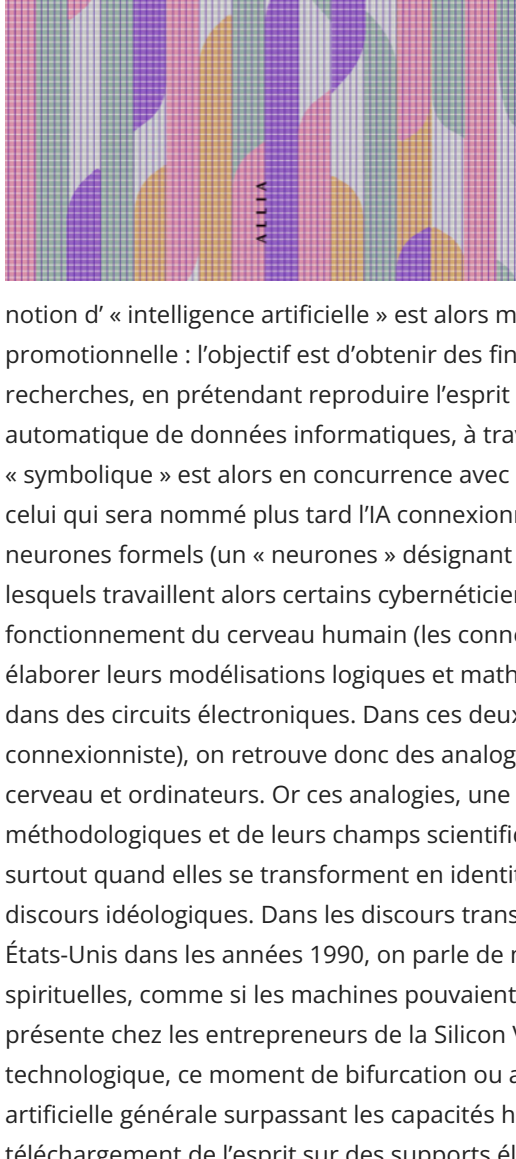


Anne Alombert : « Les termes d'intelligence artificielle ou de machine pensante sont trompeurs »

PHILOSOPHIE SACHA CORNUEL-MERVELLE 7 JANVIER 2026 0



Anne Alombert est maître de conférences à l'université Paris-8 Vincennes. Ses essais portent sur les rapports entre nos vies et la technologie moderne, mais aussi sur les conséquences anthropologiques de l'utilisation de l'Intelligence Artificielle (IA). L'auteur revient sur son dernier ouvrage, *De la bêtise artificielle* (Allia, 2025), dans lequel elle décrit avec brio les ambivalences de cet outil qui est à la fois un remède et un poison.



PHILITT : Au début de votre ouvrage, vous évoquez la genèse de « l'intelligence artificielle ». Pourriez-vous revenir sur son origine ?

Anne Alombert : Dans le livre, je reviens brièvement sur l'histoire du programme technoscientifique de l'IA, dont on situe traditionnellement l'origine lors de la conférence de Dartmouth : des chercheurs en informatique, logique, mathématiques, ingénierie, sciences cognitives proposent alors de reproduire mécaniquement des opérations considérées comme caractéristiques de l'intelligence (comme le langage, le raisonnement, le jeu, etc.), grâce à des machines informatiques et électroniques. La

