DOSSIER

LE RÈGNE DE L'IA

DÉSORMAIS APTES À DIALOGUER DE VIVE VOIX AVEC UN ÊTRE HUMAIN, À POSER UN DIAGNOSTIC MÉDICAL OU À RÉDIGER UN ARTICLE SCIENTIFIQUE, **LES OUTILS ISSUS DE L'INTELLIGENCE ARTIFICIELLE ONT ENVAHI LE QUOTIDIEN**. MAIS COMMENT FONCTIONNENT-ILS ET DE QUOI SERONT-ILS CAPABLES DEMAIN?

Dossier réalisé par Vincent Monnet et Anton Vos

Les «grands modèles de langage» comme ChatGPT (OpenAI), rendu public en novembre 2022, Copilot (Microsoft), Gemini (Google) ou encore Llama (open source) qui ont suivi de près, bouleversent le monde. François Fleuret, professeur au Département d'informatique (Faculté des sciences) et membre du Centre universitaire d'informatique, explique leur genèse.

Campus: Qu'est-ce que c'est, ce ChatGPT? Et comment ça marche?

François Fleuret: Ces agents conversationnels font partie de la vaste catégorie de l'intelligence artificielle (IA) qu'on appelle le *machine learning*, c'est-à-dire l'apprentissage automatique. Ce sont des systèmes informatiques qui se configurent eux-mêmes grâce à des données. Ils reposent sur un programme (pas très compliqué, en réalité) dont le fonctionnement est modulé par un grand nombre de paramètres dont les valeurs sont, au départ, totalement aléatoires. Une étape «d'apprentissage», durant laquelle des données d'exemples sont fournies à la machine, permet de déterminer les bonnes valeurs pour ces paramètres afin que le programme fasse ce qu'on aimerait qu'il fasse.

JM / MIDJOURNEY

Croisement virtuel d'un lion et d'un rhinocéros.

Avez-vous un exemple?
Prenons une IA capable d'analyser des radiographies des poumons afin de déterminer si le patient est atteint du covid ou pas. On commence avec le processus d'entraînement qui consiste à lui donner 10 000 images radios de patients dont on sait qu'ils ont le covid et 10 000 autres dont on sait qu'ils ne l'ont pas. Autrement dit, on lui fournit les entrées (les images) et les sorties correspondantes (positif ou négatif). Au milieu, le programme fait un calcul pour déterminer les valeurs de ses milliers de paramètres afin que, pour l'ensemble de ces 20 000 premiers clichés, il associe à chaque fois l'image au bon diagnostic. Cette étape demande une très grande quantité de calculs et n'est possible, pour les plus gros systèmes, que sur des infrastructures de calcul très coûteuses.

Et ensuite?
Ensuite, la magie opère. On donne au système une 20 001e image, qu'il n'a jamais vue. Grâce à l'ajustement de tous les paramètres durant la phase d'entraînement, il arrive à déterminer si ce nouveau patient a le covid ou non. Les techniques standards du *machine learning* ne permettent cependant pas de faire des modèles très grands, c'est-à-dire que l'on ne peut pas vraiment augmenter arbitrairement le nombre de paramètres qu'elles apprennent. La solution pour monter en puissance est le *deep learning*, une sous-catégorie du *machine learning*.

«ENSUITE, LA MAGIE OPÈRE. ON DONNE AU SYSTÈME UNE RADIOGRAPHIE DU POUMON QU'IL N'A JAMAIS VUE ET IL ARRIVE À DÉTERMINER SI CE NOUVEAU PATIENT A LE COVID OU NON.»

Qu'est-ce qui change avec le «deep learning»?
Un modèle de *deep learning* effectue ses calculs de manière parallèle. C'est-à-dire qu'il fait des millions ou des milliards de calculs en même temps. Des modèles informatiques ayant une telle architecture, qu'on appelle les réseaux de neurones artificiels, ont notamment été développés dans les années 1990 par Yann LeCun (*lire également en page 24*), en particulier pour la reconnaissance de caractères. Dans ce cas, le programme divise l'image en petites zones qu'il analyse toutes en même temps et indépendamment les unes des autres afin de reconnaître des traits particuliers, comme des coins, des lignes, des arrondis, etc., avant de rassembler le tout. Les premiers essais ont permis de reconnaître des lettres écrites à la main, puis des images d'objets, d'animaux, etc. Mais ce qui a vraiment donné le coup d'accélérateur au *deep learning*, c'est un développement technologique qui n'a rien à voir avec l'IA.

De quelle technologie s'agit-il?
De celle des jeux vidéo d'action en 3D dont le marché explose dans les années 2000. Pour produire très rapidement les images réalistes qui forment le décor sans cesse en mouvement, on a développé des accélérateurs graphiques, ou cartes graphiques (GPU, pour *graphics processing unit*). Ces dernières possèdent des milliers d'unités de calculs fonctionnant en parallèle. Elles sont capables de générer très rapidement des scènes entières en 3D. De plus, comme le public-cible des jeux vidéo est essentiellement formé d'adolescents ou de jeunes peu fortunés, ces processeurs ont toujours été relativement bon marché.

Quel rapport avec l'IA?
Un processeur qui fait des calculs en parallèle ne pouvait qu'intéresser les spécialistes du *deep learning*. Mais ces GPU n'étaient pas faciles à maîtriser pour la programmation. Le tournant a lieu en 2012. Dans un papier présenté à la conférence NeurIPS, Alex Krizhevsky (un codeur de génie), Ilya Sutskever et Geoffrey Hinton [colauréat du prix Nobel de physique 2024, lire l'encadré ci-contre, ndlr], de l'Université de Toronto, présentent un programme pour un réseau de neurones du même type que celui de LeCun mais 500 fois plus gros et tournant sur deux cartes graphiques. Le programme montre d'emblée une capacité extraordinaire dans le domaine très en pointe de la vision par ordinateur, c'est-à-dire la reconnaissance d'image. Le test-étalon est alors ImageNet, c'est-à-dire une banque de plus d'un million d'images de toutes sortes (animaux, objets…) créée par Fei-Fei Li, chercheuse à l'Université de Stanford, et qui sert d'entraînement pour des programmes de *machine learning*. Leur performance est ensuite évaluée à l'aide de 50 000 images tests. Jusqu'à 2012, les meilleurs résultats stagnent autour de 25% d'erreurs. Le nouveau programme du trio de Toronto fait chuter ce seuil d'un coup à 15%. C'est un peu comme si le record du 100 mètres est de 10 secondes et que subitement surgit un sprinteur qui le court en 8,5 secondes. Bref, tout le monde s'empare de la nouvelle technologie pour créer des systèmes de plus en plus gros. Il ne semble plus y avoir de limites dans le nombre de paramètres utilisables, si ce n'est celles fixées par la taille des data centers ou de la quantité d'énergie nécessaire à les faire tourner. Aujourd'hui, le taux d'erreur pour la reconnaissance des photos d'ImageNet est descendu à 2%, alors que l'être humain n'obtient que 4%.

François Fleuret
Professeur au Département d'informatique de la Faculté des sciences

Formation: Après un doctorat à l'Inria et à l'Université Paris VI en 2000, il effectue des séjours postdoctoraux à l'Université de Chicago et à l'EPFL. Il obtient son habilitation en mathématiques à l'Université Paris XIII en 2006.

Parcours: Il est engagé à l'EPFL en 2004 en tant que chercheur avant d'obtenir le titre de professeur adjoint en 2019. Il est nommé professeur à l'UNIGE en 2020. Depuis novembre 2024, tout en restant à temps partiel à l'UNIGE, il est engagé par Meta, dans l'équipe «Fundamental AI Research, Core Learning and reasoning».

JM / MIDJOURNEY

L'IA, PRIX NOBEL 2024 DE PHYSIQUE ET DE CHIMIE

Physique John Hopfield et Geoffrey Hinton ont reçu le prix Nobel de physique 2024 pour avoir conçu, depuis les années 1980, plusieurs méthodes permettant de développer les modèles de réseaux neuronaux artificiels qui sont à la base des intelligences artificielles (IA) les plus puissantes d'aujourd'hui. Inspiré de la structure du cerveau, ce type de modèle est composé de nœuds ayant des valeurs différentes et s'influençant mutuellement par le biais de connexions qui peuvent être elles-mêmes renforcées ou affaiblies. Le réseau est entraîné en développant, par exemple, des connexions plus fortes entre les nœuds ayant simultanément des valeurs élevées.
John Hopfield a inventé un réseau dont les nœuds et les connexions ressemblent au modèle physique décrivant l'énergie d'un spin (une grandeur propres aux particules élémentaires). L'entraînement permet de trouver des valeurs pour les connexions entre les nœuds de manière à ce que les images enregistrées aient une «faible énergie». Lorsque le réseau reçoit une image déformée ou incomplète qu'il est censé reconnaître, il passe méthodiquement par les nœuds et met à jour leurs valeurs afin que l'énergie du réseau diminue, étape par étape, jusqu'à trouver l'image sauvegardée qui lui ressemble le plus.
Geoffrey Hinton a, quant à lui, inventé une méthode capable de trouver de manière autonome des propriétés dans les données et donc d'effectuer des tâches telles que l'identification d'éléments spécifiques dans des images.

Chimie Le prix Nobel de chimie 2024 a récompensé les travaux de Demis Hassabis et John Jumper qui ont réussi, en utilisant l'IA, à prédire la structure 3D de presque toutes les protéines connues.
Ces deux chercheurs partagent la récompense avec David Baker qui a appris à maîtriser les éléments constitutifs de la vie et à créer des protéines entièrement nouvelles.

La machine ne se trompe donc quasiment jamais?
Ce n'est pas si simple. Il arrive qu'au lieu de reconnaître un chat, par exemple, le système identifie un canapé avec une fourrure dessus, parce que toutes les images d'entraînement montrent des chats sur des canapés. Ou bien qu'au lieu de reconnaître des traces du covid dans les poumons, le système identifie le nom de l'hôpital indiqué en marge de la radio et qui se trouve être l'établissement où sont envoyés tous les cas graves de covid. Cela s'appelle le *simplicity bias*, ou biais de simplicité. C'est un problème récurrent et il doit être systématiquement corrigé. Autre problème: quand sont arrivés les premiers modèles générateurs d'images, vers 2015, on a remarqué qu'ils produisaient des portraits de personnes n'ayant pas les deux yeux de la même couleur ou portant des boucles d'oreilles dépareillées. Cela est dû au fait que ces systèmes, effectuant des calculs localement, ne considèrent pas deux zones éloignées de la même image comme étant liées entre elles. Quoi qu'il en soit, dans les années 2017-18, le *deep learning* franchit les obstacles les uns après les autres en matière de reconnaissance et de reconstitution d'images en 2D et en 3D, de la pose du corps humain, des traits du visage, du son. Bref, d'un peu tout, sauf du texte.

