L' intelligence artificielle. Une question de point de vue ?

Didier Peters

Introduction

« *Ce qui fait donc que de certains esprits fins ne sont pas géomètres, c'est qu'ils ne peuvent du tout se tourner vers les principes de géométrie. Mais ce qui fait que des géomètres ne sont pas fins, c'est qu'ils ne voient pas ce qui est devant eux et qu'étant accoutumés aux principes nets et grossiers de géométrie, et à ne raisonner qu'après avoir bien vu et manié leurs principes, ils se perdent dans les choses de finesse où les principes ne se laissent pas ainsi manier. On les voit à peine, on les sent plutôt qu'on ne les voit, on a des peines infinies à les faire sentir à ceux qui ne les sentent pas d'eux mêmes. Ce sont choses tellement délicates, et si nombreuses, qu'il faut un sens bien délicat et bien net pour les sentir et juger droit et juste selon ce sentiment, sans pouvoir le plus souvent le démontrer par ordre comme en géométrie, parce qu'on n'en possède pas ainsi les principes, et que ce serait une chose infinie de l'entreprendre. Il faut tout d'un coup voir la chose d'un seul regard, et non pas par progrès de raisonnement, au moins jusqu'à un certain degré. Et ainsi il est rare que les géomètres soient fins et que les fins soient géomètres, à cause que les géomètres veulent traiter géométriquement ces choses fines et se rendent ridicules, voulant commencer par les définitions et ensuite par les principes, ce qui n'est pas la manière d'agir en cette sorte de raisonnement. Ce n'est pas que l'esprit ne le fasse, mais il le fait tacitement, naturellement et sans art, car l'expression en passe tous les hommes, et le sentiment n'en appartient qu'à peu d'hommes.* ».
Pascal, Géométrie-Finesse II – Fragment n° 1 / 2.

L'intelligence artificielle (IA) est devenue omniprésente dans nos sociétés modernes, façonnant de manière significative la façon dont nous travaillons, interagissons et vivons au quotidien. Cette technologie, qui consiste en des systèmes informatiques capables d'imiter certaines fonctions cognitives humaines, a progressé à pas de géant au cours des dernières décennies.

Au-delà des rêves les plus fous, l'intelligence artificielle (IA) a trouvé d'innombrables applications pratiques dans notre quotidien. Des ordinateurs portables aux téléphones portables, des applications diverses aux voitures autonomes, des logiciels de correction orthographique aux traducteurs en ligne, des systèmes de reconnaissance vocale et faciale aux moteurs de recherche et aux jeux vidéo en ligne, l'IA est omniprésente. Les compagnies d'assurance et les prestataires de soins de santé peuvent désormais prédire l'état de santé futur des individus et leurs potentiels problèmes. Des programmes peuvent anticiper et organiser le trafic routier et aérien, contribuer à la découverte de nouveaux médicaments et à l'élaboration de plans de traitement pour les patients, analyser les marchés économiques, identifier les fraudes financières, composer de la musique, améliorer la production automobile et agricole, détecter les fausses nouvelles et le spamming, et bien plus encore.

Des robots intelligents peuvent accomplir des tâches que les humains ne savent ou ne veulent pas faire, comme les missions de sauvetage, les voyages dangereux dans l'espace ou sous l'eau, ou encore des calculs dont la rapidité dépasse l'entendement humain. Certains donc envisagent l'IA comme une solution aux plus grands défis de l'humanité : les maladies, la faim, le changement climatique, quand il ne s'agit pas de créer des compagnons et des aides robotiques pour les personnes âgées ou handicapées, et même des amis pour les individus isolés ou confinés à domicile.

Cette omniprésence de l'IA suscite à la fois fascination et appréhension, quand elle ne soulève pas des questions éthiques, sociales et politiques complexes s'agissant de l'impact qu'elle a sur nos vies. D'une part, l'IA étend le domaine du possible et promet de transformer radicalement nos modes de travail, nos modes de communication et nos interactions sociales.

D'autre part, elle soulève des préoccupations légitimes en matière de confidentialité des données, de sécurité, de biais algorithmique, de perte d'emplois et d'équité sociale. Les inquiétudes, légitimes, associées à sa montée en puissance, sont à l'origine de nombreux débats sur la responsabilité, la transparence et la réglementation de cette technologie.

Nous proposons dans cet article de partir de la notion d'intelligence, pour ensuite aborder plus particulièrement l'intelligence artificielle. Nous verrons que le problème de l'intelligence est, selon nous, doublement mal posé. D'une part que l'intelligence en tant que telle résiste à toute définition simple et univoque, et que, d'autre part, le concept d'intelligence artificielle résulte de ce qu'Alfred North Whitehead a nommé le sophisme du concret mal placé, consistant à prendre la carte pour le territoire. Nous proposerons finalement une approche différente s'appuyant sur la métaphysique processuelle.

Photo by Icons8 Team on Unsplash

Une question d'intelligence ?

La première question que nous devons nous poser est une question de définition. Qu'entendons-nous tout d'abord par intelligence ? Les définitions qui lui sont données sont multiples. Citons, à titre d'exemples :

- L'intelligence est une aptitude mentale très générale qui implique notamment l'habileté à raisonner, à planifier, à résoudre des problèmes, à penser abstraitement, à bien comprendre des idées complexes, à apprendre rapidement et à tirer profit de ses expériences. L'intelligence ne se résume pas à l'apprentissage livresque, ni à une aptitude scolaire très circonscrite, ni aux habiletés spécifiquement reliées à la réussite des tests mentaux. Au contraire, elle reflète cette habileté beaucoup plus étendue et profonde à comprendre son environnement, à « saisir un problème », à « donner un sens » aux choses ou à imaginer des solutions pratiques. (Françoys Gagné, Serge Larivée).
- Selon le Larousse, l'intelligence est :

  - L'ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle : Les mathématiques sont-elles le domaine privilégié de l'intelligence ? Test d'intelligence.

  - L'aptitude d'un être humain à s'adapter à une situation, à choisir des moyens d'action en fonction des circonstances : Ce travail réclame un minimum d'intelligence.

  - Une personne considérée dans ses aptitudes intellectuelles, en tant qu'être pensant : C'est une intelligence supérieure.

  - La qualité de quelqu'un qui manifeste dans un domaine donné un souci de comprendre, de réfléchir, de connaître et qui adapte facilement son comportement à ces finalités : Avoir l'intelligence des affaires.

  - La capacité de saisir une chose par la pensée : Pour l'intelligence de ce qui va suivre, rappelons la démonstration antérieure.
- Si nous nous tournons vers Wikipédia, l'intelligence est l'ensemble des processus trouvés dans des systèmes, plus ou moins complexes, vivants ou non, qui permettent d'apprendre, de comprendre ou de s'adapter à des situations nouvelles. La définition de l'intelligence ainsi que la question d'une faculté d'intelligence générale ont fait l'objet de nombreuses discussions philosophiques et scientifiques. L'intelligence a été décrite comme une faculté d'adaptation (apprentissage pour s'adapter à l'environnement) ou au contraire, faculté de modifier l'environnement pour l'adapter à ses propres besoins. Dans ce sens général, les animaux, les plantes (intelligence primaire faite d'[instinct] et de [réflexes] conditionnés) ou encore certains outils informatiques (apprentissage automatique, intelligence artificielle) font preuve d'intelligence. L'acquisition de la parole articulée et de l'écriture, qui aident au développement du raisonnement, font de l'intelligence humaine la référence.