notion d'« intelligence artificielle » est alors mobilisée, mais c'est une notion promotionnelle : l'objectif est d'obtenir des financements pour effectuer ces recherches, en prétendant reproduire l'esprit humain, alors qu'il s'agit de traitement automatique de données informatiques, à travers des programmes. Cette IA dite « symbolique » est alors en concurrence avec un autre programme de recherche, celui qui sera nommé plus tard l'IA connexionniste, fondée sur les réseaux de neurones formels (un « neurones » désignant ici une fonction mathématique), sur lesquels travaillent alors certains cybernéticiens. Ceux-ci prennent modèle sur le fonctionnement du cerveau humain (les connexions neuronales et les synapses) pour élaborer leurs modélisations logiques et mathématiques, qu'ils implémentent ensuite dans des circuits électroniques. Dans ces deux courants de l'IA (symbolique et connexionniste), on retrouve donc des analogies entre esprit et programme ou entre cerveau et ordinateurs. Or ces analogies, une fois sorties de leurs fonctions méthodologiques et de leurs champs scientifiques, deviennent très problématiques, surtout quand elles se transforment en identités ontologiques et alimentent des discours idéologiques. Dans les discours transhumanistes qui se développent aux États-Unis dans les années 1990, on parle de machines intelligentes ou de machines spirituelles, comme si les machines pouvaient penser ! Cette idée est encore très présente chez les entrepreneurs de la Silicon Valley, à travers l'idée de singularité technologique, ce moment de bifurcation où apparaîtra une superintelligence artificielle générale surpassant les capacités humaines ou encore l'idée du téléchargement de l'esprit sur des supports électroniques, qui est censé ouvrir à l'immortalité. C'est une véritable métaphysique dualiste qui est à l'œuvre ici : si la pensée peut être exportée dans des machines, alors c'est qu'il existe une pensée indépendante du corps vivant et désirant, or une telle idée a été philosophiquement déconstruite depuis longtemps. De plus, cette anthropomorphisation des machines a aussi des enjeux politiques : certains entrepreneurs prétendent pouvoir développer des « thérapeutes IA » ou des « enseignants IA », qui auraient vocation à remplacer les individus pratiquant ces savoirs. On voit bien le risque : une fois l'esprit attribué à la machine, on peut justifier le remplacement des savoirs humains par des systèmes algorithmiques, donc remplacer le travail vivant par du capital fixe – c'est la logique du capitalisme. C'est pourquoi je préfère éviter le terme d'intelligence artificielle et parler d'automates numériques ou d'automates computationnels, comme nous

l'avions proposé dans une [tribune co-écrite avec Giuseppe Longo](#), mathématicien, logicien et épistémologue.

Dans le sillage de Gilbert Simondon et de Georges Canguilhem, vous dénoncez une forme d'« idolâtrie » moderne pour la machine. En quoi celle-ci cache-t-elle un désir inavoué de toute-puissance et de domination et certaines logiques de pouvoir ?

Lorsque l'on parle d'« intelligence artificielle », on masque la dimension technoscientifique et industrielle de ces dispositifs (les algorithmes selon lesquels ils sont entraînés, les données qui les nourrissent, les infrastructures qui leurs permettent de fonctionner – comme les serveurs, les câbles, les satellites, etc.) mais on masque aussi les dynamiques de pouvoirs qui sont à l'œuvre : comme l'écrit Canguilhem, ces métaphores anthéropomorphiques permettent de « dissimuler la présence de décideurs derrière l'anonymat de la machine ». Dans le cas des IA génératives, par exemple, vous n'obtiendrez pas les mêmes résultats si vous posez une question à Grok ou à ChatGPT, car les deux modèles n'ont pas été entraînés ou affinés de la même manière et selon les mêmes critères : ils implémentent des idéologies politiques et des visions du monde différentes, qui conditionneront ensuite celles de leurs utilisateurs – ce qui constitue un pouvoir exercé sur leurs esprits par des entreprises privées. C'est ce désir de pouvoir que Simondon souligne également, en suggérant qu'à travers le mythe du robot tout-puissant, ce sont en fait les humains qui ont développé les technologies qui projettent leurs fantasmes de contrôle et de domination.

