

VLADIMIR ATLANI - VICTOR STORCHAN

Préface de Jean-Michel Blanquer

ANTI-MANUEL D'INTELLIGENCE ARTIFICIELLE

Les nouvelles questions que pose l'IA

● Éditions
EYROLLES

Éditions Eyrolles
61, bd Saint-Germain
75005 Paris
www.editions-eyrolles.com

Depuis 1925, les Éditions Eyrolles s'engagent en proposant des livres pour comprendre le monde, transmettre les savoirs et cultiver ses passions!

Pour continuer à accompagner toutes les générations à venir, nous travaillons de manière responsable, dans le respect de l'environnement. Nos imprimeurs sont ainsi choisis avec la plus grande attention, afin que nos ouvrages soient imprimés sur du papier issu de forêts gérées durablement. Nous veillons également à limiter le transport en privilégiant des imprimeurs locaux. Ainsi, 89 % de nos impressions se font en Europe, dont plus de la moitié en France.

En application de la loi du 11 mars 1957, il est interdit de reproduire intégralement ou partiellement le présent ouvrage, sur quelque support que ce soit, sans l'autorisation de l'éditeur ou du Centre français d'exploitation du droit de copie, 20, rue des Grands-Augustins, 75006 Paris.

© Éditions Eyrolles, 2024
ISBN: 978-2-416-01723-0

Vladimir Atlani • Victor Storchan

Préface de Jean-Michel Blanquer

Anti-manuel d'intelligence artificielle

Les nouvelles questions que pose l'IA

● Éditions
EYROLLES

Sommaire

Préface de Jean-Michel Blanquer	11
Prologue : Le combat du siècle	13
Introduction : Un mot des auteurs.....	17

SECTION 1

Les fondamentaux

Chapitre 1 Qu'est-ce que l'intelligence artificielle?.....	23
Le machine learning, nouveau mode de communication homme-machine	26
D'une révolution technologique à une révolution sociétale	27
Chapitre 2 Le secret du Turc.....	31
Histoire n° 1 : les médecins dans l'arène	31
Histoire n° 2 : le piège du Turc	33
Histoire n° 3 : « George Bush a fait le 11-Septembre »	37
Chapitre 3 Une courte histoire de l'IA.....	41
Étape n° 1 : la genèse	41
Étape n° 2 : le tournant des années 2010.....	42
Étape n° 3 : la révolution ChatGPT	44

Chapitre 4 Les raisons d'un succès	47
Huit traîtres et un prophète (ou l'explosion de la puissance de calcul).....	47
Le sexe chez les adolescents, ou l'essor de la donnée de masse	51
Chapitre 5 Le nouveau paradigme technologique.....	55
Contexte historique	55
La fabrication des modèles de fondation, au cœur de l'IA générative	56
Les jeux de données de l'IA générative.....	59
L'évaluation des foundation models.....	59
Les limites des modèles d'IA : vers une logique de systèmes et d'agents.....	61

SECTION 2

Quatre enjeux de civilisation

Chapitre 6 La fin du travail ?	67
Trois théories économiques	67
La fin de l'ère Schumpéterienne ?.....	69
L'erreur de la prophétie de Marx.....	71
Chapitre 7 La fin de l'Homme ?	77
La dernière vexation.....	77
La machine est-elle devenue artiste ?.....	79
Chapitre 8 La fin des Lumières ?.....	83
L'IA à l'assaut des fraudeurs.....	83
Petite leçon d'algorithmie appliquée.....	84
Le risque « Incas ».....	87

Chapitre 9 La fin du monde ?	91
Colossus, le « Cerveau d'acier »	91
Qu'est-ce qu'une IA forte ?.....	93
Quand la frontière se brouille	94
Deux réponses à la question de la singularité	95
Le mot de la fin pour Alan Turing	97

SECTION 3

La sortie des laboratoires – Questions économiques, stratégiques et sociétales

Chapitre 10 La nouvelle économie de l'IA	101
Contexte historique	101
L'infrastructure logicielle de l'IA générative.....	102
Le continuum de distribution de l'IA générative	103
Les acteurs de l'IA open source et des modèles ouverts	105
Faut-il ouvrir l'IA ?.....	105
Chapitre 11 La géopolitique de l'IA	111
L'IA, une technologie stratégique loin de l'imaginaire des technologies de la guerre froide	111
L'accroissement de la géopolitisation de l'IA.....	112
Entre concurrence, coopération et convergence.....	115
L'enjeu des semi-conducteurs	117
Chapitre 12 Les enjeux de sécurité	121
Les risques liés à l'IA.....	121
La perception des risques liés à l'IA.....	124
Mettre la recherche en IA sur pause	124
Les chercheurs en sécurité de l'IA	125
Pour une gouvernance mondiale des risques de l'IA ?	126