Nous pourrions en citer d'autres ; mais nous pouvons constater que sa définition n'est pas univoque. Si le terme « intelligence » est régulièrement utilisé dans de nombreux domaines d'activité totalement différents (la psychologie, la pédagogie, l'informatique)comme l'intelligence émotionnelle, l'intelligence logico-mathématique, l'intelligence économique, et, bien évidemment l'intelligence artificielle, en fonction des domaines dans lesquelles elle s'exerce, sa définition peut varier.

L'intelligence semble résister à nos tentatives de la caractériser en termes clairs. Elle ne semble pas être une fonction définie, à l'image d'une fonction biologique ou mécanique, que nous pourrions associer à un mécanisme, aussi complexe soit-il. Elle se manifeste dans chaque situation particulière par le déploiement de capacités qui pourraient être reproduites artificiellement, nous y reviendrons, donnant ainsi l'impression que l'intelligence n'est rien d'autre que l'ensemble de ses capacités et qu'en les reproduisant toutes, nous en arriverions à une intelligence artificielle.

Quelle intelligence artificielle ?

Un consensus existe quant aux définitions de l'intelligence artificielle. Deux types sont généralement distingués : la faible et la forte.

L'IA faible, également connue sous le nom d'IA étroite, fait référence aux systèmes d'intelligence artificielle qui sont conçus et entraînés pour effectuer des tâches spécifiques ou résoudre des problèmes particuliers. Contrairement à l'IA forte ou à l'intelligence artificielle générale (IAG), l'IA faible ne possède pas de conscience, de conscience de soi ou de capacité à comprendre le contexte au-delà de son champ d'application limité et prédéfini. Elle fonctionne plutôt dans les limites des tâches pour lesquelles elle a été programmée et entraînée.

Les systèmes d'IA faible ont été développés pour gérer des tâches spécifiques, les plus connues étant la traduction, les assistants virtuels, les systèmes de recommandations, la reconnaissance d'images, les chabots ou encore les véhicules autonomes. Ils excellent dans ces tâches, mais ne peuvent pas généraliser leurs connaissances à d'autres domaines non apparentés, un système développé pour la reconnaissance d'images ne sera pas en mesure d'effectuer des traductions. Par exemple, une IA faible formée pour recommander des films ne peut pas passer automatiquement au diagnostic de conditions médicales.

Une IA faible ne possède ni la conscience de soi ni la capacité de réfléchir sur ses actions. Elle traite des entrées et produit des sorties basées sur des algorithmes et des données sans les comprendre. Elle n'est pas autonome, son développement et ses améliorations potentielles dépendent des données qui lui sont fournies.

L'IA faible représente l'état de la plupart des technologies actuelles d'IA. Elle excelle dans des tâches spécifiques, mais il lui manque les capacités cognitives plus larges de l'intelligence humaine. Bien qu'elle ait des applications pratiques

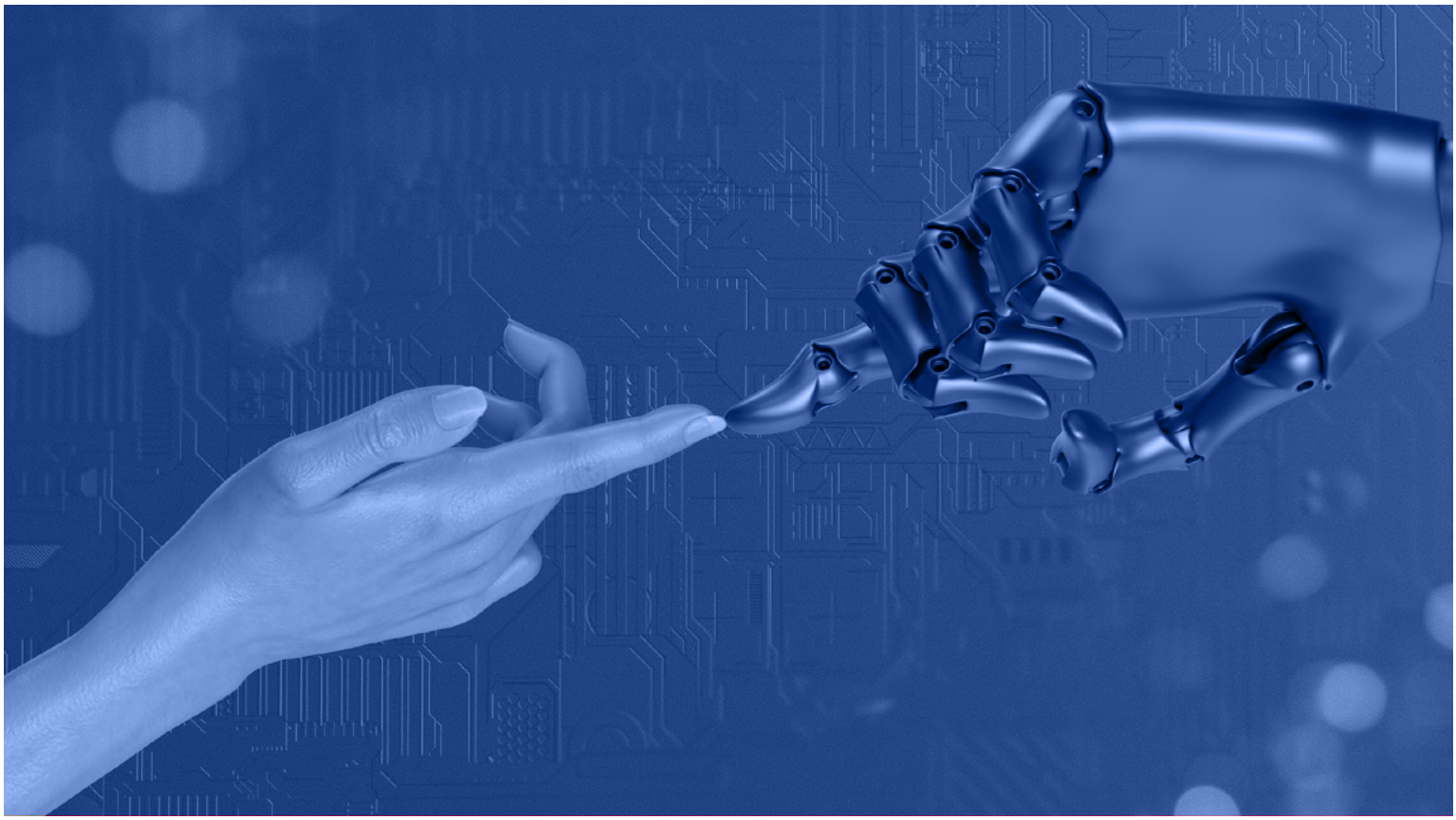
significatives et puisse grandement améliorer l'efficacité et la productivité dans divers domaines, elle fonctionne dans des limites prédéfinies et ne possède pas la conscience de soi ou l'intelligence générale envisagées pour les futurs systèmes d'IA.

