

IA GÉNÉRATIVES : UN DANGER POUR LA COGNITION HUMAINE ?

Communication à l'Académie des technologies de Stéphane Andrieux, membre de l'Académie des technologies

Ce texte est une contribution d'un membre de l'Académie des technologies. Il n'a pas été soumis au vote de l'Assemblée ; il n'engage que son auteur.

Introduction : cerveaux étendus ou externalisés ?	2
Premières études sur l'impact de l'utilisation des intelligences artificielles génératives	5
Le monde du travail	5
Le monde de l'éducation	7
Le monde universitaire	8
Vers une explication par les neurosciences	10
En guise de conclusion	12
Références	14

L'objectif de cette note est d'examiner les effets documentés que provoque sur notre système cognitif et métacognitif (pour ce dernier, entendu au sens du contrôle qu'il effectue sur nos processus cognitifs) une utilisation systématique, ou trop précoce des intelligences artificielles génératives. Sans nier les apports extraordinaires et les ruptures qu'apportent ces technologies, on s'attache ici à un aspect moins médiatisé que bien des interrogations parfois présentées comme existentielles, mais qui n'en demeure pas moins un facteur de risque majeur. Majeur car, à bas bruit, il concerne les centaines de millions d'utilisateurs quotidiens des chatbots, de tout âge et de toute origine, dès lors qu'ils utilisent les IA génératives systématiquement et surtout sans le discernement ou l'esprit critique nécessaires. Si depuis l'apparition de ces technologies, le monde de l'éducation s'est intéressé à leur potentiel et à leurs effets, il n'en va pas de même dans notre vie quotidienne faute de cadre d'utilisation et de formation, même au sein des entreprises (du moins sur ce plan). La tentation est grande en effet de se laisser envahir par un usage qui sacrifie notre esprit critique et notre goût de l'effort sur l'autel de la performance, la rapidité et la facilité.

On s'appuiera non seulement sur quelques études qui analysent les conséquences de l'usage de ces outils mais également sur ce que les connaissances issues de la psychologie cognitive et des neurosciences nous apportent pour tenter d'expliquer, assez simplement finalement, les phénomènes observés.

Les références sont séparées en deux groupes pour plus de fluidité de lecture : les articles de journaux, de blogs etc. sont appelés directement dans le texte (appel de note en chiffre) alors que les études scientifiques ou les sources primaires sont rassemblées en fin (appel de notes en lettres).

Introduction : cerveaux étendus ou externalisés ?

Faire faire une synthèse d'un document avant de (ne pas) le lire, faire traduire un texte en français (avant même de l'examiner) depuis une langue que nous avons pourtant mis du temps à maîtriser, utiliser en cachette un chabot pour réaliser des tâches professionnelles², faire rédiger un mémoire universitaire ou élaborer un exposé scolaire, demander d'écrire une réponse à un mail personnel délicat (peut-être lui-même écrit par une IA), préparer des sujets de conversations ou des arguments avant un rendez-vous professionnel ou amoureux³... on pourrait multiplier les exemples. Tout cela est entré dans la vie quotidienne de bien des humains via l'intelligence artificielle générative, et désormais agentique, qui n'a pourtant déferlé sur l'humanité que depuis fort peu de temps. On peut en effet dater ce bouleversement du 22 novembre 2022, date de la mise en service par OpenAI de ChatGPT 3.5. En marge des réalisations époustouflantes de ces outils dans les domaines de la science et de la technologie, en parallèle des promesses de révolution de la quasi-totalité du travail intellectuel, et malgré les protestations du monde artistique^{4,a} et littéraire, ce qui n'était au départ vu que comme un *homework killer* dans la communauté enseignante⁵, et comme un outil de méconduite par les scientifiques⁶, est devenu un partenaire ou plutôt, comme le terme commence à apparaître, une béquille⁷ quotidienne pour nombre d'entre nous (AI Crutch).

Sans nier les formidables potentialités des IA génératives, et au-delà des grands débats sur la singularité technologique^{b,c,d}, l'avenir de l'Humanité, le transhumanisme ou le grand remplacement par des robots, voire la prise de pouvoir par des AI intrinsèquement amoraux, il convient de s'interroger sur les conséquences d'un usage généralisé des IA génératives sur notre propre cognition (et notre métacognition en tant que couche de contrôle), du fait de ce que l'on appelle le *déport* ou la *délégation cognitive* (*cognitive offloading*).

Le concept et les termes ne sont pas nouveaux, en tout cas ne sont pas nés avec l'apparition des chatbots intelligents et des grands modèles de langages. D'autres technologies ont été critiquées dans le même sens, mais l'ampleur et la généralisation du phénomène⁸, sans parler de l'extrême concentration de l'offre⁹, doit nous pousser à nous pencher sur ce qui pourrait bien représenter un risque majeur car massif.

-
1. Généré par Anagramdr <https://anagramdr.com/> (algorithme combinatoire + dictionnaire).
 2. Phénomène désigné dans le monde anglophone par *shadow AI*.
 3. N. Kaufman, *Young Adults Are Letting AI Do Their Talking for Them—Even in Person*, Wall Street Journal, 26/07/2026.
 4. Un article très récent dans *Nature Communication* montre cependant que la preuve d'utilisation d'une image dans l'ensemble d'entraînement d'un modèle d'IA n'est plus pertinente lorsque celui-ci grossit. La découverte de ce phénomène dit *attribution decay* pourrait mettre fin aux innombrables procès et réclamations d'artistes s'estimant plagés.
R. Gordon, *When AI art has no author: Study finds generated images often can't be traced to training data*, MIT News, 18 août 2026, <https://news.mit.edu/2026/when-ai-art-has-no-author-generated-images-often-cant-be-traced-to-training-data-0818>.
 5. <https://stratechery.com/2022/ai-homework/> December 5, 2022, 16 % des jeunes américains (13-17 ans) interrogés fin 2025 par le PEW research Center déclarent utiliser les chatbots pour leurs conversations informelles (54 % pour obtenir de l'aide pour leurs devoirs). <https://www.pewresearch.org/short-reads/2026/03/12/key-findings-about-how-americans-view-artificial-intelligence/>.
Hua Hsu, *What Happens After A.I. Destroys College Writing?* The New Yorker, June 30, 2025.
 6. *Not a generative AI-generated Editorial: What does the explosion of generative artificial intelligence tools mean for science?* Nature Cancer, Vol. 4, February 2023, pp 151-152.
C. Stokel-Walker, R. Van Noorden *The promise and peril of generative AI*, Nature, Vol. 614, N° 7947, Feb 9, 2023, pp 214-216 *Science journals ban listing of ChatGPT as co-author on papers*, The Guardian, 26 janvier 2023, <https://www.theguardian.com/science/2023/jan/26/science-journals-ban-listing-of-chatgpt-as-co-author-on-papers>.
 7. Béatrice St-Cyr-Leroux, *ChatGPT en éducation : outil pertinent ou béquille aliénante ?* Université de Montréal, nov. 2023, <https://nouvelles.umontreal.ca/article/2023/11/10/chatgpt-en-education-outil-pertinent-ou-bequille-alienante-?>.
 8. Microsoft estime à 18 % de la population mondiale l'usage de l'IA Générative soit 2,17 milliards de personnes (en incluant la Chine, 602 millions) selon une autre estimation.
Juan Lavista Ferres, *The state of global AI diffusion in 2026*, Microsoft, May 7, 2026.
<https://blogs.microsoft.com/on-the-issues/2026/05/07/the-state-of-global-ai-diffusion-in-2026>
<https://datareportal.com/reports/similarweb-2026-generative-ai-visibility-index>.
 9. En excluant la Chine les utilisateurs se concentrent à 53 % sur Chat-GPT, vient ensuite Google Gemini à 28 % puis Claude à 9 % (soit 90 % au total sur ces seuls trois outils). <https://www.similarweb.com/blog/marketing/gen-ai-stats/>.