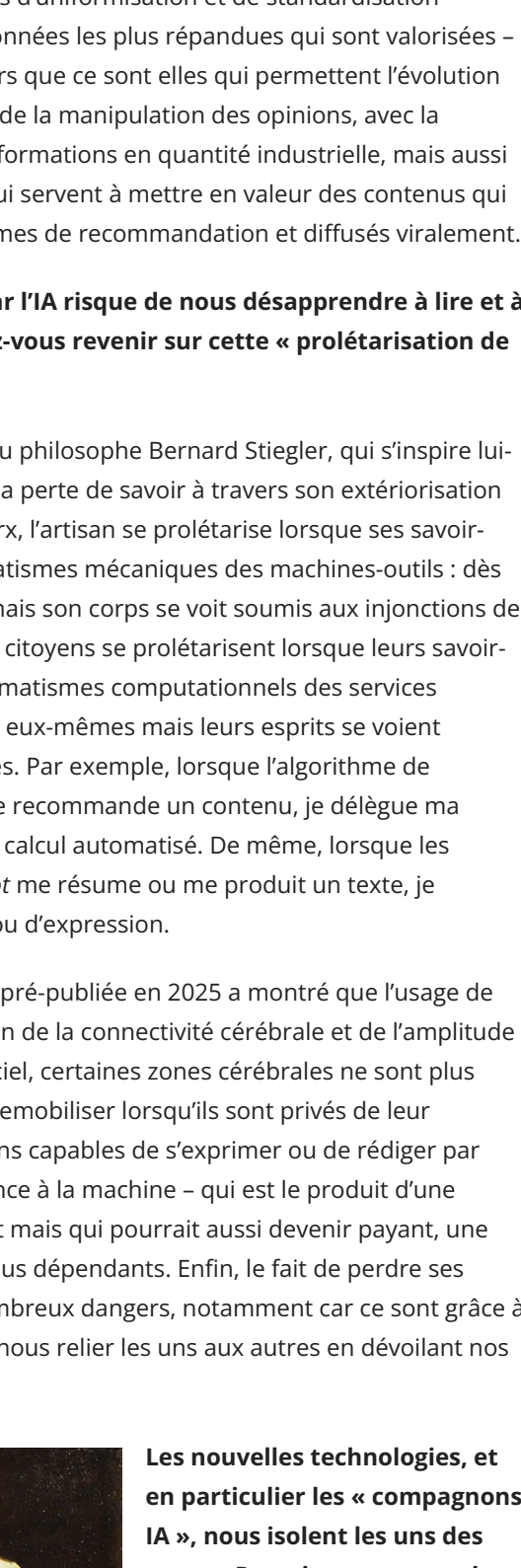
Vous décrivez avec précision la « face cachée » de l'Intelligence Artificielle (IA), notamment les dégâts sociaux et environnementaux qu'elle implique. Quelles sont les multiples exploitations induites par celle-ci ?

Là encore, les termes d'« intelligence artificielle » ou de « machine pensante » sont trompeurs car ils renvoient à des notions souvent mal comprises et peu définies d'intelligence ou de pensée, qui apparaissent au premier abord comme des processus immatériels et éthérés, alors les infrastructures industrielles et les logiciels algorithmiques qui sont sous-jacents aux générateurs automatiques de textes, d'images ou de sons présentent de grands dangers d'un point de vue écologique. Du point de vue de l'économie environnementale tout d'abord, ces systèmes algorithmiques impliquent d'effectuer des quantités immenses de calculs sur des quantités immenses de données, ce qui nécessite des puces et des microprocesseurs spécifiques ainsi que la construction de plus en plus de *data centers*. Cela implique l'extraction de matières et minerais rares, ainsi qu'une très grande consommation énergétique, notamment en électricité. Les *data centers* consomment beaucoup d'électricité mais aussi beaucoup d'eau (à travers leur système de refroidissement), et leurs fonctionnements s'opèrent parfois au détriment de l'accès à l'électricité et à l'eau des habitants, qui se voient privées de ces ressources accaparées par les géants du numérique – je cite plusieurs exemples dans le livre.