L'intelligence artificielle forte, également connue sous le nom d'intelligence générale artificielle (AGI), se réfère, quant à elle, à une forme d'intelligence artificielle capable de comprendre, d'apprendre et d'appliquer des connaissances d'une manière qui ne se distingue pas de l'intelligence humaine. Contrairement à l'IA faible, conçue pour des tâches spécifiques, l'IA forte a la capacité d'accomplir toute tâche intellectuelle qu'un être humain peut réaliser, avec la potentialité de manifester une conscience de soi, une conscience et une compréhension émotionnelle.

Elle pourrait généraliser les connaissances en appliquant ce qu'elle a appris dans un domaine pour résoudre des problèmes dans un autre, contrairement à l'IA faible, qui est limitée à la tâche pour laquelle elle a été conçue. Elle possèderait des capacités cognitives similaires à celles des humains, telles que le raisonnement, la résolution de problèmes, la pensée abstraite et la compréhension de concepts complexes. L'IA forte aurait une conscience de soi, c'est-à-dire une conscience de sa propre existence, associée à des capacités de compréhension et de réflexion sur ce qu'elle pense ou entreprend. Elle serait en mesure d'apprendre de ses expériences, de s'adapter à de nouvelles situations, et de s'améliorer avec le temps sans intervention humaine. Elle pourrait, en outre, comprendre et générer un langage humain de manière contextuellement appropriée et significative, permettant une communication naturelle, empathique et fluide avec les humains.

Ses applications seraient sans limites. Elle serait en mesure de diagnostiquer des maladies, développer des plans de traitement et fournir des soins personnalisés en comprenant des données médicales

Photo by Igor Omilaev on Unsplash



complexes et les antécédents des patients. Elle pourrait servir de tuteurs personnalisés, adaptant les méthodes d'enseignement aux besoins et styles d'apprentissage individuel des étudiants, mener des recherches avancées, générer des hypothèses et analyser des données complexes pour faire des découvertes révolutionnaires dans divers domaines scientifiques. Elle pourrait aussi intervenir dans des processus de prise de décision complexes par les vastes quantités de données qu'elle serait en mesure de traiter.

L'usage du conditionnel reste encore de rigueur. Car, bien qu'elle détienne un immense potentiel pour révolutionner divers domaines et améliorer la vie humaine, elle nécessitera non seulement des avancées technologiques, mais aussi une réflexion approfondie sur son impact sur la société et l'humanité dans son ensemble. En d'autres termes, cette IA forte reste, au jour d'aujourd'hui, hypothétique.

L'histoire et la technologie informatique de l'IA sont des sujets qui dépassent le cadre de cet article, mais il nous semble néanmoins important d'en dresser un aperçu. L'idée de créer des êtres artificiellement intelligents existe depuis des siècles et des siècles. Les Grecs et les Chinois de l'Antiquité avaient des mythes sur les robots, les Égyptiens de l'Antiquité construisaient des automates, les vieilles horloges européennes avaient des coucous mécaniques et les jouets ambulants à remontoir existent depuis des siècles. La science-fiction, depuis le Frankenstein de Mary Shelley jusqu'à Isaac Asimov ou Philip K. Dick, s'est intéressée, elle aussi, à la création d'êtres artificiels.

L'intelligence artificielle repose depuis longtemps sur l'idée que le processus de la pensée humaine peut être mécanisé. Nombreux sont ceux qui, au cours de l'histoire, ont pensé que l'esprit humain était essentiellement un ordinateur fonctionnant selon une logique symbolique et qu'il était possible, du moins en théorie, de reproduire l'esprit et la pensée humaine en manipulant des symboles, par exemple à l'aide d'un ordinateur. Le mathématicien et philosophe Gottfried Leibniz avait imaginé un langage symbolique qui réduirait les arguments philosophiques à des calculs mathématiques, de sorte qu'il serait possible de répondre aux débats philosophiques comme un comptable calcule les finances. L'étude de la logique mathématique formelle menée au début du XXe siècle par des mathématiciens et des philosophes tels que Bertrand Russell, Alfred North Whitehead et David Hilbert a également beaucoup influencé les premiers travaux sur l'intelligence artificielle.

Avec l'introduction d'ordinateurs capables de manipuler des nombres, les mathématiciens ont réalisé qu'ils pouvaient manipuler des symboles et ils se sont demandé s'ils pouvaient réellement créer un cerveau artificiel. Les premières recherches fondamentales sur les cerveaux artificiels ont eu lieu dans les années 1930 à 1950, les scientifiques ayant découvert que le cerveau humain fonctionnait avec des impulsions électriques, à l'instar des signaux numériques. Alan Turing a théorisé le fait que tout calcul pouvait être décrit numériquement, laissant entrevoir la possibilité de fabriquer un cerveau électronique. Walter Pitts et Warren McCulloch du M.I.T. montrèrent comment les réseaux neuronaux artificiels pouvaient exécuter des fonctions logiques, et Alan Turing écrivit un célèbre article en 1950 qui spéculait sur la possibilité de créer des machines capables de penser.

Une conférence organisée en 1956 au Dartmouth College introduisit le nom d'intelligence artificielle

et définit sa mission. Selon John McCarthy, la conférence devait

> « *procéder sur la base de la conjecture selon laquelle chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence pouvait en principe être décrit avec une telle précision qu'une machine pouvait être fabriquée pour le simuler.* » (dartmouth.edu).

Depuis lors, les informaticiens ont essayé une grande variété de techniques et d'idées pour créer une intelligence artificielle, qu'il s'agisse d'une IA faible ou d'une intelligence générale artificielle.

Les pionniers de l'IA Marvin Minsky et McCarthy ont utilisé une approche symbolique de l'IA qui a dominé la recherche et la pratique de l'IA des années 1950 aux années 1980. Cette approche consistait à créer des règles logiques, étape par étape, afin que l'ordinateur exécute ses tâches, comme le pensaient les chercheurs, de la manière logique dont les humains les exécutaient : étape par étape. D'autres informaticiens introduisirent des heuristiques, ou règles empiriques, pour simplifier ce processus. En substance, ces systèmes d'IA symboliques tentaient de représenter les connaissances humaines sous la forme de faits et de règles logiques.

De nombreux financements commerciaux et gouvernementaux furent consacrés à ce domaine, et de nombreux scientifiques utilisant ces méthodes affirmèrent que l'intelligence artificielle générale était à portée de main[1]. Malgré ces avancées, les limitations technologiques et les attentes non satisfaites conduisirent à une réduction du financement et de l'intérêt pour la recherche en IA, période que d'aucuns appelèrent « l'hiver de l'IA ».

L'introduction de nouvelles techniques d'apprentissage automatique, comme les réseaux de neurones artificiels et l'algorithme de rétropropagation, permit de surmonter certaines limitations des systèmes précédents. En 1997, Deep Blue, développé par IBM, battit Garry Kasparov, le champion du monde d'échecs, montrant la capacité des IA à exceller dans des tâches spécifiques.

La recherche et le développement des réseaux neuronaux artificiels permirent, grâce à l'augmentation de la puissance de calcul et de capacité de stockage des ordinateurs, permettant la manipulation d'un volume de plus en plus important de données, d'aborder l'intelligence artificielle sous un angle très différent. En effet, un réseau neuronal est une tentative de simulation d'un cerveau basée sur les neurones biologiques du cerveau humain. Elle effectue le traitement de l'information en profondeur, appelé apprentissage profond (deep learning). Elle connaît un grand succès dans des domaines tels que l'interprétation des images et des sons, les services personnalisés tels que ceux proposés par Amazon et Google, les moteurs de recherche, la cartographie en ligne, l'analyse des données médicales pour traiter les maladies et découvrir de nouveaux médicaments, et l'identification des fraudes financières.

