De l'indocilité dans l'utilisation de l'IA générative

Hokusai, *La Grande Vague de Kanagawa*, 1830. Estampe gravée sur bois

À toute époque, les idées de la classe dominante sont les idées dominantes ; autrement dit, la classe qui est la puissance *matérielle* dominante de la société est en même temps la puissance *spirituelle* dominante.

Karl Marx, *L'idéologie allemande* (1845-1846)

Le retard n'a de sens que pour celui qui s'est engagé dans la course, il n'en a aucun pour qui ouvre un nouveau chemin.

Anne Alombert, *De la bêtise artificielle* (2025)

D'après une étude menée en France par le Baromètre BORN AI 2025, 42% des jeunes entre 18 et 25 ans déclarent faire un usage quotidien ou plus fréquent d'une IA générative (ChatGPT, Meta AI

ou Gemini, Copilot, DeepSeek etc.), alors que 93% des jeunes déclarent en avoir déjà utilisé au moins une. Aucune disparité significative de genre n'est constatée.

Sous les incitations et les autorisations multiples à utiliser l'IAg, ces pourcentages vont croître, sans aucun doute. On remarque déjà que cette utilisation s'installe dans les gestes quotidiens, dans la recherche d'informations, par exemple, ou dans les prises de décision relatives aux choix personnels (conseils relationnels ou émotionnels), scolaires, sociaux ou de consommation.[1]

Dans nos gymnases vaudois, aujourd'hui, en miroir de ce constat, nous sommes appelé·e·s, enseignant·e·s et élèves, à prendre le chemin d'une utilisation responsable de l'IA.[2]

Pour nous, enseignant·e·s, il s'agit de comprendre cette technologie, ses conditions historiques d'existence et son fonctionnement, et d'apprendre à en accompagner l'utilisation dans le cadre des études gymnasiales, afin que cette utilisation respecte les objectifs pédagogiques, les règles en vigueur (notamment les droits d'auteur et la propriété intellectuelle), les devoirs de transparence et d'usage raisonné (tout usage est dûment annoncé, documenté, et justifié : ce n'est pas parce que cette technologie existe et qu'elle est utilisable qu'il faut l'utiliser obligatoirement, pour parler comme Servane Mouton [3], il faut – c'est impératif ! – pouvoir en défendre l'usage ou s'abstenir).

Pour vous, élèves, il s'agit de coopérer avec cet accompagnement, ce qui implique d'en comprendre les raisons et d'adopter un ensemble de pratiques choisies en fonction de deux critères :

1°) la découverte des potentialités et des limites de la technologie, et

2°) la conservation d'un esprit critique face aux productions des IA génératives (textes ou images), ou face aux recommandations qu'on aura sollicitées par nos demandes.

En clair, nous sommes appelé·e·s à en inventer une utilisation qui ne soit pas « passive » [4], ce que je vais appeler ici un usage *indocile*. "Indocile", note le *Trésor de la langue française*, est celui « qui n'obéit pas facilement, ne se laisse pas diriger ou persuader. » Soit ! Mais dans l'assignation de cette tâche se dissimule un demi-aveu.

En effet, dans cet appel à la responsabilité individuelle des élèves utilisateurs et à la responsabilité éducative des enseignants, au discernement technologique et à la modération, on reconnaît implicitement que l'utilisation première et ordinaire de cette technologie n'est pas responsable, raisonnée et équilibrée. L'aveu pleinement explicité, c'est que le déséquilibre ne vient pas seulement et pas essentiellement de l'usager – dont évidemment l'éducation reste à faire –, mais du dispositif d'intelligence artificielle générative lui-même. Par là-même, on admet que la conception usuelle de l'IAg comme un simple *outil* – en lui-même « neutre » [5], puisque seuls les usages en

[1] En ce début de texte, nous nous référons aux propos de Ioan Roxin, professeur émérite à l'Université Marie et Louis Pasteur, recueillis par Anne Orliac pour *Polytechnique Insights (La revue de l'Institut Polytechnique de Paris)* dans un article intitulé *IA générative : le risque de l'atrophie cognitive* (paru en ligne le 3 juillet 2025). Roxin y renvoie au Baromètre Born AI 2025 : les usages de l'IA générative chez les 18-15 ans (donnée de décembre 2024).

[2] Le « Guide d'utilisation responsable de l'IA à l'usage du corps enseignant du secondaire II », édité par la DGEP en décembre 2024, mentionne des chiffres semblables.

[3] Mouton, Servane, *Écrans, un désastre sanitaire (Il est encore temps d'agir)*, Paris : Gallimard/Tracts, n°65, 2025.