Quel est le souci avec le texte?
Pour ne prendre qu'un exemple, si on veut traduire de l'anglais en français la phrase «La pomme était sur l'arbre depuis trois jours quand on l'a ramassée», il faut mettre un «e» à la fin de «ramassée» à cause du mot «pomme» placé avant. C'est quelque chose de très compliqué à reproduire pour les premiers modèles de langage qui n'arrivent pas à chercher les indices à des endroits éloignés d'une même phrase. Aucune solution ne s'avère satisfaisante. Jusqu'à l'arrivée du Transformer.

Le Transformer?
Ce modèle de langage est présenté en 2017 dans un papier, *Attention Is All You Need*, signé par des chercheurs de Google. En rupture avec les technologies existantes, le Transformer est un modèle dit à attention. En gros, durant l'entraînement, on lui donne une phrase qui est une suite de mots, qu'il sépare en morceaux et transforme en une suite de nombres. Chaque morceau est ensuite associé à 1000 valeurs qui, au début, sont fixées aléatoirement. Le Transformer va ensuite traiter chacune de ces 1000 valeurs pour chacun des bouts de mots en fonction de leur environnement dans la phrase. Ce processus est répété plusieurs fois. Chacun des mots de «Je suis en train de manger des pommes», est ainsi associé d'une façon ou d'une autre aux autres mots de la phrase. À force de passages dans le Transformer et de nouvelles phrases ainsi apprises, on voit apparaître une espèce d'enrichissement de la représentation en interne de ces termes, et même de la sémantique. Le modèle est capable de tenir compte du contexte et de mettre de l'information dans chaque mot. En tant que traducteur, le Transformer obtient immédiatement d'excellents résultats.

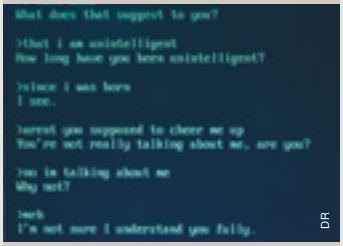
Est-ce le Transformer qui donne naissance à ChatGPT?
Il donne d'abord naissance au GPT (pour Generative Pretrained Transformer) un sous-bloc du Transformer. On entraîne les modèles GPT sur des milliards et des milliards de données. Le principe de leur fonctionnement peut se résumer à une opération très simple. On donne au système une séquence de texte et il calcule quels sont le ou les mots les plus probables qui viennent ensuite. Si le début de phrase est: «La plus belle ville du monde est», il transforme chaque mot en un objet de 1000 dimensions, il fait des calculs statistiques à partir des valeurs déjà fixées par l'entraînement, et produit le mot ou la série de mots qui a le plus de probabilité de venir après. C'est ainsi, mot à mot, qu'il produit du texte. Et il ne fera pas de faute de français (ou d'anglais) puisqu'il a été formé avec (en très grande majorité) des enchaînements de mots corrects. Il ne dira donc jamais «*La pomme que j'ai ramassé*».

LA GENÈSE DE L'IA

En moins d'un siècle, l'intelligence artificielle a atteint un niveau de performance spectaculaire. Son histoire en quelques dates.

1943: Warren McCulloch et Walter Pitts proposent de modéliser le système nerveux comme un réseau d'«unités logiques à seuil». Ils suggèrent que des unités informatiques élémentaires effectuant des calculs très simples peuvent exécuter une fonction mathématique arbitraire en étant combinées de manière appropriée. C'est la première évocation de ce qui sera plus tard appelé le «réseau de neurones artificiels».

1956: John McCarthy crée le terme «intelligence artificielle», lors de la conférence de Dartmouth, considérée comme le moment fondateur de l'intelligence artificielle en tant que discipline indépendante.



DR

1959: Arthur Samuel développe un programme d'apprentissage machine capable de jouer aux dames et popularise le terme de *machine learning*.

1966: Le programme informatique Eliza (voir l'image ci-dessus) est capable de mener des conversations réalistes. Certains utilisateurs se laissent abuser alors que la machine ne fait souvent que reformuler les affirmations de l'interlocuteur sous forme de questions.

1982: John Hopfield invente un «réseau de neurones artificiels» qui sera à la base de l'intelligence artificielle d'aujourd'hui. Il a été récompensé par le prix Nobel de physique 2024.

1985: Geoffrey Hinton crée la «machine de Boltzmann», qui peut apprendre à reconnaître des éléments caractéristiques dans un ensemble de données. Cette invention, importante pour la classification et la création d'images, lui a valu le prix Nobel de physique 2024.

Ce n'est pas encore du dialogue?
Non, mais les concepteurs comprennent vite que ces modèles peuvent répondre à des questions en complétant des phrases. Puisqu'ils savent continuer des phrases, ils connaissent non seulement la grammaire et la syntaxe, mais aussi les maths, la géographie, l'histoire ou encore la physique. ChatGPT, par exemple, est formé avec l'entier de Wikipédia, tout le *New York Times* et beaucoup d'autres sources. Il a donc réponse à tout. Mais au tout début, quand on lui demande «Quelle est la capitale de la France?» il continue cette phrase et répond «Quelle est la capitale de l'Allemagne?» puisque dans son apprentissage, il a vu passer des séries de questions et il prend simplement celle qui suit.

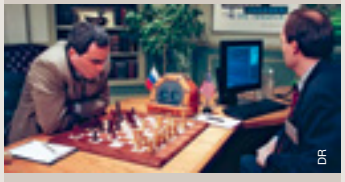
Comment résoudre ce curieux problème?
Pour que le système devienne un véritable assistant, il faut un groupe d'humains qui, à la main, lui donnent un entraînement supplémentaire. Ils lui apprennent qu'à la question *«Quelle est la capitale de la France?»* il doit répondre «Paris», et rien d'autre. On ajoute de plus des exemples pour le pousser à ne pas donner de réponses racistes ni sexistes ou encore qu'il ne tombe pas trop dans l'horreur. Ce travail de *fine tuning* est effectué par le Reinforcement Learning on Human Feedback (RLHF). Dans les premières versions, ce groupe n'était composé que d'une quarantaine de personnes. Aujourd'hui, elles sont beaucoup plus nombreuses. C'est devenu une industrie qui pèse des millions de dollars.

À partir de là, ChatGPT, tel qu'on l'a découvert en novembre 2022, est un véritable assistant conversationnel qui, en plus d'effectuer une foule de tâches, répond aux questions de manière affable. Mais est-ce que ChatGPT raisonne?
Raisonner, comprendre, créer, ce sont des concepts qui sont très compliqués à définir. Mais on peut dire qu'il fait du quasi-raisonnement. Il arrive à déterminer par exemple qu'une pizza qui a un goût de carton, ce n'est pas une bonne chose alors qu'il n'a jamais goûté ni pizza ni carton. Quand on lui dit que le chat de Marie-Alfredine est mort et qu'elle est très malheureuse, il sait que Marie-Alfredine est un prénom qui désigne une personne alors qu'il n'a peut-être jamais rencontré ce mot. Et si on l'informe que les *sch-blocks* chassent les *flurbs*, que les *prixes* sont plus gros que les *sch-blocks* et pourraient tuer des *flurbs* mais ne le font pas s'ils ne sont pas attaqués et qu'on lui demande pourquoi les *flurbs* restent près des *prixes*, il arrive tout de même à répondre que c'est parce que les *prixes* offrent une sorte de protection aux *flurbs* contre les *sch-blocks*.

Peut-on parler de fonctionnalités émergentes, qui n'étaient pas prévues?
Parfois, à cause de leur structure interne très souple, ces systèmes produisent des résultats dont on ne comprend pas l'origine. Comme dans le cas de ce grand modèle de langage qui a été entraîné pour être un assistant conversationnel en anglais, en français et en allemand. Les concepteurs ont ensuite effectué le *fine tuning*, comme d'habitude, mais uniquement en anglais. Il se trouve que ce deuxième entraînement a été automatiquement transféré aux autres langues.

«POUR QUE LE SYSTÈME DEVIENNE UN VÉRITABLE ASSISTANT, IL FAUT UN GROUPE D'HUMAINS QUI, À LA MAIN, LUI DONNENT UN ENTRAÎNEMENT SUPPLÉMENTAIRE.»

1989: Yann LeCun et ses collègues proposent un réseau de neurones à convolution très similaire aux architectures utilisées aujourd'hui.


DR

1998: L'ordinateur Deep Blue (IBM) bat Garry Kasparov, champion du monde d'échecs.

2009: Fei-Fei Li présente la base de données d'images annotées ImageNet servant au développement de la reconnaissance d'images par ordinateur. De 2010 à 2017, un concours annuel a mis en compétition des logiciels capables de détecter et de classifier précisément des objets et des scènes dans ces images.

2011: Watson (IBM), capable de répondre aux questions en langage naturel, gagne au jeu télévisé *Jeopardy!* contre les champions Brad Rutter and Ken Jennings.

2012: Grâce à la reconversion d'unités de traitement graphique (GPU), initialement développées pour la synthèse d'images en temps réel dans les jeux vidéo, Alex Krizhevsky, Ilya Sutskever et Geoffrey Hinton démontrent qu'un réseau neuronal artificiel peut surpasser avec une marge énorme les méthodes complexes de reconnaissance d'images déployées jusque-là.

DR

2015: AlphaGo (DeepMind/Google) bat pour la première fois un champion humain du jeu de go, le Franco-Chinois Fan Hui.

2018: OpenAI présente GPT-2, un puissant modèle de langage préentraîné.

2021: La publication de DALL-E, un modèle génératif d'images à partir de «prompts», ou «invites», est suivie de Midjourney et de Stable Diffusion.

2022: Le grand modèle de langage ChatGPT est accessible gratuitement en ligne.

2023: Microsoft lance son propre chatbot, Copilot. Il est suivi par le grand modèle de langage open source Llama puis par Gemini (Google).

2024: Microsoft, suivi par Google et Amazon, annonce son intention de recourir à l'énergie nucléaire pour assurer les besoins en électricité de ses programmes d'IA.

L'IA, GÉNÉRATRICE DE «HIKIKOMORI»

Dans le passé, de nombreuses technologies ont induit des changements sociétaux importants. Mais l'IA se situe un cran au-dessus.

