

Déchargement cognitif et IA générative

Mémo interne

Objet — Faire le point sur les travaux qui documentent les effets de l'IA générative sur l'apprentissage, et en tirer des orientations utilisables pour la conception des enseignements et des évaluations.

Jean-François Van de Poël — Conseiller IA enseignement

Date — 18 août 2026 · **Version** — 1.1 · **Licence** — CC BY-NC-SA 4.0

En bref

En juin 2025, une étude du MIT Media Lab a mis un mot sur une inquiétude que partageaient intuitivement les enseignant·es : la « dette cognitive » que contracterait celui ou celle qui délègue à une IA le travail de penser. Le retentissement a été considérable, la prudence des auteur·rices beaucoup moins relayée.

Ce mémo a pour but de replacer ce travail dans son cadre théorique — la délégation cognitive, étudiée bien avant l'IA — et de le confronter à cinq études récentes. Il en ressort une conclusion moins spectaculaire mais plus utile que celle qui a circulé dans la presse : ce n'est pas le recours à l'IA qui dégrade l'apprentissage, c'est le fait de déléguer précisément l'opération que la tâche visait à faire travailler. La variable sur laquelle nous pouvons agir n'est donc ni l'interdiction ni la règle générale, mais la conception des tâches.

1. Le point de départ

En juin 2025, huit chercheur·ses conduit·es par Nataliya Kosmyna déposent sur arXiv une étude intitulée *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*. Le protocole est solide et son originalité tient au croisement des mesures.

Cinquante-quatre participant·es, recruté·es dans plusieurs institutions de la région de Boston, rédigent des dissertations au fil de trois séances. Un premier groupe travaille avec ChatGPT, un deuxième avec un moteur de recherche, un troisième sans aucun outil. Une quatrième séance renverse les conditions : les utilisateur·rices de ChatGPT rédigent sans outil, celles et ceux qui n'en avaient aucun passent au LLM. Dix-huit personnes participent à cette dernière étape, l'ensemble s'étalant sur quatre mois. Les chercheur·ses enregistrent l'activité cérébrale par électroencéphalographie, analysent les textes par traitement automatique du langage, les font noter par des enseignant·es et par un agent évaluateur, et mènent un entretien après chaque séance.

Trois résultats ressortent.

- D'abord, la connectivité cérébrale décroît à mesure que le support externe augmente : réseaux les plus étendus chez le groupe sans outil, engagement intermédiaire avec le moteur de recherche, couplage le plus faible avec le LLM.
- Ensuite, le sentiment d'être l'auteur·rice du texte suit la même pente, au plus bas dans le groupe LLM.

- Enfin, les participant-es de ce groupe peinent à citer un passage du texte qu'ils venaient d'écrire quelques minutes plus tôt.

Trois précautions s'imposent avant toute reprise de ces résultats.

- Il s'agit d'un préprint, non évalué par les pairs, et les auteur-rices demandent explicitement que leurs conclusions soient traitées comme préliminaires.
- L'échantillon est réduit et concentré sur quelques établissements voisins, l'équilibre de genre imparfait ; le sous-groupe de dix-huit personnes de la quatrième séance, qui a nourri l'essentiel des commentaires, est particulièrement fragile.
- L'étude ne porte enfin que sur ChatGPT et sur une seule tâche, la dissertation, sans distinguer les étapes que sont la recherche d'idées, la rédaction et la révision.

L'équipe a par ailleurs demandé publiquement que l'on cesse d'employer, en rendant compte de son travail, des termes comme *brain rot*, dommages, atrophie ou passivité, absents de l'article. La demande mérite d'être respectée, y compris dans nos propres communications.

Un commentaire critique a depuis été publié (Stanković et al., 2026). Ses auteur-rices saluent la constitution du jeu de données et la qualité des chaînes d'analyse, tout en invitant à une lecture plus prudente de plusieurs résultats, notamment sur le traitement des signaux EEG et la reproductibilité. Il vaut la peine de citer les deux textes ensemble.

2. Ce qu'est la délégation cognitive

L'étude du MIT ne surgit pas de nulle part. Elle prolonge un courant de recherche bien établi, celui du déchargement cognitif ou de la délégation cognitive, dont la référence principale est un article de synthèse de Risko et Gilbert paru en 2016 dans *Trends in Cognitive Sciences*.