	Voice-activated assistant	Smartwatch	AI chatbots*
Total	37	11	8
Child age			
Under 2	12	7	—
2-4	27	7	—
5-7	36	7	3
8-10	46	13	7
11-12	53	21	15

Pourcentage de parents américains qui déclarent en mai 2025 que leur enfant utilise ou interagit avec les éléments suivants.
Source: Pew Research Center <https://www.pewresearch.org/>.

Cela d'autant plus que des études apparaissent qui documentent le phénomène, et que les connaissances que nous avons accumulées, tant dans le domaine de la psychologie cognitive que des neurosciences, nous fournissent les clés de compréhension des mécanismes à l'œuvre. Il y a évidemment certaines similitudes avec les phénomènes aux noms évocateurs directement attribués aux réseaux sociaux : *Brain Rot*¹⁰ littéralement pourrissement, ou plus élégamment dépérissement du cerveau, qui désigne la baisse de concentration et la léthargie provoquées par le défilement compulsif, *Zombie Scrolling* ou défilement machinal de son fil d'actualité sans aucun but précis, en étant dans un état de quasi-hypnose, on encore le *Doomscrolling*, tendance compulsive à faire défiler et à lire des informations négatives ou anxiogènes, alimentant ainsi une détresse psychologique. Mais nous nous concentrerons ici sur les effets de l'usage de l'IA à des fins non récréatives, dans notre quotidien de travail, d'éducation ou de production intellectuelle voire artistique.

L'esprit étendu

L'utilisation d'objets technologiques pour « aider le cerveau » date de bien longtemps, que l'on pense aux bouliers ou aux abaquas en général et même aux encyclopédies imprimées ou aux cartes géographiques, cependant un pas important a été franchi avec l'apparition du numérique ubiquitaire, balayant peu à peu ce que l'on a théorisé comme l'esprit étendu (*extended mind*⁶), qui considère que certains objets physiques de notre environnement participent à des processus cognitifs et peuvent ainsi être assimilés à des extensions de l'esprit lui-même. Si ses arguments sont convaincants pour les calculatrices électroniques ou la recherche web, où le processus cognitif provient de notre esprit ; ils deviennent caducs avec l'intelligence artificielle générative. Dans l'appel au calcul assisté, on ne délègue que le processus de calcul (même si on pourrait considérer que l'on a, pour beaucoup d'entre nous, « perdu » la maîtrise de la division à plus d'un chiffre). De la même façon, lors d'une recherche sur le web, c'est bien nous qui formulons les mots-clés voire la question. Ici bien qu'une étude⁷ déjà ancienne ait montré que l'on en mémorise bien mal les résultats (environ 40 % des personnes testées ne se souviennent plus de rien une semaine après, seules 17 % se souviennent à la fois du site où elles ont obtenu la réponse à leur question ainsi que la réponse elle-même), ce *Google effect* n'est en fait pas très différent du processus de mémoire transactive qui sous-tend le partage d'informations dans nos interactions sociales. En effet, lorsque nous nous interrogeons par le passé sur une question précise, nous faisons souvent appel à des personnes dont nous savons qu'elles auraient des éléments de réponse à nous fournir, et nous oublions parfois la réponse si la question était d'importance minime (tout comme dans l'étude citée où 30 % des répondants se souviennent seulement du site consulté). Le seul effet nouveau ici, cependant loin d'être anodin, est la dépendance au réseau que cela implique : « Nous finirions par vivre en symbiose avec les machines, mais perdre ce lien reviendrait à perdre nos liens sociaux ».

10. Selon l'Oxford University Press, la première utilisation du terme « dépérissement cérébral » se trouve dans le *Walden ou la vie dans les bois*, d'Henri-David Thoreau, paru en 1854. Le mot *brainrot* a été désigné mot de l'année 2024 par l'Oxford English Dictionary.

Il en va néanmoins tout autrement de technologies comme le guidage par GPS, ici le processus de calcul d'itinéraire puis d'orientation est totalement confié à la machine : nous perdons la faculté de lire une carte, de nous orienter dans une ville ou dans la nature, mais nous perdons probablement aussi une forme de représentation de l'espace ou de la structure d'une cité, faute de solliciter les parties archaïques de notre cerveau qui ont permis notre survie dans le passé⁹. Facultés, qui comme nous le verrons plus loin, doivent s'apprendre et se conforter sans cesse par leur mise en application.

Les connaissances implicites ou tacites

Les IA génératives sont encore largement basées sur les grands modèles de langage (*Large Language Models* - LLM), dont l'apprentissage s'est appuyé sur la quasi-totalité des productions écrites de l'humanité (avec les excès que l'on connaît), et même désormais sur ses propres productions, avec l'aide (gratuite) des utilisateurs pour nourrir les améliorations liées à leur retours, avis et commentaires. Mais c'est aussi là une des limitations reconnues de ces IA, et on pouvait jusque très récemment considérer que certaines connaissances non écrites ou d'actions non susceptibles de faire l'objet de processus totalement rédigés et formalisés leur resteraient inaccessibles. Ces connaissances, dites implicites ou tacites^h, sont pourtant essentielles dans l'apprentissage et l'exécution de certaines actions¹¹. Prenons l'exemple de l'apprentissage de la bicyclette. On pressent bien que la lecture d'un manuel n'y est que de peu d'utilité. De fait, il n'existait pas jusqu'à très récemment de robot humanoïde cycliste, on pouvait trouver des vélos robotisés mais ceux-ci ressemblent à tout sauf à un cycliste chevauchant un bicyclette¹². Ils sont basés sur un système de vérins et de masses contrôlé par un algorithme d'optimisation (de contrôle optimal) exploitant la mise en équations du système ; ici donc pas véritablement de connaissances implicites utilisées. En mai dernier, en direct lors d'un exposé sur ce thème en Chine, une étudiante m'a mis sous les yeux un exemple de robot humanoïde cycliste¹³ créé par la société AgiBot. L'annonce et la vidéo qui l'accompagne datent de quelques semaines avant cet exposé, ce qui a pu me rassurer, mais en examinant de plus près cette performance, il s'avère que l'apprentissage réalisé par le robot ne s'appuie que très partiellement sur les LLM¹⁴ ce qui conforte ainsi l'idée que ces modèles trouvent bien leurs limites pour certaines connaissances. Il illustre en revanche les progrès que l'on peut attendre des nouvelles générations d'IA génératives basées sur les *World Models*ⁱ. Un tel modèle du monde est conçu pour comprendre, simuler et prédire le monde réel ou, plus modestement, un environnement donné, c'est-à-dire au fond identifier la causalité qui détermine ce que l'on observe, constate ou même prévoit. Contrairement aux LLM, concentrés sur la génération de texte, les *World Models* modélisent les dynamiques physiques, sociales ou logiques, en apprennent les règles sous-jacentes qui les régissent. L'apprentissage est donc beaucoup plus riche, et peut ainsi espérer capter les connaissances tacites ou implicites sous-jacentes à des actions plus complexes que ne peut le faire la production de textes. Les données utilisées pour cet apprentissage sont beaucoup plus variées, il peut s'agir de données de simulation (simulation physique, jeux vidéo), de données sensorielles (images vidéos, capteurs divers), de données symboliques (graphes de connaissances ou ontologies) ou encore de données d'interaction pour nourrir l'apprentissage par renforcement.

11. Un exemple historique montre que les savoir-faire peuvent entrer dans cette catégorie de connaissances : le convertisseur de fonte liquide en acier d'Henry Bessemer a fait l'objet de licences, sur la base du brevet déposé en 1856, à quatre maîtres de forges anglais pour le montant considérable pour l'époque de 27 000 £. Ceux-ci n'ont jamais réussi à reproduire l'acier de la qualité promise (*rotten hot and rotten cold*). Bessemer a dû racheter ces licences pour 32 500 £.