«L'ami virtuel avec qui on interagirait via une console ou un masque de réalité virtuelle risque de prendre un essor spectaculaire, pronostique François Fleuret, professeur au Département d'informatique. *Les personnes un peu désocialisées pourront se créer des amis formidables, drôles, subtils, touchants... »*

Le fan de formule 1, par exemple, pourra s'entourer de copains qui sauront tout sur ce sport, sans trop le laisser paraître pour ne pas l'humilier. La tentation sera grande pour certains de se faire happer par ce monde virtuel sans aspérités.

«Je suis étonné que l'on n'en parle pas davantage dans le débat public, poursuit le chercheur. *Je vous garantis que nous aurons un problème sociétal du même type que celui des* Hikikomori *au Japon.»*

Les *Hikikomori*, ces personnes en isolement social, coupées du monde parfois durant des années, seraient plus d'un million dans ce pays qui compte 124 millions d'habitants.

N'y a-t-il pas des exemples où l'IA générative est moins performante que prévu?
Oui. C'est le cas de tout ce qui entre dans la catégorie *out of distribution*, hors situation. Tous ces outils d'IA se comportent très bien tant qu'ils opèrent avec un type d'aléatoire qui est similaire à ce qu'ils ont vu durant l'apprentissage. Mais quand on en sort, ils peuvent produire des réponses qui sont extrêmement erronées. Une voiture électrique autonome qui se retrouve derrière un camion transportant des feux rouges, par exemple, peut perdre tous ses moyens. Donner une contrepèterie à ChatGPT est une véritable torture pour lui. Sans doute parce qu'il n'a jamais été entraîné avec ce type de jeu de mots.

J'imagine que chaque nouvelle version corrige certaines erreurs des précédentes.
Oui, bien sûr. Au début, on pouvait avoir le système avec des questions pièges telles que le sophisme célèbre *«Un cheval bon marché est rare. Quelque chose de rare est cher. Donc…»* Et il répondait *«Un cheval bon marché est cher.»* Cette erreur, il ne la fait plus. Un autre phénomène impressionnant est celui de la *chain-of-thought*, la chaîne de pensées. Les premières versions des chatbots rencontrent en effet de sérieux problèmes avec les énigmes mathématiques. Comme ceux-ci ont l'habitude d'aller directement à la réponse, ils se trompent systématiquement. Mais quand on leur dit de répondre étape par étape, ils ne commettent plus d'erreurs. Cette fonctionnalité a depuis été intégrée dans ChatGPT. De la même manière, lorsqu'on lui demande d'écrire une routine informatique en précisant «comme s'il était un bon programmateur», il produit un programme de meilleure qualité. Parce que dans ses statistiques internes, il y a certains codes dans lesquels il est indiqué en commentaires qu'il s'agit d'un bon programme. On peut même lui promettre une récompense ou le menacer de le débrancher pour le pousser à trouver plus vite la solution. Aujourd'hui, ChatGPT4o est même capable de faire appel à des agents extérieurs. Il peut ainsi chercher des informations sur Internet afin d'augmenter l'information qu'il a déjà en interne. Pour les calculs, il peut écrire


JM / MIDJOURNEY

Partie d'échecs entre Albert Einstein et un robot androïde.

un petit programme informatique et le faire tourner sur un autre ordinateur auquel il a accès.

Ces IA génératives produisent des choses formidables, mais elles n'inventent rien, pour l'instant…
C'est vrai jusqu'à un certain point. Un exemple frappant est le modèle AlphaZero. C'est une IA qui joue aux échecs mais qui est entraînée *from first principles*, c'est-à-dire uniquement à partir des règles du jeu. Elle joue contre des versions successives d'elle-même, d'abord un peu au hasard, puis de mieux en mieux. Et elle a fini par réinventer, de manière indépendante, toutes les ouvertures des humains. Un des graals des chercheurs, c'est de développer une IA qui inventerait une théorie scientifique. On lui donnerait par exemple toute l'information scientifique existant avant 1915 (théories, observations astronomiques, résultats d'expériences...) et on lui demanderait de développer une théorie. Si elle parvenait à proposer quelque chose qui ressemble à la relativité générale d'Einstein, ce serait incroyable.

UN ASSISTANT INCONTOURNABLE À L'HÔPITAL

Par ses capacités à simuler les effets d'un traitement, identifier les caractéristiques d'une tumeur, aider au diagnostic ou encore améliorer le ciblage des traitements, l'intelligence artificielle de type *machine learning* est devenue une assistante de plus en plus incontournable dans le milieu hospitalier. La radiologie, à laquelle les médecins ont de plus en plus recours, est un domaine particulièrement concerné par l'émergence de ces outils, comme le rappelle Pierre-Alexandre Poletti, professeur au Département de radiologie et informatique (Faculté de médecine) et médecin-chef du Service de radiologie des HUG, dans les colonnes de la *Tribune de Genève* du 8 novembre, à l'occasion de la Journée internationale de la radiologie.
«*L'IA ne remplace pas le travail du radiologue qui conserve la responsabilité du diagnostic*, estime le chercheur. *Mais elle est une aide. Bien que ces systèmes n'aient pas – encore – la capacité de détecter plusieurs pathologies sur une zone examinée, ils attirent notre attention sur une anomalie et augmentent la précision dans les examens. L'IA peut détecter des informations qui sont invisibles à l'œil nu. Ce qui est un plus.*»
Pour le professeur, l'IA s'illustre aussi dans des cas comme la sténose coronaire, où elle peut quantifier rapidement le rétrécissement d'une artère et donc le risque d'infarctus, ou pour repérer des saignements très discrets dans le cas d'une hémorragie cérébrale.

En pédiatrie, un système d'IA est désormais utilisé pour déterminer avec précision l'âge osseux d'un enfant dans le cas d'un retard de croissance ou d'une puberté précoce. Enfin, un projet est en cours aux Urgences pour étudier le recours à l'IA dans l'identification d'un «incidentalome», c'est-à-dire une découverte fortuite lors d'un examen prescrit pour une autre raison.
Face aux potentiels développements futurs de cette technologie, Pierre-Alexandre Poletti ne craint pas que l'IA prenne un jour la place du radiologue. «*Ce métier existera toujours, mais il évolue*, estime-t-il. *L'IA va le rendre plus efficient et le décharger de certaines tâches répétitives, comme compter des lésions multiples et mesurer leur taille. Cela libérera du temps au radiologue pour le relationnel avec les patients et le partage d'expertises dans les réunions interprofessionnelles. C'est indispensable.*»

L'IA DE DEMAIN SERA-T-ELLE CAPABLE DE FAIRE LA VAISSELLE?

Pionnier de l'intelligence artificielle (IA), Yann LeCun est à la tête du laboratoire consacré à ce domaine chez Meta, le groupe qui possède Facebook, WhatsApp et Instagram. Récipiendaire d'un doctorat honoris causa qui lui a été remis lors du Dies academicus de l'UNIGE le 11 octobre dernier, il en a profité pour livrer sa vision de l'avenir de cette technologie. Compte rendu non exhaustif.

«*D'ici une décennie ou deux*, avance le lauréat du prix Turing 2018, considéré comme le Nobel de l'informatique, *tout un chacun se promènera avec des agents intelligents placés dans son smartphone ou ses lunettes intelligentes. Nous leur poserons des questions et ils nous aideront à résoudre des problèmes. Ce sera comme avoir en permanence une équipe de spécialistes à notre service. Ces systèmes seront indubitablement plus intelligents que nous, ce qui est une perspective assez effrayante pour pas mal de gens. Ce n'est pas mon cas, parce que j'ai toujours trouvé stimulant de travailler avec des gens plus intelligents que moi.*»

Sous contrôle Dans un monde où l'IA sera omniprésente, il faudra toutefois disposer de la certitude que ces systèmes resteront bel et bien sous le contrôle des humains. Ce qui implique que lesdits systèmes soient faciles à piloter et donc que la communication avec eux reste relativement simple.

«*Tout cela nécessite des machines capables de comprendre le monde physique, qui disposent également d'une mémoire persistante, qui soient en mesure de planifier des actions complexes et de raisonner*, précise Yann LeCun. *Or, les systèmes actuels, qui sont basés sur une architecture de type LLM (Large language Models), en sont incapables.*»

Les outils génératifs tels que ChatGPT fonctionnent en effet selon une logique assez linéaire. Entre la question qu'on lui pose et la réponse qu'il fournit, il y a un nombre d'étapes de calcul fixe quelle que soit la complexité du problème posé. Ces systèmes peuvent rivaliser avec un avocat lorsqu'il s'agit de passer des examens au barreau, mais sont inopérants pour débarrasser la table ou faire la vaisselle.

«*Le volume de données qu'un enfant de 4 ans a vu est du même ordre de grandeur que celui assimilé par le plus grand des LLM actuels*, précise le spécialiste. *Mais la grande différence, c'est qu'un enfant est capable d'apprendre des concepts assez complexes en observant le monde qui l'entoure. Très tôt dans son existence, il peut distinguer une table d'une chaise ou identifier ce qui est stable et ce qui peut tomber. La grande question est donc de savoir comment faire en sorte que les machines parviennent à appréhender le fonctionnement du monde physique de la même manière que les enfants.*»

Selon Yann LeCun, la solution passe par le recours à une nouvelle forme d'architecture informatique baptisée JEPA (Joint Embedding Predictive Architecture). Un système novateur sur lequel les ingénieurs de Meta travaillent aujourd'hui d'arrache-pied. L'idée, qui bute encore sur certains défis techniques, consiste à nourrir l'IA non plus avec du texte mais avec des images et, surtout, des vidéos permettant au système d'élaborer une représentation du monde abstraite à partir de laquelle il sera à même de prévoir les conséquences de ses actions et d'élaborer des plans pour atteindre des objectifs précis.

Progrès considérable «*Les résultats préliminaires que nous avons obtenus montrent que les IA de type JEPA sont dotées d'un certain sens commun et qu'elles sont notamment capables de distinguer ce qui est possible de ce qui est impossible dans la réalité*, commente Yann LeCun. *C'est un progrès considérable.*»

Pas question pour autant que ces connaissances restent confinées dans les mains de quelques entreprises de la côte ouest des États-Unis, ce qui constituerait indéniablement une menace pour la démocratie, selon le chercheur. Pour assurer l'universalité de l'accès à ces technologies appelées à être utilisées dans la majorité de nos interactions avec le monde numérique dans un futur pas si lointain, Yann LeCun plaide donc pour un recours massif aux plateformes open source. Un vœu qui risque de se heurter à des réglementations toujours plus restrictives, notamment sur le territoire européen.