Aujourd'hui, l'IA symbolique et les techniques des réseaux neuronaux artificiels sont toutes deux utilisées. Elles ont toutes deux leurs propres limites et fonctionnent mieux dans des domaines différents. L'IA symbolique a permis de créer l'ordinateur capable de battre n'importe qui aux échecs, tandis que l'IA des réseaux neuronaux artificiels à réflexion profonde a produit la technologie capable d'identifier votre visage et votre voix sur votre téléphone portable.

[1] Par exemple, en 1970, Marvin Minsky, du M.I.T : "*Dans trois à huit ans, nous aurons une machine dotée de l'intelligence générale d'un être humain moyen*" (référence : BBC.com).

Une des avancées majeures de l'IA faible est ce qui est appelé l'IA générative. Celle-ci démontre d'impressionnantes compétences dans la création de contenu original, mais elle reste malgré tout limitée à des tâches spécifiques, celles pour lesquelles elle a été formée. Des modèles comme GPT-4[2] peuvent générer du texte cohérent et pertinent en réponse à des prompts, sans toutefois comprendre véritablement le sens du texte qu'ils produisent. Les GANs[3] peuvent, eux, créer des images réalistes à partir de données d'entraînement, mais ils n'ont pas de compréhension de ce qu'est une image en termes de perception humaine.

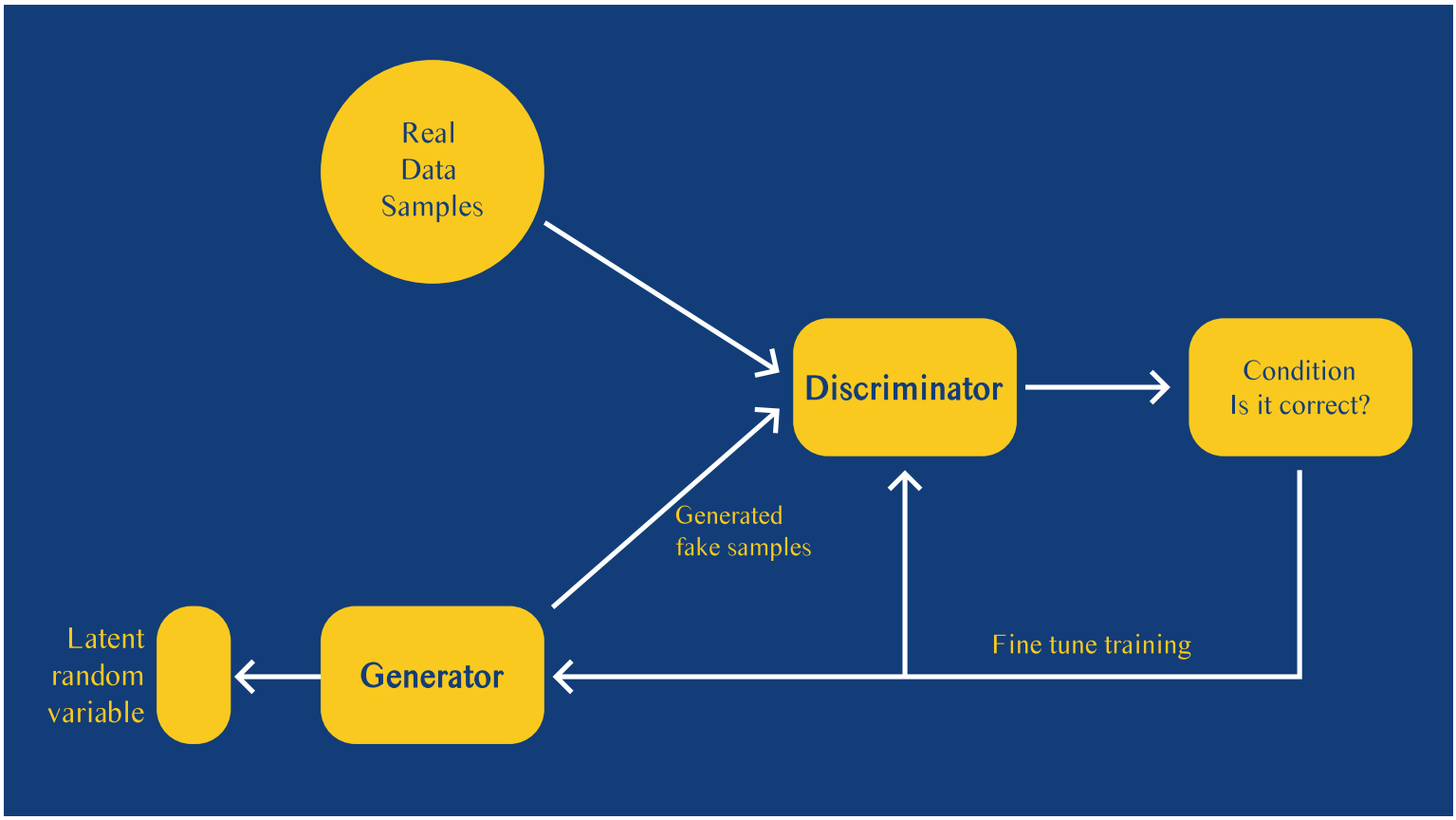
[2] Les transformeurs, comme GPT-4 (Generative Pre-trained Transformer 4, développé par OpenAI) sont « entraînés » sur d'énormes corpus de données textuelles et peuvent produire des textes cohérents et contextuellement pertinents en réponse à des invites données. GPT-4 est un LLM (Large Language Model – Modèle Massif de Langage) d'environ 1,8 trillion de paramètres.

[3] Les **GANs** ou Réseaux Génératifs Antagonistes se composent de deux réseaux neuronaux en compétition : un générateur, qui crée des échantillons de données, et un discriminateur, qui tente de distinguer les échantillons générés des échantillons réels. Le but est de faire en sorte que le générateur produise des échantillons si réalistes que le discriminateur ne puisse pas les distinguer des échantillons réels.

Une question de point de vue

> « *Nous pensons que les hommes et les autres animaux sont comme des machines d'un point de vue scientifique parce que nous pensons que les seules méthodes fructueuses pour l'étude du comportement humain et animal sont les méthodes applicables au comportement des objets mécaniques également. Ainsi, notre principale raison de choisir les termes en question était de souligner que, en tant qu'objets de recherche scientifique, les hommes ne diffèrent pas des machines.* ». Rosenblueth, Arturo, and Norbert Wiener. « Purposeful and Non-purposeful Behavior. », p. 326.

Le domaine de l'intelligence artificielle (IA) se développe rapidement et a des implications significatives pour la société et l'humanité. Alors que l'IA continue de façonner notre monde et d'influer sur notre vie quotidienne, il est essentiel de procéder à une évaluation critique de ses fondements philosophiques et de veiller à ce que son développement s'aligne sur les valeurs et les objectifs de l'humanité.



