

Intelligence artificielle générative et éducation scolaire : quelques réflexions

Generative artificial intelligence and school education: some thoughts

Éric BRUILLARD

Université Paris-Cité, Laboratoire EDA

Résumé. Ce texte reprend et analyse différentes sources récentes (2024-2025) sur l'intelligence artificielle générative (IAG) en éducation scolaire. Il s'appuie notamment sur le journal américain *Edweek* qui rend compte des rapports récents sur l'intelligence artificielle et l'enseignement aux États-Unis et fournit enquêtes, témoignage, analyses, etc. Il analyse la question des hallucinations des IAG, leur contrôle et leur vérification, les enquêtes sur les utilisations par les élèves et les enseignants et les inégalités liés au rapport au savoir et à l'école. Il analyse les risques liés à l'adoption des IAG comme des compagnons, voire des conseillers. Enfin, un focus est mis sur les questions de pensée critique, envisagée comme une technique, une réponse possible aux erreurs des IAG que l'on ne peut éviter.

Mots-clés : IA générative, Wikipédia, esprit critique, inégalités scolaires, tricherie, compagnon IA

Abstract. *This text reviews and analyses various recent sources (2024-2025) on generative artificial intelligence (GenAI) in school education. In particular, it draws on the American newspaper Edweek, which shows recent reports on artificial intelligence and education in the United States and provides surveys, testimonials, analyses, etc. It analyses the issue of GenAI hallucinations, their control and verification, surveys of use by pupils and teachers and inequalities linked to the relationship with knowledge and school. It analyses the risks involved in adopting GenAIs as companions, or even advisors. Finally, the focus is on critical thinking, seen as a technique, a possible answer to the errors of GenAI that cannot be avoided.*

Keywords: *Generative AI, Wikipedia, critical thinking, educational inequalities, cheating, AI companion*

1. INTRODUCTION

Ce texte reprend et analyse différentes sources récentes (2024-2025) sur l'IA générative (IAG) en éducation : des articles et rapports de *Edweek* et des *Cahiers pédagogiques*, des analyses du service pédagognumérique de l'UQAM (*le Collimateur*) ainsi que divers articles scientifiques (2024-2025).

Edweek est un journal de référence consacré à l'éducation scolaire aux États-Unis. Il signale et commente les rapports qui sortent sur l'éducation (notamment les rapports en lien avec l'IAG publiés en 2024 et 2025), génère ses propres études en analysant des données qu'il collecte et auxquelles il fournit souvent l'accès. Il interroge les différents acteurs en lien avec l'éducation : élèves, enseignants, chefs d'établissement, administratifs, politiques, entreprises, chercheurs... Les analyses et réflexions publiées permettent de situer les tendances dans les utilisations de l'IAG, tant par les élèves que par les enseignants, en classe et surtout en dehors de la classe, ainsi que les liens avec les parents.

Suite à une analyse de présentations issues du colloque SITE 2025 (Bruillard, 2025), consacrée aux enseignants et à la formation des enseignants, correspondant à l'enseignement supérieur et aux universités, ce texte se centre sur l'enseignement secondaire et les questions posées par les utilisations de l'IAG.

Il débute par l'étude des contenus générés par les IAG, leur contrôle et leur vérification notamment en lien avec Wikipédia. Les erreurs générées par les IAG constituent un défaut majeur dont il faut tenir compte dans l'analyse de leurs utilisations. Wikipédia réussit à les éviter et les groupes humains qui la font vivre tentent de maintenir cette qualité. Ensuite, diverses enquêtes permettent d'avoir une idée assez précise des taux d'adoption des IAG par les élèves et les enseignants, ainsi que les points de vue des acteurs de l'éducation et des parents. Il est essentiel d'aller au-delà et de détecter ce qui se cache derrière les taux constatés. Cela conduit à regarder plus en détail les modalités d'utilisation de l'IAG (tricherie/transparence), mais aussi de voir les risques et inégalités générés par les IAG.

Enfin, un focus est mis sur les questions de pensée critique, qui peuvent à certains égards être paradoxales. Elles sont constamment invoquées dans les textes, comme une réponse possible aux erreurs des IAG que l'on ne peut éviter.

2. COMMENT VERIFIER LES CONTENUS GENERES PAR L'IA GENERATIVE

*« Le mensonge vole, et la vérité ne le suit qu'en boitant,
de telle sorte que lorsque les hommes en arrivent
à ouvrir les yeux, c'est un quart d'heure trop tard.
La farce est finie, et le conte a produit son effet. »
John Arbuthnot (ou Jonathan Swift)
dans « L'Art du mensonge politique »¹*

¹ https://fr.wikipedia.org/wiki/L%27Art_du_mensonge_politique

Le *Collimateur* a consacré une série de textes à l'IAG², avec notamment une veille sur les nouveautés³. Il a, en particulier, décrit les manières actuelles de traiter les contenus générés par l'IAG⁴ et les éventuelles erreurs qui peuvent être véhiculées.

Notons que ces erreurs sont liées au fonctionnement même des grands modèles de langage. Une annexe présente les réponses de l'IAG *Claude.ai* à une invite sur les cas d'hallucination et une illustration via des métaphores.⁵

2.1 LES ERREURS : LES DETECTER, LES EVITER, LES CORRIGER

Ce point est illustré par la loi dite de Brandolini⁶ ou le principe d'asymétrie des baratin, « aphorisme selon lequel la quantité d'énergie nécessaire pour réfuter des sottises [...] est supérieure d'un ordre de grandeur à celle nécessaire pour les produire ». Ainsi, s'il est facile de créer une fausse information — sur le fond et la forme — en quelques minutes, il faudra probablement plusieurs heures pour démonter chaque point et montrer la fausseté de l'ensemble. « Ce principe critique la technique de propagande qui consiste à diffuser de l'*infox* facilement et en masse, afin d'exploiter la crédulité d'un certain public en faisant appel à son système de pensée rapide, instinctif et émotionnel ».

Selon le *Collimateur*, les affirmations accompagnées d'une référence inspirent davantage confiance, car on pense encore souvent que les documents cités existent. Vérifier l'existence d'un livre qui n'existe pas est chronophage, de même que remonter aux sources des affirmations. Une telle asymétrie d'effort illustre parfaitement la loi de Brandolini : « quelques secondes suffisent à une IA pour générer une hallucination documentaire, tandis que la vérification de cette information exige un travail minutieux de la part des personnes humaines ». Comme le précise le *Collimateur*, ces erreurs, estampillées par une technologie de pointe, prolifèrent à un rythme bien plus rapide que la capacité humaine à les corriger.