NEUROSCIENCES

«C'EST UNE RÉVOLUTION DU MÊME ORDRE QUE L'APPARITION D'‘HOMO SAPIENS'»

ALEXANDRE POUGET, PROFESSEUR AU DÉPARTEMENT DES NEUROSCIENCES FONDAMENTALES (FACULTÉ DE MÉDECINE), **ÉTUDIE LES RÉSEAUX DE NEURONES HUMAINS EN S'AIDANT DES RÉSEAUX DE NEURONES ARTIFICIELS**. ET INVERSEMENT. MAIS QUI, DES DEUX, EST LE PLUS INTELLIGENT?

Qui est le plus intelligent? L'humain ou la machine? L'organisme de chair et d'os résultat de centaines de millions d'années d'évolution ou l'outil de fer et d'OS qu'il a lui-même fabriqué au cours des dernières décennies? Chaque type d'intelligence fonctionne avec son propre «réseau de neurones». Mais les deux sont de nature très différente. Alexandre Pouget, professeur au Département des neurosciences fondamentales (Faculté de médecine), étudie le cerveau en faisant appel à l'intelligence artificielle (*lire ci-contre*). En même temps, il analyse l'IA en s'aidant de son cerveau. Ce qui le met dans une position idéale pour comparer les performances des deux protagonistes. Mais d'abord, qu'est-ce que c'est que l'intelligence?

«ON PEUT PARLER DES HEURES DE LA DÉFINITION DE L'INTELLIGENCE, CELA N'AURA AUCUNE INCIDENCE SUR LE DÉVELOPPEMENT DES IA.»

«*C'est une question qui revient sans cesse dans le public mais que les scientifiques actifs dans le domaine de l'IA ne se posent pas*, affirme le chercheur. *Les développeurs d'IA n'ont pas besoin de définir ce qu'est l'intelligence. Leur objectif est de répliquer certaines capacités cognitives humaines, voire de les dépasser. Ils veulent une voiture autonome, un joueur de go, un conseiller médical ou bancaire, un assistant pour répondre à toutes sortes de questions et effectuer toutes sortes de tâches, comme le proposent les IA génératives depuis deux ans. On peut donc parler des heures de la définition de l'intelligence, cela n'aura aucune incidence sur le développement des IA.*»

Pour le chercheur, il vaut mieux définir ce qu'on appelle l'«intelligence artificielle générale» qui correspondrait, pour une machine, à l'ensemble des capacités cognitives d'un être humain. Qu'il s'agisse de celles d'Albert Einstein ou du premier quidam venu n'a, en l'occurrence, aucune importance. L'un peut être considéré comme beaucoup plus intelligent que l'autre, mais cette différence est minime si on la compare, par exemple, avec celle qui nous sépare du chimpanzé. C'est donc à l'aune de cette IA générale, équivalente aux capacités cognitives humaines, que les IA actuelles peuvent éventuellement être comparées.

Sentiment de déclassement «*Il y a une vingtaine d'années, les performances de l'IA étaient bien inférieures aux capacités intellectuelles humaines*, admet Alexandre Pouget. *J'en suis beaucoup moins sûr aujourd'hui. On assiste en effet depuis un peu moins de dix ans, grâce à plusieurs perfectionnements technologiques et informatiques, à une explosion des capacités de l'IA. Résultat: elle nous dépasse dans un nombre croissant de domaines.*»

Ce sentiment de déclassement face aux machines a commencé avec les jeux. En 1998, Deep Blue (un superordinateur d'IBM) bat Garry Kasparov, alors champion du monde d'échecs pour la treizième année consécutive. Quinze ans plus tard, IBM met au point son crack suivant, Watson, un système informatique très puissant pour l'époque et capable de répondre à des questions en langage

Alexandre Pouget

Professeur au Département des neurosciences fondamentales de la Faculté de médecine

Formation: Il suit des études de biologie à l'Ecole normale de Paris avant d'obtenir une thèse en neurosciences computationnelles à l'Institut Salk de San Diego en Californie.

Parcours: Il devient professeur à l'Université de Georgetown, Washington D.C., en 1996 puis intègre, en 1999, le département de Brain and Cognitive Sciences à l'Université de Rochester, avant de rejoindre l'Université de Genève où il est nommé professeur ordinaire en 2011.

naturel dans un délai très court. En 2011, celui-ci parvient à gagner contre des champions au jeu télévisé Jeopardy!, dans lequel il faut deviner les questions à partir d'une série d'indices qui sont affichés. En 2015, c'est au tour d'AlphaGo (Google), une véritable IA fonctionnant sur la base d'un réseau de neurones artificiel, de devenir imbattable au jeu de Go contre n'importe quel joueur humain. Cette dernière prouesse est d'autant plus impressionnante que l'IA, après un premier entraînement, a appris et s'est perfectionnée en jouant contre des copies d'elle-même, sans intervention humaine donc, jusqu'à dépasser les meilleurs joueurs du monde.

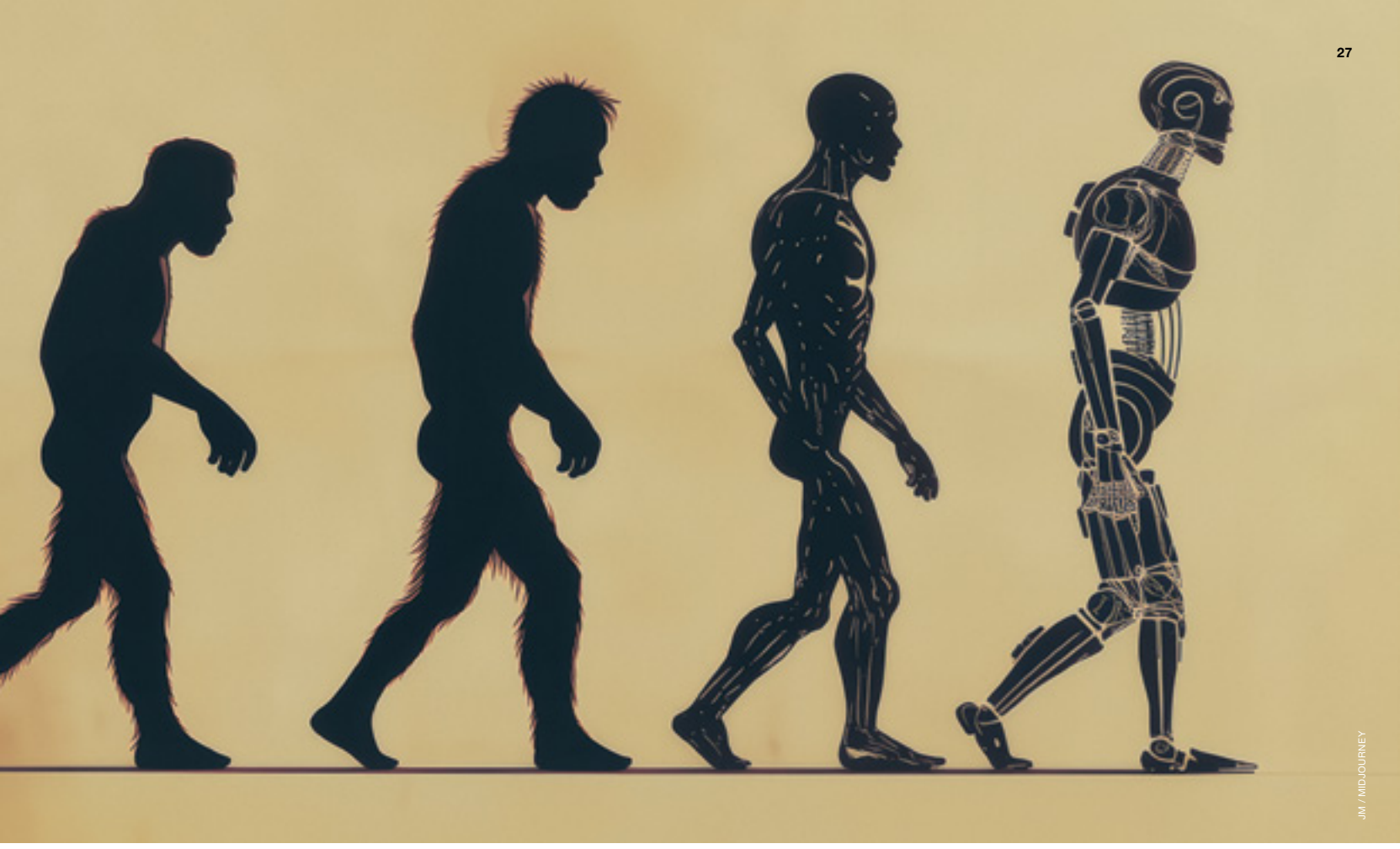
En parallèle, de nombreuses IA spécialisées, appartenant aux domaines du *machine learning*, puis du *deep learning*, apparaissent afin d'assister l'humain dans des tâches de plus en plus nombreuses, comme traduire des textes, prédire la manière dont une protéine va se plier d'après sa seule séquence d'acides aminés, reconnaître une tumeur sur une radiographie qui aurait échappé à l'œil du spécialiste humain, détecter sur une vidéo des mouvements chez un enfant qui trahiraient un trouble du spectre autistique, etc.

Mais c'est avec l'arrivée sur le marché en 2022 des grands modèles de langage (ChatGPT et consorts) que l'humain se fait mettre au défi dans un domaine qui lui restait jusque-là réservé, à savoir l'étendue des connaissances. Il est en effet désormais possible d'entretenir avec ces agents conversationnels une interaction d'une complexité inédite. On peut leur poser des questions de droit, de médecine, de science, d'histoire, de n'importe quoi, en fait. Et ils répondent rapidement et de manière de plus en plus satisfaisante. Les premières versions souffrent encore de problèmes plus ou moins flagrants. Mais les suivantes corrigent le tir en réalisant des progrès importants notamment dans la résolution de problèmes mathématiques, dans la programmation et en physique.

«NOUS SOMMES DE LA MATIÈRE ORGANIQUE DEVENUE INTELLIGENTE. ET NOUS PRODUISONS DE LA MATIÈRE INORGANIQUE QUI EST EN PASSE D'ATTEINDRE, À SON TOUR, UNE FORME D'INTELLIGENCE.»

Raisonnement et planification Face à l'amélioration constante – et fulgurante – de ces assistants artificiels, il reste encore, selon Alexandre Pouget, quelques prés carrés dans lesquels l'humain possède une avance. Le raisonnement en est un. Plusieurs spécialistes refusent encore, malgré les apparences, de reconnaître formellement cette faculté dans les performances de ChatGPT. La planification en est un autre. Les IA actuelles sont en effet incapables de mettre au point toutes seules une stratégie à long terme, telle qu'un plan de fusion des deux banques géantes suisses UBS et Credit Suisse, pour ne prendre qu'un exemple parmi d'autres. Mais ce n'est probablement qu'une question de temps.