L'intelligence artificielle (IA) et la science des données (data science), en particulier l'apprentissage automatique et ses applications contemporaines qui permettent la classification, la prédiction, la prise de décision et la manipulation automatisées dans de nombreux domaines de l'activité humaine et de la société, ont suscité de nombreuses controverses au cours de la dernière décennie : l'impact éthique et sociétal (potentiel) de l'IA suscite de nombreuses inquiétudes. En philosophie et dans les domaines académiques connexes, cela s'est accompagné d'une vague de publications sur les aspects éthiques et politiques de l'IA. Toutefois, la nature philosophique de l'IA a fait l'objet de moins d'attention.

Nous l'avons déjà noté, la nature de l'intelligence est l'un des fondements de la philosophie de l'IA. « Que signifie être intelligent ? » Cela amène les ingénieurs et les développeurs de l'IA à se demander si l'intelligence n'est qu'une question de traitement de l'information ou si elle peut être constituée de bien d'autres choses encore. Alors que la plupart des philosophes défendent l'exigence de conscience et de conscience de soi, d'autres affirment que la certitude de l'intelligence de l'IA peut être prouvée par sa capacité à exécuter avec succès des tâches qui nécessitaient auparavant une véritable intelligence. Cette question ne mériterait-elle pas mieux ?

Un sophisme du concret mal placé[4] ?

[4] Traduction littérale de "fallacy of misplaced concreteness".

Le « sophisme du concret mal placé » est un concept philosophique introduit par Alfred North Whitehead dans son ouvrage « Process and Reality ». Ce sophisme se produit lorsque des concepts ou des modèles abstraits sont traités comme s'il s'agissait de réalités concrètes, entraînant des malentendus voire des erreurs de raisonnement. Les concepts abstraits sont des simplifications ou des généralisations qui nous aident à comprendre et à décrire des réalités complexes. Ils ne sont pas les réalités elles-mêmes, mais plutôt des outils de réflexion sur ces réalités. Par exemple, le concept d'« arbre » est une abstraction qui renvoie à une idée générale de ce qu'est un arbre, englobant de nombreux types d'arbres différents aux caractéristiques variées. L'erreur se produit lorsque ces abstractions sont traitées à tort comme si elles avaient la même existence concrète que les phénomènes qu'elles décrivent. Cette erreur de positionnement conduit à une compréhension déformée de la réalité. Par exemple, l'utilisation d'une carte (une abstraction) comme s'il s'agissait du territoire réel peut entraîner des erreurs de navigation, car la carte ne peut jamais saisir pleinement les complexités du monde réel.

En science, les modèles et les théories sont des abstractions utilisées pour expliquer et prédire des phénomènes. Traiter ces modèles comme des représentations exactes de la réalité peut conduire à des conclusions erronées, car tous les modèles ont des limites et des simplifications. De même, les modèles économiques utilisent souvent des hypothèses simplifiées pour prédire le comportement du marché. Lorsque les décideurs politiques considèrent ces modèles comme des représentations parfaitement exactes de la réalité économique, ils risquent de mettre en œuvre des politiques qui ne tiennent pas compte des complexités du monde réel. Dans la vie de tous les jours, les stéréotypes sont un exemple courant de cette erreur.

Lorsque les gens appliquent des caractéristiques généralisées à tous les membres d'un groupe comme si ces caractéristiques étaient des faits concrets, ils ignorent les variations et les complexités individuelles au sein de ce groupe.

Il est essentiel de comprendre ce sophisme, car il nous aide à rester conscients des limites de nos abstractions et de nos modèles. En reconnaissant que ces outils sont des simplifications, nous pouvons les utiliser plus efficacement et éviter les pièges d'une interprétation trop littérale. Cette prise de conscience favorise une réflexion plus nuancée et plus précise dans diverses disciplines, de la science et de l'économie à la prise de décision quotidienne. Ce sophisme du concret mal placé nous rappelle que si les abstractions et les modèles sont utiles, ils ne doivent pas être confondus avec les réalités complexes qu'ils représentent.

Nous avons vu que les derniers développements de l'IA sont basés sur une représentation neuronale du cerveau. Cependant, les neurosciences en tant que discipline, comme toute autre science, sont tributaires de certains modes de traitement de leurs objets, qui projettent les choses sur un plan de simplicité plus élevé.

Cette réduction est logique. La méthodologie séculaire, consistant à utiliser des mécanismes artificiels comme exemples de processus naturels a eu tendance à éluder la distinction entre les produits de la technologie et les objets qui n'ont pas été fabriqués par la main de l'homme, a toujours été tolérable dans la mesure où il y avait gain d'efficacité à supposer que l'objet de la recherche était le modèle rationalisé, et non la chose concrète.

Si les philosophes de l'esprit attendent de la science qu'elle les renseigne sur la nature de ces objets de recherche : qu'est-ce que la mémoire, la vision, la compréhension et la conscience ? Leur préoccupation n'est pas de savoir comment les choses se passent avec le modèle — qu'ils considèrent trop souvent comme un support transparent et non déformant — mais dans notre cerveau et nos processus cognitifs. Si une abstraction d'un modèle, imposée par la nécessité pratique, est prise pour une découverte sur la façon dont le cerveau est constitué, nous nous trouvons dans une situation typique de sophisme du concret mal placé, avec les risques d'erreur d'interprétation qu'elle suppose.

Les opinions philosophiques « classiques » sur le potentiel de l'intelligence artificielle sont le produit de l'erreur qui accompagne l'interprétation littérale des modèles vaguement « inspirés du cerveau », paradigme dominant actuel de la recherche en IA, dans le but d'obtenir des performances cognitives semblables à celles de l'homme pour des tâches prédéfinies telles que la reconnaissance d'objets, la production de langage ou le jeu. Nombre de ces systèmes experts, qui sont des réseaux de neurones artificiels (RNA) formés pour exceller dans l'une de ces tâches, ont atteint des capacités surhumaines. Dans le même temps, les RNA ont montré une série d'échecs surprenants qui semblent découler de leur nature d'expert, c'est-à-dire de leur manque d'intelligence générale, de bon sens élémentaire. La question est de savoir si la technologie actuelle d'apprentissage automatique, lorsqu'elle sera portée à une taille de réseau suffisante, permettra d'obtenir une intelligence générale semblable à celle de l'homme, avec les caractéristiques qui font actuellement défaut aux IA actuelles. L'incapacité des RNA à obtenir les capacités associées à l'intelligence générale, telles que la sensibilité et la capacité d'appliquer les connaissances acquises à des

situations fondamentalement nouvelles, n'est pas surprenante si l'on considère l'interprétation analogique des modèles informatiques.

Rappelons-nous qu'en philosophie, la théorie computationnelle de l'esprit affirme que les processus cognitifs (y compris la perception, le contrôle moteur et l'affectivité) sont essentiellement des processus computationnels. Être une créature cognitive, c'est avoir un cerveau qui met en œuvre des calculs affinés par la sélection naturelle pour soutenir un comportement intelligent. Cette théorie est un moyen populaire de naturaliser l'esprit — de montrer comment l'esprit est le résultat d'événements physiques ordinaires — en postulant que, tout comme un ordinateur fabriqué est un morceau de matière orchestré de manière à réaliser des exploits cognitifs tels que la logique et l'arithmétique, le cerveau soutient la cognition par l'assemblage de ses parties matérielles en un système computationnel. L'intelligence animale n'est donc pas plus mystérieuse que le fonctionnement de n'importe quelle machine élaborée. La conscience est alors le résultat de certains types de calculs particuliers qui se produisent dans le cerveau des créatures disposant d'une conscience, et si ces calculs étaient découverts, ils pourraient en principe être mis en œuvre dans une machine, ce qui donnerait un système doté de la même forme de conscience que l'animal.

