Le Monde
SAMEDI 4 OCTOBRE 2025

De quelle façon l'IA altère-t-elle notre pensée ?

ENTRETIEN

La philosophe **Anne Alombert** et l'entrepreneur **Jonathan Bourguignon** croisent leurs regards sur l'intelligence artificielle, son impact sur notre manière de penser et les risques qu'elle soulève

On s'émerveille de leur intelligence, on rit de leur bêtise. Le succès des agents conversationnels comme ChatGPT ou Perplexity ne se dément pas : ces systèmes appuyés sur l'intelligence artificielle (IA) générative comptent des millions d'utilisateurs fidèles. Mais que fait-on, exactement, lorsqu'on sollicite quotidiennement ces grands modèles de langage ? Et peut-on penser avec eux ? La philosophe Anne Alombert, maîtresse de conférences à l'université Paris-VIII et autrice de *De la bêtise artificielle* (Allia, 144 pages, 8,50 euros) en débat avec Jonathan Bourguignon, auteur d'*Internet, année zéro* (Divergences, 2021) et cofondateur de Squadra, un studio de création d'assistants intelligents.

Par rapport aux précédentes innovations comme l'écriture, l'imprimerie ou, plus récemment, les moteurs de recherche, l'IA est-elle d'une nature extraordinairement différente ?

Jonathan Bourguignon : Chaque découverte qui a un lien avec nos capacités cognitives soulève des inquiétudes, c'est vrai depuis l'invention de l'écriture. Lorsque le livre de poche est arrivé en France, en 1953, il se trouvait des gens pour prédire que la production de masse d'une littérature bas de gamme allait ravager les esprits – et le même argument était déjà utilisé contre l'imprimerie à la fin du XVe siècle, ou contre Internet dans les années 1990. Dans les trois cas, le public avait soudainement accès à une masse de connaissances inédites, là où les critiques considéraient qu'il valait mieux privilégier l'accès de tous à un nombre restreint de textes importants. Les algorithmes et l'IA générative n'échappent pas à ce schéma, mais ils génèrent aussi une nouvelle inquiétude, celle d'une dépossession.

La « première » version de Google présentait encore, pour une même demande, une information identique à tous les utilisateurs. Lorsque l'entreprise a changé de modèle économique et mis en place un système d'enchères publicitaires, dans lequel des annonceurs paient pour diffuser leurs publicités, le succès de ce système est devenu dépendant de la capacité de Google à prédire le comportement des internautes grâce à l'analyse des données de trafic. Cette rencontre d'Internet avec le *machine learning [« apprentissage de la machine »]* a ouvert la voie au Web 2.0, caractérisé par les plateformes numériques et les algorithmes de prédiction et de recommandation – comme ceux qu'on retrouve aujourd'hui sur Facebook, Netflix ou bien d'autres services. Désormais, plus besoin de chercher un contenu ou une information : ils s'imposent à nous. Et l'arrivée de l'IA générative, qui permet de générer du contenu et formate spontanément l'information pour l'adapter à nos demandes, est bien sûr une nouvelle rupture fondamentale.

Anne Alombert : Le point commun des technologies que vous avez mentionnées (l'écriture, l'imprimerie, les moteurs de recherche et l'IA), c'est que leur usage implique une délégation de certaines de nos capacités intellectuelles, psychiques, mentales. On se représente à tort la pensée comme un processus immatériel, qui se produirait dans le ciel des idées, alors que c'est une activité tout aussi corporelle et incarnée que les autres, qui nécessite notamment des outils auxquels nous déléguons des capacités. A travers l'écriture, puis le livre, on délègue la mémoire : plus besoin de se remémorer par nous-mêmes les savoirs. Avec les technologies d'enregistrement analogique comme la photographie, la phonographie, la télévision, on délègue la mémoire des sons et des images. Avec le cinéma, on délègue l'imagination, là où le livre nous obligeait encore à produire des images mentales propres à chacun. Avec le numérique, nous déléguons de nouvelles capacités. Aux algorithmes de recommandation, notre capacité de jugement et de décision : plus besoin de chercher et de choisir de regarder tel ou tel contenu. Et aux IA génératives, notre capacité d'expression. Ce n'est plus moi qui m'exprime avec mes propres mots, qui fabrique mes propres images, mes propres sons : les machines le font à ma place.