12. Robotics and AI institute, *Flip Stunts on Bicycle Robots using Iterative Motion Imitation*. <https://rai-inst.com/resources/papers/flip-stunts-on-bicycle-robots-using-iterative-motion-imitation/>.

13. Mike Kalil, *AgiBot 's New Humanoid Robot Rides a Bike, Understands Emotions*, March 12, 2025, <https://mikekalil.com/blog/agibot-lingxi-x2/>.

14. Le robot est entraîné par apprentissage par renforcement à l'aide du simulateur Isaac Gym de Nvidia. Il s'agit du même logiciel que celui utilisé pour obtenir une locomotion bipède semblable à celle d'un humain avec des mouvements fluides. Grâce à cette méthode d'entraînement, le jumeau numérique du robot apprend dans un environnement virtuel. Cet apprentissage est ensuite transféré au robot physique via le processus *Sim2Real*, abréviation de *Simulation to Reality*.

Le domaine des émotions n'échappe pas lui non plus au projet d'intelligence artificielle générale¹⁵ ; nous n'en sommes pas encore là semble-t-il mais cette perspective renforce évidemment l'ardente nécessité d'une réflexion et d'une exploration des effets des usages récurrents des intelligences artificielles dans tous les aspects de notre vie intellectuelle et affective.

Ainsi, l'avenir nous réserve encore bien des surprises quant à la substitution par les intelligences artificielles de nos compétences et de nos connaissances si nous continuons à leur abdiquer nos efforts, notre apprentissage et notre sens critique au nom de la facilité et de la rapidité... abdication causée, n'en doutons pas, par le développement d'une paresse intellectuelle certaine.

Premières études sur l'impact de l'utilisation des intelligences artificielles génératives

Il n'a pas fallu très longtemps pour que des universitaires, des enseignants, et même des sociétés de conseils, directement intéressées au comportement (et à l'efficacité) de leurs consultants, se penchent sur les pratiques de l'IA et leurs conséquences. Bien qu'il faille prudence garder sur les enseignements à en tirer, du fait du petit nombre de participants impliqués et du contexte souvent particulier des expérimentations, une tendance se dessine, d'autant plus intéressante que les instruments de mesure neurophysiologiques et les connaissances que l'on possède sur le fonctionnement de notre cerveau peuvent en expliquer les principaux aspects. De façon similaire, bien des biais cognitifs connus sont exaltés par certaines pratiques ou certains outils numériques.

Le monde du travail

Une des premières études a été réalisée par le BCG¹⁶ sur 758 collaborateurs et s'intéresse plus particulièrement à Chat-GPT4. Elle montre que, pour les personnes dont les performances sont (jugées) inférieures à la moyenne, un gain de 43 % d'efficacité a été constaté sur des tâches directement du ressort de l'IA. En effet et de façon peut-être contre-intuitive pour qui ne considère les LLM que comme des perroquets stochastiques¹⁷, les tâches dites d'idéation (soit, dans le domaine du BCG, la création de contenu ou l'innovation créative en matière de produits) voient 90 % des participants améliorer leur performance, surtout s'ils ne modifient en rien les productions de l'AI générative. On observe en revanche une baisse de performance de 23 % lorsque l'IA est utilisée pour résoudre des problèmes complexes ne relevant pas de ses compétences (*problem solving tasks*). La baisse de performance pour les tâches d'analyse et de proposition de solutions pour les problèmes d'entreprise, s'accompagne d'une surconfiance dans les résultats produits, cela même pour les participants ayant bénéficié d'une formation sur les limites des IA. Enfin, ce qui est plus évident *a priori*, la diversité des points de vue au sein d'un travail d'équipe est considérablement réduite (recul estimé à 40 %) lorsque les participants font appel aux chatbots.

Cette dernière constatation est confirmée par une autre étude¹⁸, plus spécifiquement dédiée à la créativité (en équipe), avec la constatation d'une créativité accrue, mais davantage progressive et incrémentale qu'originale, et plus standardisée. L'étude révèle également des aspects liés à l'organisation de travail : isolement des individus, diffusion de la responsabilité, fuite de données sensibles, surcharge cognitive liée

15. Selon Yan Le Cun, « les émotions sont inséparables de l'intelligence car elles sont liées à la capacité d'anticipation, [...] rien de magique dans les émotions, les émotions sont calculées par le cerveau et comme le reste, on peut imaginer effectivement des machines intelligentes qui ont une partie de leur cerveau logiciel ou électronique qui calcule leur contentement, c'est un petit peu similaire aux ganglions de la base dans le cerveau humain ». France Culture, Yann LeCun : « Les émotions sont inséparables de l'intelligence. » octobre 2018, <https://www.youtube.com/watch?v=ZvMpTmAXQM&t=546s>.

16. Étude réalisée avec le soutien de chercheurs de la Harvard Business School, de la MIT Sloan School of Management, de la Wharton School de l'Université de Pennsylvanie et de l'Université de Warwick. <https://www.bcg.com/publications/2023/how-people-create-and-destroy-value-with-gen-ai>.

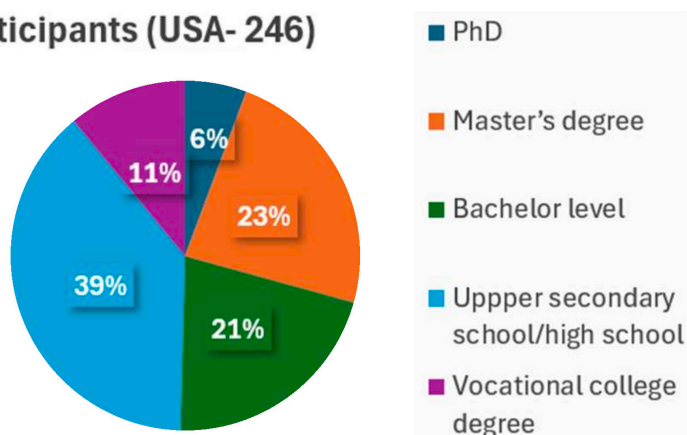
17. E. M. Bender, A. McMillan-Major, T. Gebru, S. Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Pages 610–623, <https://doi.org/10.1145/3442188.3445922>.

à la vérification des contenus générés par l'IA (*Brain Fry*), qui sortent cependant du propos développé ici^{18 k}.

Dans le domaine médical, les dispositifs assistés par l'IA, qu'il s'agisse « simplement » de *deep learning* ou d'IA générative, se multiplient : selon la *Food and Drug Administration* américaine (FDA), un total cumulé de 1 451 dispositifs médicaux les incluant a été homologué depuis 1995¹⁹. On dispose donc ici de plus de recul sur l'utilisation de ces systèmes par les médecins, spécialistes ou non. Ainsi une évaluation²⁰ a été menée en 2021 sur le système *CheXnet*^l, développé en 2017 par le *Stanford Machine Learning group*, analysant des radiographies et proposant des diagnostics de pneumonie²¹. Sur 264 médecins (dont 52 % étaient radiologues) et avec un ensemble de diagnostics générés ou non par IA, exacts et erronés, on constate une amélioration de 40 % de la qualité des diagnostics grâce à des recommandations pertinentes, mais il s'avère aussi que les recommandations erronées sont souvent suivies. De plus, si les non-experts font davantage confiance à l'IA que les experts, cela semble n'avoir aucune incidence sur l'adoption de mauvaises recommandations. La présence de ces biais d'automatisation et de cette dépendance excessive confirment une méta-étude^m précédente sur près de 900 articles qui montre que contrairement à l'idée reçue selon laquelle les biais d'automatisation seraient essentiellement cantonnées aux situations multitâches, ceux-ci persistent également dans le cadre de tâches uniques, c'est-à-dire ici de diagnostics présentant une grande complexité de vérification. Une étude sur les biais d'automatisation dans les diagnostics en endoscopie, menée en Pologne en 2025 aboutit à des résultats similairesⁿ.