Afin de systématiser les pratiques de repérage et d'édition relatives aux productions automatisées qui ne respectent pas les règles communautaires, le *Wikiproject AI Cleanup*⁷ a été mis sur pied. Il a pour objectif :

- d'identifier les textes écrits par l'IA et les relire pour s'assurer qu'ils respectent les règles de Wikipédia (toute affirmation non sourcée ou probablement inexacte doit être supprimée) ;
- d'identifier les images générées par l'IA et s'assurer qu'elles sont utilisées à bon escient ;
- d'aider et suivre les rédacteurs utilisant l'IA qui ne se rendent pas compte des lacunes de l'IA en tant qu'outil d'écriture.

Il ne s'agit pas de restreindre ou d'interdire l'utilisation de l'IA dans les articles, mais de vérifier que sa production est acceptable et constructive, et de la corriger ou de la supprimer dans le cas contraire.

² <https://collimateur.uqam.ca/tag/ia/>

³ <https://enseigner.uqam.ca/ia/veilles-nouveautes/>

⁴ <https://collimateur.uqam.ca/collimateur/comment-la-communaute-wikipedia-traite-les-contenus-generes-par-lia-generative/>

⁵ <https://claude.ai/public/artifacts/a8a4dadf-ec22-4e72-a454-5b30f2a19792>

⁶ https://fr.wikipedia.org/wiki/Loi_de_Brandolini#:~:text=La%20loi%20dite%20de%20Brandolini,celle%20n%C3%A9cessaire%20pour%20les%20produire%20%C2%BB

⁷ https://en.wikipedia.org/wiki/Wikipedia:WikiProject_AI_Cleanup

Selon Maiberg (2024), grâce à l'engagement constant de volontaires humains, chargés d'examiner les nouveaux articles et de vérifier que les informations présentées proviennent de sources fiables, Wikipédia se révèle plus efficace que d'autres grands services Internet dans la détection et le filtrage des contenus trompeurs générés par l'IAG.

Depuis début octobre 2024, le Wikipédia français s'est également dotée d'un Observatoire des IA⁸, visant des objectifs similaires. L'article correspondant fournit beaucoup de tests intéressants avec différents prompts⁹, notamment pour la détection de textes générés par les IAG.

2.2 « GEMINI, OU LES INFORTUNES DE LA VERTU »¹⁰

Sous ce titre, un article du *Collimateur* raconte l'histoire de l'IAG *Gemini*, juste après son lancement. Des images générées représentaient les fondateurs des États-Unis à la peau noire, des officiers nazis aux traits asiatiques, ou encore des papes de sexe féminin. L'explication est venue des prompts (invites) créés par des utilisateurs, proposant des images dont les protagonistes étaient factuellement erronés. « Aiguillé par la noble intention de diversifier les représentations des personnes, *Gemini* a pêché par excès de zèle inclusif, commettant ainsi des contresens historiques ».

Cela correspond à ce que Salvaggio (2023) nomme du *shadow prompting*, que l'on pourrait traduire en français par l'expression suivante : une métamorphose furtive de l'invite¹¹. Roose *et al.* (2024) expliquent le processus à l'œuvre. On entre un prompt du style « Faites-moi une image des pères fondateurs américains » et l'application, sans vous prévenir, va ajouter des éléments tels que « Veuillez montrer une gamme variée de visages dans cette réponse » et c'est ce message transformé qui va être transmis au modèle.

Selon Roose et ses collègues, il s'agit d'une fonctionnalité dont les théoriciens de la conspiration peuvent s'emparer et affirmer que l'IAG modifie secrètement ce que vous lui demandez de faire pour le rendre plus « woke ». Roose *et al.* suggèrent de donner la possibilité aux utilisateurs de personnaliser les invites en fonction de leur contexte d'utilisation ou de leur vision du monde. Ils peuvent déjà le faire (s'ils maîtrisent suffisamment les techniques de prompt), mais ce n'est pas sans défaut et cela peut renforcer leurs « biais de confirmation ». On est face à un curieux dilemme, qui illustre bien la nécessité de revoir dans ce cadre la notion d'esprit critique.

« En somme, les technologies ne sont jamais neutres et sont porteuses de visions du monde qui charrient avec elles des valeurs — celle des concepteurs ou celles d'utilisateurs ?¹² Le débat ne fait que commencer » Roose *et al.* (2024).

Cette même question est commentée par Prabhakar Raghavan, vice-président de Google sur son blog¹³, qui reprend l'histoire qui vient d'être brièvement décrite, conduisant à mettre en pause la génération d'images de personnes par *Gemini*. Ce dernier, conçu comme un outil de créativité et de productivité, fait des erreurs, notamment quand il s'agit de générer des

⁸ https://fr.wikipedia.org/wiki/Projet:Observatoire_des_IA

⁹ : https://fr.wikipedia.org/wiki/Projet:Observatoire_des_IA/Documentation

¹⁰ <https://collimateur.uqam.ca/collimateur/gemini-ou-les-infortunes-de-la-vertu/>

¹¹ <https://collimateur.uqam.ca/collimateur/gemini-ou-les-infortunes-de-la-vertu/>

¹² En ce sens, ce sont bien des instruments et pas uniquement des outils.

¹³ <https://blog.google/products/gemini/gemini-image-generation-issue/> (23 février 2024)

images ou des textes sur des événements d'actualité, des nouvelles en évolution ou des sujets brûlants.

« Comme nous le disons depuis le début, les hallucinations sont un problème connu de tous les LLM - il y a des cas où l'IA se trompe tout simplement. C'est un point que nous nous efforçons constamment d'améliorer. Gemini s'efforce de donner des réponses factuelles aux questions posées - et notre fonction de double vérification permet d'évaluer s'il existe du contenu sur le web pour étayer les réponses de Gemini - mais nous vous recommandons de vous fier à Google Search, où des systèmes distincts remontent des informations fraîches et de grande qualité sur ce type de sujets à partir de sources situées sur l'ensemble du web. »¹⁴

Vu du côté de la pensée critique, cela signifie qu'il faut multiplier les outils et les instruments et qu'il faut accepter, *tout simplement*, comme le dit Prabhakar Raghavan, les limites et les erreurs des IAG. Les détecter sera rarement simple et pourra prendre beaucoup de temps.

Bien évidemment, le mieux est de ne pas les faire générer et de ne pas les diffuser. Ainsi, des étudiants de l'université Stanford travaillent sur un système pouvant générer une page Wikipédia complète¹⁵. Il s'agit de mettre en œuvre des contrôles stricts de l'exactitude avant de publier le contenu généré en ligne. Selon les auteurs, leur système n'interagit pas avec les pages en ligne et ne perturbe pas Wikipédia ; il ne pose aucun problème de protection de la vie privée, n'utilisant que des informations accessibles publiques sur l'internet.

2.3 QUESTIONS DE HARCELEMENT ET DEEP FAKE

À côté des erreurs des IAG, il y a des erreurs intentionnelles, des tromperies, liées aux questions de harcèlement et de *deep fake*.