La délégation cognitive y désigne le recours à une action physique ou à une ressource extérieure pour modifier les exigences de traitement d'une tâche et réduire ainsi la demande cognitive qu'elle impose. Le phénomène est banal : incliner la tête pour lire un texte pivoté, programmer un rappel, prendre des notes, poser une alarme. Risko et Gilbert distinguent ce que l'on délègue à son propre corps, le doigt qui suit une ligne, le geste qui structure la pensée, de ce que l'on délègue au monde, des listes aux moteurs de recherche.

Les systèmes génératifs relèvent de la seconde catégorie mais y occupent une place particulière. Les outils antérieurs prenaient en charge le stockage ou la récupération de l'information ; ceux-ci prennent en charge la formulation, la structuration d'un argument, la production du sens. C'est ce déplacement qui justifie de rouvrir le dossier plutôt que d'appliquer sans examen les conclusions établies sur le GPS ou sur Google.

Le point le plus utile pour nous concerne la métacognition. La décision de déléguer repose sur une évaluation, souvent implicite, de ses propres capacités comparées à celles du dispositif. Cette évaluation est faillible, et l'usage de la stratégie retenue reconfigure ensuite le jugement que la personne porte sur ses capacités. Deux conséquences en découlent. La délégation est fréquemment sous-optimale : on délègue ce qu'on aurait eu intérêt à traiter soi-même, et l'inverse. Et la boucle se renforce d'elle-même, puisque plus on délègue, plus on tend à sous-estimer ce dont on serait capable. C'est précisément ce mécanisme que la métaphore de la dette cherche à nommer.

Il faut également rappeler que la littérature documente des bénéfices, ce qui rend intenable une lecture purement alarmiste. Sparrow, Liu et Wegner ont montré dès 2011 que l'anticipation d'un accès futur à l'information dégrade son encodage tout en améliorant la mémoire du chemin pour la retrouver —

l'effet Google. Dahmani et Bohbot ont observé en 2020, chez cinquante conducteur·rices, qu'un usage cumulé plus important du GPS allait de pair avec de moins bonnes performances de mémoire spatiale en navigation autonome. Mais Storm et Stone avaient établi en 2015 qu'enregistrer une première liste améliore la mémorisation de la suivante, en libérant des ressources et en réduisant l'interférence. Déléguer réalloue des ressources ; cela ne les détruit pas.

En français, plusieurs traductions circulent — déchargement cognitif, délestage cognitif, externalisation cognitive. Nous retenons ici « délégation cognitive ».

3. Une objection à prendre au sérieux

Tout ce qui précède pourrait laisser croire que l'IA a fait naître un problème. Tim Fawns, de l'université Monash, rappelle qu'il n'en est rien, et le fait avec une brutalité utile.

Écrire relève de la délégation cognitive, dit-il en substance, comme dessiner, utiliser un livre, un téléphone ou un ordinateur, parler à quelqu'un, employer le langage, vivre dans une maison. La délégation va de bout en bout ; être humain consiste en grande partie à déléguer sa cognition. La délégation cognitive *est* la cognition humaine. Il ajoute qu'il reste évidemment légitime de s'en inquiéter, mais à une condition : savoir distinguer les formes qui posent un problème de celles qui n'en posent pas, et dire quand, pourquoi et comment.

L'objection déplace la question. Elle invalide l'idée que déléguer constituerait en soi une dégradation, et elle interdit d'inférer d'une baisse d'activité cérébrale une perte d'apprentissage ; une charge plus faible peut aussi bien signaler une automatisation réussie qu'un désengagement. Elle disqualifie surtout l'usage du concept comme argument massue contre l'IA ; Fawns va jusqu'à le qualifier, ainsi mobilisé, de paresseux.

Ce qu'elle impose en retour est plus exigeant. La question n'est plus de savoir combien un·e étudiant·e délègue, mais quelle opération précise il ou elle délègue, à quel moment de son parcours, si la compétence concernée est déjà constituée ou en cours de construction, et si cette délégation sert l'objectif d'apprentissage ou s'y substitue. Ces questions ne se posent qu'au niveau d'une tâche donnée, avec un objectif donné. Aucune politique générale ne peut y répondre.

Le rappel de Fawns ne disqualifie donc pas l'étude du MIT ; il en précise la portée. Ce que Kosmyna et ses collègues documentent n'est pas la nocivité de la délégation en général, mais quelque chose de beaucoup plus circonscrit : dans une tâche où l'opération déléguée était exactement celle que la tâche devait faire travailler, la délégation a produit un désengagement mesurable, une appropriation dégradée et une mémoire lacunaire du texte. C'est un résultat sur l'alignement entre ce que l'on délègue et ce que l'on veut apprendre.