Commettre le sophisme du concret mal placé, c'est tomber dans la tentation de réifier les abstractions qui font le succès de la modélisation scientifique. Il semble que l'interprétation littérale des modèles neuro-informatiques, ainsi que la théorie computationnelle de l'esprit qui l'accompagne, soit coupable de cette substitution du cerveau, avec tous ses détails concrets, par une version mathématiquement précise et simplifiée de certains de ses processus. En succombant au sophisme du concret mal placé, nous oublions non seulement que le modèle est une abstraction, mais aussi qu'en décidant de procéder à certaines simplifications, nous risquons de ne pas laisser de place dans le cadre explicatif aux caractéristiques mêmes pour lesquelles une explication peut être recherchée par la suite. Ces modélisations, fondées sur des analogies sélectives entre machine et organisme, partent du principe que les différences entre la source de l'analogie et la cible de l'analogie peuvent être ignorées pour un ensemble circonscrit d'objectifs prédictifs et explicatifs. Il n'y a cependant aucune garantie que cette hypothèse se vérifie lorsque le cadre est étendu pour tenter d'expliquer des caractéristiques supplémentaires de la cible au-delà de la portée initiale de l'analogie.

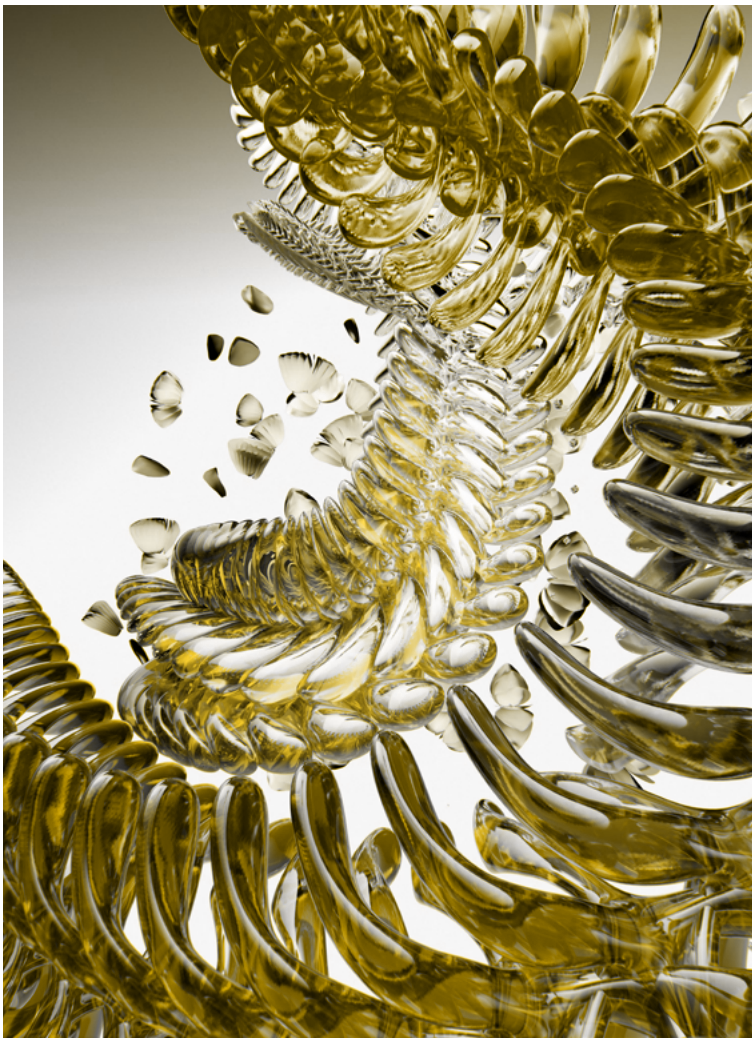


Illustration de l'intelligence artificielle (IA). Cette image explore l'IA générative et la manière dont elle peut renforcer la créativité des humains. Elle a été créée par Zünc Studio dans le cadre du projet Visualising AI lancé par Google DeepMind.

Une approche processuelle

Nous avons l'habitude de penser aux technologies en termes d'objets. Lorsque nous pensons à la technologie, nous imaginons des objets matériels comme un marteau ou des objets immatériels comme des données et des logiciels. Lorsque nous pensons à l'IA, nous imaginons un ordinateur, un robot, un logiciel, une voiture autonome … Cette façon de voir l'IA, qui s'inscrit dans une longue tradition de la métaphysique occidentale depuis Platon, voyant le monde comme une collection d'objets ou de substances, se reflète dans la philosophie de la technologie et n'a pas fortement évolué depuis.

Cependant, dans la métaphysique occidentale, nous trouvons également une autre tradition, la philosophie du processus, selon laquelle le monde n'est pas une collection d'objets, mais un processus du devenir (plutôt que d'être). Cette tradition, qui s'inspire de la doctrine du flux radical d'Héraclite (panta rhei : tout coule), a été développée dans l'idéalisme allemand (Hegel) et le pragmatisme (James, Dewey, Mead, Peirce), et est également présente dans une certaine mesure chez Heidegger, mais elle a trouvé son élaboration la plus connue dans les philosophies du processus de Bergson et de Whitehead, qui ont ensuite influencé des philosophes contemporains tels que Deleuze, Simondon et Latour.

Image de Whitehead à l'université de Harvard, vers 1924

Le philosophe français Henri Bergson a soutenu que l'intelligence individuelle a émergé dans un processus d'évolution qui exprime une force vitale, qu'il dénomma l'élan vital. Il a utilisé le terme « durée » pour parler du temps : il distinguait entre le temps tel que nous le vivons (le temps vécu) et le temps de la science, conçu comme des constructions spatiales discrètes. Cette durée n'est pas seulement le temps subjectif, psychologiquement vécu, elle est aussi quelque chose de réel, qui peut être expérimenté ou transformé en quelque chose de spatial. Il n'y a pas d'abord un temps objectif, puis notre expérience de ce temps objectif : le temps ne peut pas être isolé de notre expérience de celui-ci. Ce que nous appelons le temps objectif est produit par nos instruments, par la technologie. La seule métaphysique dont nous avons besoin, a soutenu Bergson, est celle qui reconnaît la durée et insiste sur l'émergence.

Le philosophe anglais Alfred North Whitehead a utilisé le terme « processus » : l'existence réelle est un processus du devenir. Alors que la philosophie occidentale a traditionnellement privilégié l'être sur le devenir, la philosophie du processus inverse ce paradigme. De plus, comme Bergson, Whitehead voulait aller au-delà de la division sujet-objet : il cherchait à fusionner le monde objectif des faits avec le monde subjectif des valeurs. Dans sa métaphysique du processus, les entités et l'expérience font, toutes deux, partie du devenir.