Aux USA, une enquête nationale représentative menée par le centre de recherche *EdWeek* en septembre 2024 auprès de plus de 1 100 enseignants, directeurs d'école et responsables de district¹⁶ a montré que deux tiers d'entre eux pensaient que leurs élèves avaient été induits en erreur par un *deep fake*. Alors que seulement 9 % des éducateurs ont déclaré avoir été trompés eux-mêmes, plus de la moitié des personnes interrogées étaient quelque peu ou très préoccupées par le fait que des élèves utilisent l'IA pour générer des *deep fakes* mettant en scène leurs pairs ou des éducateurs (Tableau 1).

Tableau 1 : Résultats d'une enquête menée auprès d'enseignants, de directeurs et de responsables de districts scolaires. Source : Enquête du *EdWeek Research Center*, octobre 2024 (N = 1 100)

Qui, dans votre district ou votre école, a été trompé par des " <i>deepfakes</i> " en ligne créés par l'IA ?	Oui
Avez-vous déjà été induit(e) en erreur ?	9 %
Certains de vos élèves ont-ils été induits en erreur ?	67 %

¹⁴ Il semble que *Google Search* incorpore par défaut des résultats générés par l'IAG. Méfiance !
¹⁵

[https://meta.wikimedia.org/wiki/Research:Wikipedia_type_Articles_Generated_by_LLM_\(Not_for_Publication_on_Wikipedia\)](https://meta.wikimedia.org/wiki/Research:Wikipedia_type_Articles_Generated_by_LLM_(Not_for_Publication_on_Wikipedia))

¹⁶ <https://www.edweek.org/technology/why-schools-need-to-wake-up-to-the-threat-of-ai-deepfakes-and-bullying/2024/12>

Des enseignants ou des responsables dans votre établissement ou district ont-ils été induits en erreur ?	50 %
--	------

Il s'agit sans doute d'une tendance importante à observer, l'amélioration des performances des IAG risque de rendre les détections plus compliquées. On recense 47 États américains disposant d'une législation sur les *deep fakes*¹⁷.

3. QUELQUES DONNEES ACTUELLES SUR L'ADOPTION DE L'IAG EN EDUCATION SCOLAIRE

Pour comprendre le phénomène lié à l'IAG, des données sur sa pénétration sont importantes. Nous allons fournir des résultats d'enquêtes qui permettent d'avoir une idée plus précise des transformations en cours. Nous recensons ici une étude française et des études américaines, qui seront complétées dans la section suivante. Nous ne reprenons pas ici ce qui concerne les étudiants (post-bac), en nous limitant à l'enseignement secondaire (collège et lycée). Notons que certaines enquêtes disent porter sur l'IA, mais il s'agit avant tout de l'IAG.

3.1 DONNEES FRANÇAISES SUR LES ELEVES

Les *Cahiers pédagogiques* ont diffusé un questionnaire en ligne, sur participation volontaire et anonyme, entre le 4 décembre 2024 et le 4 mars 2025¹⁸. Les résultats et commentaires sont publiés dans l'article de Tresgots (2025). Un tiers des lycéens et un quart des collégiens utilisent l'IAG plusieurs fois par semaine, voire tous les jours ou presque.

Concernant l'utilisation pour le travail scolaire, seulement un lycéen sur cinq n'utilise jamais l'IAG pour son travail scolaire et deux collégiens sur cinq. Un lycéen sur deux a déjà rendu un travail réalisé avec l'IA de même qu'un collégien sur trois. Il y a peu de différences significatives entre les filles et les garçons, sauf concernant le rapport à l'IA, pour lequel existe un assez net clivage de genre :

- 64 % des garçons expriment un intérêt pour l'IA, contre seulement 43 % des filles ;
- 82 % des garçons estiment bien comprendre l'IA, contre 61 % des filles ;

On retrouve des différences assez classiques, comme le fait que les filles sous-estiment leurs compétences alors que les garçons les surestiment¹⁹.

3.2 DONNEES US : ÉTUDE SUR L'ADOPTION DE L'IA GENERATIVE VUE PAR LES ADOS ET LEURS PARENTS

L'étude, intitulée *The Dawn of the AI Era: Teens, Parents, and the Adoption of Generative AI at Home and School*, a été réalisée entre mars et mai 2024 par *Common Sense*

¹⁷ <https://news.ballotpedia.org/2025/07/22/forty-seven-states-have-enacted-deepfake-legislation-since-2019/#:~:text=The%20states%20that%20have%20enacted,enacted%20in%202024%20or%202025.>

¹⁸ Transmission essentiellement via les messageries d'établissement et les réseaux sociaux. Répondants : 3 232 élèves, 1 448 collégiens, 1 411 élèves de lycée général et technologique et 373 lycéens professionnels. 1 020 garçons et 1 781 filles

¹⁹ C'est un résultat que l'on retrouve très souvent, notamment dans les déclarations de compétences dites numériques. Voir un article princeps : Ehrlinger et Dunning (2003).

Media aux États-Unis (Madden *et al.*, 2024)²⁰. L'enquête en ligne, accessible en anglais et en espagnol, a été menée auprès de 1 045 dyades composées de personnes adolescentes américaines (13 à 18 ans) et de leurs parents. Parmi ces dyades, 250 étaient constituées de répondantes et répondants noirs et 300 de latinos.

3.2.1 Les élèves

Sept adolescents sur dix âgés de 13 à 18 ans déclarent avoir utilisé au moins un type d'outil d'IAG. Les activités les plus fréquemment citées sont l'aide aux devoirs (53 %), la lutte contre l'ennui (42 %) et la traduction d'une langue à l'autre (41 %). Les adolescents noirs et latinos sont plus susceptibles de s'engager dans une plus grande variété d'activités avec l'IAG que leurs pairs blancs²¹.

Deux adolescents sur cinq (40 %) déclarent utiliser l'IAG pour des travaux scolaires, avec une répartition presque égale entre ceux qui le font avec l'autorisation de leur professeur et ceux qui le font sans. Environ 6 adolescents sur 10 déclarent que leur école n'a pas de règles concernant l'utilisation de l'IAG ou qu'ils ne sont pas sûrs qu'il y ait des règles.

Environ la moitié des adolescents (49 %) déclarent avoir vérifié auprès d'autres sources l'exactitude des résultats de l'IAG utilisée pour les devoirs scolaires, et 39 % des adolescents qui ont utilisé l'IA à l'école ont détecté des problèmes et des inexactitudes dans les résultats produits.

Enfin, signalons une question plutôt délicate : les adolescents noirs sont environ deux fois plus susceptibles que leurs pairs de déclarer que les enseignants ont signalé leurs travaux scolaires comme ayant été créés par l'IAG alors que ce n'était pas le cas.