4. Ce que dit la recherche récente

Cinq travaux publiés entre 2024 et 2026 permettent de préciser le tableau. Ils sont présentés ici du plus général au plus concluant.

Gerlich (2025) a interrogé 666 personnes au Royaume-Uni, par questionnaire et entretien. Il observe une corrélation négative entre la fréquence d'usage des outils d'IA et les capacités de pensée critique, et montre que cette relation passe par la délégation cognitive : celle-ci est le mécanisme, non un simple parallélisme. Les participant·es les plus jeunes présentent la dépendance la plus forte et les scores les plus faibles. L'auteur parle de « paresse cognitive » pour désigner le recul de la propension à s'engager dans une réflexion approfondie. C'est l'étude qui a installé la délégation comme variable médiatrice dans le débat public. Elle reste corrélationnelle, repose en partie sur des mesures déclaratives, et une

correction a été publiée en septembre 2025. Gerlich a d'ailleurs testé ensuite l'hypothèse inverse auprès de 150 participant-es, avec ou sans accompagnement structuré de l'usage de ChatGPT, son argument étant qu'un usage délibéré peut renforcer la pensée critique.

Lee et ses collègues (2025), à CHI, ont interrogé 319 travailleur-ses du savoir sur 936 situations réelles d'usage professionnel. Leur résultat quantitatif est net et contre-intuitif : la confiance placée dans l'IA prédit un moindre exercice de la pensée critique, tandis que la confiance dans ses propres compétences en prédit un exercice accru. Mais l'apport principal est ailleurs. L'IA ne supprime pas la pensée critique, elle en déplace le lieu : de la collecte d'information vers sa vérification, de la résolution de problème vers l'intégration des réponses produites, de l'exécution de la tâche vers son pilotage. L'effet dépend de l'enjeu — sur les tâches exigeant de l'exactitude, les participant-es déclarent fournir davantage d'effort critique avec l'IA que sans ; sur les tâches routinières menées sous pression, ils en fournissent moins. Un constat mérite ici une attention particulière : les personnes les moins assurées de leurs compétences se contentent généralement d'accepter ce que produit la machine. C'est exactement la position de nos étudiant-es.

Zhai, Wibowo et Li (2024) ont conduit la revue systématique de référence, selon la méthode PRISMA, sur quatorze articles issus de quatre bases. Ils définissent la surdépendance comme le fait d'accepter les recommandations d'un système sans les questionner, situation qui survient lorsque la personne peine à évaluer sa fiabilité. Leur conclusion : cette surdépendance affecte les capacités cognitives, les individus privilégiant les solutions rapides aux solutions lentes. L'originalité de l'article tient à la chaîne qu'il établit — les défauts du système, hallucinations, biais, opacité, alimentent la surdépendance, laquelle dégrade ensuite la cognition. La base documentaire reste étroite et antérieure aux travaux marquants de 2025 et 2026, mais l'article épargne de reconstruire l'état de l'art.

Bastani et ses collègues (2025), dans *PNAS*, apportent la démonstration la plus parlante pour un public enseignant. Dans un lycée turc, sur quatre séances de 90 minutes couvrant environ 15 % du programme de mathématiques du semestre, près de mille élèves ont été réparti-es en trois conditions : un tuteur reproduisant l'interface standard de ChatGPT, un tuteur dont les consignes visaient à préserver l'apprentissage, et aucun accès. L'accès à GPT-4 améliore les performances pendant les exercices. Mais lorsqu'on le retire, les élèves de la première condition réussissent moins bien que celles et ceux qui n'y avaient jamais eu accès : dix-sept pour cent de baisse. Les garde-fous du second tuteur neutralisent largement cet effet. Les auteur·rices parlent d'une béquille : quand elle fonctionne pendant la pratique, l'élève réussit moins bien seul ensuite. Deux systèmes bâtis sur le même modèle produisent donc des effets opposés selon qu'ils livrent la réponse ou étayent la démarche. Le contexte reste unique et des facteurs confondants ont été signalés — niveau initial, motivation, effet enseignant, milieu social. On notera que le titre publié est plus prudent que celui de la version de travail, qui affirmait simplement que l'IA générative peut nuire à l'apprentissage.