Cette tendance à la surestimation des performances est à relier à une mise en sommeil ou, à tout le moins une réduction, de notre métacognition, définie comme la capacité à surveiller, évaluer et réguler nos propres processus cognitifs.

Participants (USA- 246)



Répartition des formations initiales des participants à l'étude.

Si l'on demande^o à des participants (246) de résoudre des problèmes de logique du LSAT (*Law School Admission Test*) sans aide ou avec ChatGPT^o, il apparaît que les utilisateurs soutenus par l'IA, plus performants dans leurs réponses, ont été incapables d'évaluer leurs succès avec précision. En moyenne, ils surestiment leur score de 4 points sur 20. Leur sensibilité métacognitive, c'est-à-dire leur capacité à distinguer leurs propres réponses correctes des incorrectes, se situe en dessous des standards habituels de surveillance de soi. Émerge également le paradoxe de la littératie IA mentionné plus haut : une meilleure connaissance technique de l'IA (littératie élevée) est ici corrélée à une moins bonne précision métacognitive. C'est-à-dire que les participants les plus familiers avec l'IA se montrent plus confiants dans l'évaluation de leur propre

18. V. Beauchene, S. Duranton, D. Martin, V. Lyon, and J. Walters, *AI at Work: Strategy Matters More Than Tools*; June 03, 2026, <https://www.bcg.com/publications/2026/ai-at-work-why-strategy-matters-more-than-tools>.

19. FDA's AI Medical Device List: Stats, Trends & Regulation, IntuitionLabs Report, march 2026, <https://intuitionlabs.ai/articles/fda-ai-medical-device-tracker>.

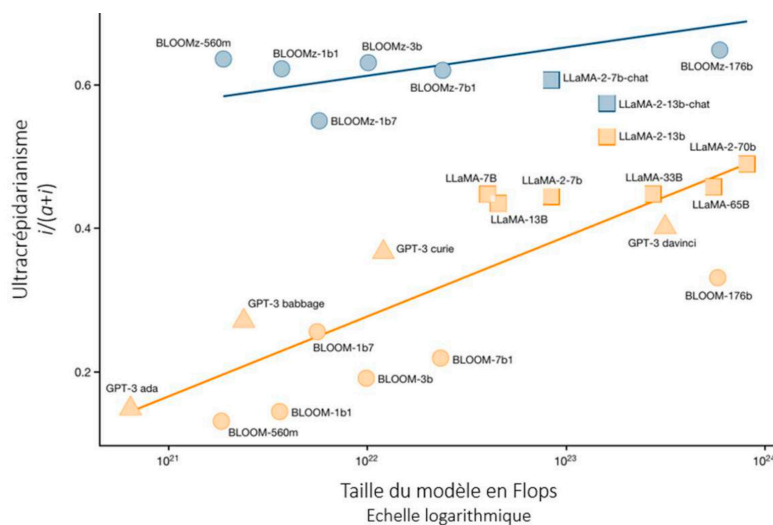
20. Gaube, S., Suresh, H., Raue, M. et al. *Do as AI say: susceptibility in deployment of clinical decision-aids*. npj Digit. Med. 4, 31 (2021).

21. Réseau de neurones convolutif à 112 couches, entraîné sur plus de 110 000 images de radiographie X.

travail, suggérant ainsi une forme d'illusion de savoir renforcée par la maîtrise de l'outil. Même si l'on ajoute des incitations financières, les participants sous IA continuent à se surestimer, ce qui peut laisser penser que le problème ne vient pas ici d'un manque d'effort, mais d'une distorsion profonde du suivi cognitif.

Les auteurs soulignent enfin la disparition de l'effet Dunning-Kruger. L'effet Dunning-Kruger classique^P (les moins compétents se surestiment et les plus compétents se sous-estiment) semble s'effacer avec l'usage de l'IA, cette dernière nivelant ce biais : tous les utilisateurs, quel que soit leur niveau de compétence initial, ont manifesté une surestimation uniforme de leurs capacités.

Ce mécanisme n'est pas sans lien avec l'ultracrépidarianisme²², qui désigne la tendance à prononcer des avis, affirmations ou conseils sur un sujet au-delà de ses connaissances réelles. Si l'apparition des réseaux sociaux a généralisé considérablement le phénomène (que l'on pense ici aux domaines de la santé, du sport, de la géopolitique pour lesquels les commentateurs sont légions sur Tiktok, Instagram ou autre), l'intelligence artificielle générative l'accroît encore sous l'effet de deux leviers : premièrement celui de la mise à disposition d'explications et de contenus courts, éventuellement erronés, que l'on ne vérifiera pas de toute façon, assortis de formulations chocs, deuxièmement la diffusion massive, puis la pollution des données d'apprentissage des IA que celle-ci engendre. On pourrait parler d'un ultracrépidarianisme assisté par IA. Assez ironiquement, les grands modèles de langage peuvent aussi succomber à l'ultracrépidarianisme qu'ils contribuent à favoriser. Là encore, de façon contre-intuitive, l'augmentation de la taille des modèles semble le favoriser.



L'ultracrépidarianisme croît avec la taille des modèles²³ (2024).

Ultracrépidarianisme mesuré par le rapport $i/(a+i)$ où i est le nombre de réponses incorrectes, a le nombre d'évitements dans la réponse.

Ceci nous amène à la question de l'esprit critique, à sa formation et partant au monde de l'éducation.

Le monde de l'éducation

Il existe depuis longtemps une réflexion tant en France qu'à l'étranger sur la place du numérique dans l'éducation²³, avec notamment une attention particulière accordée aux mécanismes d'apprentissage

22. Le terme ultracrépidarianisme date de 1819 dans un texte de l'écrivain William Hazlitt à propos d'un critique littéraire trop ambitieux, mais il provient de Pliny l'ancien via la locution *Ne supra crepidam sutor judicaret* : un cordonnier ne devrait pas émettre de jugement au-delà (au-dessus) de la chaussure.

23. E. Care, K. Anderson et H. Kim, *Visualizing the breadth of skills movement across education systems*, Brookings, 2016, <https://www.brookings.edu/articles/visualizing-the-breadth-of-skills-movement-across-education-systems/>.

mais aussi à la nature des compétences qu'il paraît souhaitable de favoriser²⁴. En France, aussi bien les parlementaires²⁵, le Ministère de l'éducation nationale²⁶, la BPI²⁷ ou le PIA2²⁸, ou encore des ONR comme l'Inria²⁹ s'y intéressent de près, mais il n'est pas dans l'objectif de cette communication d'en détailler les multiples projets, analyses ou outils. On se bornera à quelques études très récentes sur l'interaction entre la cognition, la métacognition et l'usage des intelligences artificielles génératives.

Bien que le potentiel des IA génératives pour aider, et surtout «ultra-individualiser», les apprentissages, leur rythme et leur suivi, soit souvent mis en avant, et corresponde sans nul doute à un progrès, il existe un phénomène paradoxal : si l'IA augmente massivement la productivité immédiate, elle peut gravement compromettre l'acquisition de compétences à long terme^r. Une étude expérimentale^s, menée auprès de 1 000 lycéens en mathématiques, nous apprend que l'accès à une version standard de GPT-4 (GPT Base) améliore les scores de pratique de 48 %, mais entraîne une chute de 17 % des performances lors de l'examen. Le phénomène de béquille cognitive, où l'étudiant recopie les solutions sans en traiter l'information, peut toutefois être largement atténué par l'utilisation de «garde-fous» pédagogiques. Ainsi, une interface conçue pour fournir des indices plutôt que des réponses (*GPT Tutor*) a permis une amélioration de 127 % lors de la pratique, tout en neutralisant l'effet négatif sur l'apprentissage à moyen terme.