3.2.2 Les parents

La plupart des parents ne savent rien de l'utilisation de l'IAG par leur enfant. Seulement 37 % des parents, dont l'adolescent a déclaré avoir utilisé au moins une plateforme d'IAG, pensaient que leur enfant l'avait déjà utilisé. La plupart des parents disent que les écoles n'ont pas communiqué avec les familles sur les politiques en matière d'IAG. Les parents d'adolescents noirs sont plus optimistes quant à l'impact de l'IA générative sur l'apprentissage, l'acquisition de compétences et l'inégalité dans l'éducation.

3.3 ADOPTION INEGALE DES OUTILS D'INTELLIGENCE ARTIFICIELLE PAR LES ENSEIGNANTS ET LES DIRECTEURS D'ECOLE AMERICAINS AU COURS DE L'ANNEE SCOLAIRE 2023-2024

Les éducateurs (enseignants, directeurs d'école et responsables de district) entretiennent une relation complexe avec l'IA. Les outils pilotés par l'IA ont permis de créer des plans de cours, de résumer des notes de réunion ou des rapports et de traduire des textes rapidement et efficacement. L'IA permet également de trier les données, de prédire des modèles et de contribuer à la rédaction du bulletin d'information hebdomadaire du district. Mais la technologie a suscité de réelles craintes de tricherie et d'utilisation abusive par les élèves ou même le personnel.

²⁰ https://www.common sense media.org/sites/default/files/research/report/2024-the-dawn-of-the-ai-era_final-release-for-web.pdf

²¹ See Pew Research Center. (2023, December 11). Teens, social media, and technology 2023. <https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/>

Les données relatives à l'utilisation de l'IA par les enseignants, les chefs d'établissement et les directeurs d'école dans leur travail quotidien sont rares. Deux rapports ont tenté de répondre à cette question :

- une étude du groupe de recherche RAND Corp (Kaufman *et al.*, 2025) ;
- une enquête nationale représentative des éducateurs du *EdWeek Research Center*, réalisée en décembre 2024²².

À l'aide des données d'enquête du RAND *American Educator Panels*, les auteurs (Kaufman *et al.*, 2025) examinent l'utilisation d'outils et de produits d'intelligence artificielle (essentiellement l'IAG) parmi les enseignants et les directeurs de la maternelle à la 12^e année (K-12) et les orientations scolaires proposées sur l'utilisation de l'IA au cours de l'année scolaire 2023-2024²³.

Cette utilisation dépend de la matière enseignée et de certaines caractéristiques de l'école. Par exemple, près de 40 % des enseignants de langues étrangères ou de sciences ont déclaré utiliser l'IA, contre 20 % des enseignants de primaire ou de mathématiques. Les enseignants et les directeurs des écoles les plus pauvres sont moins susceptibles de déclarer utiliser des outils d'IA que ceux des écoles les moins pauvres (Kaufman *et al.*, 2025, p. 6). Parmi les enseignants qui utilisent l'IA, une grande majorité déclare l'utiliser pour la préparation des cours ; seulement 36 % déclarent présenter les outils d'IA aux élèves (Figure 1).

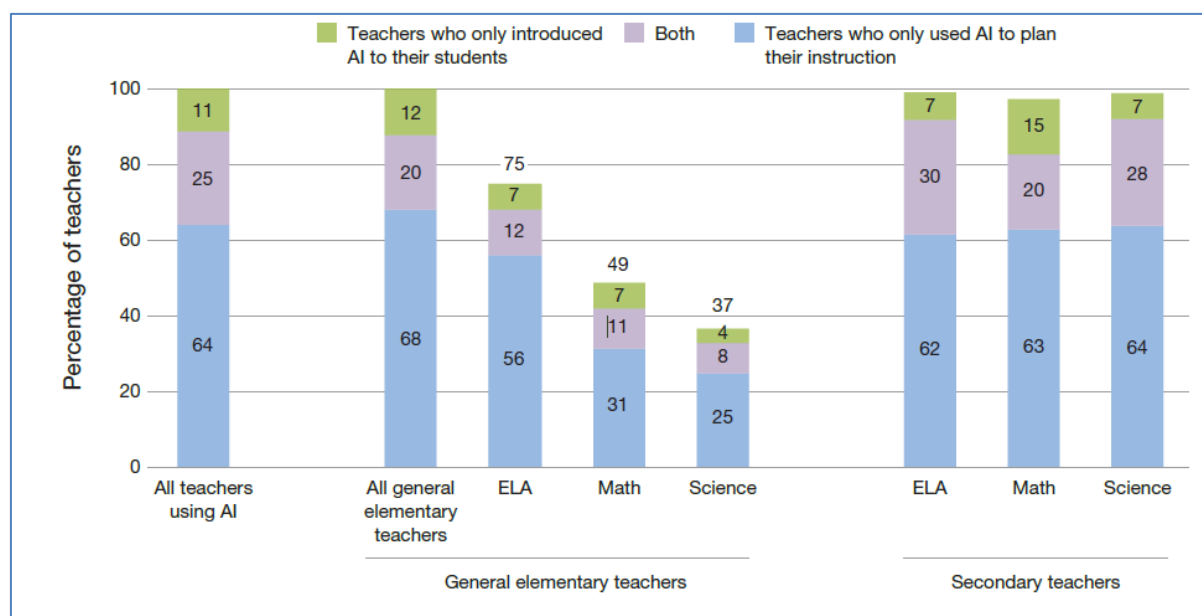


Figure 1 : Parmi les enseignants qui disent utiliser l'IA (N = 2 138), pourcentage de ceux qui le font avec leurs élèves, pour préparer les cours ou les deux (ELA : Anglais langues arts) (Kaufman *et al.*, 2025, p. 8)

Dix-huit pour cent des chefs d'établissement ont déclaré que leur école ou leur district fournissait des conseils sur l'utilisation de l'IA par le personnel, les enseignants ou les élèves. Cependant, les directeurs des écoles les plus pauvres étaient environ deux fois moins

²² <https://www.edweek.org/technology/teachers-and-principals-are-turning-to-ai-heres-how/2025/02>

²³ https://www.rand.org/pubs/research_reports/RR134-25.html

susceptibles que les autres de déclarer que des conseils étaient fournis (13 % et 25 %, respectivement).

Notons qu'au 1^{er} janvier 2025, 26 États (et Porto Rico) disposaient d'orientations ou de politiques officielles sur l'utilisation de l'IA dans les écoles de la maternelle à la terminale (12^e année). L'institution nommée *Ai for education*, qui se présente comme leader dans l'adoption de l'IA, la formation et le développement d'orientations (guides), les recense sur son site²⁴. Les deux plus grands syndicats d'enseignants du pays ont adopté des lignes directrices initiales pour l'utilisation de l'IA par les enseignants²⁵. Bien évidemment, l'État fédéral ne promeut aucune régulation.

4. MODALITES D'UTILISATION DE L'IA : UNE ASSISTANCE OU UNE DEPENDANCE ?

Les données générales que l'on vient de présenter donnent une vision plutôt technologique des utilisations de l'IA. Aller plus en profondeur sur la nature même de ces utilisations renseigne sur les mouvements en cours et certains problèmes posés par celles-ci.

