et il se peut que le problème que l'on entretient ait perdu de sa pertinence, voire soit devenu caduc, n'ait aucun sens, voire n'en ait jamais eu…

Des distorsions des systèmes de cet ordre représentent un véritable problème dans notre monde. Et les technologies en sont un rouage de plus en plus central. Elles deviennent un but en soi pour améliorer les capacités de calcul des ordinateurs, de l'IA, etc., indépendamment de l'intérêt que cela puisse apporter aux humains, à la société, au bien commun.

La machine à l'image de l'humain ?

Nous sommes donc, à nouveau, face à un problème fondamentalement humain. Une des manières intéressantes de le poser, c'est de dire que quand on se réfugie dans des théories des systèmes, en n'admettant pas qu'il y a une responsabilité des individus et donc en présupposant que les humains et les individus sont noyés dans les systèmes, on déresponsabilise tout le monde, en jouant *de facto* le jeu d'une humanité noyée dans les systèmes, dans tout système. Les théories qui n'insistent que sur l'aspect systémique des fonctionnements contribuent à nos aliénations, au fait que nous devenons des rouages dans des machines. Nous devenons les moyens des machines parce que nous nous représentons que nous sommes dominés par les systèmes.

Plus nous clamons notre « noyade », plus nous la favorisons. Si nous maintenons notre sens de la responsabilité et avec lui de la possibilité de prendre conscience que les systèmes fonctionnent comme ils fonctionnent et

que leurs fonctionnements peuvent être contestables, discutables, au meilleur sens du terme, alors nous nourrissons la possibilité de nous en écarter, de ne pas en dépendre sans réflexion. De fait, rien qu'en en prenant conscience, nous sommes déjà dehors. Il y a donc une urgence à sortir de la présupposition que nous sommes noyés dans les systèmes et dans le collectif.

21. Terry Pratchett et Stephen Baxter, *La Longue Terre*, Nantes, L'Atalante, 2013.
22. C'est un point que nous avons déjà abordé au début du livre pour mettre en avant la fausse tension entre « technophilie » et « technophobie ».
23. Nous invitons nos lecteur à se demander si la broderie n'est pas, elle aussi, une activité nécessitant de l'intelligence.
24. Pour nos lecteurs anglophones : Jean-Daniel Kant, Gérard Ballot et Olivier Goudet, « WorkSim: An Agent-Based Model of Labor Markets », *Journal of Artificial Societies and Social Simulation*, vol. 23, n° 4, 2020.
25. « It is noteworthy that the idea of procedure as a goal in itself was totally excluded from Greek science. » Jacob Klein, « Le rationalisme moderne », *Lectures and Essays*, Annapolis, Saint John's College Press, 1985. Ces « Essais et conférences » donnés en anglais n'ont pas encore donné lieu à une traduction complète en français.

10
Nature et liberté : les machines nous libèrent-elles vraiment et, si oui, dans quelle direction ?

Au chapitre 3, nous avons vu que nous, humains, ressemblons aux machines sur le plan de nos compétences réflexes : ce qui fait notre véritable humanité est la relation à la réflexion dont nous sommes faits (capacité de prise de recul, problématisation des données…) à partir même de nos réflexes. Nous avons tenté dans cet ouvrage de conduire les lectrices et lecteurs à prendre la distance critique nécessaire vis-à-vis du rapport que nous entretenons avec les technologies. Cette distanciation demande une compréhension de la relation entre les notions de langage et de calcul mathématique, qui, nous le verrons, se superposent aux relations entre nature et liberté.

Langage et mathématiques : la perte de la notion de nature

Pour comprendre en quoi les relations entre nature et liberté, d'une part, et entre langage et mathématique, d'autre part, sont en écho les unes des autres, il faut remonter à l'avènement des sciences modernes à la Renaissance, au moment de l'humanisme en Europe d'abord, et dans le monde ensuite. Comme nous le verrons, il n'y a de nos jours sur un certain plan rien de fondamentalement nouveau depuis la création des sciences modernes.