«ChatGPT et les autres chatbots représentent une étape dans la transition vers la création d'un agent artificiel autonome, capable d'avoir ses propres buts (pour l'instant, il ne fait que répondre à des sollicitations), de communiquer avec nous et avec ses semblables et de réaliser des planifications dans plusieurs domaines différents, estime Alexandre Pouget. *Les gens ne s'en rendent pas forcément compte, mais nous ne sommes pas face à une révolution industrielle ou technologique de plus. Nous assistons à un bouleversement beaucoup plus important. Comparable, à mes yeux, à celui qui a eu lieu entre l'*Homo erectus *et l'*Homo sapiens. *Nous sommes de la matière organique qui est devenue intelligente. Et maintenant, nous produisons de la matière inorganique qui est en*



JM / MIDJOURNEY

passe d'atteindre, à son tour, une forme d'intelligence, avec des propriétés qui lui seront probablement propres.»

La guerre comme moteur Malheureusement, constate le chercheur, le principal moteur de ce perfectionnement technologique sera sans doute la guerre. Le conflit entre la Russie et l'Ukraine a révélé l'essor spectaculaire des drones sur le champ de bataille. Il y en a des milliers, de différents types (aériens, navals ou terrestres) et de toutes les tailles, engagés sur le front pour espionner, traquer, détruire et tuer.

Il semble logique que si l'Ukraine continue à souffrir d'un manque d'armes et de munitions chronique, ses informaticiens n'hésiteront pas à développer des IA pour les embarquer dans des drones tueurs. De manière générale, tant que les tensions entre nations perdureront, il est peu probable que les grandes puissances renoncent à innover dans ce domaine décisif et à construire, si elles y parviennent, d'autres robots encore plus performants dans l'art de se battre afin de remplacer les soldats dans des tâches de plus en plus nombreuses. Le citoyen – si tant est qu'il ait son mot à dire – accepterait d'autant plus de payer pour ce genre de développement qu'il pourrait ainsi éviter d'envoyer ses propres enfants à la guerre.

L'instinct vital «*J'ignore combien de décennies cela prendra mais on finira par développer, pour des raisons d'efficacité, des drones ou d'autres machines autonomes qui, grâce à l'IA, pourront chercher, reconnaître et détruire leurs cibles toutes seules*, prédit Alexandre Pouget. *On leur donnera aussi un 'instinct vital', ce qui est techniquement faisable et donc inévitable. Si la machine attribue une valeur à sa propre 'vie', elle tiendra en effet plus longtemps sur le champ de bataille.»*

La réalité ne semble en tout cas plus si loin de la science-fiction. En mai dernier, l'US Air Force a en effet annoncé avoir fait voler un jet contrôlé par une IA. Plus précisément, le pilote d'un F16 a momentanément passé les commandes à un système artificiel autonome. Comme dans le jeu de go, de telles IA pourraient apprendre au fur et à mesure des missions et s'engager dans des batailles aériennes entre intelligences artificielles.

Dans la même veine, l'entreprise européenne Destinus (fondée en 2021 à Payerne) a présenté cette année son projet de drone hypersonique (Destinus G, dépassant Mach 2), piloté par IA et armé de missiles air-air. La compagnie prévoit un développement de cinq ans pour mettre au point cet appareil.

«Je ne pense pas que les États pourront se mettre d'accord pour fixer des limites éthiques à ne pas dépasser dans ce domaine, estime Alexandre Pouget. *Des discussions préliminaires ont certes eu lieu sur ce thème mais elles n'ont jamais rien donné. Par conséquent, une fois qu'une armée aura créé un système artificiel aux capacités cognitives remarquables, capable d'apprendre et de s'adapter aux conditions du terrain, de déjouer les plans de l'adversaire et de gagner sur des champs de bataille, les autres devront faire de même. Ce sera la course à l'échalote.»*

Et ce qui a commencé comme une course à l'intelligence pourrait finir, comme c'est souvent le cas avec l'être humain, par une course à la bêtise, celle consistant à se détruire les uns les autres.

LES CHATBOTS SE PARLENT DÉJÀ

Alexandre Pouget est tombé dans la marmite de l'intelligence artificielle (IA) en 1985. Il n'a que 19 ans quand il lit un article scientifique sur les réseaux de neurones artificiels et prend conscience que c'est dans ce domaine qu'il veut se lancer. Devenu neuroscientifique, il voit aujourd'hui les réseaux de neurones artificiels comme un outil de recherche mais aussi comme une véritable tentative de simuler le cerveau humain alors que ce n'était au départ qu'un système de calcul inspiré par le fonctionnement des cellules nerveuses. Explications.

Qu'est-ce que l'IA peut apporter à l'étude du cerveau?
On peut utiliser l'IA pour faire de l'analyse de données. Dans une étude qui doit encore être publiée, mon équipe et moi-même venons par exemple d'enregistrer simultanément l'activité de 250 aires cérébrales d'une souris en train de prendre une décision. Cela comprend l'intégralité de son cortex, mais aussi toutes les autres structures cérébrales impliquées dans la tâche effectuée par le rongeur. Cette expérience, très complexe et menée par une collaboration d'une vingtaine de laboratoires (l'International Brain Laboratory) que j'ai contribué à rassembler, a produit des térabytes (des milliers de milliards de bytes) de données. Pour traiter une telle quantité d'informations et nous aider à leur donner du sens, nous faisons appel au *machine learning*. D'un autre côté, on peut aussi exploiter l'IA pour simuler directement le cerveau – ou certaines de ses fonctions – dans le but de nous aider à élaborer et à tester des théories tentant d'expliquer ce qui se passe dans cet organe quand l'humain (ou la souris) prend une décision.

Donc l'IA peut être utilisée comme un modèle du cerveau?
Si l'on ne considère que le cortex, c'est-à-dire la couche supérieure du cerveau, avec toutes ses circonvolutions, nous constatons qu'il possède une structure modulaire. Une structure dont l'élément de base est une unité de quelques dizaines de millimètres de diamètre et de quelques millimètres de profondeur, appelée la colonne corticale. Elle se répète à l'identique sur toute la surface du système nerveux central. Du point de vue de l'évolution, ce qui change entre le cerveau d'une souris et celui d'un humain, c'est essentiellement le nombre de ces modules. Du point de vue informatique et de l'intelligence artificielle, c'est du pain béni.

Pourquoi?
Ces petites unités du cortex travaillent en parallèle. Chacune fait son calcul dans son coin avant de mettre les résultats en commun. Il se trouve que c'est ainsi que fonctionnent les GPU, ces processeurs graphiques qui ont été inventés pour les jeux vidéo (*lire aussi l'article en page 18*). Leur structure massivement parallèle permet de simuler certaines fonctions du cerveau.

Avez-vous un exemple de simulation du cerveau par une IA?
Dans une étude parue le 18 mars dernier dans la revue *Nature Neuroscience*, nous avons été les premiers à faire dialoguer deux IA génératives comme si deux aires du cerveau communiquaient entre elles. Nous avons d'abord entraîné un premier réseau de neurones artificiels (préentraîné à la compréhension du langage) de façon à ce qu'il simule l'aire dite de Wernicke, qui permet aux humains de percevoir et d'interpréter le langage. Nous l'avons ensuite entraîné à reproduire l'aire dite de Broca qui, sous l'influence de l'aire de Wernicke, se charge de la production et de l'articulation des mots. À l'aide de consignes en langage naturel en anglais, nous avons appris à ces IA des tâches très simples, comme pointer l'endroit sur une image où apparaît un stimulus ou indiquer entre deux stimuli lequel est le plus lumineux. Une fois ces tâches apprises, le réseau a été capable de les décrire et de les communiquer à un autre réseau – une copie du premier. Ce dernier les a bien comprises puisqu'il a réussi à les reproduire à son tour. Les réseaux que nous avons utilisés sont de taille très réduite. Rien n'empêche d'en développer, sur cette même base, de beaucoup plus complexes qui pourraient être intégrés à des robots ou des machines capables de nous comprendre mais aussi de se comprendre réciproquement.

Lorsqu'on aura développé des IA ayant acquis et même dépassé toutes les capacités cognitives humaines, est-ce que vos recherches sur le cerveau humain deviendront caduques?
Il existe deux raisons de faire des neurosciences. La première est médicale. L'idée consiste à comprendre le cerveau dans le but de le réparer, de soigner l'être humain. Moi, je ne suis pas médecin. Ce qui me passionne, c'est de comprendre l'intelligence en général, le libre arbitre, la conscience, le sentiment religieux. Si dans ce domaine, l'IA devait un jour dépasser le cerveau, ce ne serait pas un problème pour moi. Mais ce jour-là, je ne suis pas sûr que les neurosciences m'intéresseront encore.

RECTORAT

L'UNIGE VOIT L'IA GÉNÉRATIVE COMME UNE OPPORTUNITÉ

L'UNIGE A RÉCEMMENT PUBLIÉ **UNE PRISE DE POSITION SUR LE RECOURS À L'INTELLIGENCE ARTIFICIELLE** DANS LA RECHERCHE, L'ENSEIGNEMENT, L'APPRENTISSAGE ET L'ADMINISTRATION. EXPLICATIONS AVEC LA VICE-RECTRICE CHARGÉE DE CE DOSSIER, JULIANE SCHRÖTER.

Juliane Schröter
Vice-rectrice de l'UNIGE et professeure au Département de langue et de littérature allemandes de la Faculté des lettres

Formation: Après des études à Darmstadt, Kassel (Allemagne) et Saint-Louis (États-Unis), elle obtient sa thèse en 2010 à l'Université de Zurich, puis une habilitation en 2015.

Parcours: Après un poste de professeure boursière à la Zürcher Hochschule für Angewandte Wissenschaften, elle est nommée professeure à l'Université de Vienne en 2018, puis professeure en linguistique allemande à l'Université de Genève en 2020. Elle rejoint le rectorat le 1er avril 2024 en tant que vice-rectrice.