Peut-on dire que les IA génératives produisent une forme de raisonnement proche de celui des humains ?

J. B. : D'abord, je veux bien entendre la définition philosophique de « raisonnement » !

A. A. : C'est impossible, elle change avec chaque philosophe !

La plupart des chercheurs technoscientifiques en IA soutiennent, depuis très longtemps, que l'on peut formaliser la pensée sur la base d'opérations logiques, c'est-à-dire la réduire à du calcul. Ces mêmes opérations logiques peuvent être traduites sous forme de signaux dans des circuits électroniques – les ordinateurs. Comme on a ainsi réduit la pensée au calcul, et qu'on a fabriqué des machines à calculer, on en conclut que les ordinateurs « pensent ». Pourtant, on peut défendre d'autres définitions, et dire que la pensée provient du désir, du rêve, de l'imagination…

Quoi qu'il en soit, il faut bien comprendre que raisonner ou réfléchir ne sont pas des processus qui adviennent seulement dans notre cerveau à travers des connexions neuronales. Cela implique toujours ce que j'appelle des « prothèses cognitives » : un crayon, une feuille de papier, une machine à traitement de texte, un boulier, une calculatrice, un supercalculateur, un *large language model [LLM, modèle capable de comprendre et de générer des textes]*… Les machines dont nous parlons, elles, n'ont ni corps vivant ni prothèse cognitive. Comparer les humains aux machines n'a donc pas grand sens. C'est un peu comme si vous me demandiez : « Est-ce qu'une feuille de papier mémorise ? », ou : « Est-ce que la chaise s'assoit ? » La chaise conditionne la manière dont je m'assois. Et les machines IA conditionnent la manière dont nous pensons : ce qui est important, c'est de savoir comment, à l'échelle individuelle et collective.

J. B. : Je vais me faire l'avocat du diable, mais, pour cela, je dois expliquer les quatre étapes du processus d'entraînement des LLM comme celui de ChatGPT. Tout commence par le modèle de fondation : à ce stade, il apprend juste à compléter du texte. La deuxième étape est l'entraînement supervisé : par des exemples, on apprend au modèle à se comporter comme un assistant serviable. La troisième est l'alignement éthique et social : pour que le modèle soit poli, respectueux, utile et sûr dans toutes situations, on lui donne des règles et des exemples de ce qu'il peut dire ou pas. La dernière étape est le renforcement avec vérification. Le modèle de fondation permettait d'explorer des chemins de pensée à travers le langage, mais sans pouvoir vérifier s'ils étaient corrects. Le renforcement avec vérification lui permet, dans certains domaines, de vérifier ses propres hypothèses et d'apprendre de ses erreurs.

Si les machines ne « raisonnent » peut-être pas, elles sont aujourd'hui capables de produire des raisonnements, avec des étapes qui mènent d'un point A à un point B, d'un problème à sa résolution, en suivant une voie originale. Les modèles se forgent une sorte d'instinct, que leur perfectionnement rend de plus en plus fiable. Désolé pour cet anthropomorphisme.

« CHAQUE DÉCOUVERTE QUI A UN LIEN AVEC NOS CAPACITÉS COGNITIVES SOULÈVE DES INQUIÉTUDES, C'EST VRAI DEPUIS L'INVENTION DE L'ÉCRITURE »
JONATHAN BOURGUIGNON

A. A. : Vous n'y êtes pour rien, on a beaucoup de difficulté à éviter ce vocabulaire. Cela commence d'ailleurs avec le mot « intelligence », qu'on ne devrait pas utiliser pour ces modèles.

Qu'entendez-vous par « instinct », Jonathan Bourguignon ?