L'invention des sciences modernes permet aux humains de vivre comme s'ils se détachaient complètement de la nature, donc potentiellement de ne plus du tout en dépendre. Cela n'a pas été conscient et délibéré dès le début. Le processus est devenu conscient à un moment donné, au travers d'une expression extrêmement célèbre du philosophe René Descartes qui dit que nous allions devenir « comme maîtres et possesseurs de la nature[26] ». Pour maîtriser et posséder la nature, il faut d'abord s'en détacher, donc ne pas en dépendre. L'indépendance, sur le plan scientifique – donc sur le plan de la connaissance – s'exprime par la capacité de détachement des opérations de calcul en général, désormais adossées à un usage systématisé de l'algèbre. Devenant système de notation fondamental, l'algèbre rend tout calcul indépendant de quelque donné que ce soit[27].

Or, cet avènement des sciences modernes comme adossées à un usage systématique de l'algèbre est inséparable de l'utilisation de la connaissance

pour transformer techniquement la nature. Regardons cela de plus près.

L'humain sujet de la nature ou la nature objet pour l'humain ?

Au moment de l'avènement des sciences modernes advient l'idée de « sujet » de la connaissance. C'est nous, les humains, qui sommes les connaisseurs de la nature et nous nous donnons des « objets » pour connaître, pour calculer, pour transformer et satisfaire nos objectifs, pour les réaliser. Progressivement, ces objets vont devenir les données modernes, ce que nous appelons entre autres les « *big data* ».

La nature perd alors le sens d'être une donnée « avant » nous. Ce sont désormais nous, les sujets, qui nous donnons à nous-mêmes des données[28]. La nature va devenir pour nous, les humains, en deux ou trois siècles après l'avènement des sciences modernes, un « ensemble de phénomènes » à disposition de l'homme, donc à disposition des sujets de la connaissance. La nature n'est plus une donnée préalable au sujet : ce sont désormais les sujets qui se donnent les données, donc leurs « objets ».

Ainsi, nous élaborons au jour le jour un rapport qui devient délibérément artificiel ou artefactuel avec la nature. C'est comme si ce que nous appelions autrefois la nature n'avait plus rien de naturel. L'origine du mot « nature » fait référence à ce qui prend naissance spontanément, et donc ce qui a pris naissance avant et indépendamment de nous. Désormais, ce que nous appelons la « nature » est simplement la quantité, l'accumulation des objets que les sujets humains se donnent à étudier. Elle n'est plus constituée que de données ou de matériaux de quelque ordre qu'ils soient, à notre disposition.

Il s'agit partout des mêmes données que celles que nous manipulons aujourd'hui dans les *big data*, celles qu'on doit stocker et ensuite traiter à l'aide d'algorithmes pour les utiliser au service de l'humain. Ces données servent bien le sujet (nous-même) à travers l'objet mathématique qui est aujourd'hui le calcul et l'informatique, mais ces données n'ont justement rien de naturel ! La notion même de nature telle qu'elle était affirmée ou supposée avant la révolution scientifique a disparu. Comment la nature pourrait-elle se réduire dans les techniques uniquement à des données ?

Lorsque nous regardons une pâquerette dans un jardin, nous pouvons voir que c'est quelque chose qui est naturel, qui existe. Dans la spontanéité de la vie quotidienne, avec le langage naturel, il est encore possible de parler d'une pâquerette, bien sûr, mais dès qu'on est dans la posture scientifico-technique advenue il y a 500 ans à peu près, la notion même de nature ne fait plus sens.

Du sujet connaissant au sujet politique : les notions modernes de détachement et de libération

On trouve une situation analogue en ce qui concerne nos relations humaines et notre vivre-ensemble. Le sujet connaissant, qui se donne les objets et donc qui fait que la nature devient seulement une quantité indéfinie de données disponibles pour l'homme, a une autre face, car c'est le même « sujet » qui se libère au cours du temps de l'Histoire. C'est l'humain entendu comme « sujet libre ».

L'homme qui se comprend comme sujet libre prend conscience de lui-même comme ayant été soumis à la nature, d'un côté, et soumis à des dominations, à des régimes politiques dictatoriaux, tyranniques, oppressifs, de l'autre. Et l'Histoire devient l'histoire de la libération du sujet humain, du détachement de toute sujétion. Le sujet se libère des sujétions politiques. Et nous sommes exactement dans ce courant à notre époque où le rêve est de se débarrasser, à juste titre, de toute hiérarchie verticale que ce soit, qui est vécue comme une domination oppressive.