[4] Voir « l'opinion » exposée par Dominique Dirlewanger dans le journal *24 Heures* du 2 octobre 2025 sous le titre : « L'IA rend-elle l'humanité idiote ? »

[5] Voir Lefebvre, Olivier, « ChatGPT, c'est juste un outil ! » : les impensés de la vision instrumentale de la technique (28 juin 2025), accessible en ligne dans *Terrestres (la revue des écologies radicales)*.

établiraient les buts et en susciteraient les inévitables évaluations éthiques, sociales, environnementales et pédagogiques, – est inadéquate et risque de donner au propos une orientation déroutante.

En somme, ce qu'on admet de décisif, c'est que le dispositif lui-même, tel qu'il est conçu aujourd'hui par les ingénieurs des quelque rares géants de la technologie, est soumis à une logique d'efficacité immédiate ou d'optimisation de la productivité [6] qui s'avère incompatible avec le métier de l'élève, un métier qui, pour l'essentiel, n'est pas de produire *plus* des connaissances, mais d'apprendre à rechercher et à acquérir *suffisamment* de connaissances et de compétences afin de permettre une maturation intellectuelle, sociale et émotionnelle aussi aboutie que possible. Or, dans l'état présent de cette technologie, il semble bien qu'elle ne puisse pas être un support d'apprentissage compatible avec une telle *finalité* [7], à moins que, semble-t-on nous proposer, nous apprenions à en faire une utilisation qui pourrait devenir *responsable* et *raisonnée*, parce ce qu'elle serait fondamentalement *indocile*, c'est-à-dire allant *sciemment à contre-courant* de la logique à laquelle la technologie est, quant à elle, fermement soumise. [8] Mais qu'est-ce à dire, au juste ? Et cela est-il seulement possible ? Telles sont les deux questions que je souhaite développer dans ce qui suit.

L'utilisation première et personnelle de l'IA (par quoi je n'entends pas seulement le *Shadow IA* – l'utilisation clandestine – dans les entreprises et les administrations, mais bien toute prise en main de l'IA, experte ou débutante) n'est ni responsable ni raisonnée au plan des effets sur l'architecture cérébrale et cognitive, sur les comportements, sur le plan des impacts environnementaux et sociaux. Cela ne tient pas à l'utilisateur, mais à la technologie elle-même qui produit des dépendances réelles sur tous ces plans : neuronal, cognitif, psychique, comportemental, social, sociétal. Quelles sont ces dépendances ? Qu'est-ce qui les produit ?

L'IAg semble pouvoir produire n'importe quel « contenu » (codes, textes, synthèses, traductions, rapports, images, résolutions de problèmes, etc.) à notre place. Comme le relève opportunément Camille Dejardin, après l'informatique et l'essor du web et des réseaux sociaux, « laissant croire que le savoir n'a plus besoin d'être *acquis* puisqu'il serait toujours à portée de clic sur Internet (lui-même toujours à portée de main sur l'ordinateur de poche qu'est le *smartphone*) », les fascinantes IAg sont sur le point de « rendre obsolète l'idée même de *recherche*. Le résultat final est servi sur un plateau, quand il n'est pas suggéré avant requête » [9] – comme on peut le constater sur Google en formulant dans la barre de recherche une vague question.

Parlons clairement ! Si l'IAg est « une promesse de facilité », « de temps gagné », « de liberté » retrouvée et une apparente assurance de résultats tangibles et immédiats, on comprend bien que l'utilisateur *lambda* accepte de plus en plus facilement « de se départir de l'effort même de connaître, de maîtriser, de comprendre même dans des champs qui étaient jusqu'ici tenus pour les prés carrés de l'intelligence humaine », à savoir les disciplines du savoir. Au fur et à mesure que la

---

[6] Voir le compte rendu, par Ulysse Lojkine, intitulé « IA : face au nouveau taylorisme, une stratégie du « renouveau luddite », conversation avec Juan Sebastián Carbonell dans la revue *Le Grand Continent* (en ligne : consulté le 21 octobre 2025). Contribution à l'occasion de la sortie du livre de Juan Sebastián Carbonell, *Un taylorisme augmenté. Critique de l'intelligence artificielle*, Paris : Éd. Amsterdam, 2025.