J. B. : Chez l'humain, que ce soit en mathématiques, en économie ou dans d'autres disciplines, l'apprentissage consiste à démontrer, puis à intérioriser des théorèmes, des lois, des procédures jusqu'à ce qu'ils deviennent naturels, presque instinctifs. Il en est de même pour apprendre à conduire ou à écrire : au début, tout demande un effort conscient, puis une partie du processus devient automatique. Eh bien, les modèles d'IA, eux aussi, renforcent leurs propres comportements : ils se créent un instinct.

A. A. : Les automatismes, c'est vrai, sont à la base de l'apprentissage. Mais les êtres que nous sommes – qui désirent, rêvent, sentent, cherchent du sens – ne produisent d'idées originales que lorsqu'ils sortent de ces automatismes, lorsqu'ils « dés-automatisent ». C'est, par exemple, le pianiste qui, après avoir fait ses gammes, propose une interprétation inédite. C'est encore le mathématicien qui, à l'instar de Jules-Henri Poincaré *[1860-1934]*, fait une découverte en posant le pied sur le marchepied d'un bus, après avoir travaillé en vain pendant des semaines sur des calculs minutieux. Les idées nouvelles s'appuient sur une accumulation d'apprentissages, mais elles émergent dans un moment de relâchement, presque par surprise.

On est très loin de la notion d'« aléatoire » qui caractérise les machines algorithmiques. Quand on lance des dés, le résultat est imprévisible. Mais il n'est pas « nouveau », car toutes les possibilités sont répertoriées. Le lancer de dé ne fait que recombiner différemment les résultats. Avec l'IA, c'est un peu la même chose : on recombine toutes les données passées, ce qui donne des réponses imprévisibles (car on a introduit de l'aléatoire dans les calculs probabilistes). Mais pas nouvelles. Ce qui est « nouveau », ce n'est pas le résultat d'une combinaison. On ne peut pas le générer à partir des calculs du passé.

J. B. : Je pense, pour ma part, que les IA peuvent générer du nouveau. Exemple : en 2016, le programme d'IA AlphaGo, développé par DeepMind, une entreprise rachetée par Google deux ans plus tôt, a pour la première fois battu le champion du monde de go *[Lee Sedol]*. Or, à la différence des échecs, le go était jusque-là considéré comme hors de portée d'une victoire des machines, en raison du nombre astronomique de combinaisons possibles. AlphaGo a surmonté cette difficulté, en suivant des stratégies originales, auxquelles jamais aucun grand maître n'avait eu recours, malgré une sagesse ancestrale transmise sur des milliers d'années.

A. A. : Oui, mais il n'a pas créé de nouveau jeu.

Une IA ne pourrait-elle pas créer un nouveau jeu ? Certes, elle s'appuierait sur les jeux qui existent déjà. Mais n'est-ce pas ce que nous faisons aussi ?

A. A. : Oui, nous nous appuyons sur l'expérience. Mais quand nous produisons du nouveau, ce n'est pas sur la base de calculs s'appuyant sur des données. Ou alors nous sommes de simples machines à calculer…

Est-ce exclu ?

A. A. : Il n'y en a aucune preuve. Où sont ces calculs, vous les avez vus ? On compare allègrement l'homme à la machine, mais on n'a aucune preuve que la pensée humaine fonctionne de cette façon. Alors, bien sûr, on peut définir la nouveauté comme « ce qui est différent des données qu'on a accumulées », mais c'est une définition très faible.

La force de ces machines, c'est qu'elles se nourrissent de toutes les données humaines ; mais le problème, c'est qu'elles se nourrissent aussi de plus en plus des données qu'elles génèrent. Il se produit un phénomène de dégradation cumulative. Un peu comme quand vous faites la photocopie d'une photocopie d'une photocopie… au bout d'un certain temps, l'image sera de moins bonne qualité. Pour cette raison, un effondrement des modèles n'est pas exclu. Il ne faut jamais oublier que si ces machines sont efficientes, c'est parce que nous les avons nourries avec des savoirs humains, donc elles simulent l'intelligence humaine. Le jour où elles seront alimentées par des bêtises, elles simuleront la bêtise.

L'IA menace d'atrophier certaines capacités humaines comme la mémoire, le raisonnement, l'autonomie de la décision. Concrètement, comment cela se passe-t-il ?