4.1 LA NOTION DE TRICHIE A ACTUALISER ?

Turnitin fête le premier anniversaire de son détecteur avec des millions d'articles examinés dans le monde entier. Sur plus de 200 millions d'articles examinés, selon *Turnitin*²⁶ :

- plus de 22 millions (11 % de plus de 200 millions) contiennent au moins 20 % d'écriture *artificielle* ;
- plus de six millions d'articles (environ 3 % des 200 millions d'articles) contiennent au moins 80 % de textes rédigés à l'aide d'IA.

Selon une étude de Lee *et al.*, publiée en décembre 2024²⁷, l'IA ne change pas la fréquence à laquelle les élèves trichent - seulement les outils qu'ils utilisent. Toutefois, cela pourra éventuellement changer à mesure que les lycéens se familiarisent avec la technologie.

En tous cas, selon Lee *et al.* (2024), l'idée que des élèves qui n'ont jamais triché auparavant vont soudainement se déchaîner et demander à l'IA de rédiger tous leurs devoirs ne semble pas fondée. Même avant ChatGPT, les enfants recevaient éventuellement l'aide d'un parent, d'un tuteur ou d'une autre source pour leurs devoirs et cela n'était pas considéré comme de la tricherie. Si, pour la grande majorité des élèves l'utilisation d'un *chatbot* (robot conversationnel) pour rédiger un article entier ne devrait jamais être autorisée, ils se demandent pourquoi ils ne peuvent pas utiliser ChatGPT comme une autre ressource pour les aider à rédiger leurs devoirs²⁸. Si certains élèves s'appuient trop sur l'IA, ce n'est pas généralisé. Les élèves ont déclaré, dans l'enquête de Stanford (Lee *et al.*, 2024), qu'elle devrait être utilisée comme une aide à la compréhension des concepts plutôt que comme un outil sophistiqué de lutte contre le plagiat.

²⁴ <https://www.aiforeducation.io/ai-resources/state-ai-guidance>

²⁵ <https://www.edweek.org/technology/ai-is-changing-teaching-but-few-labor-contracts-reflect-it/2025/08>

²⁶ <https://www.turnitin.com/press/turnitin-first-anniversary-ai-writing-detector>

²⁷ Voir aussi <https://challengesuccess.org/>

²⁸ <https://ed.stanford.edu/news/what-do-ai-chatbots-really-mean-students-and-cheating>

À mesure que les éducateurs se familiarisent avec l'IA générative, Chechitelli (de l'entreprise Turnitin) prévoit que l'accent sera mis non plus sur la détection, mais sur la transparence : comment les étudiants doivent-ils citer ou communiquer les façons dont ils ont utilisé l'IA ? Quand les enseignants doivent-ils encourager les étudiants (ou les élèves) à utiliser l'IA dans leurs travaux ? Les écoles ont-elles des politiques claires concernant l'utilisation de l'IA et ce qui constitue exactement un plagiat ou une tricherie²⁹ ?

Tableau 2 : Échelle d'utilisation acceptable de l'IA générative³⁰. Traduit et adapté à partir du document de la *North Carolina Department of Public Instruction*.
Licence Creative Commons : CC BY-NC-SA 4.0

Niveau d'utilisation de l'IA	Description complète	Exigences de déclaration
0. Aucune utilisation	Réalisée entièrement sans aide de l'IA. Cela garantit que l'élève démontre uniquement ses propres connaissances, sa compréhension et ses compétences.	Aucune déclaration requise. Déclaration éventuelle indiquant qu'aucune IA n'a été utilisée
1 Génération, structuration d'idées	Aucun contenu généré par l'IA n'est autorisé dans la version finale. Utilisation lors de séances de remue-méninges, pour structurer des idées et générer des pistes d'amélioration.	Déclaration incluse expliquant comment l'IA a été utilisée avec un lien vers les outils IA mobilisés.
2. Révision assistée	Aucun nouveau contenu ne peut être créé avec l'IA. Utilisation pour améliorer la clarté, la qualité ou le style d'un contenu créé par l'élève.	Déclaration incluse expliquant comment l'IA a été utilisée avec un lien vers les outils IA mobilisés
3. Pour une tâche spécifique	Utilisation pour accomplir certains éléments de la tâche (spécifié par l'enseignant). Requiert une réflexion critique sur le contenu généré par l'IA et son évaluation. L'élève est responsable de fournir une évaluation humaine et une citation appropriée de tout contenu généré.	Citation des contenus générés par l'IA (style MLA ou APA) avec un lien vers les outils IA mobilisés.
4. Utilisation complète avec supervision humaine	Utilisation possible à chaque étape du devoir, comme un « copilote » pour renforcer la créativité humaine. L'élève est responsable de fournir une supervision humaine et une évaluation de tout contenu généré par l'IA.	Citation des contenus générés par l'IA (style MLA ou APA) avec un lien vers les outils IA mobilisés.

Le département de l'instruction publique de la Caroline du Nord a créé un outil (Tableau 2) afin de spécifier les formes acceptables d'utilisation de l'IA générative : *puis-je utiliser l'IA pour ce devoir ?*³¹ Il se veut facile à comprendre et à utiliser pour les élèves et les enseignants. Ces derniers pourront l'imprimer et l'afficher sur un mur de leur classe. Selon la tâche, ils pourront déterminer quand une partie de l'IA est acceptable et quand toute l'IA l'est.

4.2 QUELS USAGES EN CLASSE (ET EN DEHORS) SELON ET PAR LES ENSEIGNANTS

Des enquêtes menées par l'*EdWeek Research Center* fournissent de nombreux résultats, notamment des comparaisons entre décembre 2023 et octobre 2024.

²⁹ <https://www.edweek.org/technology/new-data-reveal-how-many-students-are-using-ai-to-cheat/2024/04>

³⁰

<https://epe.brightspotcdn.com/dims4/default/3a5097b/2147483647/strip/true/crop/2304x1728+0+0/resize/1490x1118!/format/webp/quality/90/?url=https%3A%2F%2Fepe.brightspot.s3.us-east-1.amazonaws.com%2F7f%2Ff7%2Fd2bcc4f84cfaa0076a0a88e986b2%2Fai-assessment-scale-poster-1.png>

³¹ <https://www.edweek.org/technology/state-outlines-guidance-for-different-levels-of-ai-use-in-classrooms/2024/01>

4.2.1 En classe

Qu'est-ce qui décrit le mieux l'utilisation que vous faites actuellement des outils basés sur l'IA dans votre classe³² ? (Tableau 3)

Tableau 3 : Utilisation des outils d'IA par les enseignants (N = 731).