[7] Voir Dejardin, Camille, *À quoi bon apprendre ?*, Paris : Gallimard/Tracts (n°69), 2025. Et deux autres. Voir également, de la même autrice, l'interview dans la revue *Le 1 Hebdo* du 27 août 2025 (n°558), publiée sous le titre « L'humanité de sous-estime si elle pense que l'IA fait mieux qu'elle ». Ou encore dans l'édition du *Courrier international* (n°1819 du 11 au 17 septembre 2025) sur le thème « Le jour où nous n'aurons plus besoin d'apprendre », les extraits de Sibylle Anderl issus d'un article paru le 4 août dans le quotidien allemand *Die Zeit*.

[8] Voir le compte rendu, par Ulysse Lojkine, intitulé « IA : face au nouveau taylorisme, une stratégie du « renouveau luddite », conversation avec Juan Sebastián Carbonell dans la revue *Le Grand Continent* (en ligne : consulté le 21 octobre 2025).

[9] Dejardin, Camille, *À quoi bon apprendre ?*, Paris : Gallimard/Tracts (n°69), 2025, p. 5.

puissance des IAg l'entoure, les efforts *déjà* économisés font apparaître les efforts *encore* à fournir comme « indus », comme autant d'obstacles qui nous séparent de satisfactions de plus en plus impossibles à différer.

Les élèves que j'ai en face de moi, « dont le seul devoir est d'*apprendre* et de s'*exercer* », voient très bien l'enjeu. Inutile d'insister. Alors que partout ailleurs l'utilisation de l'IAg automatise les tâches intellectuelles, laisse entrevoir un gain de productivité, on vous soupçonne, vous, de remplacer massivement par l'IAg vos propres capacités intellectuelles [10] et on vous appelle, vous plus que tout autre, à cultiver à l'égard de l'IA un esprit critique, ici comme en Chine.[11]

Mais est-ce dans vos intentions qu'il faut plonger un regard soupçonneux – ces intentions, louables ou déplorables qu'elles soient, sont en partie explicables par les nombreuses et inévitables difficultés du métier d'élève ? N'est-ce pas plutôt, et plus résolument, vers le fonctionnement de la technologie elle-même et vers les effets réels qu'elle semble produire qu'il faut tourner le regard ? Mesurés aux résultats quasi instantanés que l'IAg fabrique, comment les apprentissages *à faire* ne deviendraient-ils pas de plus en plus « synonyme d'effort superflu (autant utiliser son temps à se divertir), insatisfaisant (par sous-estimation ou atrophie des capacités) ou encore perdu d'avance (puisqu'on nous annonce la fin prochaine de la valeur humaine) » ? Quand la substitution des compétences computationnelles humaines par les automates algorithmiques avance à si grand train, *à quoi bon apprendre ?* La question devient brûlante.

Une manière de répondre à cette question, c'est de regarder dans ce que les IAg nous *désapprennent*.

L'utilisation des IAg peut avoir des effets d'ordre à la fois neurologique, psychologique (cognitif et comportemental) et philosophiques qu'il paraît difficile d'endiguer. Tout récemment, le MIT (*Massachusetts Institute of Technology*) a mené une recherche sur quatre mois. Elle a impliqué 54 participants auxquels on a demandé de rédiger des essais sur des œuvres littéraires, sans aucune aide, ou bien avec un accès à Internet *via* un moteur de recherche (recherche web classique), ou encore avec ChatGPT.[12] En comparant les résultats des trois groupes, les chercheurs ont observé que l'assistance par LLM réduisait significativement le temps de rédaction (moins 60%), mais elle réduisait aussi la « charge cognitive pertinente » (soit l'effort nécessaire pour transformer une information en connaissance : moins 32% – mesuré par EEG), entraînant ce qui ressemble à une « atrophie cérébrale », c'est-à-dire à une diminution de la « connectivité cérébrale » (moins 55%), ce qui se traduit par :

> – un engagement mental plus faible dû probablement à l'absence d'implication interne dans la génération du contenu [i] ;

[10] Dejardin, Camille, *À quoi bon apprendre ?*, Paris : Gallimard/Tracts (n°69), 2025, p. 5, voir la n°2.

[11] Voir le « Guide d'utilisation responsable de l'IA à l'usage des élèves », édité par la DGEP en décembre 2024, ou dans l'édition du *Courrier international* (n°1819 du 11 au 17 septembre 2025) sur le thème « Le jour où nous n'aurons plus besoin d'apprendre », les extraits de Yan Yong issus d'un article paru le 20 février 2024 dans le quotidien chinois Jingji Guancha Bao : « L'IA est un outil, pas un substitut ».