Liu et ses collègues (2026) fournissent la pièce qui manquait : une preuve causale. Trois expériences randomisées, 1 222 participant-es, en raisonnement mathématique puis en compréhension de lecture. Dans la première, l'assistant disponible en barre latérale est retiré sans préavis après douze problèmes ; le taux de résolution tombe alors à 0,57 contre 0,73 dans le groupe témoin, avec un taux d'abandon supérieur. La deuxième réplique le résultat en écartant les biais de sélection et d'interface, la troisième l'étend à un autre domaine. Ce qui est atteint n'est pas seulement la performance mais la persévérance, et les effets apparaissent après une dizaine de minutes d'interaction seulement. Les auteur·rices avancent que l'IA habitue à attendre des réponses immédiates et prive de l'expérience de la difficulté surmontée par soi-même — ce qui compte, puisque la persévérance est l'un des meilleurs prédicteurs de l'apprentissage à long terme. Il s'agit d'un préprint, portant sur des tâches courtes et fermées, sans mesure au-delà de la séance.

Pris ensemble, ces travaux autorisent une affirmation prudente mais ferme. L'usage non encadré d'un assistant génératif dégrade la performance ultérieure en autonomie, par un mécanisme de délégation. Mais l'ampleur de l'effet dépend entièrement de la manière dont l'interaction est conçue et du maintien d'un rôle actif de vérification et de pilotage.

5. Ce que vivent les enseignant-es et les étudiant-es

Reste à savoir comment cette question se pose concrètement dans nos établissements. Une étude australienne y répond avec précision.

Corbin, Dawson, Nicola-Richmond et Partridge ont interrogé, fin 2023 et début 2024, dix-neuf étudiant-es et douze enseignant-es d'une grande université, dans des disciplines variées. Sans que personne ne la suggère, la métaphore de la ligne à tracer entre usage acceptable et inacceptable a surgi spontanément dans plus de la moitié des entretiens. Le titre de l'article reprend l'exaspération d'un-e enseignant-e jugeant cette ligne absurde.

Faute de cadre, chaque étudiant-e se construit son propre système éthique, souvent élaboré et strictement personnel. Les analogies mobilisées sont révélatrices : Grammarly, le correcteur orthographique, c'est-à-dire des technologies dont la frontière est déjà socialement stabilisée. S'y ajoute une charge permanente, celle de deviner la position de chaque enseignant-e pour calibrer son usage.

Du côté enseignant, le coût est d'un autre ordre. Les participant-es évoquent une anxiété persistante lors des corrections, des traumatismes éthiques, l'épuisement, l'absence d'espace collectif pour en parler et l'absence de reconnaissance de ce travail supplémentaire. Plusieurs déclarent n'avoir rien modifié à leurs évaluations, faute de savoir quoi faire.

La proposition des auteur-rices constitue le renversement le plus utile de tout ce dossier. La structure même des évaluations externalise sur les étudiant-es une décision aux conséquences pédagogiques lourdes. Plutôt que de clarifier la ligne puis de la faire respecter, mieux vaut supprimer la nécessité même de la tracer, en intégrant les attentes dans la conception de la tâche.

Leur modèle, le *Dynamic Educational Boundaries Model*, tient en trois principes. Les frontières doivent être inhérentes à la tâche plutôt qu'énoncées à côté d'elle : autoriser l'IA signifie l'intégrer délibérément là où elle sert l'objectif, l'interdire signifie concevoir une tâche que le recours à l'IA priverait de sens, comme une soutenance orale. Ce qui est acceptable varie ensuite selon les disciplines : une analyse de données assistée correspond aux pratiques professionnelles en ingénierie mais contredit l'objectif d'un cours de statistique où le calcul manuel est le cœur de l'apprentissage. Enfin, la charge pesant sur le corps enseignant relève de la responsabilité institutionnelle — formation, dépôts partagés d'exemples, mentorat entre pairs, espaces de discussion. Les auteur-rices présentent ce cadre comme initial et évolutif, en insistant sur le fait qu'aucune solution universelle ne conviendra.

6. Conclusion

Trois constats se rejoignent, et c'est leur convergence qui fait la solidité de l'argument.

Fawns montre que la question « où placer la ligne ? » est mal posée : la délégation étant constitutive de la cognition humaine, aucune ligne générale ne peut être tracée. Corbin et ses collègues montrent le coût réel de cette question mal posée, renvoyée à des étudiant-es et à des enseignant-es qui la portent seul-es. Kosmyna, Bastani et Liu montrent que certaines délégations ont bel et bien un coût d'apprentissage, ce qui interdit de conclure que la question ne se pose pas.