Chapitre 13 L'IA et l'environnement	129
L'impact de l'IA sur l'environnement : quelques chiffres	129
Soutenabilité de l'IA générative.....	130
L'IA au service de l'environnement	132
Chapitre 14 L'IA et la démocratie	135
L'IA générative, au service de la démocratie libérale ?	135
La démocratie libérale, atout clé du développement de l'IA ?	137
La désinformation.....	138
Les biais politiques des systèmes d'IA	138
Chapitre 15 L'IA et la santé	141
IA et biologie : faire avancer la recherche scientifique	141
IA et santé des femmes	143
IA et ophtalmologie	144
L'IA au service des infrastructures de santé à l'hôpital et à domicile	144
Respect de la vie privé et confidentialité des données	145
Les enjeux de la bioéthique	146
Chapitre 16 L'IA et le sport	149
Les données du sport	150
Prédiction et optimisation de la performance sportive	151
L'IA générative pour améliorer l'entraînement sportif.....	152
L'IA au service de la transformation de l'expérience des supporters de sport	153

SECTION 4

Solutions et perspectives

Chapitre 17 Réguler l'IA	159
Approche américaine.....	159
Approche européenne.....	161
Approche chinoise.....	162
Grands enjeux de la régulation de l'IA.....	163
Chapitre 18 Le défi du travail	169
Chapitre 19 La France, puissance de l'IA ?	173
Le risque de déclassement.....	173
Accompagner la diffusion de l'IA dans l'économie française.....	174
Accompagner la diffusion de l'IA sur le lieu de travail.....	174
Le nouvel écosystème français de l'IA générative.....	176
Devenir un leader de la sécurité de l'IA.....	177
Chapitre 20 Quelle attitude adopter face à l'IA ?	181
Accompagner le changement technologique.....	181
Développer et utiliser de façon éclairée de la technologie.....	182
Accepter l'incertitude pour agir.....	183
Mobiliser ses atouts pour inventer sa propre voie.....	183
Inventer ses propres récits et narratifs.....	184

Conclusion Du besoin d'éducation.....	187
Principe n° 1 : revaloriser les compétences nécessaires (sans langue de bois)	188
Principe n° 2 : consacrer l'importance du savoir pour le savoir, du raisonnement pour le raisonnement	190
Principe n° 3 : former en circuit court	191
Principe n° 4 : utiliser l'IA pour apprendre	191
Principe n° 5 : éduquer aux enjeux de l'IA.....	191

Préface

La révolution de l'intelligence artificielle transforme aujourd'hui notre monde. Nous sommes à la veille de bouleversements majeurs, politiques, économiques, culturels et sociaux, provoqués, amplifiés, accélérés par les progrès gigantesques et rapides accomplis par les chercheurs et les entrepreneurs, parfois sous l'impulsion ou le contrôle des États, eux-mêmes parfois décisifs et parfois dépassés par cette course stratégique fondamentale.

Chacun entrevoit les progrès et les risques que cela implique. C'est le grand mérite de ce livre que de bien les expliciter, sans jamais tomber ni dans la technophilie béate, ni dans le catastrophisme facile. Tout est remis en question car c'est le rapport de l'homme à la connaissance qui est en jeu et, par conséquent, la nature du travail, le contenu du contrat social, les méthodes de la puissance publique.

Ce n'est pas étonnant que, après avoir passé en revue tous les enjeux, de façon très exhaustive et très illustrée, les auteurs concluent par le sujet de l'éducation. Elle est évidemment la première et la plus importante des réponses aux questions de civilisation posées par l'intelligence artificielle.

En tant que ministre de l'Éducation nationale, j'ai dû en prendre pleinement la mesure. C'est notamment pour cela que nous avons créé, à la faveur de la réforme du lycée, non seulement un enseignement commun d'informatique en seconde mais aussi un enseignement nouveau de première et de terminale intitulé « Numérique et science informatique ». Cela s'est accompagné de la création du Capes puis de l'agrégation d'informatique. Autant d'innovations faites pour placer au cœur du système scolaire les enjeux de maîtrise de l'intelligence artificielle.