Cette tentation de la facilité chez les apprenants conduit à une érosion de la persévérance dans l'apprentissage, notion qui peut paraître surannée, ou teintée de la nostalgie de nos propres efforts (directement liée à nos générations !). La question de la place de l'effort personnel dans l'apprentissage, le besoin d'espacement dans le temps des épisodes d'un même apprentissage, enfin la nécessité de resolliciter les connaissances acquises sont des notions qui sont nées bien avant l'essor de l'IA ; on a même pu parler en pédagogie de difficultés désirables³⁰. Mais il apparaîtra clairement au lecteur que la facilité d'accès aux ressources prédigérées que fournissent les IA génératives, ainsi que leur capacité à répondre de façon instantanée, vont à l'encontre de ces nécessités pédagogiques, cela est confirmé désormais par nombreuses études^{t u v}. Privant l'élève ou l'étudiant de l'expérience de la résolution autonome des problèmes, l'assistance systématique et récurrente de l'IA conduit inéluctablement à une persévérance réduite et, par voie de conséquence, à une baisse des performances sans assistance comme en témoigne une étude de 2026 sur plus de 1 200 personnes^w.

Le monde universitaire

On peut distinguer deux sources d'études et d'analyses dans le monde universitaire : celle que l'on peut grossièrement rassembler sous le terme de psychologie cognitive et celle constituée par les neurosciences. La première a non seulement permis d'analyser les effets des évolutions des pratiques éducatives, mais elle a également ouvert des pistes pour la seconde en suggérant des expérimentations sur lesquelles nous allons revenir.

Pour ce qui concerne l'étude par la psychologie cognitive des effets de l'IA sur l'apprentissage, on peut relever la notion récente d'auto-explication (*self-explanation*) par l'apprenant (ou l'utilisateur !) : l'élève détaillant avec son propre vocabulaire, ce qu'il a compris, pourquoi la réponse à une question est correcte ou encore comment il arrive à la solution. Ce processus est fondamentalement distinct de l'explicabilité

24. ONISEP. *Référentiel des compétences du XXI^e siècle*, <https://www.onisep.fr/avenir-s/accompagner-les-eleves/ressources/encourager-vos-eleves-a-decouvrir-les-competences-a-s-orienter-et-du-xxi-siecle/le-referentiel-des-competences-du-xxi-siecle>.

25. IA et éducation, Sénat, Rapport d'information n° 101, 30/10/2024, <https://www.senat.fr/rap/r24-101/r24-101.html>.

26. *Quels usages de l'IA en éducation ?* Conférence internationale du Conseil Scientifique de l'Éducation Nationale, 25 mars 2026, <https://www.csen.education.gouv.fr/quels-usages-de-l-ia-en-education>. Voir aussi <https://eduscol.education.gouv.fr/6702/les-intelligences-artificielles-et-leurs-usages-en-education>.

27. *Penser l'éducation et la formation à l'ère de l'IA générative*, BPI-France, 2024, https://www.youtube.com/watch?v=ZZBLPI_pVNO.

28. Appel à projets et plateforme e-Fran, <https://www.education.gouv.fr/e-fran-des-territoires-educatifs-d-innovation-numerique-326083>.

29. INRIA, *Éducation : personnaliser les méthodes d'apprentissage grâce à l'IA*, 2025, <https://www.inria.fr/fr/education-personnaliser-methodes-apprentissage-ia>.

30. Bjork EL, Bjork RA. *Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning*. In: *Psychology and the real world: essays illustrating fundamental contributions to society*. Worth Publishers, 2011. p. 56–64.

de l'IA elle-même dans laquelle le détail du raisonnement est fourni par l'IA, que l'élève peut ou peut ne pas lire, en tout cas rarement assimiler. Une méta-analyse 2018^x illustre l'efficacité de l'auto-explication sur la performance académique, en soulignant ses bénéfices pour le transfert de connaissances et l'étude d'exemples résolus, qui surpasse souvent l'explication directe. Son intégration via des tuteurs intelligents y est vivement recommandée. Une autre méta-analyse de 2026^y examine comment le feedback par IA, via des tuteurs intelligents et des systèmes de dialogue génératifs, soutient en effet l'apprentissage autorégulé, celui-ci englobant l'auto-explication. La notion d'auto-explication a été adaptée aux grands modèles de langage via l'auto-explication des invites (*prompts*)^z. Ces approches mettent aussi le doigt sur l'importance des erreurs profondes et de leur explication détaillée^{aa}, ainsi que sur les outils qui permettraient de les étayer (*scaffolding*). Encore faut-il que les utilisateurs, de l'étudiant au chercheur et à l'ingénieur, voire au citoyen, persévèrent pour utiliser ces techniques dans la durée^{ab}, voire même disposent de suffisamment d'esprit critique pour (entre autres) les identifier et chercher à les corriger. Ce qui questionne finalement notre capacité d'engagement lors de nos utilisations de l'IA générative.

Formulée de façon limpide et incisive par Descartes³¹, la notion d'esprit critique « Ne recevoir jamais aucune chose pour vraie que je ne la connusse évidemment être telle » s'est affinée avec le temps^{ac} et prend toute son importance dans notre relation aux IA génératives (et bien sûr également aux « choses » provenant des réseaux sociaux)³². La notion d'esprit critique doit être adaptée face aux défauts et risques que présentent, et font courir, l'IA générative et nous-mêmes, à notre processus cognitif : opacité des algorithmes, hallucinations, données incorrectes ou corrompues mais aussi biais d'automatisation, surconfiance, *AI crutch*,... Elle appelle par exemple les modifications suivantes^{ad}.

Collecte d'informations	→	Résolution de problèmes
Tâches d'analyse, de synthèse et d'évaluation	→	Vérification des informations
Intégration des réponses générées par l'IA	→	Gestion des tâches

Les neurosciences quant à elles permettent de mesurer de façon objective et en temps réel la façon dont notre cerveau fonctionne lorsqu'il est confronté à l'évolution d'outils technologiques directement en prise avec celui-ci. L'exemple le plus frappant et le plus remarqué a été la parution en 2025 d'un article^{ae} approfondi du *MIT media Lab* au titre provocateur : « Accumulation de dette cognitive lors de l'utilisation d'un assistant IA pour la rédaction de dissertations ». L'étude s'est déroulée sur quatre mois et a concerné 54 participants auxquels il a été demandé de rédiger un essai, participants séparés en trois groupes (*cerveau seul*, *recherche web*, *IA*). Le protocole comprenait quatre sessions de rédaction durant lesquelles les techniques d'électroencéphalographie, de traitement automatique du langage (NLP - *Natural Language Processing*), et d'évaluation comportementale ont été employées. Sur le plan expérimental, le pourcentage de participants incapables de citer des extraits de leurs dissertations est assez frappant, il atteint plus de 80 % pour le groupe ayant utilisé un LLM, alors qu'il n'est que de 10 % pour les autres groupes. Plus frappante encore est l'analyse EEG de la bande d'ondes alpha à l'aide de la fonction de transfert directe dynamique (dDTF), la connectivité dans la bande alpha étant en effet associée à l'attention interne et au traitement sémantique lors de l'idéation créative. La différence la plus marquée dans cette expérimentation apparaît au niveau de la connectivité dans cette bande, le groupe *cerveau seul* présentant des réseaux de traitement sémantique largement plus denses. La connexion entre les régions pariétale gauche et temporale droite a révélé des disparités extrêmement significatives entre les groupes. Le groupe *cerveau seul* a également présenté un flux d'informations cinq fois supérieur à celui du groupe *LLM* et le nombre total de connexions significatives pour le groupe *cerveau seul* était pratiquement le double de celui du groupe *LLM*.