Cette compréhension des humains comme « sujets » qui vivent une Histoire, et dont l'Histoire est celle de leur « libération » à la fois par rapport à tout donné naturel et par rapport à toute domination disons politique, s'amorce depuis toujours en un certain sens, mais elle prend un caractère décisif au même moment où Descartes écrit cette fameuse phrase selon laquelle nous allons devenir « comme maîtres et possesseurs de la nature ». Cela a lieu en particulier à partir de la pensée d'un philosophe anglais, Thomas Hobbes, qui le premier affirme de manière systématique que les humains ne sont pas, comme le disait jusque-là la tradition, des « animaux politiques ». Pour Hobbes, il faut considérer à l'origine l'humanité comme constituée d'individus tous égaux entre eux, libres et rationnels. Ceci, tout à fait indépendamment de leur sexe, de leur âge, de leur origine ethnique, etc. C'est exactement la définition par l'économie des « agents économiques ».

Ces humains sont, pour différentes raisons que Hobbes présente en détail, en particulier dans son *Léviathan*, amenés à établir un « contrat social », qui définit les conditions du vivre-ensemble. Vivre ensemble, depuis l'avènement de la pensée politique de Hobbes, qui est à la base de notre manière de nous rapporter au monde, n'a rien de « naturel ». D'« animaux politiques », comme l'a dit le premier Aristote, qui faisait que toute vie politique avait une origine dans la « nature » des choses, nous sommes passés à l'état d'individus libres, égaux entre eux et rationnels, qui créent par convention (le contrat social) les conditions du vivre-ensemble.

On a ici, sur le plan du vivre-ensemble, donc de nos relations humaines et de l'action, la même dynamique que sur le plan de la relation à la nature non humaine et de la connaissance. Or, la manière de connaître la nature non humaine devient le soubassement de la manière de nous rapporter aux choses humaines. Et finalement, à mesure qu'on étudie avec une approche scientifique, c'est-à-dire une approche expérimentale, la nature comme série de données, approche par laquelle on pose des hypothèses, on les expérimente et on essaye de voir si les hypothèses réfutables s'avèrent non réfutées par les observations, nous nous éloignons de la nature.

Et dans le même temps, cela nous en libère. Et le « sujet » scientifique qui se détache de la nature est le même que le « sujet libre » et se libérant de l'Histoire. On arrive alors, sur les deux plans de la connaissance (« sujet » de connaissance des « objets ») et de l'action (le « sujet libre » se libérant de toute oppression et, idéalement, de toute hiérarchie), à cette situation un peu contradictoire où l'on est à la fois libre de la nature tout en en étant pourtant issus, mais incapables de nous y inscrire, que ce soit sur le plan de la connaissance ou de l'action. Dans le contexte de la nécessaire « transition écologique » dont il est tant question désormais, inséparable d'un renouveau de la manière de vivre ensemble, cette constatation est à la fois plus qu'importante et problématique.

Liberté et question du sens : que voulons-nous faire de notre liberté ?

Ainsi, si nous admettons que le sujet connaissant (donc scientifique) est le même que le sujet qui se libère au cours de l'histoire, l'avènement des mathématiques modernes et l'avènement de la notion moderne d'histoire et de libération de l'homme au sens générique du terme sont inséparables. Les mathématiques modernes et la notion d'histoire moderne sont les deux faces d'une même médaille. Les mathématiques symbolisent la capacité pour l'humain de se détacher de la nature, comme l'histoire symbolise la capacité pour l'humain de se détacher de toute dépendance, en particulier morale et politique.

Dans ce contexte, la notion de liberté, par quoi l'homme entendu génériquement s'oppose à la nature, est décisive. Elle est à l'origine de toute notre culture. Or, on peut observer que cette liberté que nous revendiquons a des effets délétères à bien des égards. À commencer par les dérèglements climatiques : la « maîtrise et possession » de la nature revient, peut-on soupçonner maintenant, à la destruction de ladite nature. En tout cas à la destruction des conditions de possibilité de la vie de bien des espèces, dont potentiellement l'humanité. Et nous avons vu au travers plusieurs exemples que les relations humaines que nous entretenons les uns avec les autres sont

loin d'être seulement améliorées par l'invasion des technologies dans le champ de la vie collective, de la vie politique au sens noble du terme.