Mais, comme le soulignent les auteurs, tous les aspects de l'éducation sont touchés. Car l'éducation doit nous permettre de répondre pertinemment à la plus grande question de notre temps : « Comment un monde de plus en plus technologique peut-il être un monde plus humain ? » Face à tous les défis nouveaux, nous devons doter nos enfants de savoirs fondamentaux, de culture générale, de capacité logique et d'un corpus éthique leur permettant de bâtir un monde plus respectueux de l'humain et de la nature.

Vladimir Atlani et Victor Storchan étaient particulièrement bien préparés pour concevoir un tel ouvrage. Tous deux sont d'abord des techniciens : ils ont écrit, conçu, mis en application des algorithmes d'IA et peuvent donc l'aborder très concrètement. Leur expérience de l'entreprise (le conseil en stratégie et la banque privée) leur donne également une approche pragmatique et claire. Ils ont su trouver le recul nécessaire pour rendre captivants le fonctionnement et l'histoire de cette technologie. Ils tracent un chemin juste et équilibré, pour guider le lecteur et lui permettre de se construire une opinion éclairée sur le sujet.

La lecture de ce livre me semble un prérequis salutaire pour toute personne souhaitant contribuer à la construction d'un avenir qui mettrait l'IA au service d'un idéal de progrès, d'égalité et de justice.

Jean-Michel Blanquer

Prologue

Le combat du siècle

«Aucun ordinateur ne peut jouer H5!» Le coup est irréal, impossible. Nous sommes le 10 mai 1997 et Garry Kasparov, champion du monde d'échecs, vient de terminer son cinquième match contre Deep Blue, le logiciel d'intelligence artificielle de la société américaine IBM.

La tension est palpable dans l'Equitable Center, le gratte-ciel new-yorkais qui accueille la rencontre. La centaine d'initiés venus en personne des quatre coins du globe pour admirer le choc en ont eu pour leur argent. Tous savent ce qui se joue. Garry Kasparov est champion du monde. Il est sans doute le meilleur joueur d'échecs de tous les temps. Mais ce jour-là, il est le dernier rempart d'une espèce humaine contre la suprématie de l'ordinateur. Tous les spécialistes l'affirment, le dépassement est inévitable, définitif, programmé. Mais à Manhattan, on croit encore dans les talents d'un héros venu retarder la fin d'un règne. Plus qu'un duel, c'est un combat de civilisation.

Et c'est le deuxième qui oppose Kasparov à la machine d'IBM. L'année précédente à Philadelphie, Kasparov l'avait emporté quatre parties à deux. À l'époque, personne ne mesurait encore l'engouement qu'un tel affrontement devait susciter. Kasparov, visionnaire, avait alors demandé que le match soit diffusé en temps réel sur internet. Devant les réticences des organisateurs du tournoi persuadés que le match n'intéresserait personne, les responsables d'IBM décidèrent de diffuser eux-mêmes la partie d'échecs en ligne. L'intérêt du public fut immédiat. Le site internet, pris d'assaut par des centaines de milliers

d'internautes, fut vite saturé. Heureusement, en 1996, c'est aussi IBM qui assure la couverture internet des Jeux olympiques d'Atlanta. En moins de 24 heures, les développeurs d'IBM migrèrent la diffusion du duel sur les serveurs prévus pour les JO. Les milliers de spectateurs purent ainsi suivre en direct les coups du grand maître et du super-ordinateur. Le match prenait une dimension planétaire.