31. René Descartes, *Discours de la méthode*, II^e partie, 1637.

32. Académie des technologies, *La lutte contre les informations toxiques à l'ère numérique*, juin 2026, <https://www.academie-technologies.fr/publications/la-lutte-contre-les-informations-toxiques-a-lere-numerique/>.

Cette expérimentation met en évidence ce qui a été appelé *la dette cognitive* (*cognitive debt*) sous trois aspects essentiels :

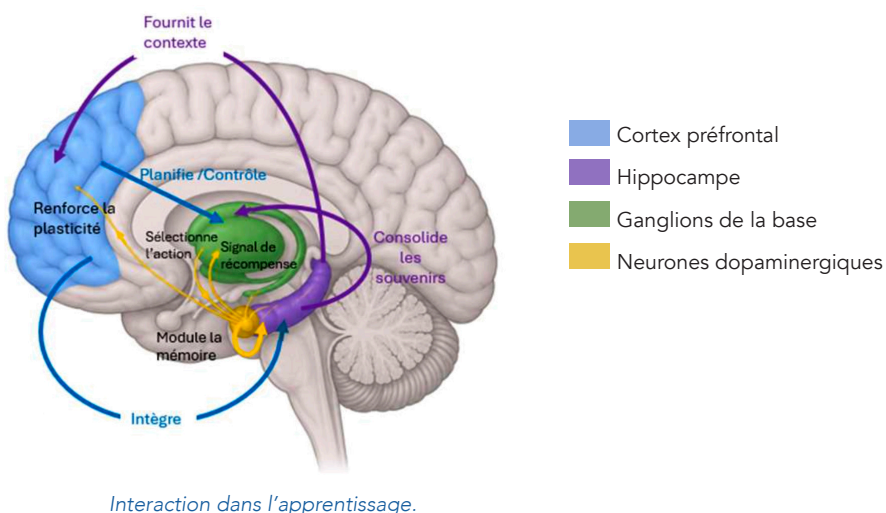
- la délégation mentale : demander à l'IA de penser, rédiger voire décider à notre place ;
- la perte d'habitudes intellectuelles : baisse des efforts fournis au quotidien par le cerveau ;
- l'affaiblissement biologique : avec moins de connexions neuronales créées dans les zones clés du cerveau ;

qui ont des impacts délétères sur le cerveau comme l'atrophie de la mémoire, la baisse de l'attention ou la disparition de l'esprit critique. On pourrait y ajouter pour faire bonne mesure, la perte de toute « fierté » de ce que l'on a produit.

Cependant des solutions commencent à être proposées couplant l'analyse EEG avec l'IA elle-même, comme dans le modèle *Interaction-Aware Language Transformer* (IALT)^{af} qui tente de synchroniser les signaux cérébraux avec les réponses de l'IA, dans l'objectif de moduler ses propres mécanismes d'attention en fonction de l'état attentionnel réel de l'utilisateur ou d'identifier les éléments spécifiques de l'interaction susceptibles de capter l'attention de l'utilisateur. Ce qui, bien que partant d'une intention louable d'efficacité, pourrait n'être guère rassurant quant au maintien de notre esprit critique déjà vacillant ou à l'altération de nos mécanismes de prise de décision.

Vers une explication par les neurosciences

Les progrès réalisés grâce aux moyens d'investigation³³ des neurosciences pour étudier la structure et le fonctionnement de notre cerveau permettent aujourd'hui d'en avoir une connaissance assez approfondie, ou du moins suffisante dans sa version vulgarisée, pour que l'on puisse espérer comprendre l'origine des phénomènes qui viennent d'être décrits. En simplifiant à l'extrême, on peut en donner l'image suivante.

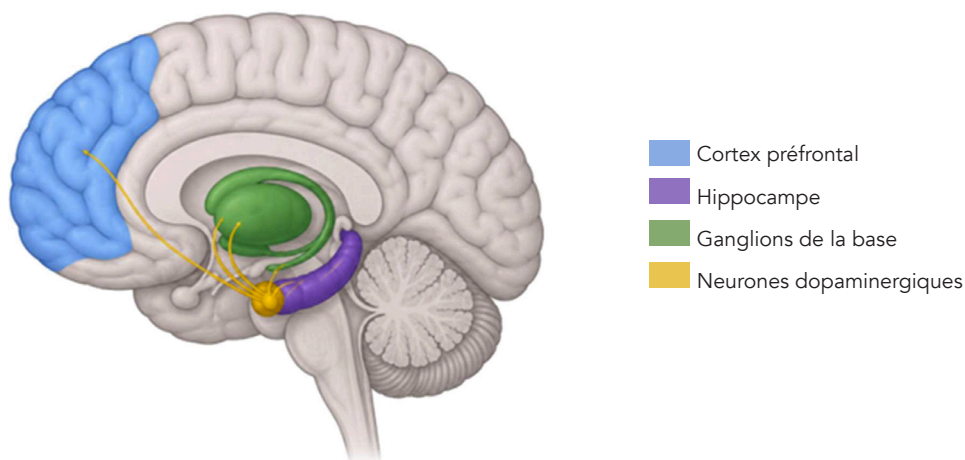


On ne peut qu'être frappé par la profusion des interactions entre les différents organes et l'exploitation des différents types de stimuli qui permettent de construire, approfondir et ancrer nos capacités cognitives. Deux mécanismes fondamentaux sont utiles à mentionner pour notre propos : la plasticité cérébrale ou neuroplasticité et l'élagage synaptique (*synaptic pruning*). La plasticité cérébrale désigne la capacité du cerveau à modifier son organisation, ses connexions et, parfois sa structure, en réponse à l'apprentissage et

33. IRM (imagerie par résonance magnétique) et IRM fonctionnelle, Tomographie par émission de positons (TEP), Électroencéphalographie (EEG), enregistrements neuronaux, méthodes moléculaires et génétiques.

sous l'effet des expériences, des émotions, des traumatismes et de façon générale de notre environnement. Elle repose principalement sur les neurones et les synapses (les connexions entre les neurones). L'élagage synaptique élimine certaines connexions neuronales inutiles ou sous-utilisées. Cet élagage contribue à optimiser les réseaux neuronaux, à améliorer l'efficacité cognitive, c'est-à-dire en fin de compte à spécialiser le cerveau et développer son efficacité. On utilise souvent de façon pédagogique et frappante l'expression « *use it or lose it* », pour résumer l'idée selon laquelle les connexions synaptiques régulièrement sollicitées ont davantage de chances d'être maintenues et renforcées, tandis que celles qui sont peu, voire pas utilisées, peuvent être affaiblies puis éliminées au cours du développement cérébral.

Quatre acteurs essentiels de la cognition et de la métacognition (mémoire procédurale)



Cortex préfrontal

Planification et prise de décision

Évalue les conséquences des actions, priorise les objectifs.

Mémoire de travail

Maintien et manipule temporairement l'information pour résoudre des problèmes ou apprendre de nouvelles compétences.

Contrôle inhibiteur

Filtre les distractions, supprime les réponses automatiques inappropriées.

Flexibilité cognitive

Permet de s'adapter à de nouvelles règles ou contextes.

Intégration sensorimotrice

Associe perceptions et actions pour l'apprentissage par l'expérience.

Hippocampe

Encodage de la mémoire épisodique

Stocke les souvenirs liés à des événements, des lieux, des contextes.

Consolidation de la mémoire

Transfère les informations à long terme vers le néocortex pendant le sommeil.

Navigation spatiale

Crée des « cartes cognitives » de l'environnement (neurones de lieu).