Comment pourrions-nous concevoir l'IA comme un processus, une expérience vécue et un devenir ? C'est difficile à imaginer, car nous avons l'habitude de voir et d'imaginer l'IA comme une chose, une substance. Par exemple, nous pouvons observer les résultats d'un modèle statistique (que nous considérons comme une chose), voir un ordinateur équipé d'un logiciel d'IA ou imaginer une voiture autonome conduite par un système d'IA. Cette façon de percevoir l'IA divise déjà le monde en un percepteur et une chose perçue. Nous pouvons également observer l'IA à différents moments. Le temps est alors construit de manière spatiale : comme une succession d'instants discrets. Au moment t1, « l'IA » fait x, au moment t2, « l'IA » fait y, et ainsi de suite. Cette façon scientifique de comprendre l'IA peut être mise en contraste avec notre expérience personnelle de la technologie. Par exemple, la conduite d'une voiture autonome — ou plutôt le fait d'être conduit par un système d'IA — peut être vécue comme un flux, plutôt que comme une succession de moments distincts. D'un point de vue traditionnel, ces différentes conceptions du temps s'opposent : il y a un fossé entre la technologie et le monde de la vie, entre la connaissance objective et la connaissance subjective. La philosophie des processus nous permet de remettre radicalement en question les fondements métaphysiques d'un tel fossé et de considérer l'IA, ainsi que notre relation à l'IA et notre expérience de l'IA, comme un processus plutôt que comme un objet, et plus précisément comme un processus qui fusionne des éléments « objectifs » et « subjectifs », la science et le monde de la vie, le temps scientifique et le temps vécu.

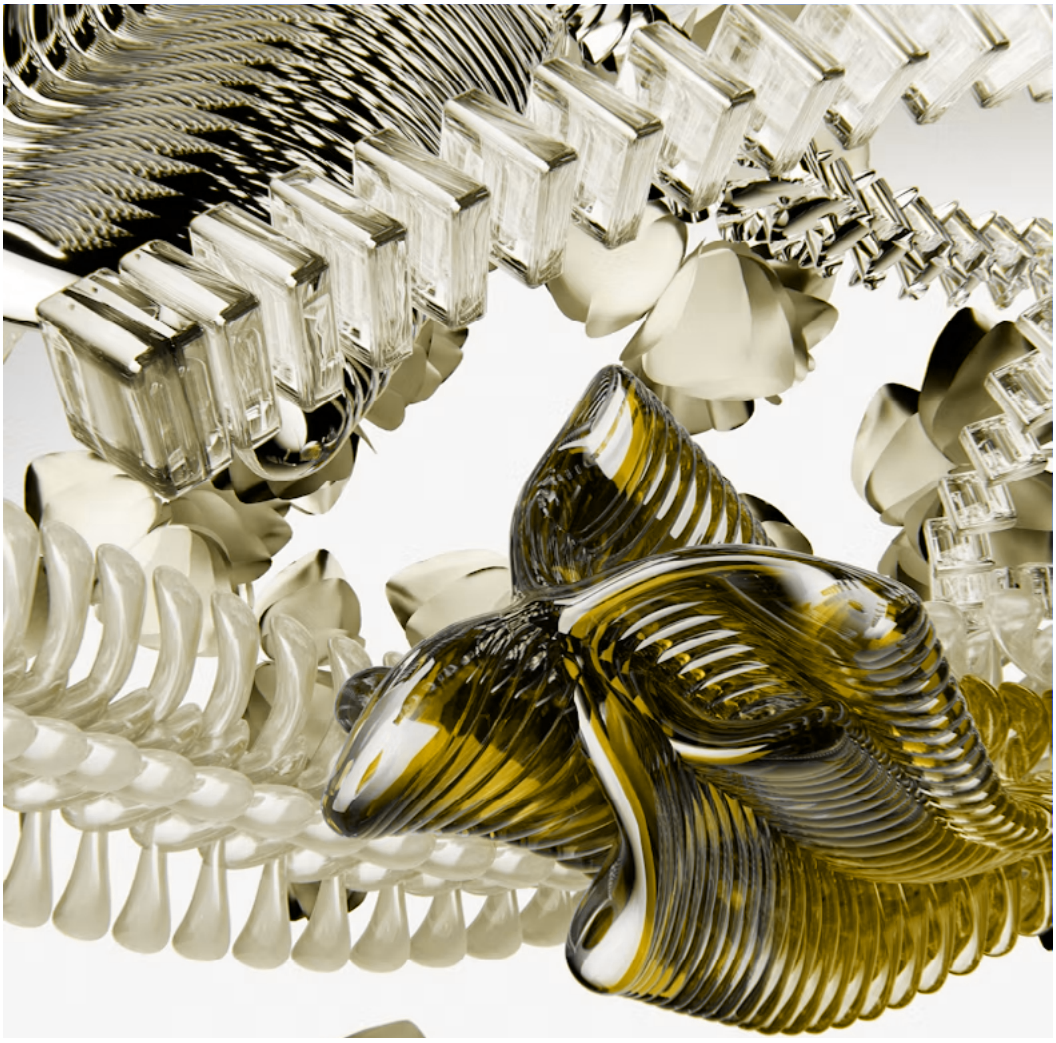


Illustration de l'intelligence artificielle (IA). Cette image explore l'IA générative et la manière dont elle peut renforcer la créativité des humains. Elle a été créée par Zünc Studio dans le cadre du projet Visualising AI lancé par Google DeepMind.

L'IA en tant que processus, plutôt qu'en tant qu'objet, fait référence à l'utilisation et au développement de processus qui se déroulent dans le temps, par exemple l'IA en tant que processus de manipulation de données. Le terme « temps » ici peut se référer à deux temps différents dans lesquels l'IA se déroule ou se développe : le temps scientifique-objectif et le temps du monde vécu, le temps vécu que Bergson conceptualisait comme la durée. Cependant, dans le processus, les deux types de temps fusionnent : l'IA dans le temps est alors à la fois mesurée/contrôlée et vécue. C'est une durée qui est à la fois expérimentée par les humains (vécue) et rendue « objective » et produite par des mesures, des technologies et des techniques de gestion. La meilleure façon de comprendre comment cela se produit est de considérer les processus de manipulation des données. Ces processus comportent diverses étapes, telles que la collecte de données, l'analyse des données, la modélisation, etc. Cette façon de percevoir/construire le processus de manipulation appartient à ce que nous pourrions appeler le temps « objectif » ou le temps scientifique. Il s'agit de gestion et de contrôle. Les étapes divisent le temps d'une manière qui le spatialise. Les différentes étapes sont des boîtes distinctes, marquant des segments de temps discrets. Mais chaque étape implique des humains, qui expérimentent, agissent et interprètent. Il n'y a pas seulement le temps façonné par le processus technologique et scientifique ; il y a aussi l'expérience humaine et l'expérience humaine du temps. Dans la pratique, à la fois, le processus technologique-scientifique et le temps vécu sont à l'œuvre. Si, conceptuellement, les deux types de temps peuvent et doivent être distingués, ils se combinent dans le processus et dans la pratique.

Ce que l'IA « est », c'est donc ce processus, voire le résultat de ce processus. Il est impossible de dire ce qu'est l'IA a priori, avant ou en dehors du processus. L'IA ne peut pas être « extraite » du temps, pas plus qu'elle ne peut être dissociée de ce que les humains font et expérimentent. Le processus peut être décrit en termes spatiaux, c'est-à-dire en termes d'étapes, mais la connaissance du processus est toujours vécue en même temps. En outre, l'IA conduit en même temps à l'émergence de sujets humains : le mesureur et le contrôleur sont le résultat du processus de mesure et de contrôle. L'utilisateur des données est façonné par le processus de manipulation des données.