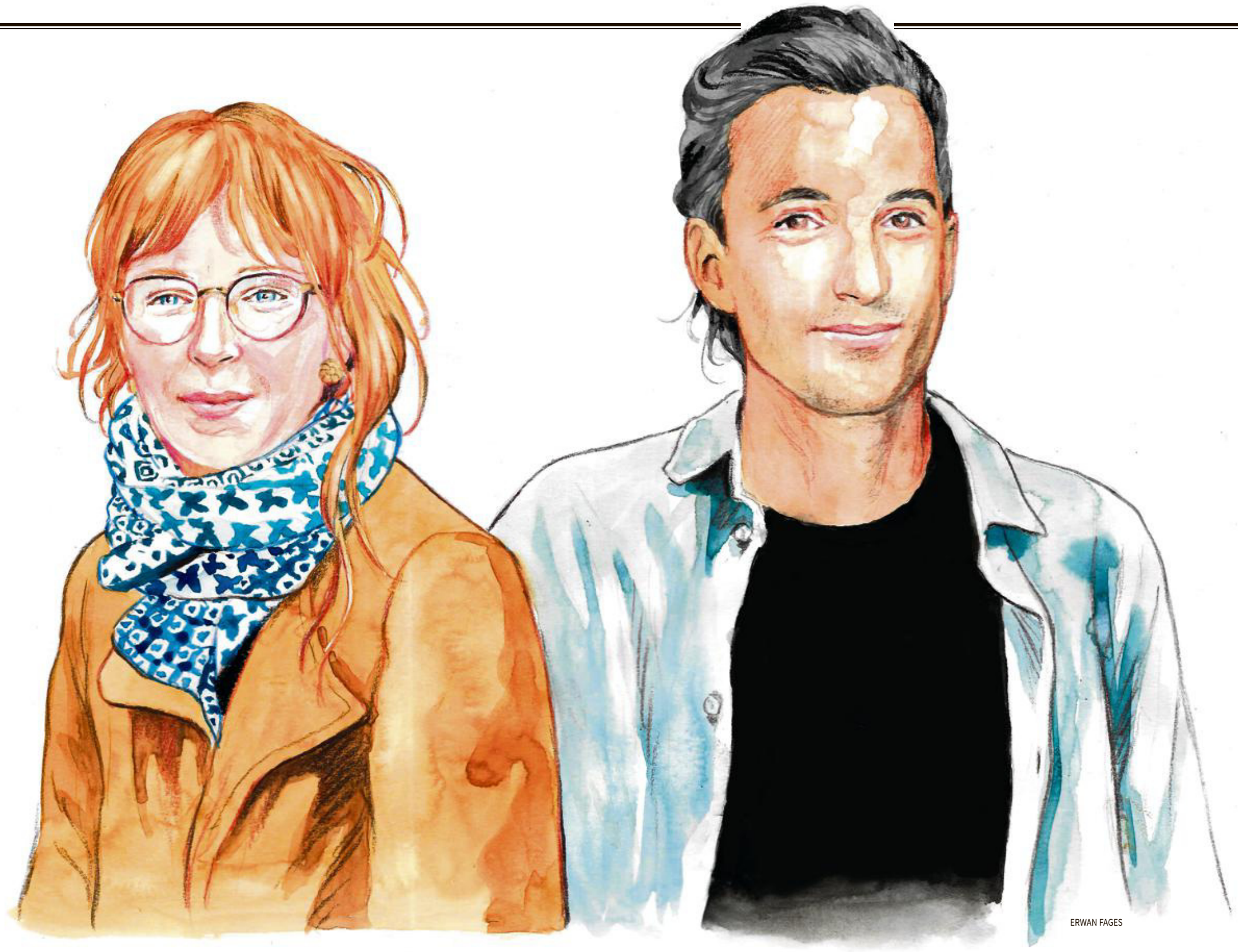
A. A. : Une récente étude du MIT *[Massachusetts Institute of Technology]* s'est penchée sur la question. Les chercheurs ont constaté que quand on utilise ChatGPT, certaines zones du cerveau sont mobilisées – essentiellement le cortex visuel – et d'autres non – spécifiquement, celles liées à la compréhension et à la production du sens. Ils en concluent que l'utilisation quotidienne de ChatGPT entraîne une « réduction de l'amplitude cognitive » (une baisse de la diversité des zones cérébrales mobilisées) doublée d'une « dette cognitive » : lorsqu'on demande aux utilisateurs de ChatGPT d'écrire un texte sans son aide, ils rencontrent de plus grandes difficultés, car ils ont du mal à réactiver les zones cérébrales qui ont été délaissées. Ils deviennent, en réalité, dépendants de l'outil.