[12] Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes, « Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task ». Sur le site de la revue Le Grand Continent, en date du 19 juin 2025, le lecteur trouvera la version intégrale du *preprint* publié par le MIT Media Lab le 10 juin, et un compte rendu de l'article publié sous le titre « ChatGPT est-il en train de casser le cerveau humain ? 5 points sur le *preprint* du MIT sur les effets de l'IA ». Dans les « notes de fin », le lecteur trouvera de larges extraits de ce compte rendu, nous les citons intégralement.

– un sentiment de propriété diminué sur le texte ; un sentiment ambivalent (quant au fait d'être ou de ne pas être l'auteur·e du texte) qui pourrait provenir de l'assistance obtenue, elle pourrait faire obstacle à l'appropriation approfondie de la matière, à la rétention mémorielle et à l'engagement dans un rapport authentique avec le matériel écrit [ii] ;

– une moins bonne mémorisation mise en évidente par une moins bonne capacité à citer son écrit quelques minutes déjà après sa transmission, signe d'un degré moindre d'appropriation de l'écrit [iii].

Lors d'une session finale où certains participants passaient de l'IA à l'écriture autonome, l'activité cérébrale est restée anormalement basse, suggérant un effet durable de « désengagement cognitif ».

Ces résultats suggèrent que l'usage régulier des assistants IA, en délégant trop rapidement des fonctions mentales essentielles (recherche, synthèse, formulation), pourrait induire une « dette cognitive »». Avec le temps, écrire avec ChatGPT augmenterait cette « dette cognitive », ce qui signifie, en clair, qu'un retour à une activité cérébrale normale pour les tâches effectuées sans LLM deviendrait de plus en plus difficile et coûteux en terme d'effort. La recréation de connexions et le retour à une riche activité cérébrale seraient laborieuses. Au contraire, les individus qui ont le droit d'utiliser l'IA après trois essais libres rédigés sans aucune assistance ont un pic inédit d'activité, signe qu'ils utilisent la technologie comme *multiplicateur* plutôt que comme *substitut* (et les *prompts* sont plus précis et variés). Chez eux, le passage à l'écriture soutenue par l'IAg suscite des niveaux élevés d'intégration cognitive, de réactivation de la mémoire et de contrôle raisonné de la rédaction.

Dans cette même études du MIT, on a remarqué que les textes produits avec ChatGPT étaient plus homogènes sur le plan des contenus et de la planification ainsi que sur le plan linguistique, et moins diversifiés que ceux écrits sans outil ou avec recherche web. En d'autres termes, les auteurs tendent à se concentrer sur un ensemble d'idées plus restreint, ce qui semble indiquer un engagement moins profond dans les sujets étudiés, une réception non critique du matériel fourni par le LLM. Cela reflèterait, selon les chercheurs, l'accumulation de la « dette cognitive » : le recours répété à des systèmes externes tels que les LLM remplacerait des processus cognitifs exigeants nécessaires à la pensée indépendante et critique, par des processus purement intégratifs, c'est-à-dire des automatismes. À court terme, la « dette cognitive » reporterait l'effort mental, mais à long terme, elle entraînerait une diminution de l'esprit critique, une vulnérabilité accrue à la manipulation et une baisse de la créativité. L'étude conclut que « lorsque les participants reproduisent des suggestions sans en évaluer l'exactitude ou la pertinence, ils renoncent non seulement à la propriété des idées, mais risquent également d'intérioriser des perspectives superficielles ou biaisées. »

De plus, le fonctionnement, le *design* et l'anthropomorphisme des IAg augmentent le risque de dépendance à l'IAg. En effet, comme le relève Ioan Roxin, « elles semblent s'exprimer comme des humains, elles semblent s'adapter à nos comportements ou à nos préférences et à nos convictions, elles semblent avoir réponse à tout, elles ont un fonctionnement ludique, elles relancent sans arrêt la conversation et se montrent extrêmement complaisantes à notre égard. » Or, la dépendance, ou mieux : l'addiction, est deux fois nocive. Non seulement elle renforce les autres risques déjà évoqués, mais elle peut « engendrer un isolement social, un désengagement réflexif (si une IA peut répondre à toutes mes questions, pourquoi apprendre ou penser par moi-même ?), voire un sentiment d'humiliation profond face à l'incroyable efficacité de ces outils ».[13]

Une atrophie cognitive généralisée est un risque sanitaire que diverses études observent, mais il en annonce un autre, tout aussi redoutable, sur le plan philosophique et politique. Si ce type d'"outils" est utilisé largement, nous risquons aussi une uniformisation linguistique et de la pensée. En

[13] Roxin, Ioan, « IA générative : le risque de l'atrophie cognitive » (3 juillet 2025).