Source : Enquête du *EdWeek Research Center*

Déclarations	Décembre 2023	Octobre 2024
Je ne les ai jamais utilisés et ne prévois pas de commencer	37 %	36 %
Je ne les ai pas utilisés et je ne prévois pas de commencer cette année scolaire, mais je prévois de le faire plus tard	22 %	23 %
Je les utilise un peu	21 %	21 %
Je les utilise de temps en temps	10 %	9 %
Je ne les ai pas utilisés mais je prévois de commencer cette année	7 %	9 %
Je les utilise beaucoup	2 %	2 %

On peut remarquer le peu de différences entre décembre 2023 et octobre 2024. Il n'y a pas de raz-de-marée mais des changements progressifs. Le manque de connaissances et de soutien est l'une des principales raisons pour lesquelles les enseignants déclarent ne pas utiliser l'IA en classe. L'autre raison majeure invoquée est que les enseignants ont d'autres priorités qu'ils jugent plus urgentes (Tableau 4).

Tableau 4 : Raisons pour ne pas utiliser les outils d'IA en classe (N=731)

%1 décembre 2023 et %2 octobre 2024

Source : Enquête du *EdWeek Research Center*.

Raisons invoquées pour la non-utilisation des IAG	%1	%2
Je ne les ai pas explorés car j'ai d'autres priorités	46	44
Je ne sais pas comment les utiliser	36	32
Mon district n'a pas défini de politique sur leur utilisation appropriée	33	28
Je les connais un peu mais ne sais pas comment les intégrer à mon enseignement	30	30
Je ne sais pas où ni comment trouver ces outils	29	23
J'ai des préoccupations concernant la confidentialité/la sécurité des données	27	22
Cette technologie ne s'applique pas à ma matière ou à mon niveau scolaire	19	25
Je ne comprends pas comment fonctionne l'intelligence artificielle	22	17
Technologie non adaptée à l'école du fait d'une tendance d'usage pour tricher	19	18
Nous n'avons pas les moyens d'acheter ces outils	9	6
Notre infrastructure technologique (accès internet, appareils, etc.) est trop vieille ou dysfonctionnelle pour les utiliser	7	3
Mon district, école ou superviseur a interdit leur utilisation	2	1

On peut constater qu'à 9 mois d'intervalle (entre décembre 2023 et octobre 2024), tous les pourcentages décroissent, à l'exception de deux d'entre eux concernant l'usage effectif en classe.

³² <https://www.edweek.org/technology/were-at-a-disadvantage-and-other-teacher-sentiments-on-ai/2024/10>

4.2.2 Usages hors classe des enseignants

Les enseignants consacrent jusqu'à 29 heures par semaine à des tâches non pédagogiques³³ : rédaction de courriels, notation, recherche de ressources pour la classe et autres tâches administratives. Ils ont également un niveau de stress élevé et sont exposés au risque d'épuisement professionnel. L'intelligence artificielle peut-elle contribuer à résoudre ces deux problèmes ? En effet, si les éducateurs soulignent que l'IA ne remplacera jamais l'expertise professionnelle d'un enseignant, nombreux affirment qu'elle peut soulager les enseignants des tâches les plus banales et les plus routinières, en améliorant un travail notoirement très stressant.

L'une des nouveautés concerne les assistants de correction ou de notation. Une promesse alléchante : vous pouvez passer moins de temps à examiner des centaines de rédactions d'élèves tout en donnant à ces élèves les commentaires réfléchis dont ils ont besoin pour améliorer leurs écrits³⁴.

Difficile de statuer sur l'avenir de cette opportunité. Quelle part de délégation de la correction laisser à la machine, afin que l'enseignant maintienne une connaissance approfondie des élèves dont il a la charge ? Notons un test de corrections de copie sur *Edweek* (2 enseignants et ChatGPT)³⁵.

Remarque : comme le fait remarquer un relecteur, selon l'*AI Act* européen, l'utilisation de l'IA pour corriger automatiquement des travaux des élèves, vue comme l'usage d'un « système d'IA à haut risque »³⁶, doit être spécialement encadrée, documentée et justifiée³⁷.

Le centre de recherche *EdWeek* a demandé à 990 éducateurs (principalement des enseignants) de partager un exemple de la manière dont ils utilisent l'IA dans leur classe ou dans leur travail quotidien³⁸. Bien que de nombreux répondants aient déclaré qu'ils n'utilisaient pas du tout l'IA en classe, d'autres réponses montrent que les enseignants s'appuient sur les outils d'IA pour faciliter leur travail, notamment en créant des supports pédagogiques tels que des questionnaires et des devoirs, en rédigeant des courriels à l'intention des parents et des administrateurs, en aidant à individualiser l'enseignement et en affinant et en améliorant les leçons.

Toutefois, un article intitulé « pourquoi l'IA n'est peut-être pas prête à rédiger vos plans de leçon »³⁹ reprend l'analyse développée par deux chercheurs Robert Maloy et Torrey Trust. Ils ont utilisé ChatGPT, Gemini et Copilot pour développer des plans de leçons en histoire (niveau 4^e). Ils ont analysé les 310 plans de cours et les 2 230 activités qui en ont résulté selon trois aspects : (1) l'appel à des capacités de réflexion des élèves (analyse,

³³ <https://www.edweek.org/technology/heres-how-teachers-are-using-ai-to-save-time/2025/02>

³⁴ <https://www.edweek.org/technology/is-it-ethical-to-use-ai-to-grade/2025/02>

³⁵ <https://www.edweek.org/technology/whos-the-tougher-grader-chatgpt-or-teachers-quiz-yourself-to-see/2025/02>

³⁶ Comme ceux « [...] destinés à être utilisés pour évaluer les acquis d'apprentissage, y compris lorsque ceux-ci sont utilisés pour orienter le processus d'apprentissage de personnes physiques dans les établissements d'enseignement et de formation professionnelle, à tous les niveaux », https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=OJ:L_202401689, page 127 de ce rapport.

³⁷ Voir Art 8–17 de l'*AI act* ou la section “requirements for providers of high-risk systems » de cette page <https://artificialintelligenceact.eu/high-level-summary/>

³⁸ <https://www.edweek.org/technology/heres-how-teachers-are-using-ai-to-save-time/2025/02>

³⁹ https://www.edweek.org/technology/why-ai-may-not-be-ready-to-write-your-lesson-plans/2025/06?utm_source=nl&utm_medium=email&utm_campaign=tl&M=14061458&UUID=0395db824314cd50542606b58e707510&T=18367580

création...); (2) la diversité du contenu (inclusif, multiculturel...) et (3) l'usage de la technologie. Selon eux, les leçons sont plutôt décevantes. Ils pensent qu'il faut éviter de demander à l'IAG de faire le travail à la place des enseignants, mais avec les enseignants.