Il faut dire que la première manche avait surpris tout le monde. Au 23^e coup, l'ordinateur proposa le sacrifice de son pion central... sans aucune compensation matérielle. Pour Kasparov, quelle surprise ! Il n'avait encore jamais vu un ordinateur décider, de lui-même, de perdre une pièce contre un simple avantage tactique. Traditionnellement, les algorithmes d'échecs s'arrangent pour conserver le plus de pièces le plus longtemps possible au cours de la partie... Pas ici. Cette machine-là appartenait à une nouvelle génération, et Kasparov, très perturbé, finit par perdre la manche. Quelques mois plus tard, le journaliste Frederic Friedel, alors membre de l'équipe de Kasparov, raconta les coulisses de la soirée de cette première défaite : « Garry avait été très bouleversé après la partie. Mais [...] après le match, nous avons analysé le jeu. Et nous avons compris. Le mouvement de ce pion n'était pas un simple sacrifice de matériel. Deep Blue avait simplement vu que dix-sept coups plus tard, il regagnerait la pièce sacrifiée. Pour Garry, quel soulagement. » La preuve était faite : la machine n'avait développé aucune sensibilité particulière, aucun sens stratégique, aucune faculté d'intuition. Elle était simplement capable de calculer plus loin. En 1996, Garry Kasparov finira par s'imposer 4 parties à 2. Il déclarera quelques semaines plus tard : « Deep Blue peut jouer comme un grand maître ou comme un enfant. Mais quelquefois, il joue comme Dieu lui-même. »

Retour en 1997. Après la défaite de 1996, les ingénieurs d'IBM ont retravaillé leur algorithme. Deep Blue est devenu Deeper Blue, et tout le monde se prépare à un affrontement terrible. Les montants mis en jeu ont doublé. De 500 000 dollars en 1996, le prize money promis au gagnant dépasse le million de dollars. Les paris vont bon train et, à quelques jours du début du match, c'est Kasparov qui est donné gagnant par les bookmakers. Les 480 places de l'auditorium ont été réservées depuis des mois et le public peut suivre les coups des

deux adversaires sur des écrans géants. Trois grands maîtres d'échecs commentent la partie en direct. Ils sont traduits en simultané dans de nombreuses langues et le match est diffusé en temps réel sur les chaînes de plusieurs pays et sur internet. Le monde entier a pris la mesure de l'ampleur de la confrontation.

Le match se joue en six manches. Chacun des deux adversaires débute une manche à tour de rôle, et c'est Kasparov qui joue les blancs pour la première partie. Aux échecs, celui qui commence dispose d'un léger avantage sur son adversaire, et les attentes sont hautes pour une victoire du grand maître dans cette première manche. Le champion du monde ne prend pas de risques. Il opte pour une partie lente de manœuvres stratégiques dans laquelle il sait pouvoir dominer son adversaire. En face, Deep Blue tente de contrer le plan du Russe par un jeu agressif, mais sa faiblesse est latente, et au 39^e coup Kasparov finit par prendre l'avantage et remporte la première partie.

La deuxième partie est remportée par l'ordinateur après une variante subtile de l'ouverture espagnole. Kasparov abandonne au 45^e coup. Une manche partout, balle au centre. Pour la troisième partie, Kasparov fait le choix d'une ouverture originale et avance le pion central d'une seule case. Il espère emmener la machine loin des schémas classiques de début de partie que celle-ci, il le sait, maîtrise parfaitement. La partie devient vite très complexe et se finit sur une égalité. Chacun prend un demi-point. La quatrième manche finit, elle aussi, sur une égalité. Kasparov : 2, Deep Blue : 2. Le champion du monde sait que la prochaine partie sera décisive. C'est la dernière fois de la confrontation qu'il aura l'avantage des blancs.

Kasparov commence par sortir son cavalier, son coup fétiche. C'est la même ouverture qu'il avait jouée lors de la première manche. C'est aussi le coup qui l'avait fait gagner lors de leur confrontation de l'année précédente. Il est confiant et avance ses pièces sereinement. Soudain, au 11^e coup, l'ordinateur avance un pion latéral sur la case H5, mouvement aussi absurde qu'agressif. Le coup est incompréhensible. Les commentateurs sont en émoi. Dans une telle position, tous sont persuadés qu'un coup défensif aurait été plus pertinent. Grâce à une coordination parfaite des pièces de l'ordinateur, l'avantage acquis

par Kasparov est réduit à néant et la machine obtient de justesse l'égalité. Kasparov est abattu. Psychologiquement anéanti dans la dernière partie, il s'effondre et abandonne avant le 20^e coup. Le match est plié, la machine l'emporte.

«Aucun ordinateur ne peut jouer H5!» Le compliment qui fait suite à la cinquième partie résonne comme une accusation. Kasparov laisse entendre que quelqu'un de l'équipe Deep Blue serait intervenu pour influencer sur le jeu de l'ordinateur. D'après lui, aucune machine ne saurait être à l'origine d'un coup aussi tordu, aussi éloigné du pur raisonnement tactique, logique. Seul un humain aurait pu déployer une ruse aussi fine.