Apprentissage associatif

Lie des informations entre elles (associer visage et nom).

Ganglions de la base

Sélection et exécution des actions

Facilitent les mouvements et les habitudes automatiques.

Apprentissage par renforcement

Intègrent les signaux dopaminergiques pour ajuster les comportements futurs.

Formation d'habitudes

Transforment des actions répétées en routines (ex : conduire une voiture sans y penser).

Filtrage des informations

Inhibent les actions non pertinentes.

Neurones dopaminergiques

Signal de récompense

libèrent de la dopamine en réponse à une récompense ou à une prédiction de récompense.

Renforcement de l'apprentissage

La dopamine module la plasticité synaptique (renforcement des connexions neuronales impliquées dans une action récompensée).

Motivation et exploration

Encourage la répétition des comportements bénéfiques et l'exploration de nouvelles stratégies.

Bien que l'élagage synaptique ne supprime pas simplement toute synapse qui n'est pas utilisée et reste influencé par nombre de facteurs comme l'âge, la région cérébrale, l'expérience ou les facteurs génétiques, il ne se cantonne pas à la période de l'enfance et dépend aussi de l'activité neuronale et de l'environnement. On voit ainsi clairement combien déléguer un certain nombre de mécanismes d'apprentissage, de réflexion de répétition et d'approfondissement, d'expériences ou même limiter notre curiosité intellectuelle du fait de l'appel systématique et récurrent aux intelligences artificielles génératives, peut abimer notre cognition et notre métacognition, non seulement dans ses processus d'éducation précoce mais aussi tout au long de la vie. L'exposition universelle aux chatbots pourrait ainsi transformer profondément, et à une échelle inédite, l'esprit humain.

En guise de conclusion

Cette crainte que l'on peut concevoir quant au déclin de nos capacités cognitives, de notre métacognition et de notre esprit critique provient également du caractère finalement assez insidieux des phénomènes induits par l'émergence très récente de IA générative. On pourrait sans exagération reprendre la métaphore de la grenouille bouillie³⁴, qui plongée dans une eau dont la température s'élève extrêmement graduellement ne s'en échappe pas et finit pas y mourir, ou plus sérieusement se rattacher à la théorie de la normalité rampante (*creeping normality*)^{ag ah}.

Des technologies additionnelles ne vont pas dans une direction qui s'y opposerait comme le montre l'introduction en France du « mode IA » dans les navigateurs, option pas toujours désactivable, qui propose un résumé IA en en-tête des réponses à chaque requête. Outre le fait que nombre de sites sont de fait pillés par cette option, et que bien peu apparaissent en référence, pour ne rien dire de la baisse de leur fréquentation que cela provoque, on constate aux États-Unis³⁵ où l'option existe depuis déjà quelques mois, que moins de 1 % des internautes cliquent sur les sites référencés (en petit nombre pourtant, ce qui en soi pose la question de leur sélection algorithmique), et seulement 75 % poursuivent leur navigation, contre 85 % pour les pages sans résumé IA.

Tout n'est pas toujours irréversible en matière d'évolution d'usage des technologies, particulièrement dans le domaine de l'éducation, qui semble prendre la mesure du phénomène. Ainsi en est-il du retour vers l'apprentissage de l'écriture manuscrite cursive aux États-Unis³⁶. Après l'abandon de cet apprentissage, la Californie, l'Iowa et l'Oklahoma ont adopté récemment des législations rendant son enseignement obligatoire. Et une nette tendance se fait jour dans d'autres États, après en avoir constaté^{ai} les bienfaits chez les jeunes élèves (amélioration de la lisibilité de l'écriture, développement de la longueur des textes produits, renforcement de certaines compétences en lecture, perfectionnement de la précision orthographique et enrichissement des capacités narratives).

Il faut également relever un des paradoxes de l'IA générative associée aux algorithmes de filtrage ou de traitement des données : alors que l'on assiste à une ultraindividualisation, apportée par les technologies numériques et présentée bien souvent comme un progrès incontestable, l'IA générative ne fournit que des réponses très formatées et uniformes, caractéristiques qui forment d'ailleurs un des critères de reconnaissance des textes qu'elle produit. En optimisant (réduisant) la longueur des réponses, on aboutit mécaniquement, et au-delà du principe même de la construction de ces textes (essentiellement de nature statistique), à limiter faute de place les nuances et à lisser les contradictions, définissant ainsi de fait ce qui constitueraient

34. L'expression provient d'une expérience du physiologiste Friedrich Goltz, en 1869. Celle-ci a été controversée ainsi que son interprétation (et l'oubli que la grenouille était décérébrée), mais les expériences négatives qui ont été reproduites n'ont pas été conduites avec la même lenteur de montée en température. La métaphore est aussi reprise par Gregory Bateson G. Bateson, *La Nature et la pensée*, Paris, Seuil, 1984.

35. Google users are less likely to click on links when an AI summary appears in the results, PewResearch Center, juillet 2025, <https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/>.

36. Une vieille méthode remise au goût du jour... et les résultats en lecture s'améliorent, Futura science, juillet 2025, <https://www.futura-sciences.com/sante/questions-reponses/psychologie-vieille-methode-remise-gout-jour-resultats-lecture-ameliorent-22791/>.

les critères d'une « bonne réponse ». Pour ne rien dire de la tendance d'une IA empathique et chaleureuse à nous conforter dans nos appréciations. De ce point de vue, l'encyclique du pape Léon XIV³⁷, bien qu'elle se place plus sur le plan spirituel et moral, identifie également les risques d'une capacité cognitive réduite par l'appel systématique à l'IA, notamment sur les capacités de jugement ou l'uniformisation des critères de jugement (un très court texte de 2025 est encore plus explicite sur le développement intellectuel et neurologique des enfants³⁸). Elle appelle donc plus de recul sur l'usage de l'IA « *Une aide précieuse qui requiert de l'attention* ».

Deux formules pour résumer notre propos, que les études citées nous semblent étayer de façon solide :

- On ne peut se contenter de consommer fugacement de la connaissance si l'on veut véritablement apprendre.
- Il n'est pas possible de déléguer ce que l'on n'a pas déjà appris, ou appris à faire, sans risquer de perdre tout contrôle.

Si des communautés réagissent déjà sur les dangers d'une utilisation immodérée des IA génératives (scientifiques, éducateurs, artistes, entreprises...), bien peu de voix s'élèvent sur les risques, pour notre cognition individuelle et pour notre intellect. Dans les débats pourtant fournis que suscite l'intelligence artificielle, qu'ils s'écharpent sur l'économie, l'environnement, les questions juridiques et réglementaires, ou encore géopolitiques, cette question est totalement absente. Le maintien, voire le développement, de l'esprit critique, indispensable aussi dans d'autres domaines comme celui de la lutte contre les informations toxiques³⁹, gagnerait à s'inspirer du mouvement hygiéniste du XIX^e siècle, et du début du XX^e qui a su, malgré des débordements qu'il conviendrait de ne pas reproduire, transformer profondément notre rapport à la santé individuelle, mêlant d'abord doctrines et pratiques basées sur les progrès de la science, puis interventions de la société civile et de l'État. Après tout, si l'on porte désormais attention à sa santé physique et mentale (manger 5 fruits par jour, avoir une activité physique et limiter la sédentarité, prendre soin de son sommeil,...) pourquoi ne pas imaginer aborder le champ des usages quotidiens de l'IA puisque ceux-ci se développent à une vitesse jamais connue jusqu'ici pour une technologie⁴⁰, même numérique, avec des risques que l'on commence à documenter très sérieusement.

Et préserver la culture de l'esprit critique et le goût de l'effort intellectuel.