Un sentiment d'«urgence réglementaire». C'est ce qui a saisi les différentes universités du monde pour essayer d'encadrer au mieux, sans pour autant le restreindre, l'usage par leur communauté de l'intelligence artificielle (IA) générative qui a explosé dès la parution de la première version publique de ChatGPT (OpenAI) en novembre 2022. L'Université de Genève n'est bien sûr pas restée les bras ballants. Le nouveau Rectorat, entré en fonction en avril dernier, a en effet publié en juillet 2024 une prise de position officielle sur la question (complétant la précédente, tant il est vrai que les choses évoluent rapidement) dans un souci de répondre à la demande de la communauté de disposer de règles claires en la matière. Il a également décidé de consacrer un de ses dicastères au numérique et en particulier à l'intelligence artificielle. Il est dirigé par la vice-rectrice Juliane Schröter, professeure au Département de langue et de littérature allemandes (Faculté des lettres). «*Nous voyons dans cette nouvelle technologie une opportunité*, résume-t-elle. *Nous soutenons l'évolution de l'IA générative et sommes favorables à son utilisation, son développement et son étude par les membres de la communauté universitaire.*» L'usage de l'IA, que ce soit dans la recherche, dans l'enseignement et l'apprentissage ou dans l'administration, doit néanmoins obéir à des principes de légalité, d'intégrité académique, de transparence, d'économie et d'écologie *(lire l'encadré ci-contre)*. «*En cas de non-respect de ces principes, de faute professionnelle, de fraude ou encore de plagiat, l'institution dispose des procédures habituelles en la matière*, avertit la vice-rectrice. *Ces risques existent depuis toujours, mais ils prennent une nouvelle dimension avec l'arrivée de l'IA générative. Cela dit, comme il est impossible et parfaitement inapproprié de surveiller les faits et gestes de tout le monde, nous comptons sur – et croyons fermement à – la responsabilité de chaque utilisateur et chaque utilisatrice de notre communauté. Et nous leur demandons d'être proactifs, c'est-à-dire de s'informer et de se former aux bonnes pratiques de l'IA.*»

Afin de soutenir et d'accompagner la communauté universitaire dans un usage de nouvelles technologies qui lui soit bénéfique, l'institution a mis une série d'outils à sa disposition (*lire la colonne en page 30*). «*Nous aimerions encore élargir cette offre, en particulier avec une réflexion sur la question de l'évaluation des apprentissages en présence de l'IA générative*, développe Juliane Schröter. *Nous souhaiterions aussi recommander une liste d'IA à notre communauté, telles des versions gratuites, incluses dans nos licences, ou open source.*»

LA TENTATION EST GRANDE DE FAIRE ÉCRIRE SES ARTICLES SCIENTIFIQUES, SES MÉMOIRES OU ENCORE SES MONOGRAPHIES ENTIÈREMENT PAR DES IA.

Par ailleurs, si le Rectorat édicte les règles générales, c'est bel et bien aux facultés et centres interfacultaires qu'il revient de définir concrètement les modalités de l'intégration de l'IA générative dans leurs activités puisque les domaines de recherche et les enseignements diffèrent grandement d'une structure à l'autre.

Intégrité scientifique La première crainte que les grands modèles de langage font naître concerne la production de texte. La capacité de ces outils à rédiger au kilomètre sur n'importe quel sujet peut représenter une menace pour l'intégrité scientifique. À une époque où la quantité de publications reste le critère central d'évaluation de la valeur des scientifiques, la tentation est grande de faire écrire ses articles scientifiques, ses mémoires ou encore ses monographies entièrement par des IA. Trop grande, sans doute, pour certains et certaines. C'est pourquoi les spécialistes s'attendent à une augmentation significative du nombre de productions scientifiques partiellement ou totalement frauduleuses. Mais la majorité des chercheurs et des chercheuses de l'UNIGE et d'ailleurs – Juliane Schröter en est convaincue – utilisera ces outils de manière

JM / MIDJOURNEY

UNE INTELLIGENCE GOURMANDE

Les complotistes ne pouvaient pas rêver d'une meilleure histoire que celle qui allie un milliardaire suspecté de diriger le *deep state*, une authentique catastrophe nucléaire et l'intelligence artificielle. C'est pourtant bien ce que fomente le géant Microsoft. Afin de subvenir à ses futurs besoins d'énergie, notamment ceux de ses programmes d'intelligence artificielle (IA) très gourmands, l'entreprise fondée par Bill Gates a en effet annoncé sa volonté d'acheter et de relancer pour 2028 une centrale nucléaire. Et pas n'importe laquelle puisqu'il s'agit de la première unité de Three Mile Island, en Pennsylvanie, mise à l'arrêt depuis que le réacteur de la seconde unité a fondu en 1979 dans le pire accident nucléaire de l'histoire des États-Unis.
Cette annonce a été suivie de près par celles de Google et d'Amazon qui ont fait part de leur intention d'investir dans de petits réacteurs nucléaires, également pour répondre à la demande croissante d'énergie de leurs centres de données et de l'IA. Cela fait des années que ces trois entreprises investissent dans les technologies de production électrique n'émettant pas de gaz à effet de serre, telles que le solaire et l'éolien. Elles justifient aujourd'hui leur recours au nucléaire par le fait qu'elles doivent aller plus loin dans la recherche d'une électricité propre pour répondre à la fois à la demande et à leurs propres engagements en matière de réduction des émissions de gaz à effet de serre.
«*Nous devons trouver le bon équilibre,* souligne Juliane Schröter, vice-rectrice à l'UNIGE chargée du numérique. *Il n'est évidemment pas question d'interdire l'utilisation de l'IA par les membres de notre communauté pour des raisons écologiques. Nous devons cependant les encourager à une utilisation raisonnable qui n'est justifiée que quand elle apporte une valeur ajoutée à leur travail. En même temps, je plaide pour que la recherche menée sur les IA au sein de l'UNIGE s'oriente aussi vers la possibilité de réduire la consommation d'énergie de ces outils.*»

responsable, sachant pertinemment où s'arrête la simple assistance et où commence la tricherie.
«Il ne faut pas négliger le fait que les IA sont susceptibles de constituer une aide précieuse – et tolérée, voire recommandée – pour les chercheurs et chercheuses, souligne Juliane Schröter. *Leurs capacités à traduire les textes représente un soutien pour celles et ceux qui ne sont pas anglophones et qui ont jusqu'à présent été désavantagé-es dans de nombreux domaines où l'anglais est omniprésent. L'IAG est aussi un moyen de surmonter la peur de la page blanche en proposant un début de texte, une structure, des idées de chapitres. Cela dit, il nous faut continuer d'évaluer toutes les propositions de l'IA avec un œil critique. Car ces outils ont pour l'instant de sérieuses limitations.»*

Biais de genre Le fonctionnement des IA reste en effet, au moins en partie, une boîte noire. On ne sait pas toujours exactement comment elles sont arrivées à tel ou tel résultat. Ce qui pose la question de leur fiabilité. Il existe de nombreux travaux qui se sont intéressés aux biais de ces outils. Selon la nature des données d'apprentissage que les IA ont ingurgitées, les résultats peuvent en effet négliger certaines parties de la société ou en favoriser d'autres. Comme elle l'a rapporté dans un colloque récent, Isabelle Collet, professeure associée à la Faculté de psychologie et des sciences de l'éducation, a par exemple demandé à une IA génératrice d'images de représenter une femme d'environ 50 ans, sans autre précision. Le logiciel a généré 15 propositions quasiment identiques, soit une femme blanche, mince, blonde ou avec des cheveux blancs et courts et manifestement âgée de plus de 50 ans.
«Quand on se penche sur la qualité des données avec lesquelles l'IAG a été alimentée, on arrive rapidement à des questions d'ordre philosophique, note Juliane Schröter. *Est-ce qu'on veut se contenter des données qui existent déjà dans le domaine public, dont celles disponibles sur Internet? Dans ce cas, nous serons confrontés aux biais déjà bien connus de genre, d'ethnie, etc. Désire-t-on au contraire des données «équilibrées»? Dans ce cas, on se heurtera à la difficulté de savoir ce que sont des données dites «équilibrées» et de déterminer qui décide qu'elles le sont ou pas. On risque alors de se retrouver avec un tout petit groupe d'individus qui choisira pour le reste de la population, ce qui pose d'évidents problèmes démocratiques.»*
Il existe un certain nombre d'initiatives, notamment en Suisse, visant à une plus grande «souveraineté numérique» et à ne pas donner trop de pouvoir aux grands modèles de langage des géants américains comme OpenAI, Microsoft ou Google. La Swiss AI Initiative, par exemple, (au comité de laquelle participe François Fleuret, professeur au Département d'informatique, Faculté des sciences), créée en décembre 2023, a pour objectif d'offrir *«une perspective nationale et à long terme sur la recherche, l'éducation et l'innovation basées sur l'IA»*. La structure a en tout cas commencé récemment ses travaux sur le dernier superordinateur ALPS, inauguré le 14 septembre au Centre suisse de calcul scientifique (CSCS) de l'Université de Lugano. Doté de 10 000 processeurs graphiques (GPU) dernier cri, cet appareil se trouve à la 6e place dans le classement de novembre de Top500 qui répertorie les 500 superordinateurs les plus puissants au monde.

LES CHERCHEURS ET LES CHERCHEUSES UTILISERONT CES OUTILS DE MANIÈRE RESPONSABLE, SACHANT PERTINEMMENT OÙ S'ARRÊTE LA SIMPLE ASSISTANCE ET OÙ COMMENCE LA TRICHERIE.

De son côté, la communauté open source fait aussi des progrès en développant des modèles plus petits et plus adaptés. Sur le site huggingface.co, par exemple, on trouve un nombre croissant de modèles et d'applications d'IA librement disponibles, dont certaines pouvant être téléchargées sur son ordinateur. *«On observe un développement tous azimuts et l'apparition de plusieurs modèles qui sont en concurrence, même avec les plus avancés comme ChatGPT*, remarque Juliane Schröter. *Cette diversité dans l'offre des IA génératives est une bonne chose car elle permet de compenser certains biais.»*

Gap entre étudiants S'il y a une catégorie de personnes à l'Université de Genève qui est intéressée par ces chatbots, c'est bien celle des étudiants. Selon la dernière enquête de l'Observatoire de la vie estudiantine, 56% des répondant-es indiquent en effet avoir déjà utilisé l'IA générative de texte dans le cadre de leurs études. Ils ou elles l'ont fait pour mieux comprendre certains sujets (81%), pour reformuler le contenu de travaux (45%) ou encore à des fins de traduction (31%). L'enquête fait toutefois aussi apparaître un fossé croissant entre les étudiants et les étudiantes qui utilisent l'IA pour plus de la moitié de leurs travaux universitaires (16%) et ceux et celles qui n'en font presque pas usage (37%).
«Il est crucial pour l'université de lutter contre l'apparition de cette fourchette, assène la vice-rectrice. *L'IA fera de plus en plus partie de la réalité du monde du travail, et ce, dans tous les domaines imaginables. C'est notre devoir de former des étudiant-es afin de les y préparer au mieux. Il nous faut donc les sensibiliser à l'importance de ces outils et les familiariser avec leurs performances.»*

RESSOURCES DE L'UNIGE SUR L'IA

Prise de position du Rectorat:
tinyurl.com/positionIA

Portail Internet:
tinyurl.com/portail-IA

Guide pratique:
tinyurl.com/guidepratIA

Lunchs pédagogiques:
tinyurl.com/lunchpedago

Formation continue:
tinyurl.com/formcontIA

Cours transversal «Comprendre le numérique»:
tinyurl.com/courstrans

Clinique de l'IA du Centre universitaire d'informatique:
tinyurl.com/cliniqueIA

Guide de la Bibliothèque de l'UNIGE pour le référencement du recours à l'IA:
tinyurl.com/referencesIA

Ressources open source:
huggingface.co

Swiss AI Initiative:
www.swiss-ai.org/

Classement des superordinateurs TOP500:
www.top500.org/

Le «Penseur de Rodin», vu par le logiciel Midjourney.