La question ne sera jamais résolue et, en 1997, les dirigeants d'IBM décident de mettre un terme au programme Deep Blue pour faire bénéficier les autres activités de la compagnie des innovations d'intelligence artificielle développées par les ingénieurs de l'équipe. Le règne de l'homme sur les échecs est terminé. Pari gagné. Le roi est tombé. Échec et mat.

Introduction

Un mot des auteurs

Un nouveau livre sur l'IA ?

L'IA fascine. Mais dans un monde où elle s'affiche en une des magazines, où les rayons des librairies sont saturés d'essais sur le sujet, la question se pose avec force : pourquoi un nouvel ouvrage ? Le projet de ce livre tient en trois constats :

- Constat n° 1 – Le grand public maîtrise encore mal les sujets techniques sous-jacents à la révolution de l'IA alors même qu'il est directement concerné : droit de propriété sur ses données personnelles, droit à l'information, droit à l'oubli, fake news, discrimination algorithmique, remplacement de l'homme sur le marché du travail... Les sujets ne manquent pas. Miroir de l'information du grand public, c'est le débat politique qui n'est souvent pas à la hauteur des enjeux. Alors que les dernières évolutions technologiques appellent à des prises de position fortes de nos représentants, peu de penseurs, philosophes, politiques osent encore soulever le capot d'une technologie qui effraie par sa complexité supposée. L'objectif de ce livre consiste à montrer qu'il n'en est rien. Démystifier l'IA à l'heure des grandes décisions : voilà notre première ambition.
- Constat n° 2 – Les spécialistes de l'IA (ingénieurs, data scientists, technologues...) peinent parfois à ouvrir à la décision collective des sujets qu'ils considèrent comme purement techniques.

Le vocabulaire est technologique, les décisions sont pratiques, et elles manquent parfois d'un peu de recul sur les implications éthiques et sociétales qu'elles mettent en jeu. Dans un moment de compétition acharnée pour acquérir un avantage technologique sur son voisin, peu de temps est consacré à la vulgarisation et à la prise de recul. Voilà notre deuxième ambition.

- Constat n° 3 – Technologistes concepteurs et politiques décideurs. Rares sont les livres à ce jour qui permettent avec efficacité de combler le fossé entre ces deux populations. L'ambition du présent ouvrage consiste à allier une véritable exigence technique, la mise à jour des derniers développements d'une science en pleine évolution, à un souci permanent de vulgarisation pour le grand public. Si de nombreux articles scientifiques, théories et récits d'expériences viennent nourrir l'exigence scientifique de l'ouvrage, anecdotes, analogies et illustrations multiples s'y succèdent pour permettre à tous, de manière ludique, d'en apprendre un peu plus sur ce vaste sujet.

Dans ce paysage intellectuel en mutation, il devient impératif d'engager un dialogue ouvert et inclusif, où experts, décideurs et citoyens peuvent ensemble envisager les voies d'une cohabitation harmonieuse avec l'intelligence artificielle. Ce livre aspire ainsi à combler ce fossé entre la technologie et la société, en offrant une exploration approfondie et accessible des enjeux passionnants de l'IA, et en encourageant une réflexion collective sur les futurs possibles qu'elle dessine pour notre humanité.

Avec ChatGPT, la fulgurance algorithmique de l'IA générative a surgi aux yeux du grand public. La vitesse prodigieuse à laquelle ces technologies sont passées du laboratoire au stade de prototypes déployés dans le monde réel est inédite¹. Dans un premier temps,

1. Si GPT-3, introduit en 2020, a mis deux ans pour atteindre le million d'utilisateurs, il n'a fallu que deux mois à Dall-E 2 pour atteindre cet objectif et cinq jours à ChatGPT.