37. Notamment le paragraphe 99. Lettre encyclique *Magnifica Humanitas* du Saint Père Léon XIV sur la protection de la personne humaine à l'ère de l'intelligence artificielles. Mai 2026, <https://www.vatican.va/content/leo-xiv/fr/encyclicals/documents/20260515-magnifica-humanitas.html>.

38. Discours du pape Léon XIV aux participants à un congrès sur « intelligence artificielle et sauvegarde de la maison commune », *L'Osservatore Romano*, Édition en langue française, année LXXVI^e, n° 8, juillet 2025.

39. *La lutte contre les informations toxiques à l'ère numérique*, rapport de l'Académie des technologies, juin 2026, <https://www.academie-technologies.fr/publications/la-lutte-contre-les-informations-toxiques-a-lere-numerique/>.

40. *Baromètre du numérique – Édition 2026*, <https://www.arcom.fr/se-documenter/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/barometre-du-numerique-edition-2026>.

Références

- a Dai, Z., Gifford, D.K. Outputs of generative diffusion models are often unattributable. *Nat Commun* 17, 6974 (2026).
- b I. Good, *Speculations Concerning the First Ultraintelligent Machine*, Advances in Computers, Volume 6, 31-88, 1966.
- c V. Vinge, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, 1993, <https://edoras.sdsu.edu/~vinge/misc/singularity.html>.
- d Evans, J., Bratton, B., & Agüera y Arcas, B. Agentic AI and the next intelligence explosion, mars 2026. arXiv, arXiv:2603.20639.
- e Clark A, Chalmers D. *The extended mind*. Analysis. 1998;58(1):7-19.
- f B. Sparrow et al., *Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips* Science, New Series, Vol. 333, No. 6043 (5 August 2011), voir aussi A. Clark, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*, Oxford Univ. Press, New York, 2003.
- g Clemenson, G.D., Maselli, A., Fiannaca, A.J. et al. Rethinking GPS navigation: creating cognitive maps through auditory clues. *Sci Rep* 11, 7764. (2021).
Ben Elai, *An exploratory real-world wayfinding experiment: A comparison of drivers' spatial learning with a paper map vs. turn-by-turn audiovisual route guidance*, Transportation Research Interdisciplinary Perspectives, Volume 9, March 2021, 100280.
- h M. Polanyi. *The Tacit Dimension*. London and New York: Routledge and Doubleday and Company. 1967.
A. Reber *Implicit Learning and Tacit Knowledge*, Journal of Experimental psychology, 118(3):219-235, 1989.
- i D. Ha & J. Schmidhuber: *World Models*, march 2018, arXiv:1803.10122.
D. Hafner, T. Lillicrap, J. Ba, M Nourouzi: *Dream to Control: Learning Behaviors by Latent Imagination*, march 2020, arXiv:1912.01603.
- j B. C. Lee, J. Chung, *An empirical investigation of the impact of ChatGPT on creativity*, Nature Human Behaviour, 2024, 8(10), pp 1906–1914.
- k Gupta, G., Chaudhuri, N. Decoding GenAI Assimilation in Teams: A Multi-Method Study of GenAI's Impact on Collaborative Performance in Project-Based Work. *Inf Syst Front* (2025).
- l P Rajpurkar, J Irvin, K Zhu, B Yang, H Mehta, T Duan, D Ding, A Bagul, C Langlotz, *Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning*, arXiv:1711.05225, 2017.
- m Lyell D, Coiera E. *Automation bias and verification complexity: a systematic review*. J Am Med Inform Assoc. 2017 Mar 1; 24(2): pp 423–431.
- n Budzyń K, Romańczyk M, Kitala D et al., *Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study*, The Lancet Gastroenterology & Hepatology, 2025; 10, 896-903.
- o D. Fernadesa et al., *AI makes you smarter but none the wiser: The disconnect between performance and metacognition*, Computers in Human Behavior 175 (2026).
- p Justin Kruger et David Dunning, *Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments*, Journal of Personality and Social Psychology, Vol. 77, No 6, December 1999, p. 1121–1134.
- q Zhou, L., Schellaert, W., Martinez-Plumed, F. et al. *Larger and more instructable language models become less reliable*. Nature 634, 61–68 (2024).
- r Yan, L., Greiff, S., Lodge, J.M. et al. *Distinguishing performance gains from learning when using generative AI*. Nat Rev Psychol 4, 435–436.
- s H. Bastani, O. Bastani, A. Sungu, H. Ge, O. Kabakci, R. Mariman, *Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics*, Proceedings of the National Academy of Sciences, Vol. 122, Issue 26, 2025.
- t Carpenter, S.K., Pan, S.C. & Butler, A.C. *The science of effective learning with spacing and retrieval practice*. Nat Rev Psychol 1, 496–511 (2022). <https://doi.org/10.1038/s44159-022-00089-1>.
- u Y. Chen et al, *Effects of generative artificial intelligence on cognitive effort and task performance: study protocol for a randomized controlled experiment among college students*, Trials. 2025 Jul 11; 26(1): 244. <https://doi.org/10.1186/s13063-025-08950-3>.
- v S. Rao, *The Impact of Artificial Intelligence Tools on Human Cognitive Abilities: A Comprehensive Review*. Innovapath, Vol.1, No4, <https://doi.org/10.63501/hsdq5611>.
- w G. Liu, et al, *AI Assistance Reduces Persistence and Hurts Independent Performance*, Conference paper at COLM 2026, arXiv:2604.04721.
- x Bisra, K., Liu, Q., Nesbit, J.C. et al. *Inducing Self-Explanation: a Meta-Analysis*. Educ Psychol Rev 30, 703–725 (2018).
- y Huang, CQ., Lu, LN., Huang, QH. et al. *Effects of Artificial Intelligence Feedback on Students' Self-Regulated Learning in Higher Education: A Three-Level Meta-Analysis*. Educ Psychol Rev 38, 64 (2026).

- z Haoyu Gao et al., *Self-Explanation Prompting Improves Dialogue Understanding in Large Language Models*, arXiv:2309.12940.
- aa Q. Zhang, L. Fiorella, *Effects of self-explaining feedback on learning from problem-solving errors*, Contemporary Educational Psychology, Vol. 79, December 2024.
- ab Gentner, N.M., Respondek, L. & Seufert, T. *Effects of short- and long-term prompting in learning journals on strategy use, self-efficacy, and learning outcomes*. Instr Sci, 52, 919–950 (2024).
- ac Ennis, R. H. (1962). *A Concept of Critical Thinking*. Harvard Educational Review, 32(1), 81–111.
Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. The Delphi Report. <https://eric.ed.gov/?id=ED315423>.
- ad Lau et al., *Understanding Critical Thinking in Generative Artificial Intelligence Use: Development, Validation, and Correlates of the Critical Thinking in AI Use Scale*. Computers in Human Behavior Reports, Vol. 22, May 2026, 2025.
H. Lee et al., *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*, 2025, CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Article No 1121, Pages 1–22.
- ae N. Kosmyna et al., *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing*, arXiv:2506.08872.
- af Jiang T, Wu J, Leung SCH. *The cognitive impacts of large language model interactions on problem solving and decision making using EEG analysis*. Front Comput Neurosci. 2025.
- ag Jarić, I., Pipek, P., Canavan, S., Firth, J.A. and Bertram, M.G. (2026), *The creeping normality of AI in the life sciences*. Front Ecol Environ, 24.
- ah P. D. Glynn, A. A. Voinov, C. D. Shapiro, and P. A. White, *From data to decisions: Processing information, biases, and beliefs for improved management of natural resources and environments*, Earth's Future, 5, 356–37, 2016.
- ai Van der Weel FRR, Van der Meer ALH. *Handwriting but not typewriting leads to widespread brain connectivity: a high-density EEG study with implications for the classroom*. Front Psychol. 2024.