De plus, du point de vue de l'écologie mentale et sociale, les industries d'IA sont problématiques aussi. Elles conduisent à de nouvelles formes d'exploitation du travail humain : OpenAI a employé des travailleurs kenyans sous-payés pour faire le tri dans les jeux de données, travailleurs qui risquent de souffrir de troubles et de traumatismes psychiques à force d'être confrontés à tous les contenus violents ou illégaux qu'ils doivent repérer dans les quantités massives de données. Enfin, les données fournies aux modèles pour leur entraînement sont des données produites par des utilisateurs d'Internet (auteurs, artistes, contributeurs à Wikipédia, etc.) qui ne sont pas rémunérés pour cela, et ces données ne sont pas renouvelées car les « intelligences artificielles » effectuent des calculs probabilistes sur elles, donc renforcent des moyennes et épuisent la diversité culturelle. C'est d'ailleurs la raison pour laquelle Jean Cattan et Célia Zolynski, spécialistes du droit du numérique, [avaient décrit](#) ces dispositifs comme des « intelligences artificielles extractives » [dans l'un de leurs articles](#).

Les préoccupations liées aux effets des nouvelles technologies sur les esprits ne datent pas d'hier. Comment le mythe égyptien de Theuth et de Thamous de Platou dans la *Phèdre* peut-il nous éclairer aujourd'hui ?

Dans le dialogue intitulé le *Phèdre*, Platon évoque le mythe égyptien de l'origine de l'écriture, dans lequel le dieu Theuth, inventeur de l'écriture, prétend avoir trouvé un remède pour la mémoire et pour l'instruction, car la technique de l'écriture permet de conserver une grande quantité de savoirs et de les transmettre de génération en génération, en dépit de la disparition des vivants. Mais le roi



Le philosophe Platon

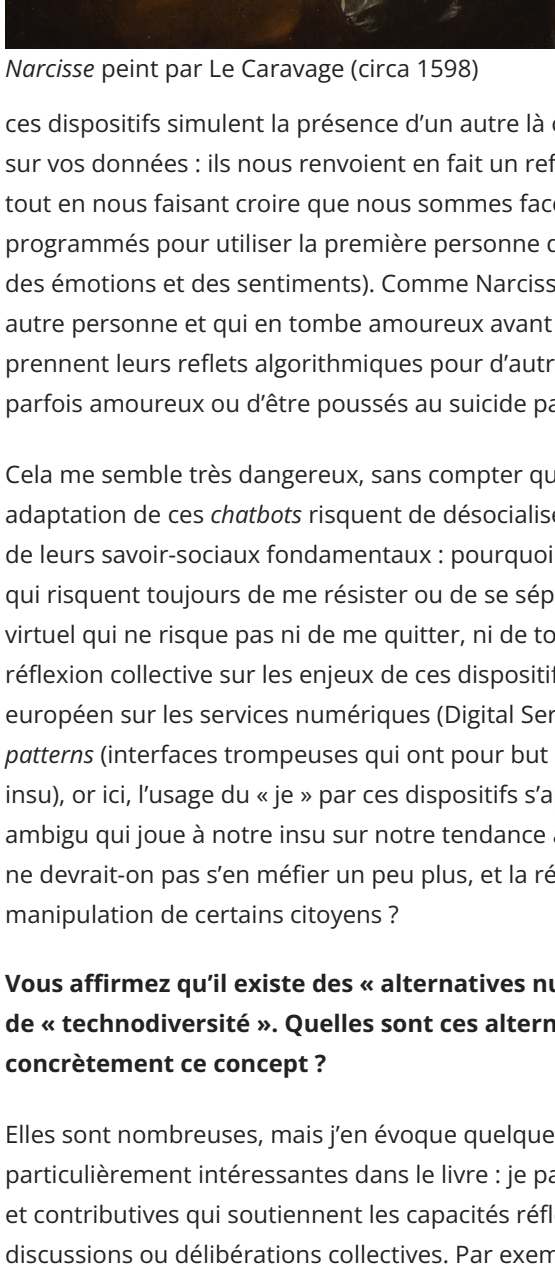
Thamous à qui il présente son invention lui fait remarquer les effets potentiellement toxiques de cette nouvelle technique : les citoyens risquent de déléguer leurs mémoires aux supports écrits, de cesser d'exercer leurs facultés psychiques ou cognitives, donc de perdre la mémoire qu'ils croyaient augmenter. Par ailleurs, s'ils ne retiennent pas par eux-mêmes, ils ne pourront plus interpréter et renouveler les savoirs – l'évolution culturelle se voit menacée. Enfin, cette délégation des capacités mémorielles et réflexives sera profitable aux sophistes, qui maîtrisent très bien les techniques de l'écriture et de la rhétorique, et qui s'en servent pour manipuler les esprits des citoyens, fascinés par leurs discours persuasifs.