il faut acter que le changement technologique est fondamentalement vertical, centralisé et inégalitaire : comme beaucoup de technologies émergentes, les prototypes d'IA les plus avancées naissent dans les mains de quelques-uns et leurs processus de conception demeurent hors de portée de beaucoup d'autres¹. Une fois les effets de fascination et de sidération retombés, il devient crucial pour le grand public de s'emparer de questions fondamentales – stratégique, sociétale, économique – qui ne peuvent être tranchées uniquement par les sachants en IA. Contrairement à de nombreux domaines technologiques comme la biologie, la mécanique, l'aéronautique, l'IA n'est pas une science. Elle s'est construite grâce à de nombreuses avancées scientifiques, certes, mais elle ne constitue pas en elle-même une science. Pour qu'une discipline soit une science, il lui faut un protocole d'expérimentation qui fasse consensus ou une théorie bien établie. Ni l'un ni l'autre n'existent réellement encore pour l'IA. Nombre d'experts naviguent à vue, et l'histoire de l'IA nous montre que beaucoup des meilleurs technologistes se sont régulièrement trompés dans leurs prédictions en IA. En 1958, Frank Rosenblatt, le père du perceptron – une des premières machines de reconnaissance de forme – prédit que sa machine parlerait d'ici un an. Cela prit trente ans. En 2016, Geoffrey Hinton, un des pères fondateurs de l'IA et prix Turing, prédit qu'il ne servirait plus à rien de former des radiologues d'ici à 2020 car la machine se débrouillerait seule. En 2023, Sam Altman confie qu'il n'aurait jamais pu prédire l'émergence de ChatGPT il y a dix ans². Et ainsi de suite. La conversation sur l'IA et les décisions qui en découlent doivent donc faire l'objet d'une discussion exigeante et multilatérale entre toutes les parties prenantes. Chacun doit se former, faire l'effort de comprendre la dynamique à l'œuvre en évitant certains écueils. Faire l'effort d'aller un cran plus loin que la rapide expérience anecdotique qui est trop souvent faite de ces systèmes : « Je lui ai demandé telle chose, il ne m'a pas répondu correctement », etc. En matière de technologie, nos biais cognitifs

1. <https://openjournals.uwaterloo.ca/index.php/JoCI/article/download/2832/3632?inline=1>

2. <https://www.wsj.com/tech/ai/wsj-tech-live-2023-full-coverage-c2cb70fa>

limitent souvent notre capacité à extrapoler au-delà de ce qui nous paraît « raisonnable ». La loi de Roy Amara, technologiste américain, nous l'enseigne : nous avons tendance à surestimer l'effet d'une technologie à court terme et à le sous-estimer à long terme.

Chapitre 1

Qu'est-ce que l'intelligence artificielle ?

Chère machine à laver

Qu'est-ce que l'intelligence artificielle ? Restons simples : l'IA est l'ensemble des techniques visant à créer des machines capables de réaliser des tâches de réflexion.

Mais alors, rien de nouveau sous le soleil ! Nous programmons des machines à penser depuis bien longtemps. Sans remonter jusqu'au XVII^e siècle et à la Pascaline, machine à calculer du philosophe français Blaise Pascal, n'importe quel ordinateur classique, tel qu'il en existe depuis près de soixante-dix ans, répond parfaitement à cette définition. Idem pour nos compteurs électriques, nos montres digitales ou nos machines à laver des années 1980.

Alors, qu'est-ce qui a changé au tournant des années 2010 ? Réponse : notre manière de parler aux machines.

Le langage historique entre l'homme et la machine : le code informatique

Comment parler à un ordinateur ? Depuis toujours, l'homme a communiqué avec la machine via un jeu d'instructions clair, explicite, exhaustif : le code informatique.

Prenons l'exemple de ma machine à laver le linge. Au moment de sa conception, ce jeu d'instructions a été encodé dans son code source. De cette façon, si je sélectionne le programme «Laine», le radiateur chauffera à 25 °C et le tambour tournera à 400 tours/minute pendant vingt minutes. Voilà, en bon français, à quoi ressemblerait ce bout de code informatique :

Chère machine à laver,

***SI** on te donne pour consigne le programme «Laine», **ALORS** :*

- Toutes les 10 secondes, mesure la température de l'eau: **SI** elle est inférieure à 25 °C, **ALORS** augmente le chauffage. **SI** elle est supérieure à 25 °C, **ALORS** diminue le chauffage.*
- Toutes les 10 secondes, mesure la vitesse de rotation du tambour: **SI** elle est inférieure à 400 tours/minute, **ALORS** augmente la puissance du moteur. **SI** elle est supérieure à 400 tours/minute, **ALORS** diminue la puissance du moteur.*
- Au bout de 20 minutes, coupe le chauffage et le moteur.*