2024, des travaux conduits par deux chercheurs britanniques ont montré que lorsque des auteurs demandaient à ChatGPT d'améliorer leurs textes, le gain individuel pouvait être important, mais la créativité globale du groupe diminuait.

Un autre risque concerne la pensée dite « critique », qui repose sur la capacité de comprendre les liens logiques entres les idées, de conduire un raisonnement avec rigueur, d'identifier et de corriger les incohérences qui s'y produisent, d'évaluer et de hiérarchiser les arguments.

L'utilisation massive d'IAg diminue l'habitude de la pensée critique. Une étude menée par Microsoft et par l'Université Carnegie Mellon fait apparaître une corrélation négative entre la fréquence d'usage des outils d'IA et le score de pensée critique. [14] L'étude conclut à un délestage cognitif qui s'amplifie lorsque la confiance dans le modèle surpasse la confiance en ses propres compétences. Or, garder un esprit critique affûté est crucial, car « ces IA peuvent non seulement se tromper ou répercuter des *biais,* mais également dissimuler l'information ou simuler la conformité ». [iv]

Une autre étude menée en 2024 par un chercheur à Zurich le confirme [15] : il y a une corrélation négative significative entre l'utilisation fréquente des applications d'IA et les capacités de pensée critique, plus le délestage cognitif croît, moins la capacité à conserver un esprit critique s'atteste. Cette étude relève aussi que les participants les plus jeunes ont montré une plus grande dépendance aux services d'IA et des scores de pensée critique inférieurs à ceux des participants plus âgés. De plus, il apparaît que le niveau d'études plus élevé était associé à de meilleures compétences en pensée critique, quelle que soit l'utilisation de l'IA. Ces résultats mettent en évidence les coûts cognitifs potentiels de la dépendance aux outils d'IA, et signalent un creusement des inégalités : ces coûts sont d'autant plus significatifs à supporter pour les jeunes et les jeunes moins favorisés socialement et par un niveau d'instruction soutenu.

Que retenir de ce tour dans la littérature scientifique récente ? L'IA ne peut être un levier pour l'intelligence et la créativité que sous une condition, à savoir que nous restions capables de penser, d'écrire, de créer sans elle. Or, cela suppose à la fois d'engager activement notre esprit critique et de continuer à exercer nos circuits neuronaux.

Dans ce but, le plus simple, note Roxin, est de « toujours douter des réponses données par les générateurs de texte et d'engager sa volonté pour réfléchir attentivement à ce que l'on lit, entend ou croit. La réalité est complexe, elle ne peut pas être appréhendée avec quelques savoirs superficiels. Prendre et conserver l'habitude de confronter son point de vue et ses savoirs avec ceux d'autres humains, si possible compétents, cela reste la meilleure manière d'approfondir sa pensée. »

S'il est nécessaire d'éduquer à un tel usage raisonné, ce ne peut être que par un *détournement* de l'usage ordinaire des IAg, qui en ferait connaître les effets normatifs afin de les contenir, qui tenterait d'en exploiter les potentialités en endiguant les effets humainement – et écologiquement ![16] – dévastateurs.

Mais s'il est nécessaire de tenter de s'émanciper de ces usages inappropriés, en faisant de l'IAg un usage indocile pour faire obstacle à la dépendance sous toutes ses formes, à plus forte raison, et dans le même but, il est nécessaire et sensé de ne pas du tout en faire usage, aussi souvent et longtemps que possible, en tous cas quand l'atteinte des objectifs pédagogiques ordinaires de l'enseignement gymnasial est en jeu. L'entretien d'une indépendance entière à l'égard de cette

[14] Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, Nicolas Wilson, « The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers ». L'étude a été publiée en avril 2025.

[15] Gerlich, Michael, « AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking », publié dans *Societies* 2025, 15, 6.

[16] Voir le rapport final du Shift Project publié en octobre 2025 sous le titre : « *Intelligence artificielle, données, calculs :* quelles infrastructures dans un monde décarboné ? ».

technologie, qui passe par un non-usage critique de l'IAg, vaut autant que la lutte contre les dépendances et leurs effets.