Ce phénomène n'est pas nouveau. On sait, par exemple, que l'apprentissage de la lecture et de l'écriture joue un rôle dans la synaptogenèse (la formation de nos synapses) : apprendre à lire et à écrire, puis pratiquer ces activités régulièrement, va exercer certaines zones de notre cerveau et pas d'autres. Mais si on ne pratique plus ces activités, le cerveau désapprend ce dont il n'a pas besoin. Utiliser ChatGPT pour faire ses devoirs, c'est prendre le risque de ne jamais développer suffisamment certaines zones cérébrales ; l'utiliser dans la vie de tous les jours, c'est prendre le risque d'amoindrir nos capacités de mémorisation, d'imagination, de décision, d'expression ou de jugement.

Ce n'est pas le seul problème : il y a aussi la question des biais. Ces machines sont entraînées par des personnes, selon certains critères, certaines valeurs. Leurs réponses portent la trace de biais idéologiques. Il y a aussi des biais statistiques : les réponses sont très dépendantes du contenu du jeu de données et ont tendance à renforcer les expressions majoritaires.

Et ces biais imprègnent-ils les humains en retour ?

A. A. : Oui, car les utilisateurs les intériorisent. Rappelez-vous la célèbre scène du film *Les Temps modernes* dans laquelle le personnage de Charlie Chaplin *[1889-1977]*, qui serre des boulons toute la journée sur une chaîne de montage, continue machinalement de les serrer une fois sorti de l'usine… La machine influence nos corps comme nos esprits. Ces biais, qu'ils soient statistiques ou idéologiques, influencent nos manières de parler et de penser. Or, ce pouvoir d'influence est dans



ERWAN FAGES

les mains des entreprises qui fabriquent et possèdent ces technologies. Donc, certes, l'arrivée de chaque nouvelle technologie suscite des paniques morales sur leurs effets potentiels. Mais quand on se penche un peu précisément sur l'histoire de ces paniques, on se rend compte qu'elles pointent, en réalité, souvent cette question du pouvoir.

Prenons Platon et son dialogue *Phèdre*, souvent cité en exemple pour sa critique de l'écriture dont l'usage se banalise alors dans la société grecque antique. Il souligne ses effets délétères sur la mémoire. On peut certes arguer que Platon n'a pas vu les potentialités incroyables offertes par l'écriture ; mais, en réalité, Platon s'inquiète moins de l'écriture que des sophistes. Cette petite poignée d'acteurs très puissants dans la société grecque utilise l'écriture et la rhétorique pour mieux réciter parfaitement de grands discours et ainsi influencer la manière dont les gens pensent. De même, les critiques formulées par les philosophes allemands Theodor W. Adorno *[1903-1969]* et Max Horkheimer *[1895-1973]* à l'égard du cinéma visent, en réalité, le pouvoir de fabriquer les images, tout entier entre les mains de quelques acteurs économiques – les producteurs hollywoodiens. Savoir si les IA génératives constituent un risque ou une opportunité pour les humains implique d'oser poser cette question politique : qui les possède, qui les entraîne ?