L'IA façonne aussi notre passé, notre présent et notre avenir. En effectuant des classifications basées sur des données historiques, les processus d'IA peuvent nous fixer dans le passé, façonnant ainsi des présents et des futurs particuliers. Par exemple, si des données historiques provenant d'entretiens d'embauche sont utilisées pour former un algorithme de recrutement, les modes de pensée du passé — y compris les préjugés potentiels — façonneront le recrutement actuel et donc l'avenir de l'entreprise et l'histoire des personnes qui sont (ou ne sont pas) embauchées. Par le biais de la prédiction, qui influence ensuite l'action humaine, les processus d'IA façonnent le présent et l'avenir. Par exemple, si l'IA prédit qu'il y aura plus de crimes dans une zone particulière, les forces de police peuvent concentrer leurs activités dans cette zone et y prévenir plus de crimes, ce qui modifie le présent et l'avenir.

An artist's illustration of artificial intelligence (AI). This image explores generative AI and how it can empower humans with creativity. It was created by Winston Duke as part of the Visualising AI project launched by Google DeepMind.

Si nous radicalisons davantage cette approche dans le sens de la philosophie des processus, il n'y a plus d'opposition entre l'IA et les humains en tant que relata fixes dans les processus ; au contraire, ils émergent du processus. Il y a d'abord l'histoire, le processus, la relation. Nous ne partons pas d'entités fixes ; ce que nous appelons « IA » et « humains » émergent du processus, ils deviennent. Ce qu'est l'« IA » devient clair dans et à travers le processus de manipulation des données ; elle ne peut pas être définie séparément de ce processus et est plutôt le résultat qu'un ingrédient ou un outil. De même, l'humain dans ce processus n'est pas fixé dès le départ, mais devient ce qu'il est à travers le processus : il commence avec l'idée qu'il est un individu autonome, peut-être, mais il est ensuite transformé en un consommateur manipulé dans et à travers le processus. L'IA en tant que (partie d'un) processus lui donne ce rôle. Et si elle résiste, proteste, etc., elle déclenche alors un nouveau processus, qui se connecte à l'histoire existante et peut, ou non, conduire à un résultat différent.

Cela signifie aussi que nous ne contrôlons pas totalement l'IA : non pas dans le sens où l'IA opère sans notre intervention, mais dans le sens où nous ne contrôlons pas entièrement les significations et les rôles qui résultent des processus d'IA. Les développeurs peuvent avoir une interprétation de ce que leur IA « est » et « fait », mais ce qu'elle devient peut être très différent puisque d'autres interprétations sont possibles et que le résultat d'un processus n'est pas toujours entièrement prévisible.

Les relations et les rôles de l'IA ne sont pas si différents de ceux du texte. Le texte est également une technologie dont nous pouvons parler, un processus et un créateur de sens. Il possède également des propriétés émergentes et nous ne contrôlons pas nécessairement les significations et les rôles qui en découlent. L'auteur ne contrôle pas entièrement le sens du texte. Cela semble également vrai pour le développeur, dont les intentions peuvent entrer en conflit avec ce que les utilisateurs (finaux) font du programme. Et comme nous le savons d'après la tradition de réflexion sur la technologie de l'écriture, de Platon à Stiegler, les technologies constituent également une sorte de mémoire. Dans le Phèdre, Platon s'inquiétait déjà du fait que les gens cesseraient d'exercer leur mémoire parce qu'ils s'appuieraient sur l'écriture. Le texte imprimé peut être considéré comme une mémoire étendue. Comme le texte, les processus de l'IA et de la science des données fixent les connaissances du passé. Une fois qu'elles sont sur la page (texte) ou dans l'ensemble de données et traitées par l'algorithme (IA), il n'y a plus de changement en temps réel. De même que dans un texte, nous pouvons nous laisser envoûter par les pensées et les histoires du passé, les processus de la science des données peuvent empêcher le changement social en perpétuant les préjugés du passé. En même temps, il n'y a pas de déterminisme. Nous pouvons proposer différentes interprétations du texte et nous pouvons modifier l'algorithme, les données et (en principe du moins) le comportement humain.

En revanche, il existe au moins trois différences avec l'IA. Premièrement, l'IA produit un autre type de connaissances : non pas des textes (disons des connaissances linguistiques), mais des nombres, en particulier des connaissances statistiques, par exemple des probabilités, des corrélations, etc. Les processus de l'IA et de la science des données ne sont donc pas une extériorisation de la mémoire humaine, mais constituent un tout autre type de mémoire. Les processus d'IA et de science des données produisent leur propre type de connaissances, qui sont ensuite mémorisées de manière technique (bases de données, modèles). Alors que le modèle platonicien de l'écriture présuppose une mémoire préexistante chez l'homme, qui est ensuite extériorisée par l'écriture et matérialisée sur le papier, l'IA et la science des données transforment la pensée et l'expérience humaines en données, et produisent des connaissances statistiques sur ces données, que les humains concernés ne possèdent pas déjà et (en particulier dans le cas des big data et des modèles complexes) ne peuvent ni posséder ni produire. L'IA crée ainsi ses propres « mémoires », qui peuvent être très différentes du contenu de la mémoire humaine, qui est basée sur l'expérience humaine et non sur des données.

En guise de conclusion

L'intelligence artificielle a bouleversé, bouleverse, et bouleversera encore notre existence. Si l'intelligence humaine résiste à nos aspirations de la circonscrire pour mieux la définir, l'intelligence artificielle, quant à elle, se divise en deux catégories : faible et forte. Tandis que l'intelligence artificielle faible est déjà intégrée dans nos vies quotidiennes, l'intelligence artificielle forte demeure encore aujourd'hui une chimère et le restera selon nous.

L'intelligence artificielle faible a gagné ses lettres de noblesse en reproduisant, voire en améliorant, certaines tâches humaines. L'intelligence artificielle forte, quant à elle, reste tributaire d'un objectif à atteindre : reproduire, voire surpasser, l'intelligence humaine. Mais cet objectif est-il seulement clairement défini ? Nous avons vu tout au long de cet article qu'il n'en est rien. Si objectif il y a, celui-ci se base sur un modèle. Il est cartographique et non territorial. Suivre son GPS peut nous amener à notre destination, mais il peut aussi nous conduire directement dans la rivière si nous n'y prenons garde !

En outre, l'approche processuelle gagnerait à être approfondie davantage. En insistant sur le devenir et en remettant en cause la distinction classique entre le sujet et l'objet, elle nous offrirait un nouvel éclairage sur le fonctionnement de l'intelligence artificielle et de l'intelligence humaine.

Bibliographie

Andler, Daniel (2023). *Intelligence artificielle, intelligence humaine : la double énigme*. NRF Essais. Gallimard.

Chirimuuta, Mazviita (2024). *The brain abstracted: simplification in the history and philosophy of neuroscience*. Cambridge, Massachusetts: The MIT Press.

Coeckelbergh, Mark (2021). *Time Machines: Artificial Intelligence, Process, and Narrative*. Philosophy and Technology 34 (4):1623-1638.

Rosenblueth, Arturo, et Wiener, Norbert (1950). *Purposeful and Non-purposeful Behavior*. Philosophy of Science 17 (4): 318–326.