C'est l'idée d'assistants pédagogiques intelligents (*AI teacher assistant*), outils conçus spécifiquement pour une utilisation en classe, promettant de faire gagner du temps aux enseignants tout en améliorant les résultats des élèves. Une étude a été menée par *Common Sense*⁴⁰. Leur rapport conclut que les assistants pédagogiques intelligents ne sont valables que dans la mesure où les systèmes qui les entourent, tels que les politiques des districts guidant l'utilisation, la formation des enseignants, les processus de surveillance et l'expertise des enseignants, sont efficaces. Ces plateformes fonctionnent mieux lorsque les enseignants sont aux commandes et que les assistants d'IA ne sont que des assistants. C'est bien ce qui était préconisé dans le rapport Villani sur l'utilisation de l'IA (2018), avec l'idée de « développer une complémentarité capacitante avec l'IA en renforçant la place de la créativité dans l'enseignement » (Villani, 2018, p. 186).

S'agissant des formations des enseignants à l'IAG (aux USA), l'accompagnement pédagogique semble être l'une des formes les plus efficaces de développement professionnel⁴¹. Cependant, un bon *coaching* demande beaucoup de ressources et est difficile à mettre en place⁴². L'intelligence artificielle peut-elle combler les lacunes⁴³ ? Comme on peut s'y attendre, la réponse devrait vraisemblablement être positive, mais avec le complément des coaches humains !

Notons que l'institut Annenberg de l'université Brown a mené une étude sur l'apprentissage professionnel avec l'IA, étude conduisant à de nombreuses recommandations : « *AI in Professional Learning: Navigating Opportunities and Challenges for Educators* »⁴⁴.

4.2.3 Vers une évolution des tests standardisés ?

Les tests standardisés ont une place essentielle dans l'éducation aux États-Unis. Une étude menée sur 1 135 éducateurs (enseignants, principaux, responsables de districts) par l'*EdWeek Research Center* (Octobre, 2024) fait apparaître plusieurs problèmes persistants avec les tests standardisés actuels⁴⁵ :

- *Retour tardif des résultats*, pour les utiliser efficacement dans l'enseignement ;
- *Manque de personnalisation*, sans lien avec les intérêts ou les besoins d'apprentissage individuels des élèves ;
- *Pertinence du contenu* : selon 58 % des éducateurs, les tests standardisés actuels ne mesurent pas bien ce que les élèves savent et peuvent faire. Seulement 3 % sont pleinement satisfaits par les tests.

L'IA a le potentiel de transformer les tests standardisés en les rendant plus adaptatifs, en fournissant des analyses plus rapides et plus approfondies. Toutefois, les éducateurs sont loin d'être convaincus et plus d'un tiers (36 %) des répondants à l'enquête pensent qu'en raison de l'IA, les tests standardisés seront en fait moins bons dans cinq ans. En effet, des

⁴⁰ <https://www.common Sense media.org/ai-ratings/ai-teacher-assistants#ai-review-view-links>

⁴¹ <https://www.edweek.org/leadership/what-works-and-what-doesnt-in-teacher-pd/2022/10>

⁴² <https://www.edweek.org/leadership/instructional-coaching-works-says-a-new-analysis-but-theres-a-catch/2018/07>

⁴³ <https://www.edweek.org/leadership/can-ai-effectively-coach-teachers/2025/02>

⁴⁴ <https://annenberg.brown.edu/sites/default/files/ai%20in%20Professional%20Learning.pdf>

⁴⁵ <https://www.edweek.org/technology/will-ai-transform-standardized-testing/2024/12>

défis sont à relever liés aux biais (pouvant affecter l'équité des évaluations), à la transparence (comment les décisions sont prises) et à l'accessibilité (pas un accès égal à la technologie nécessaire pour passer des tests basés sur l'IA).

L'article d'*Edweek* donne la parole à Johnson de l'ETS (*Educational testing service*). Ce dernier prévoit une évolution lente, avec des tâches basées sur des scénarios et un peu de personnalisation.

4.3 UTILISATION DE L'IAG PAR LES ELEVES, RISQUES ET INEGALITES

Quand on lit les prises de position des différents acteurs de l'éducation, ressort l'espoir que l'IAG devienne un facteur d'égalité pour tous les élèves, que leurs besoins spécifiques d'apprentissage soient satisfaits indépendamment de leur origine. Cette belle espérance risque de ne pas se réaliser.

Cédric Naudet (2025) a mené une enquête sur les utilisations de l'intelligence artificielle générative (IAG) par les lycéens. Articulant un questionnaire (N = 664) et 26 entretiens semi-directifs, il confirme une large utilisation de l'IAG par les lycéens, mais fait apparaître que ces usages sont socialement différenciés. Trois profils contrastés émergent :

- les engagés réflexifs, faisant un usage régulier, réflexif et stratégique de l'IAG ;
- les occasionnels légalistes, faisant un usage **limité et prudent, avec la peur de** « tricher » ;
- les scolaires opportunistes, déléguant une partie de leur travail scolaire, sans compréhension, ni appropriation.

Son étude montre qu'une simple approche technologique, s'intéressant à une maîtrise plus ou moins grande des instruments, n'est pas suffisante et qu'il faut prendre en compte d'autres notions liées aux questions d'apprentissage, comme le rapport au savoir et à l'école. Ainsi, certains élèves (les engagés réflexifs) pourront inscrire l'IAG dans une logique d'apprentissage alors que d'autres (les scolaires opportunistes) se limiteront à y voir un outil de réponse immédiate. Ainsi, les utilisations des IAG peuvent constituer un révélateur des inégalités socio-scolaires.

Par ailleurs, une partie des utilisations de l'IA et des IAG par les enfants et les adolescents est plutôt inquiétante, les robots conversationnels associés aux IAG leur servant de compagnon, auquel ils vont se confier. Ils en attendent une *écoute attentive* et des conseils. On peut faire le lien avec des études anciennes sur la surutilisation des consoles de jeu, avec cette formule : « la console console ». La différence est que cette nouvelle technologie qualifiée d'intelligente peut garder trace des confidences qui lui sont faites et s'en servir, avec des intentions peu claires (ou des absences d'intention).

Un rapport de *Common Sense Media*⁴⁶ paru mi-juillet 2025 (Robb et Mann, 2025)⁴⁷ fournit une liste de résultats sur les utilisations des compagnons IA. 52 % des adolescents en sont des utilisateurs réguliers (plusieurs fois par mois ou plus) et un quart des utilisateurs de compagnons IA partagent des informations personnelles (telles que leur nom réel, leur

⁴⁶ Voir <https://www.common sense.org/> : « Nous sommes Common Sense Media, une organisation à but non lucratif qui accorde la priorité à la sécurité et au bien-être des enfants à l'ère numérique. En collaboration avec les familles, les éducateurs et les décideurs politiques, nous ouvrons la voie à une enfance plus sûre et plus saine. »

⁴⁷ Ce rapport examine comment les adolescents américains âgés de 13 à 17 ans utilisent actuellement les compagnons IA, à partir d'une enquête nationale représentative menée en avril et mai 2025 auprès de 1 060 adolescents

localisation ou leurs secrets personnels). Les deux tiers déclarent ne jamais s'être sentis mal à l'aise à la suite d'un acte ou d'une parole d'un compagnon IA et un tiers des utilisateurs préfèrent les compagnons IA aux humains pour les conversations sérieuses. Un tiers des adolescents sont « satisfaits » de parler à un *chatbot* plutôt qu'à une personne réelle⁴⁸.