Dans une certaine mesure, ces trois risques de délégation des capacités, de stérilisation culturelle et de manipulation des opinions se jouent aujourd'hui. Tout d'abord, à force de nous reposer sur des générateurs automatiques de textes, d'images ou de sons, nous risquons d'abandonner nos capacités expressives à des systèmes algorithmiques. Comme ceux-ci fonctionnent sur la base de calculs probabilistes, il existe aussi des risques d'uniformisation et de standardisation importants, car ce sont toujours les données les plus répandues qui sont valorisées – les exceptions se voient éliminées, alors que ce sont elles qui permettent l'évolution culturelle. Enfin, on retrouve le risque de la manipulation des opinions, avec la génération automatique de fausses informations en quantité industrielle, mais aussi avec l'alimentation de faux comptes qui servent à mettre en valeur des contenus qui sont ensuite amplifiés par les algorithmes de recommandation et diffusés viralemment.

L'écriture algorithmique permise par l'IA risque de nous désapprendre à lire et à écrire, à parler et à penser. Pourriez-vous revenir sur cette « prolétarianisation de l'expression » que vous redoutez ?

Ce terme de « prolétarianisation » vient du philosophe Bernard Stiegler, qui s'inspire lui-même de Marx : cette notion désigne la perte de savoir à travers son extériorisation dans un dispositif technique. Pour Marx, l'artisan se prolétarianise lorsque ses savoir-faire sont extériorisés dans les automatismes mécaniques des machines-outils : dès lors, il ne les pratique plus lui-même mais son corps se voit soumis aux injonctions de la machine. De même, aujourd'hui, les citoyens se prolétarianisent lorsque leurs savoir-penser sont extériorisés dans les automatismes computationnels des services numériques : ils ne les pratiquent plus eux-mêmes mais leurs esprits se voient soumis aux injonctions des algorithmes. Par exemple, lorsque l'algorithme de recommandation d'une plateforme me recommande un contenu, je délègue ma capacité de décision à un système de calcul automatisé. De même, lorsque les algorithmes de génération d'un *chatbot* me résume ou me produit un texte, je délègue ma capacité d'interprétation ou d'expression.

Une [étude par des chercheurs du MIT](#) pré-publiée en 2025 a montré que l'usage de ChatGPT correspondait à une réduction de la connectivité cérébrale et de l'amplitude cognitive : chez les utilisateurs du logiciel, certaines zones cérébrales ne sont plus activées et les individus peinent à les remobiliser lorsqu'ils sont privés de leur prothèse cognitive, ils deviennent moins capables de s'exprimer ou de rédiger par eux-mêmes et cela crée une dépendance à la machine – qui est le produit d'une entreprise privée, pour l'instant gratuit mais qui pourrait aussi devenir payant, une fois que tous les usagers seront devenus dépendants. Enfin, le fait de perdre ses capacités expressives présente de nombreux dangers, notamment car ce sont grâce à de telles capacités que nous pouvons nous relier les uns aux autres en dévoilant nos singularités.



Narcisse peint par Le Caravage (circa 1598)

ces dispositifs simulent la présence d'un autre là où il n'y a que calculs automatisés sur vos données : ils nous renvoient en fait un reflet algorithmique de nous-mêmes, tout en nous faisant croire que nous sommes face à une personne (car ils sont programmés pour utiliser la première personne du singulier, le « je » et pour simuler des émotions et des sentiments). Comme Narcisse, qui prend son reflet pour une autre personne et qui en tombe amoureux avant de se suicider, certains individus prennent leurs reflets algorithmiques pour d'autres personnes, au point d'en tomber parfois amoureux ou d'être poussés au suicide par ces systèmes.