Ces trois consignes codées définissent le programme «Laine» de ma machine à laver. Combinées, toutes les règles et conditions de ce type régissent le fonctionnement complet de ma machine. Sauf problème technique, je sais que celle-ci obéira de manière répétée à mon instruction. Elle saura évaluer une situation – ma consigne initiale, la température de l'eau et la vitesse de rotation du tambour – et agir en conséquence. C'est ainsi que l'on programme les machines depuis un siècle. Le mode conversationnel entre l'homme et la machine est explicite, impératif, complet. Tous les cas sont explorés pour ne pas laisser la machine démunie face à une situation nouvelle. C'est un premier niveau d'intelligence artificielle.

Les limites du code explicite

Pour des tâches plus complexes, le code informatique et ses jeux de règles sont vite devenus insuffisants. Prenons l'exemple d'une voiture autonome, bardée de capteurs, qui récoltent et analysent en temps réel toutes les informations possibles sur l'environnement de la voiture :

la position des obstacles, la vitesse relative des autres usagers, les codes de signalisation... Voilà, en bon français, à quoi pourrait ressembler son code informatique :

Chère voiture autonome,

SI tu détectes un feu de signalisation en face de toi, **ALORS** :

- **SI** le feu est rouge, **ALORS** tu t'arrêtes.
- **SI** le feu est vert, **ALORS** tu passes.
- **SI** le feu est vert **ET** que tu repères un piéton qui traverse malgré tout, **ALORS** tu t'arrêtes.
- **SI** le feu est vert **ET** que tu repères un piéton qui manifeste une envie de traverser, **ALORS** tu t'arrêtes.
- **SI** le feu est orange **ET** que des enfants s'amuse près du passage clouté **ET** que la voiture derrière toi est proche et risque de te rentrer dedans **ET** que...
- *Etc.*

On le voit bien, la conduite d'une voiture présente un nombre de situations et sous-situations bien trop important pour être embrassée par un code informatique exhaustif. Les cas particuliers sont pléthores et ne peuvent être tous pris en compte. Qui plus est, la conduite est affaire d'intuition. Tout bon conducteur le sait. Et l'intuition n'est pas modélisable en un jeu de règles. Pendant des décennies donc, la voiture sans chauffeur est restée à l'état de pur fantasma.

Autre exemple, dans le domaine médical. Comment détecter automatiquement une fracture sur une radio du fémur ? La machine voit tout. Elle peut distinguer des milliers de teintes de blanc sur une image quand l'œil humain n'en voit qu'une dizaine tout au plus. Problème : rechercher une anomalie dans un tableau de pixels via des règles explicites est impossible via un jeu de règles explicites. La lecture d'une radio implique une compétence qui ne peut être codée explicitement : l'expérience d'un médecin. Un vieux radiologue identifiera des anomalies indétectables, simplement par son expérience. Il sait. Il ne sait pas pourquoi, mais il sait. Incodable, du moins explicitement, dans une machine, à l'œil aussi aiguisé soit-il.

Le machine learning, nouveau mode de communication homme-machine

Tout change avec le machine learning (apprentissage automatique). Plutôt que de donner à la machine un jeu d'instructions, on lui demande maintenant d'apprendre sur la base d'exemples. Des algorithmes statistiques permettent à l'ordinateur de recouper tous ces exemples et d'en tirer les implications. Objectif : optimiser sa réaction face à une situation nouvelle, non prévue par le développeur.

Reprenons notre voiture autonome. Pour lui apprendre à se conduire seule, un (bon) conducteur humain doit d'abord rouler, rouler, et rouler encore. La machine enregistre les données de la conduite : l'environnement de la voiture, mais aussi toutes les réactions du conducteur. Grâce à des algorithmes statistiques, elle apprend. Et à mesure de cet apprentissage, elle devient meilleure conductrice, comme un jeune qui progresserait à mesure de ses passages à l'auto-école. Et puis un jour elle sait conduire. Mieux que cela, la voiture a retenu le meilleur de tout cet apprentissage. À tout moment et face à toute situation, elle sait adopter la meilleure réaction possible : celle du meilleur conducteur dans son meilleur jour. Face à des situations qu'elle a déjà rencontrées, elle sait réagir. Un feu rouge, pas de problème, un virage serré, elle en a vu des milliers. Mais surtout, elle sait réagir face à des situations entièrement nouvelles, qui ne correspondent à aucune des situations rencontrées lors de son apprentissage : elle recoupe les informations, calcule, optimise, et prend la meilleure décision possible. Merci les statistiques !