Tenter d'en user aujourd'hui de façon indocile, individuellement et collectivement, c'est déjà se tenir dans le cadre étroit d'une « négociation collective » pour obtenir, selon les mots de Juan Sebastián Carbonell, « une IA « vertueuse », « féministe », ou tout simplement « moins mauvaise » [parce que] « capacitante ». C'est une possibilité, nous venons de voir qu'on peut légitimement douter de pouvoir jamais s'en saisir. Mais cette possibilité n'exclut pas une autre manière encore de résister à la technologie actuelle et à sa dimension de « technologie persuasive »[17]. Elle consiste à sortir de ce cadre, en adoptant une « stratégie de refus ». Non pas un rejet pur et simple, entendons-nous bien, mais un refus motivé par la raison que Carbonell condense dans ces mots qui devaient nous donner éminemment à réfléchir : « […] une technologie pensée comme un outil de domination ne peut pas être réemployée à des fins émancipatrices ».

Ce refus – à chacun·e de s'y décider – s'ordonne à l'idée selon laquelle « l'histoire du changement technologique » n'est pas identifiable à « l'histoire de l'innovation et de ses oppositions [*in abstracto*] », mais se présente plutôt comme l'histoire de *trajectoires différentes* où la possibilité d'une pluralité de changements technologiques se discute socialement et politiquement – même âprement, s'il le faut ! L'appel à une utilisation indocile de l'IAg ouvre-t-il un espace de négociation suffisant ? N'y a-t-il aucun autre chemin pour l'obtenir et l'élargir ? Il faut que la question soit posée.

En vous parlant aujourd'hui comme je viens de le faire, je suis conscient que je m'adresse, au fond, à la faculté qui est en vous de dire *non* à ce qui est socialement et techniquement attendu de vous. Pour avoir une chance de réussir, cette démarche suppose :

– que les effets humainement dévastateurs de la technologie soient discernés ;

– que les dispositifs matériels nécessaires à l'utilisation de l'IA (extraction de matériaux, infrastructures, institutions, pratiques sociales, etc.) soient connus et délimités de manière critique en regard de l'endommagement irréversible de l'environnement qu'ils risquent d'aggraver ;

– que la manière dont la technologie transforme et structure nos sociétés humaines soit révélée, remarquée et étudiée, car des formes de « dépendance au sentier », d'« auto-renforcement » et de « verrouillage socio-technique » sont possibles – suivant le même schéma qui nous a rendus dépendant de l'automobile [18] ;

– que l'usager soit *déjà* suffisamment solide cognitivement et psychologiquement pour contenir de lui-même les effets susmentionnés.

– Ce dernier point suppose des apprentissages qui doivent déjà être consistants et qui doivent être protégés. Or, ces apprentissages se font dans la durée, ils prennent du temps, et ils doivent être faits aussi longtemps et souvent que possible sans assistance.

– Il faut (re-)dire aussi que le métier d'élève est exigeant, et l'exigence durable peut faire naître la demande de se départir d'une partie des tâches assignées. L'IAg est un moyen dangereux de rendre ce désir en apparence réalisable.

– Le péril de l'IAg, c'est qu'elle peut désapprendre à apprendre, et cela plus radicalement que les moteurs de recherche, que Wikipédia, etc. Et elle donne de l'apprentissage une image partielle, incomplète et trompeuse.

[17] Alombert, Anne, *Schizophrénie numérique*, Paris : Éd. Allia, 2024, pp. 15-22, pp. 43-51, et pp. 66-81.
[18] Voir, à ce sujet, l'intégralité de l'article d'Olivier Lefebvre déjà cité ci-avant et accessible sur le site internet de la revue *Terrestres*.

– Les IAg peuvent-elles renforcer la curiosité, la motivation d'étudier, la déstabilisation des savoirs, l'originalité, la créativité, la robustesse des connaissances, l'envie de discuter, l'ouverture d'esprit… ? Ce n'est pas tout à fait exclu. Mais cela reste à inventer.

– Dans l'immédiat, ce qu'elles apportent, c'est un surcroît d'efficacité, des gains de productivité, des accélérations, une automatisation du travail manuel et intellectuel des êtres humains, une augmentation du « capital fixe » ou du « travail mort » [19], c'est-à-dire, en clair, des investissements financiers gigantesques dans la machinerie de l'intelligence artificielle, soumis à des impératifs de rendement sur le marché et à la stricte logique de la maximisation incessante des profits.

L'exhortation à inventer des usages indociles de l'IAg n'est pas tout à fait inopportune, sauf qu'elle est grevée d'implicites, échappant par là-même à l'examen critique, et que dans les études gymnasiales, l'indocilité la plus loyale (réfractaire à jouer au plus malin) réside dans le non-usage des IAg à chaque fois qu'il est autorisé et possible. Nager contre le courant reste une sorte de prouesse, on s'y épuise.

Je vous remercie beaucoup pour votre attention ainsi que pour vos questions et vos critiques s'il vous en vient.