J. B. : On peut remarquer que cette question politique se posait déjà avec le tournant technologique précédent, celui des algorithmes de recommandation. Netflix, par exemple, a commencé par proposer une plateforme dont l'algorithme recommandait des films et des séries à ses usagers ; puis il s'est mis à produire et à recommander ses propres films ; jusqu'à se trouver aujourd'hui dans une situation dominante sur toute la production et la diffusion audiovisuelle mondiale ! Le scandale Cambridge Analytica a révélé que les algorithmes de recommandation de Facebook ont été instrumentalisés par des acteurs politiques. Le fait que la quasi-totalité des grands modèles d'IA générative soient adossés à des géants du numérique et que seules quelques exceptions comme DeepSeek ou Mistral échappent à cette logique pose évidemment problème.

L'IA menace d'altérer certaines capacités humaines, mais peut-elle aider à en développer d'autres ?

J. B. : Une étude récente d'OpenAI dresse un bilan des trois premières années d'utilisation de ChatGPT. Elle montre que les usages se répartissent principalement entre l'aide à la rédaction de texte (24 %), la recherche d'information (24 %) et la résolution de problèmes personnels ou professionnels (29 %). Quand vous faites une recherche avec ChatGPT, vous n'êtes pas passif : vous formulez une question, puis ChatGPT vous pousse à la formuler mieux. Il vous apprend à bien poser des problèmes. Vous pouvez aussi utiliser ChatGPT pour confronter des points de vue, il suffit de lui demander d'exposer le point de vue adverse. Une dialectique très socratique s'instaure entre vous et la machine : vous entraînez celle-ci, mais elle vous entraîne aussi.

L'IA se place entre l'humain et le monde. Ce n'est pas sans danger, car nous ne sommes qu'aux balbutiements des nouvelles interfaces hommes-machine. Dans les prochaines années, des dispositifs portables comme les Meta Ray Ban *[des lunettes qui peuvent afficher des informations en réalité augmentée sans obstruer la vision]* ou des implants de type Neuralink *[entreprise fondée par Elon Musk en 2016 qui développe des interfaces cerveau-ordinateur]* pourraient façonner une intermédiation totale entre nos processus cognitifs et le monde extérieur, démultipliant les risques que nous avons déjà cités.

Les IA génératives peuvent-elles nous aider à apprendre ?

J. B. : Je le pense. L'IA ne se limite pas aux assistants conversationnels comme ChatGPT ou Le Chat (de Mistral AI) : elle sert aussi de base à d'autres produits plus ciblés, notamment dans l'éducation. On peut citer des start-up comme Scolibree, qui construisent des agents IA pour aider les écoliers et les collégiens à mémoriser des acquis. Ils peuvent aider chaque enfant à identifier ses forces et ses faiblesses, concevoir un programme d'exercices personnalisés, valider la consolidation des acquis, sous la supervision des parents.

A. A. : Vous voulez automatiser les profs !

J. B. : Le problème est qu'il existe des classes surchargées et des élèves en difficulté, qui ont besoin de soutien périscolaire. De tels dispositifs viennent indirectement en aide aux professeurs.

A. A. : Déléguer l'éducation des futurs citoyens à des entreprises privées qui, pour faire du profit, vont remplacer les professeurs par des robots, je ne suis pas sûre que ce soit le meilleur remède. On peut aussi recruter davantage de professeurs. C'est une question politique. Si l'on considère que le rôle de l'école est d'acquérir et de valider des compétences individuelles, on peut, bien sûr, automatiser l'apprentissage. Mais pas si l'on considère que c'est un service public qui sert à donner aux futurs citoyens des clés pour comprendre le monde, penser par eux-mêmes, apprendre à vivre avec les autres, offrir des modèles idéaux dans lesquels les enfants peuvent se projeter… Ce dilemme est valable pour bien d'autres services publics.

J. B. : Il faut être vigilant sur la façon dont sont déployés ces instruments dans les services publics. Mais dans toutes les sociétés, il y a des laissés-pour-compte et je pense que ces agents peuvent avoir de la valeur pour aider les plus désavantagés – par exemple en aidant à améliorer une orthographe, à rédiger une lettre de motivation correspondant aux attentes, voire en fournissant une assistance médicale…

Autre exemple concret : on peut imaginer un « compagnon » installé sur le smartphone des enfants et des adolescents. Il ne bloque rien, mais il prévient l'enfant et l'aide à prendre du recul quand il s'aventure vers des contenus pornographiques, quand il croise des « fake news », quand certains échanges relèvent du harcèlement. Si un vrai risque apparaît pour l'enfant, il peut éventuellement alerter les parents. C'est une approche plus didactique que le simple contrôle parental d'aujourd'hui.