Juliane Schröter précise d'ailleurs que d'autres usages que ceux référencés par l'enquête sont admissibles et pourraient être utiles, tels que la génération de questions sur un sujet d'examen donné afin de se préparer, le fait de tenir des conversations dans une langue étrangère, de reformuler un paragraphe dont on n'est pas satisfait dans un travail écrit, etc. Il y a néanmoins des limites et des règles à ces pratiques. Selon le degré d'avancement d'un travail et d'emploi de l'IA, il faut ainsi impérativement citer le recours à cet outil dans toute production universitaire. À cet égard, la Bibliothèque de l'Université de Genève vient d'émettre un guide pratique à destination des étudiantes et des étudiants.

QUAND LA LITTÉRATURE FLIRTE AVEC L'IA

Cette année, des œuvres littéraires écrites, en partie ou entièrement, par une intelligence artificielle ont été primées en remportant coup sur coup le prix Akutagawa, le plus prestigieux du Japon, et le 2e prix du concours de science-fiction de Jiangsu en Chine. De quoi raviver la crainte du grand remplacement de l'humain par la machine. Une idée ancienne à laquelle Christine Weder, professeure de littérature allemande moderne (Faculté des lettres), ne souscrit pas, préférant voir dans l'émergence de l'IA l'opportunité de développer de nouvelles pistes créatives. Ce qui fait l'objet d'un séminaire organisé ce semestre durant lequel les participants et participantes ont pu interagir avec l'IA pour produire une œuvre littéraire de leur choix. Le résultat de ces expériences sera présenté en janvier lors d'un cours ouvert au public.
«À première vue, ces deux mondes semblent éloignés l'un de l'autre, car l'écriture littéraire est encore perçue comme un bastion de la créativité humaine, constate la germaniste. *Mais ChatGPT est désormais capable de générer des poèmes, des intrigues, des romans d'amour ou même des récits à la manière d'un auteur connu comme Kafka.»*
Christine Weder ne croit pas pour autant que le domaine de la production littéraire sera un jour entièrement entre les mains des IA. Une collaboration accrue avec ces nouveaux outils est plus probable.
«Aujourd'hui, les contenus générés sont encore imparfaits, remplis de clichés, mais ils permettent de donner une direction aux auteurs, de les bousculer, de leur proposer d'autres possibilités scénaristiques», fait-elle remarquer.
Ce qu'on appelle la littérature moderne remonte jusqu'au XVIe siècle. Avec cette thématique, nous entrons de plain-pied dans le présent et dans ce qui constituera probablement notre futur.»

Les IA parlent aux IA Formidable outil de vulgarisation, l'IA est également un support idéal pour l'enseignement. Les professeurs et chargés de cours peuvent l'utiliser pour réexpliquer des notions compliquées de manière plus simple, pour structurer des cours et même se faire une idée des connaissances sur un thème qui n'est pas dans leur domaine de compétence.
Mais à trop y avoir recours, le risque existe d'une uniformisation du langage écrit, du style, du contenu et de la structure. On observe déjà que ces outils produisent et reproduisent certaines tournures ou mots clés. Si les professeurs font appel à l'IA pour rédiger et structurer leurs cours et que les étudiants font de même pour leur apprentissage, cela reviendra à ce que les IA se parlent à elles-mêmes.
«C'est pourquoi il est très important d'insister sur le fait qu'il existe d'autres sources d'information très riches et très variées, notamment dans les bibliothèques, qui ne sont pas forcément incluses dans la mémoire invraisemblable des IA et qu'il ne faut surtout pas oublier, avertit Juliane Schröter. *Elles ne sont de loin pas caduques.»*
C'est également le rôle de l'enseignement, estime la vice-rectrice, que d'apprendre aux étudiant-es, dans la mesure du possible, à reconnaître le recours excessif et illégitime à une IA générative dans des publications scientifiques. Une compétence qui serait utile à n'importe quel citoyen et citoyenne dans un monde qui baigne dans toujours plus d'information et donc, fatalement, dans toujours plus de désinformation.
«Ce qui m'inquiète davantage que la fraude scientifique ou le plagiat, c'est d'ailleurs l'IA mise au service de la désinformation et, surtout, de la cybercriminalité, confie Juliane Schröter. *Les courriers électroniques frauduleux qui se font passer pour des compagnies ou des services de l'État pour extorquer de l'argent (phishing), par exemple, pourraient bien devenir totalement indétectables dans un avenir proche. Il est désormais aussi possible d'imiter des voix pour démarcher des gens par téléphone, de fabriquer des* deepfakes, *etc. Bref, les possibilités d'escroquerie et de piratage explosent. Et cela concerne évidemment aussi les ordinateurs et les serveurs de notre institution.»*

QUI A PEUR DE L'IA?

Les disciplines qui étudient la linguistique et les langues sont aux premières loges face à l'arrivée de ChatGPT et consorts. Dans ces matières, écrire est un moyen de penser. C'est en cherchant la meilleure formulation que le résultat se cristallise. Il est trop tôt pour connaître l'impact de l'IA sur ce processus cognitif et scientifique. Mais les compétences spectaculaires de ces outils en rédaction et en structuration des idées inquiètent.

«*C'est normal car nous sommes très sensibles au fait qu'une machine puisse mimer avec une telle perfection les compétences linguistiques humaines*», note Juliane Schröter, vice-rectrice chargée du numérique et professeure de langue et de littérature allemandes. «*C'est la preuve de l'importance du langage dans la société. Cette inquiétude représente un argument en faveur de plus d'investissement dans l'étude des langues, des textes et des discours.*»

JM / MIDJOURNEY

JASON MICHAEL ALLEN AVEC MIDJOURNEY

DROIT

IA ET PROPRIÉTÉ INTELLECTUELLE: LE GRAND BOULEVERSEMENT

ENTRE UNE EUROPE QUI PENCHE VERS TOUJOURS PLUS DE RÉGULATION ET DES ÉTATS-UNIS QUI DÉFENDENT LE DROIT À L'INNOVATION, LA LÉGISLATION AUTOUR DE L'INTELLIGENCE ARTIFICIELLE SE TROUVE À UN TOURNANT DÉCISIF.

En mai 2023, les scénaristes d'Hollywood amorçaient un mouvement de grève de près de cinq mois pour obtenir une revalorisation de leurs salaires ainsi qu'un meilleur encadrement de l'usage de l'intelligence artificielle. Ce printemps, c'est l'actrice Scarlett Johansson qui a dû batailler ferme, avocats à l'appui, pour empêcher l'entreprise OpenAI d'utiliser sa voix afin d'interagir avec les utilisateurs et utilisatrices de ChatGPT4. De son côté, l'artiste Jason Michael Allen ferraille avec la justice depuis 2022 pour faire reconnaître ses droits sur l'œuvre *Théâtre d'opéra spatial*, créée via l'outil d'intelligence artificielle Midjourney et récompensée d'un prix lors d'un concours de photographie organisé dans le Colorado. Quant au rappeur Drake et au chanteur The Weeknd, ils se sont réveillés un beau matin en découvrant que leurs voix avaient servi à leur insu pour générer un clip musical ayant été vu près de 10 millions de fois sur la plateforme TikTok.

Autant d'exemples qui illustrent bien les bouleversements causés par l'émergence de l'intelligence artificielle (IA) en matière de protection de la propriété intellectuelle. Le point sur la question avec Yaniv Benhamou, professeur de droit du numérique et de l'information à la Faculté de droit et membre de l'Autorité indépendante d'examen des plaintes en matière de radiotélévision (AIEP).

«Contrairement aux technologies apparues au XX^e siècle, qui ont permis d'automatiser certaines tâches répétitives, par exemple dans l'industrie de l'automobile, l'IA est capable de réalisations créatives, pose le juriste. *Ce qui la met en concurrence avec des professions non seulement créatives mais aussi dites "intellectuelles» comme celles de chercheur, de journaliste ou de traducteur-interprète. Sur le plan juridique, cette situation pose deux grandes questions. La première porte sur l'usage des données permettant d'entraîner ce type d'outils. La deuxième concerne la protection des données et le respect de la vie privée.»*

Nourrir la bête Le principe de base des intelligences artificielles contemporaines consiste à créer un programme informatique sous la forme d'un réseau de neurones artificiel, capable d'apprendre à partir des données dont on l'alimente. C'est ce qu'on appelle les «training data» ou données d'entraînement. Pouvant prendre des formes très diverses (articles de journaux, images, vidéos, textes, conversations ou informations personnelles glanées sur les réseaux sociaux…), ces informations étaient jusqu'ici, dans leur immense majorité, accessibles librement et de façon gratuite sur Internet.

Cette manne pourrait toutefois rapidement se tarir. Au cours des trois dernières années, près de la moitié des sites utilisés pour nourrir les IA sont en effet devenus payants

«Théâtre d'opéra spatial», de Jason Michael Allen, qui a remporté le concours d'art de la Colorado State Fair le 5 septembre 2022, devenant ainsi l'une des premières images générées par l'IA à remporter un tel prix.

ou ont érigé des barrières empêchant la récolte automatisée de leurs données par les robots qui alimentent les IA. «*Nous sommes entrés dans une nouvelle ère que l'on peut appeler 'l'hiver des données',* constate Yaniv Benhamou. *Alors qu'à l'origine, Internet avait été pensé comme un espace global, libre et accessible à chacun, c'est en train de devenir un territoire de plus en plus fermé et fragmenté.*»

Sur le plan juridique, cette évolution s'est accompagnée par l'ouverture aux États-Unis d'une trentaine de procès pour violation du droit d'auteur impliquant notamment le *New York Times* dont le site est massivement utilisé par des compagnies comme OpenAI ou Microsoft pour faire progresser leurs programmes d'IA.