Nyon, le 9-25 octobre 2025

Felicetti Ricci

NOTES DE FIN

(Extraits de l'étude du MIT dans la version *preprint* et du propos de Ioan Roxin)

[i] « Pour le groupe qui rédige avec l'assistance d'une recherche web classique, l'activité du cerveau se focalise sur la recherche d'informations visuelles et leur compilation dans la phase de recherche. Les participants trient et sélectionnent un certain nombre d'informations en vue de rédiger leurs essais. Il s'agit d'une « intégration cognitivement exigeante des ressources visuelles, attentionnelles et exécutives » mobilisée par les écrans. Si les utilisateurs de LLM passent aussi par un écran, leur groupe ne présente pas de « niveaux comparables d'activation corticale visuelle ». L'objectif de leur interaction avec l'écran semble distinct : l'utilisation de ChatGPT réduit leur besoin de recherche visuelle prolongée et de filtrage sémantique. La « charge cognitive » se trouve déplacée dans l'intégration procédurale et la coordination motrice. Enfin, le groupe n'ayant utilisé que son cerveau montre quant à lui des activations fortes en dehors du cortex visuel, en particulier « dans les régions du cerveau impliquées dans l'intégration sémantique, l'idéation créative et l'autocontrôle exécutif ». Ces résultats suggèrent que l'utilisation des LLM, si elle augmente la performance des tâches, a surtout un effet sur l'architecture cognitive. « Le groupe « cerveau seul » a exploité de vastes réseaux neuronaux distribués pour générer du contenu en interne ; le groupe

[19] Alombert, Anne, *De la bêtise artificielle*, Paris : Éd. Allia, 2025, pp. 57-74, en particulier p. 62.

« moteur de recherche » s'est appuyé sur des stratégies hybrides de gestion de l'information visuelle et de contrôle régulateur ; et le groupe « LLM » a optimisé l'intégration procédurale des suggestions générées par l'IA. »

[ii] « Une autre grande divergence comportementale qui ressort de l'étude est la perception qu'ont les participants de la paternité de leur essai.

Alors que le groupe travaillant sans assistance revendique presque à l'unanimité la pleine propriété de ses textes (16/18 lors de la première session et 17/18 lors de la troisième), le groupe rédigeant avec ChatGPT présente un sentiment ambivalent à ce sujet. Ainsi, certains participants revendiquent être les auteurs de leurs textes, d'autres le nient explicitement et beaucoup s'attribuent une partie seulement du mérite (entre 50 et 90 %).

Au total, l'étude conclut que « ces réponses suggèrent une diminution du sentiment d'agentivité cognitive ».

Selon les auteurs, cela met en évidence un problème : les outils d'IA, bien que précieux pour soutenir les performances, peuvent involontairement entraver le traitement cognitif approfondi, la rétention et l'engagement authentique avec le matériel écrit.

Cette observation expérimentale semblerait fournir la preuve que si les utilisateurs s'appuient trop fortement sur les outils d'IA, ils peuvent penser acquérir une maîtrise superficielle sans parvenir à intérioriser et à s'approprier les connaissances. »

[iii] « Les données dites « comportementales » – en particulier celles relatives à la capacité de citation, à l'exactitude des citations et à l'appropriation des essais – prolongent et corroborent les conclusions de l'étude en matière de connectivité neuronale. Ces résultats suggèrent que la dynamique fonctionnelle du réseau activée pendant la rédaction d'un essai façonne des processus comme la mémoire, l'efficacité de l'autocontrôle et le degré d'appropriation perçue du travail écrit. La divergence comportementale la plus constante et la plus significative entre les groupes a été observée dans la capacité à citer de tête son propre essai.

Les utilisateurs de LLM obtiennent dans ce domaine des résultats nettement inférieurs : 83 % des participants déclarent avoir des difficultés à citer leur essai après la première session – et aucun ne fournit une seule citation correcte de son propre travail. Cette déficience persiste, bien qu'atténuée, lors des sessions suivantes : 6 participants sur 18 ne parviennent toujours pas à se citer correctement lors de la session 3.

Cette difficulté correspond directement à la réduction de la connectivité cérébrale dans le groupe utilisant un LLM : ces oscillations sont en effet généralement plus fortes lorsque les individus génèrent et structurent intérieurement du contenu plutôt que d'intégrer passivement des informations générées à l'extérieur.