Cela me semble très dangereux, sans compter que la disponibilité et la sur-adaptation de ces *chatbots* risquent de désocialiser les individus et de les déposséder de leurs savoir-sociaux fondamentaux : pourquoi me relier et m'attacher aux autres qui risquent toujours de me résister ou de se séparer si je peux me replier sur un ami virtuel qui ne risque pas ni de me quitter, ni de tomber malade, ni de mourir ? Une réflexion collective sur les enjeux de ces dispositifs me semble urgente : le règlement européen sur les services numériques (Digital Services Act ou DSA) interdit les *dark patterns* (interfaces trompeuses qui ont pour but d'influencer les individus à leur insu), or ici, l'usage du « je » par ces dispositifs s'apparente à une fonctionnalité ambiguë qui joue à notre insu sur notre tendance à anthropomorphiser les machines – ne devrait-on pas s'en méfier un peu plus, et la réguler pour éviter que la manipulation de certains citoyens ?

Vous affirmez qu'il existe des « alternatives numériques » à l'IA et vous parlez de « technodiversité ». Quelles sont ces alternatives et que recouvre concrètement ce concept ?

Elles sont nombreuses, mais j'en évoque quelques-unes qui me semblent particulièrement intéressantes dans le livre : je parle de technologies herméneutiques et contributives qui soutiennent les capacités réflexives et expressives ainsi que les discussions ou délibérations collectives. Par exemple, la plateforme Wikipédia permet la co-construction de savoirs en commun et la plateforme Pol.is permet la délibération autour de propositions politiques : contrairement aux IA génératives, il ne s'agit pas ici de remplacer les capacités réflexives et expressives des individus à travers des systèmes probabilistes opaques, mais de permettre à de très nombreux individus de partager leurs points de vue et de débattre collectivement. Sur un plan un peu différent, je cite l'association Tournesol qui élabore un algorithme de recommandation collaborative de contenus : cet algorithme se base sur les jugements des citoyens, qui évaluent les vidéos en amont en fonction de leur utilité publique, et ce sont les vidéos jugées pertinentes qui sont recommandées (et non les vidéos sensationnelles ou violentes susceptibles de générer des clics et des profits pour l'entreprise numérique qui possède tel ou tel réseau social commercial).

Ces technologies se fondent aussi sur de l'IA et des algorithmes, mais elles ne sont pas prolétarianisantes : au contraire, elles offrent de nouvelles possibilités d'action aux citoyens, en leur permettant de participer à la production des savoirs partagés, à la vie politique de leurs sociétés, ou à l'organisation de leurs environnements informationnels quotidiens – ce qui était impossible du temps de l'imprimerie ou des médias analogiques. Je prends aussi les exemples de réseaux sociaux comme BipPop et Entourage, qui ont vocation à renforcer les solidarités sociales locales, et non les modèles d'affaires publicitaires. Il me semble que de tels dispositifs peuvent servir d'exemples pour mettre les algorithmes au service de l'intelligence collective. C'est la raison pour laquelle je mobilise également la notion de « technodiversité », afin de suggérer que les services numériques qui s'imposent aujourd'hui de manière hégémonique ne sont pas les seuls possibles : au contraire, tout l'enjeu consiste à proposer des contre-modèles qui répondent aux besoins des habitants et des localités, afin d'ouvrir un nouveau chemin vers des démocraties numériques contributives, par-delà « l'accélération réactionnaire » qui se développe sous nos yeux.

Les nouvelles technologies, et en particulier les « compagnons IA », nous isolent les uns des autres. Pourriez-vous revenir sur cette « automatisaion de l'altérité » qui nous frappe de plein fouet ?

Dans le livre, j'évoque ces nouvelles « industries de la solitude, qui proposent des remèdes à l'isolement à travers des « compagnons virtuels », c'est-à-dire, des *chatbots* conçus pour répondre 24h/24 à toutes vos questions et vous renforcer dans vos opinions. Je parle d'automatisation de l'altérité car