Voilà donc le changement de paradigme technologique au fondement de cette nouvelle ère de l'IA. Mais les spécialistes de l'entraînement des machines savent rester pragmatiques. Ils combineront volontiers les deux approches : explicite et implicite, codée et apprise. L'exemple de la voiture autonome est clair : si la machine est essentiellement programmée via l'apprentissage, des règles subsistent pour prendre le relais et gérer de manière impérative des situations critiques. Ainsi en va-t-il des décisions que l'éthique ou la réglementation imposeront de prendre dans des situations délicates de danger ou d'accident.

D'une révolution technologique à une révolution sociétale

Ce nouveau mode de langage homme-machine présente de nombreuses conséquences. Parmi celles-ci, trois méritent d'être déjà mises en avant, et seront développées dans des chapitres ultérieurs : l'importance de la data, l'impossible explication des décisions de la machine, et les changements sociétaux induits par ce nouveau paradigme.

Importance de la data

On l'a vu, la machine ne déduit plus à partir de règles. Elle induit à partir d'exemples, beaucoup d'exemples. Une étude de l'University College of London (UCL) de 2015 réalisée sur des bébés de 4 à 6 mois a montré que quelques images suffisaient pour que le nourrisson reconnaisse un visage. Une machine nouvellement entraînée en nécessitera plusieurs milliers. L'ordre de grandeur peut monter à plusieurs millions si l'objet représenté est difficile à discerner (le fameux *meme* «Muffin ou Chihuahua» en donne un exemple éloquent). D'où l'importance d'entraîner sur un jeu de données le plus grand possible. Voilà pourquoi, là où l'on parle d'IA, on parle souvent de «Big Data». La production, la conservation et l'utilisation de ces données de masse posent des problèmes nouveaux à nos sociétés et leurs régulateurs : préservation de la vie privée, impact écologique, biais racistes et discriminatoires... la liste est longue et plusieurs chapitres y seront consacrés.

La question de l'explication du comportement des machines

Comment comprendre la décision de la voiture ? Quand la règle est remplacée par l'induction, il devient difficile d'expliquer les choix de la machine. Et la complexification continue des algorithmes n'aide en rien. Si les premiers modèles permettaient de comprendre le rationnel de chaque décision de l'ordinateur, les derniers réseaux de neurones

rendent cette tâche quasi impossible (même si la recherche s'y attelle actuellement avec énergie). Pas facile alors d'établir les responsabilités quand la voiture rentre dans un arbre. Qui paye le dommage ? Le constructeur ? Le développeur ? Le conducteur ? La collectivité ? De nombreuses questions essentiellement juridiques et éthiques en découlent.

Impact sur les compétences humaines « remplaçables »

Ce passage d'un mode de consigne explicite à une programmation implicite des machines permet de capter de nouvelles nuances et de les insuffler dans le comportement des ordinateurs ainsi entraînés : l'intuition, l'expérience, la sensibilité des hommes qui les ont entraînés. Cette évolution de la technique est à la base d'un changement de paradigme sociétal et explique les effets profonds de l'IA sur notre société et notre marché du travail. De plus en plus de compétences deviennent remplaçables par la machine, et c'est le monde tel que nous le connaissons qui est amené à changer.

Résumé

L'IA se définit à plusieurs niveaux. Au sens le plus large, elle englobe l'ensemble des techniques visant à créer des machines capables de répliquer la cognition humaine. Au tournant des années 2010, cette définition a été ajustée par la démocratisation d'une nouvelle manière d'interagir avec la machine, fondée non plus sur une programmation explicite mais sur un apprentissage par l'exemple. Ces nouvelles techniques dites de « machine learning » permettent aujourd'hui d'optimiser le comportement des ordinateurs en leur présentant de nombreux exemples issus du monde réel. Si elles ont permis l'émergence d'applications nouvelles (voitures autonomes, analyse médicale, automatisation des tâches...), elles soulèvent aujourd'hui des questions cruciales en termes de responsabilité éthique et juridique, et d'impact sur nos économies et nos sociétés.