La réponse ne se situe donc ni dans la règle générale, ni dans l'interdiction, ni dans l'autorisation. Elle se situe dans la conception de la tâche, seul niveau où l'on peut juger si l'opération déléguée est celle que l'on voulait faire travailler. C'est le résultat le plus robuste dont nous disposons : chez Bastani, deux systèmes fondés sur le même modèle produisent des effets opposés selon qu'ils livrent la réponse ou étayent la démarche.

Quatre orientations en découlent pour notre travail.

La première concerne la conception des évaluations. Plutôt que de multiplier les chartes et les règlements, il s'agit d'aider les équipes à formuler, pour chaque tâche, quelle opération intellectuelle elle doit faire travailler — et à en déduire ce qui peut être délégué. C'est un travail d'accompagnement, pas de normalisation.

La deuxième concerne les compétences à développer chez les étudiant·es. Si l'IA déplace l'effort vers la vérification, l'intégration et le pilotage, ce sont ces trois gestes qu'il faut enseigner et évaluer explicitement. Le résultat de Lee sur les personnes peu assurées de leurs compétences, qui se contentent d'accepter ce que produit la machine, indique où porter l'effort en priorité.

La troisième concerne le jugement métacognitif. Savoir quand ne pas déléguer relève de la même compétence que savoir déléguer. Cela s'apprend, se travaille et s'évalue, et cela vaut mieux que de tenter d'interdire.

La quatrième concerne les enseignant·es. Le coût émotionnel documenté par Corbin et ses collègues n'est pas un effet secondaire : c'est un enjeu institutionnel qui appelle des espaces de discussion, des ressources partagées et une reconnaissance de la charge de travail que représente la refonte des évaluations.

Un mot enfin sur la solidité des preuves, car nous aurons à la défendre. Le corpus disponible reste jeune. Deux des travaux les plus cités sont des préprints, deux autres sont corrélationnels ou déclaratifs, et l'ensemble porte sur des tâches courtes, dans des contextes restreints. Nous ne pouvons pas encore dire ce que produit un usage prolongé sur des années, ni comment ces effets se comportent dans les tâches complexes et ouvertes qui font l'essentiel de l'enseignement supérieur. Cette prudence n'affaiblit pas les orientations proposées : elles reposent moins sur l'ampleur des effets mesurés que sur un constat qui, lui, est stable — ce qui compte n'est pas l'outil mais l'usage que la tâche rend possible.

Déclaration d'usage de l'IA

La recherche bibliographique a été menée avec l'appui de **Google Scholar** et de **Consensus** pour l'identification et le repérage des travaux cités. La rédaction a bénéficié de l'assistance de **Claude** (Anthropic) pour la mise en forme et la formulation du texte.

L'ensemble des choix éditoriaux relève de l'auteur : sélection et hiérarchisation des sources, structure argumentative, interprétation des résultats, formulations retenues et conclusions. Les références citées ont été vérifiées, et l'auteur en assume l'entière responsabilité.

Références

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122.
- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.
- Corbin, T., Dawson, P., Nicola-Richmond, K., & Partridge, H. (2025). 'Where's the line? It's an absurd line': towards a framework for acceptable uses of AI in assessment. *Assessment & Evaluation in Higher Education*, 50(5), 705–717. <https://doi.org/10.1080/02602938.2025.2456207>
- Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, 10, 6310. <https://doi.org/10.1038/s41598-020-62877-0>
- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity Matters More than Cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016.
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006> (Correction : *Societies*, 15(9), 252.)
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*. arXiv:2506.08872. <https://arxiv.org/abs/2506.08872>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of Generative AI on Critical Thinking. *CHI '25*, ACM, art. 1121. <https://doi.org/10.1145/3706598.3713778>
- Liu, G., Christian, B., Dumbalska, T., Bakker, M. A., & Dubey, R. (2026). *AI Assistance Reduces Persistence and Hurts Independent Performance*. arXiv:2604.04721.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS). *Journal of University Teaching and Learning Practice*, 21(6).
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google Effects on Memory. *Science*, 333(6043), 776–778.
- Stanković et al. (2026). *Comment on: Your Brain on ChatGPT*. arXiv:2601.00856.
- Storm, B. C., & Stone, S. M. (2015). Saving-Enhanced Memory. *Psychological Science*, 26(2), 182–188. <https://doi.org/10.1177/0956797614559285>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11:28. <https://doi.org/10.1186/s40561-024-00316-7>