Faut-il miser sur la responsabilisation individuelle des citoyens dans l'usage des IA génératives et des plateformes guidées par des algorithmes de recommandation ?

J. B. : Puisque, à mes yeux, il y a un bon usage de ces modèles, il faut enseigner ce bon usage. On peut « s'augmenter » en jouant au ping-pong avec ces modèles, mais à la condition de connaître leur fonctionnement, et donc de savoir ce que l'on peut en attendre.

A. A. : On peut certes estimer qu'il appartient à chacun d'adopter un usage réfléchi de ces dispositifs et de veiller à ne pas en devenir dépendant. Mais, à mes yeux, c'est faire peser une responsabilité considérable sur des individus déjà soumis à de fortes pressions économiques. Il serait plus fécond d'utiliser l'IA pour créer des « plug-in » que l'on pourrait brancher sur un réseau social ou une plateforme de contenus pour modifier, par exemple, les paramètres de l'algorithme de recommandation. Cela permettrait de redonner un peu de contrôle aux utilisateurs sur ces boîtes noires, et donc sur ce qu'ils voient, ce qu'ils lisent, ce qu'ils pensent. Cette question du paramétrage des algorithmes est un enjeu politique et économique considérable, récemment mis en avant par le rapport de la commission d'enquête sur TikTok.

Qu'il s'agisse des plateformes ou des IA génératives, nous assistons à une concentration du pouvoir dans les mains de quelques entreprises privées qui ont pour but de rendre leurs utilisateurs accros. Lorsque ChatGPT vous parle à la première personne et fait tout pour vous séduire, en vous demandant constamment si vous voulez préciser votre question, en louant votre sagesse, en admirant votre progression, cela a un but : que vous lâchiez le moins souvent possible ce service. Il n'y a pas de situation plus profitable pour l'entreprise que celle où vous tombez amoureux de son produit. C'est inédit dans l'histoire du capitalisme ! Ces stratégies commerciales qui engendrent de l'addiction à ces services numériques sont d'autant plus dangereuses qu'elles agissent comme des dispositifs de désocialisation.

Quels garde-fous peuvent être mis en place pour limiter l'impact négatif de ces dispositifs sur notre psyché ?

A. A. : On pourrait exiger que les chatbots n'utilisent pas le pronom « je » pour éviter les projections anthropomorphiques et la dépendance émotionnelle. Il semble aussi nécessaire de renforcer les apprentissages fondamentaux (lecture, écriture, calcul, raisonnement complexe…), ce qui requiert d'interdire l'usage de ces machines avant un certain âge. Cela permettrait aussi de s'assurer que ceux qui les utilisent ont le recul critique nécessaire pour le faire.

Mais il faut, dans le même temps, renforcer, dans les programmes d'éducation, ce qui relève de la culture technique. Il ne s'agit pas seulement d'apprendre à utiliser correctement ces technologies, mais de donner aux citoyens la possibilité d'en comprendre le fonctionnement et d'avoir un avis éclairé sur celui-ci, pour ne pas être cantonnés au statut de consommateurs passifs.

A l'époque où l'espace public reposait principalement sur la presse écrite, l'école formait les citoyens à la lecture et à l'écriture ; à l'heure où cet espace est façonné par les algorithmes, elle doit les former aux médias numériques. ■

PROPOS RECUEILLIS PAR MARION DUPONT ET PASCAL RICHÉ

« SAVOIR SI LES IA GÉNÉRATIVES CONSTITUENT UN RISQUE OU UNE OPPORTUNITÉ POUR LES HUMAINS IMPLIQUE D'OSER POSER CETTE QUESTION POLITIQUE : QUI LES POSSÈDE, QUI LES ENTRAÎNE ? »

ANNE ALOMBERT