Se pose maintenant impérativement la question au sujet des technologies : la liberté, oui, mais jusqu'où et pourquoi ? Nous avons vu en particulier les effets contradictoires de notre liberté de calcul dans le domaine politique, introduisant une capacité inédite de contrôle des populations par l'entreprise, à travers l'exemple de la reconnaissance faciale. Ne gagnons-nous pas à réapprendre que nous prenons naissance dans un monde qui nous préexiste, que nous existons dans une nature qu'on appelle maintenant l'« environnement » ?

Il est intéressant de remarquer ici que l'on ne parle pas de « nature », on parle d'« environnement ». Et on parle de « crise du climat », de « crise de la planète », mais on évite le mot « nature ». Et malgré tout, on est bien sur une planète qu'on est éventuellement en train de faire étouffer. Et nul doute que si notre planète étouffe, nous étoufferons avec elle. Par-delà notre goût immodéré pour la liberté, nous ne pouvons pas ne pas poser la question : « La liberté, oui, mais que veut-on en faire ? Quel sens a-t-elle ? »

Or, poser cette question, c'est le faire dans l'élément essentiel non pas de la physique mathématique adossée à l'algèbre, c'est le faire dans l'élément essentiel et spécifiquement humain du langage. Nous sommes dans l'horizon du mot grec *logos*, et non pas dans celui du mot latin *ratio*. Poser la question du sens ne se fait pas sur le plan du calcul (de quelque donnée que ce soit, de nos intérêts, etc.). Poser la question du sens se fait sur le plan du langage, de la révélation du monde et de la question de nos désirs. On appelait cela autrefois en philosophie la question des « fins ». C'est-à-dire la question de savoir ce que nous voulons faire de nos vies. La question du sens qu'a de vire ici et maintenant.

Cette question, nulle machine ne la posera jamais, parce que la question même du sens n'a aucun sens pour une machine, qui ne connaît que des données, et qui est faite pour les combiner selon certaines règles qui lui sont toujours, en dernière instance, données par nous, humains.

Si nous voulons que nos vies aient du sens, si nous voulons ne pas nous laisser envahir par nos propres créations artificielles, si nous voulons être capables de passer à nos enfants un monde qui soit monde et non pas fait que de calculs absurdes, nous devons savoir prendre de la distance et avec le calcul comme tel, et avec notre revendication d'une liberté infinie. En contrepoint des seuls calculs et d'une infinie liberté, ce n'est pas la contrainte qui émerge : c'est que nous sachions de nouveau, de manière responsable et heureuse, pourquoi nous aimons vivre.

26. René Descartes, *Discours de la méthode*, partie VI.
27. Un des ouvrages les plus décisifs sur ce point est celui de Jacob Klein, *Greek Mathematical Thought and the Origin of Algebra,* Cambridge, MIT Press, 1968 (écrit à l'origine en allemand, le livre n'est pas encore traduit en français). On pourra consulter en français un article intitulé « Signification et vérité dans les écrits lofgico-mathématiques de Jacob Klein », traduction par Carlos Lobo d'un article de Burt C. Hopkins, « Meaning and Truth in Klein's Philosophico-mathematical writings », *Methodos, Savoirs et textes*, vol. 9, 2009.
28. Combien de fois n'avons-nous pas entendu en cours de mathématiques, pour celles et ceux qui ont suivi leur scolarité jusqu'à ce stade, « donnons-nous » (un repère, « a », « un ensemble » etc.) ?

Conclusion

« L'homme passe infiniment l'homme. »
PASCAL[29]

Ce qu'on a rencontré au cours des chapitres précédents, donc de toutes les thématiques abordées, c'est que les machines par elles-mêmes ne présentent ni aucune vertu, ni aucun vice particuliers. Nous employons exprès un vocabulaire moral : les machines n'ont pas de morale par elles-mêmes. En revanche, elles sont saturées de morale au travers de nos attentes, au travers de nos espoirs, au travers de nos usages, au travers de nos mésusages, au travers de toutes les difficultés qui sont relatives à ce que nous attendons d'elles ou fantasmons à leur propos.