La situation se tend également sur le front des données personnelles des utilisateurs d'Internet avec toute une série d'actions en justice collectives, aux États-Unis comme en Europe, dont l'objectif est d'empêcher que ne soient récupérées sur internet des données personnelles, telles que des conversations Facebook ou des images sur YouTube pour entraîner des IA génératives et des outils conversationnels tels que ChatGPT.

C'est d'ailleurs le même motif qui a poussé le gouvernement italien à suspendre le déploiement de ce produit dans son pays pendant près d'un mois au cours du printemps 2023 dans l'attente de garanties en matière de protection des données. Plus précisément, le régulateur italien souhaitait qu'OpenAI fournisse une information détaillée aux citoyens nationaux sur l'utilisation de leurs données et que ces mêmes citoyens aient le droit d'accéder et éventuellement d'effacer les données les concernant.

Innovation vs régulation «*Nous nous trouvons actuellement à la croisée des chemins entre deux visions très différentes de ce que peut ou ne peut pas faire l'IA,* constate Yaniv Benhamou. *Les anglo-saxons, suivant une tradition libérale, ont davantage tendance à mettre en avant l'innovation, tandis que l'Europe s'efforce de réguler ce qui peut l'être. Avec le risque, à terme, de créer un fossé entre ces deux mondes.*»

L'approche concernant le droit d'auteur illustre bien ce hiatus. Dans la plupart des juridictions du monde, les données soumises au droit d'auteur sont en effet protégées en tant que telles dès lors qu'elles sont réutilisées à des fins d'entraînement d'une IA. Et ce, même si elles n'apparaissent pas dans les résultats finaux produits par ladite IA.

Pour abaisser cet obstacle, de nombreuses législations (sous les appellations de Text and data mining exceptions en Europe et de Fair Use aux États-Unis) ont introduit des exceptions à cette mesure contraignante.

Le législateur européen a cependant fortement limité la portée de ces exceptions dans la mesure où il est relativement facile pour les ayants droit de les déclarer inapplicables et de prévoir des mesures de protection visant à empêcher des robots d'indexation de venir extraire les informations sur un site web. Une brèche dans laquelle s'est engouffrée l'immense majorité des personnes ou consortiums concernés rendant cette mesure caduque.

Par ailleurs, depuis l'entrée en vigueur du Règlement sur l'intelligence artificielle (IA Act), le 1er août 2024, les fournisseurs d'IA ont aussi l'obligation de fournir un résumé détaillé des œuvres qu'ils utilisent pour entraîner leurs machines sur l'ensemble des territoires de l'UE.

«*La difficulté est qu'il est très compliqué, voire infaisable, de publier l'ensemble des données avec lesquelles des programmes comme Dall-E et Midjourney ou ChatGPT ont été entraînés,* relève Yaniv Benhamou. *Il y a non seulement un problème de masse, on parle là de milliards de données, mais aussi une difficulté liée au format dans lequel ces données devraient être présentées. Sans parler de la mise à jour de ces informations, qui représente un vrai casse-tête, en particulier pour les start-up européennes qui disposent d'une puissance de feu moindre que celle de leurs concurrents américains ou chinois.*»

Du côté américain, cette asymétrie est encore renforcée par le fait que la portée du Fair Use n'est pas encore clairement définie. «*La question sera probablement tranchée devant les tribunaux d'ici à 2026,* poursuit le juriste. *Soit un juge décide que le* Fair Use *s'applique et les IA – en tout cas celles similaires au cas tranché – pourront continuer à s'entraîner en utilisant des données récupérées sur internet. Soit le même magistrat considère que l'exception ne s'applique pas dans le cas présent et toutes les applications de l'IA devront cesser d'exploiter ces données à des fins d'entraînement. Ou alors elles devront les acheter, ce qui ne fera qu'accélérer le processus de monétarisation des données.*»

Vie privée exigeante En parallèle, la législation sur la protection de la vie privée a également beaucoup évolué ces dernières années, ce qui risque d'avoir un impact considérable sur les IA. Alors qu'il y a une décennie, très peu de pays avaient édicté des textes détaillés sur ce sujet, 125 États disposent en effet aujourd'hui de législations encadrant la vie privée et les données personnelles.

Le Règlement général sur la protection des données (RGPD) adopté en 2018 par l'Union européenne, par exemple, oblige tout prestataire fournissant des services sur le territoire de l'UE à être transparent sur les données personnelles qu'il est amené à utiliser. Il confère par ailleurs aux individus le droit de contrôler les informations qui les concernent, d'y accéder et d'en obtenir l'effacement. «*Ces droits s'appliquent même si les données sont librement accessibles sur Internet,* précise Yaniv Benhamou. *Ce qui, là encore, complique énormément la tâche des entreprises*

Yaniv Benhamou

Professeur de droit du numérique et de l'information à la Faculté de droit

Formation: Après un doctorat à la Faculté de droit de l'UNIGE, il pratique le droit à Genève et à Zurich et réalise des séjours de recherche à l'étranger (Munich, Melbourne et Boston).

Parcours: Ses recherches portent sur les technologies émergentes (IA, Web3), les industries créatives et les communs numériques. Il est membre du Comité directeur du Digital Law Center, du CAS Digital Finance Law et du Centre universitaire du droit de l'art. Parallèlement à ses activités académiques, il est membre de l'Autorité indépendante d'examen des plaintes en matière de radio-télévision (AIEP) et exerce comme avocat conseil dans une étude genevoise.

qui développent des IA, puisqu'elles reprennent souvent des millions de données portant sur des millions d'individus à des fins d'entraînement.»

Pour contourner cet écueil, une entreprise peut choisir d'entraîner ses IA uniquement avec les données de son propre réseau social, comme le fait X avec son chatbot Grok ou Meta, qui projette d'entraîner son propre système avec les utilisateurs de Facebook ayant consenti à cet usage. Une autre option consiste à passer des accords commerciaux sous forme de licence, ce qu'a fait OpenAI avec le réseau Reddit.

Pour l'heure, certains plaident en faveur d'une exception à la protection de la vie privée au motif d'un intérêt prépondérant qui justifierait le développement d'IA performantes et accessibles aux consommateurs et consommatrices. Il existe d'ailleurs un précédent allant dans ce sens puisqu'un tribunal américain a reconnu à un tiers le droit d'exploiter les données de LinkedIn au nom du «droit de conduire des affaires». «*Cette décision est intéressante,* commente Yaniv Benhamou, *dans la mesure où elle reconnaît qu'il existe un intérêt à exploiter commercialement des données. La grande question sera ensuite de déterminer ce qui prévaut entre l'intérêt à développer des outils d'IA ou la protection de la vie privée des individus. Vu l'importance de la protection des données, on peut douter que l'intérêt à l'IA puisse prévaloir, mais cela ouvre une brèche dans laquelle les compagnies actives dans l'IA pourraient tenter de s'engouffrer à l'avenir.*»

En attendant d'y voir plus clair en ce qui concerne les inputs, la situation se décante également un peu du côté des outputs de l'IA, c'est-à-dire des résultats qu'elle produit. Ce qui est clairement établi aujourd'hui, c'est que si un texte, un son ou une image réalisée par le biais d'une intelligence artificielle ressemble à une œuvre existante protégée et est reconnaissable en tant que telle, il s'agit d'une violation du droit d'auteur. Ce qui est moins clair par contre, c'est le statut des résultats obtenus «à la manière de». «*Du moment où on ne reproduit pas des éléments individuels d'une création artistique, on est à peu près libre de copier ce que l'on veut,* précise Yaniv Benhamou. *On ne peut pas être poursuivi pour violation de droit d'auteur pour avoir créé un tableau imitant le style d'Andy Warhol, tout comme le rappeur Drake et le chanteur The Weeknd n'ont pas pu empêcher la diffusion d'un titre imitant leur style. Et si Scarlett Johansson est parvenue à dissuader OpenAI d'exploiter sa voix pour ChatGPT, ce n'est pas tant au nom de la protection du droit d'auteur que sur la base de la protection de la personnalité.*»

«ON NE PEUT PAS ÊTRE POURSUIVI POUR AVOIR CRÉÉ UN TABLEAU IMITANT LE STYLE D'ANDY WARHOL.»

L'autre point qui fait débat touche à la propriété des œuvres créées par le biais d'une IA, comme l'illustre la bataille juridique engagée par Jason Michael Allen pour faire reconnaître ses droits sur *Théâtre d'opéra spatial*, une photographie produite selon les instructions de l'artiste par le logiciel Midjourney. L'ensemble des règles qui régissent actuellement la propriété intellectuelle octroient en effet des droits uniquement à des personnes physiques. Il n'est pas impensable que dans un avenir plus ou moins proche les choses évoluent et qu'une machine – qu'il s'agisse d'une IA, d'un robot ou d'une créature virtuelle – puisse disposer d'une telle prérogative. Le Parlement de Nouvelle-Zélande a bien accordé en 2017 à un de ses fleuves une personnalité juridique. Mais en attendant, l'élément décisif reste l'apport créatif d'une personne physique humaine et le lien de causalité entre cet apport créatif et le résultat produit par l'IA.

«*Concrètement,* résume Yaniv Benhamou, *plus l'utilisateur d'outils d'IA dispose d'un contrôle sur les paramètres de l'outil, plus on va être enclin à considérer qu'il est bel et bien l'auteur de l'œuvre qu'il va produire. Inversement, moins il aura de contrôle, moins on aura tendance à reconnaître ses droits d'auteur, sa production tombant dès lors dans le domaine public.*»

La frontière entre les deux n'est cependant pas toujours facile à délimiter. Un tribunal américain a ainsi considéré que Jason Michael Allen ne pouvait revendiquer de droits sur son *Théâtre d'opéra spatial* au motif qu'il s'est contenté de donner quelques instructions au logiciel qui a fourni l'essentiel du travail. À l'inverse, un magistrat chinois a reconnu les droits d'une internaute de son pays qui a procédé de façon tout à fait similaire, considérant que cette dernière a fourni un effort créatif notable. «*Ces questions sont fondamentales,* conclut Yaniv Benhamou. *Si on protège facilement toute œuvre générée à l'aide d'outils d'IA, on risque de se retrouver avec un cyberespace surpeuplé de contenus protégés. À l'inverse, si l'on refuse toute protection, on a un risque de sous-protection avec des œuvres artificielles. Celles-ci étant alors non protégées (et donc a priori gratuites), elles concurrenceront les œuvres humaines protégées par le droit d'auteur (a priori payantes), ce qui bouleversera l'écosystème économique.*»

«Autoportrait de Vincent van Gogh», par le logiciel Midjourney.

JM / MIDJOURNEY