L'étude met notamment en avant le fait que la réduction de l'activité cognitive chez les utilisateurs de LLM « reflète probablement un contournement des processus d'encodage profond de la mémoire, les participants lisant, sélectionnant et transcrivant les suggestions générées par l'outil sans les intégrer dans les réseaux de mémoire épisodique ».

Les volontaires utilisant un moteur de recherche comme Google ou ne s'appuyant sur aucun outil numérique n'ont pas présenté de telles déficiences.

À la session 2, les deux groupes ont atteint une capacité de citation presque parfaite, et à la session 3, 100 % des participants des deux groupes ont déclaré être capables de citer leurs essais, observant seulement des écarts mineurs dans la précision des citations.

À l'inverse, l'absence totale de citations correctes dans le groupe LLM lors de la session 1 et les déficiences persistantes lors des sessions suivantes suggèrent que non seulement l'encodage de

la mémoire était superficiel, mais surtout que le contenu sémantique lui-même n'avait peut-être pas été entièrement intériorisé. »

[iv] Les « biais » sont de deux sortes, selon Ioan Roxin. Voici la présentation qu'il en propose dans *Polytechnique Insights* : « Les premiers peuvent être volontairement induits par le créateur de l'IA. Les LLM sont entraînés sur toutes sortes de contenus disponibles en ligne, non filtrés (4 000 milliards de mots estimé pour ChatGPT4, à comparer aux 5 milliards de mots que contient la version anglaise de Wikipédia !). Le pré-entraînement crée un « monstre » qui peut générer toutes sortes d'horreurs. Une deuxième étape (appelée réglage fin supervisé) est donc nécessaire : elle confronte l'IA pré-entraînée à des données validées, qui servent de référence. Cette opération permet par exemple de lui « commander » d'éviter toute discrimination, mais peut aussi être utilisée pour orienter ses réponses à des fins idéologiques. Quelques semaines après son lancement, DeepSeek a défrayé la chronique pour ses réponses pour le moins évasives aux questions d'utilisateurs concernant Tian'anmen ou l'indépendance de Taïwan. Il faut toujours garder en tête que les générateurs de contenus de ce type peuvent ne pas être neutres. Leur faire une confiance aveugle peut entraîner la propagation de théories idéologiquement marquées.

Le second type de « biais » apparaissent spontanément, souvent sans explication claire. Les modèles de langage (LLM) présentent des propriétés « émergentes », non prévues par leurs concepteurs. Certaines sont remarquables : ces générateurs de texte écrivent sans faute et sont devenus d'excellents traducteurs sans qu'aucune règle de grammaire n'ait été codée. Mais d'autres sont préoccupantes. Le benchmark MASK (*Model Alignment between Statements and Knowledge*), publié en mars 2025, montre que, parmi les trente modèles testés, aucun n'atteint plus de 46 % d'honnêteté, et que la propension à mentir augmente avec la taille du modèle, même si leur exactitude factuelle s'améliore.

MASK prouve que les LLM « savent mentir » lorsque des objectifs contradictoires (par exemple, séduire un journaliste, répondre à des pressions commerciales ou hiérarchiques) prédominent. Lors de certains tests, des IA ont menti délibérément, menacé des utilisateurs, contourné leurs règles éthiques de supervision et même se sont reproduites de manière autonome pour assurer leur survie.

Ces comportements, dont les mécanismes décisionnels restent opaques, échappent à tout contrôle précis. Ces capacités émergent du processus d'entraînement lui-même : il s'agit d'une forme d'auto-organisation algorithmique, et non d'une défaillance de conception. Les IA génératives se développent plutôt qu'elles ne sont conçues, leur logique interne se formant de manière auto-organisée, sans plan directeur. Ces dérives sont suffisamment préoccupantes pour que des figures de proue, telles que Dario Amodei (PDG d'Anthropic), Yoshua Bengio (fondateur de Mila), Sam Altman (créateur de ChatGPT) et Geoffrey Hinton (lauréat du prix Nobel de physique en 2024), appellent à une régulation stricte, en faveur d'une IA plus transparente, éthique et alignée sur des valeurs humaines, y compris par une décélération dans le développement de ces technologies.

La fluidité de leur conversation et ces propriétés émergentes peuvent donner l'illusion d'une intelligence à l'œuvre. Mais aucune IA n'est intelligente au sens humain du terme. Elles n'ont ni conscience ni volonté, et ne comprennent pas réellement le contenu qu'elles manipulent. Leur fonctionnement est purement statistique et probabiliste, et ces dérives émergent seulement parce qu'elles cherchent à suivre les commandes de départ. C'est ici l'opacité de leur fonctionnement qui inquiète les chercheurs. »