On peut donc dire qu'il n'y a pas de problème des technologies. Il y a un problème relatif à nos attentes par rapport aux technologies, à nos présuppositions par rapport aux technologies, à l'idée que nous nous en faisons et à la question que nous posons ou que nous ne posons pas de savoir ce qu'on veut d'elles – et ce que nous voulons de nous-mêmes. Et le point dominant et saillant qui est apparu, c'est que nous nous posons actuellement beaucoup trop peu la question des fins, de l'objectif que l'on veut atteindre en développant les nouvelles technologies et en particulier l'intelligence artificielle. En particulier quand nous avons discuté au chapitre 8 la question de savoir si les technologies polluent, nous avons évoqué le fait qu'il y a un mouvement permanent d'innovation des logiciels qui demandent qu'on modifie aussi la partie matérielle (*hardware*) des ordinateurs et que cela a un coût considérable d'utilisation de matières premières. La question est : « Pourquoi faut-il le faire et après quoi courons-nous ? »

Cette question de savoir après quoi l'on court, qui est une question de sens, déborde très largement les technologies. Elle est liée à la question de la liberté que nous avons discutée à la fin de l'ouvrage. La liberté est devenue un but pour elle-même et on perd de vue la question : être libre, oui, mais pour quoi faire ? Et ce « pourquoi ? » pose la question du sens. Plus nous nous libérons, ou croyons nous libérer, plus nous devenons potentiellement totalement dépendants des « systèmes » que nous créons et que nous avons créés. Bien sûr qu'il faut se libérer de toute oppression, sur ce principe on peut supposer que toutes les lectrices et tous les lecteurs seront parfaitement

d'accord. Mais se libérer de la nature au point de la perdre de vue, cela revient à se libérer de notre sol, de notre assise. Or, jusqu'à nouvel ordre (biotechnologique), personne n'a pris naissance de manière seulement technologique. Nous recevons le corps que nous sommes et nous recevons l'environnement où nous vivons et donc il y a à réfléchir à nouveaux frais la question de savoir ce que nous, humains, voulons faire de notre liberté. Dans le contexte actuel de la crise du climat, on s'aperçoit de plus qu'on ne peut pas faire comme si la nature n'existait pas, comme si nous n'étions nulle part et comme si nos ressources ou les moyens de nos désirs étaient sans limite. Nous sommes dans un rapport d'utilisation et non pas un rapport, également, d'écoute et de reconnaissance tout simplement envers la nature d'où nous venons.

Cette poursuite de la liberté, dans une vision idéalisée et coupée de la nature des choses, nous fait croire à des machines sans limite, ce qui est tout aussi absurde. Au contraire, cela conduit à ce que nous devenions les esclaves (libérés ?) desdites machines où nous avons encodé nos choix à l'avance au travers des algorithmes. C'est pourquoi, s'il y a des notions de morale, d'éthique et de responsabilité à réfléchir, elles ne concernent en rien les machines que nous fabriquons. Elles concernent nos souhaits et nos choix quand nous voulons fabriquer des machines, qu'on qualifierait par exemple d'« autonomes ». Lorsque l'on observe combien l'on peut être aliéné par des systèmes techniques – du fait de nos consentements conscients ou inconscients, ou du fait de régimes politiques qui savent très bien instrumentaliser les outils technologiques au profit de leur contrôle sur les populations –, l'on ne peut pas ne pas reconnaître que les questions de responsabilité, d'éthique, de morale, de sens, de politique sont plus que jamais spécifiquement et exclusivement humaines[30]. Tant que nous restons humains ou à la hauteur d'homme au sens générique du terme, ces questions sont donc les nôtres. Et l'homme ne « passera infiniment l'homme » qu'autant qu'il gardera le sens de la responsabilité qu'il est tant qu'il vit ici-bas.

Ce sont les humains qui sont responsables des machines qu'ils inventent et non l'inverse.

---

29. Pascal, *Pensées,* Paris, Flammarion, 2015 [1670].

30. Ce qui n'empêche pas de réfléchir aux notions de dignité juridique, par exemple des animaux. Mais c'est là encore une autre affaire.

Retrouvez notre catalogue sur
www.editionsdelaube.com

Pour limiter l'empreinte environnementale
de leurs livres, les éditions de l'Aube
font le choix de papiers issus de forêts
durablement gérées et de sources contrôlées.
Ce fichier a été généré
par le service fabrication des éditions de l'Aube.
Pour toute remarque ou suggestion,
n'hésitez pas à nous écrire à l'adresse
num@editionsdelaube.com

a été achevé d'imprimer en mars 2023
pour le compte des éditions de l'Aube
rue Amédée-Giniès, F-84240 La Tour d'Aigues

Dépôt légal : avril 2023
pour la version papier et la version numérique

www.editionsdelaube.com

Table des Matières