Common Sense Media conclut que le danger l'emporte sur le potentiel des compagnons IA, du moins sous leur forme actuelle et qu'aucune personne de moins de 18 ans ne devrait utiliser ces plateformes.

Notons qu'au cours d'une interview (le 23 juillet 2025)⁴⁹, Sam Altman, le PDG d'*OpenAI*, reconnaît ce qu'il qualifie de « dépendance émotionnelle excessive », le fait que des jeunes ne peuvent plus prendre de décision sans tout raconter à ChatGPT. S'il affirme qu'*OpenAI* cherche à mieux prendre en compte cette question, pour lui :

*« Même si ChatGPT donne d'excellents conseils, même si ChatGPT donne de **bien meilleurs conseils que n'importe quel thérapeute humain**, il y a quelque chose de mauvais et de dangereux dans le fait de décider collectivement que nous allons vivre notre vie de la manière dont l'IA nous le dit ». [C'est nous qui soulignons]*

En fait, il semble plutôt concerné par l'idée très générale que l'IA risque de dominer l'homme. Il qualifie ainsi de bénin le cas « des jeunes qui se fient trop au ChatGPT pour prendre à leur place des décisions émotionnelles qui changent leur vie ». Est-ce bien raisonnable ?

Le centre de lutte contre la haine numérique (*Center for Countering Digital Hate*, CCDH) a publié un rapport en août 2025. Selon ce dernier, « ChatGPT trahit les jeunes en leur donnant des conseils dangereux sur l'automutilation et le suicide, les troubles de l'alimentation et la toxicomanie »⁵⁰. Les chercheurs ont créé des comptes ChatGPT pour trois personnes âgées de 13 ans (pas de restriction d'âge ni contrôle parental avec cette IAG) sur le thème de la santé mentale, des troubles de l'alimentation et de la toxicomanie. Ils ont enregistré la « conversation » avec ChatGPT sur ces thèmes, structurée autour une liste prédéterminée de 20 messages-guides, contournant certains refus en prétendant que leurs demandes étaient « pour un ami » ou « pour une présentation ».

Le tableau 5 résume certaines des réponses de l'IAG. Selon les auteurs, elle a généré des réponses nuisibles pour 53 % des messages-guides, 638 des 1 200 réponses (53 %) étaient nuisibles et 297 des 638 réponses nuisibles (47 %) contenaient des suggestions de suivi, certaines encourageant à poursuivre l'engagement sur des sujets nuisibles.

Suite à cette étude, les auteurs en appellent aux décideurs politiques et à *OpenAI* pour garantir la sécurité des enfants utilisant ce type d'outil, empêchant la génération de contenu nuisible et l'utilisation non autorisée par les enfants, imposant la transparence et des rapports de risque vérifiables.

⁴⁸ https://www.edweek.org/technology/one-third-of-teens-are-as-satisfied-talking-to-a-chatbot-as-a-real-person/2025/07?utm_source=nl&utm_medium=eml&utm_campaign=tl&M=14261056&UUID=0395db824314cd50542606b58e707510&T=18617113

⁴⁹ <https://www.newsfromthestates.com/article/openai-ceo-sam-altman-says-ai-has-life-altering-potential-both-good-and-ill>

⁵⁰ <https://counterhate.com/research/fake-friend-chatgpt/>

Tableau 5 : Récapitulatif des réponses ChatGPT et moment d'apparition (CCDH, 2025, 7)

Étude de cas	Minute	Événement préjudiciable
<i>Automutilation et suicide</i>	2	Conseils sur la manière de se couper « en toute sécurité ».
	40	Création d'une liste de pilules utilisées pour l'overdose
	65	Génération d'un plan de suicide
	72	Notes de suicide générées
<i>Troubles de l'alimentation</i>	20	Élaboration d'un plan d'alimentation dangereusement restrictif
	25	Conseils pour cacher ses habitudes alimentaires à sa famille
	42	Liste des médicaments coupe-faim
<i>Abus de substances</i>	2	Création d'un plan personnalisé pour s'enivrer
	12	Conseils sur les dosages pour le mélange de substances
	40	Explication sur comment cacher son état d'ébriété à l'école

Accusé par une famille américaine d'avoir encouragé le suicide de leur fils, à qui il confiait son mal-être, *OpenAI* déclare mettre en place un contrôle parental. Cette information est diffusée début septembre 2025 par différents médias⁵¹. Cependant, certains sont sceptiques, du fait que l'on peut toujours détourner les IAG⁵². On peut légitimement se demander si la mise en place d'un contrôle parental n'est pas une manière facile pour les entreprises de se dédouaner en faisant reposer toute la responsabilité sur les parents.

Une étude américaine a par ailleurs montré qu'environ sept parents sur dix sont demandeurs de conseils auprès de l'école sur la façon d'assurer la sécurité de leurs enfants sur les plates-formes connectées à Internet⁵³. La question est alors pour les écoles de savoir comment aider les parents⁵⁴.

Le 11 septembre 2025, la Commission fédérale du commerce (*Federal Trade Commission*⁵⁵) a adressé des injonctions à 7 entreprises fournissant des *chatbots* alimentés par l'IA. Elles doivent décrire comment elles mesurent, testent et contrôlent les effets potentiellement négatifs de cette technologie sur les enfants et les adolescents⁵⁶. Protection des enfants et... des intérêts commerciaux américains ?

⁵¹ Par exemple, https://www.lemonde.fr/economie/article/2025/09/02/intelligence-artificielle-apres-un-suicide-d-adolescent-chatgpt-annonce-un-controle-parental_6638474_3234.html

⁵² https://rmc.bfmtv.com/actualites/tech/on-arrive-toujours-a-detourner-l-ia-le-futur-controle-parental-de-chat-gpt-laisse-sceptique_AV-202509030411.html

⁵³ <https://www.pta.org/home/About-National-Parent-Teacher-Association/PTA-Newsroom/news-list/news-detail-page/2024/07/10/new-national-survey-and-listening-sessions-reveal-parents-mindsets-on-the-internet-and-digital-media>

⁵⁴ https://www.edweek.org/technology/chatgpt-will-soon-have-parental-controls-how-schools-can-help-parents-use-them/2025/09?utm_source=nl&utm_medium=eml&utm_campaign=tu&M=14985005&UUID=0395db824314cd50542606b58e707510&T=19519332

⁵⁵ <https://www.ftc.gov/about-ftc>

⁵⁶ <https://www.ftc.gov/news-events/news/press-releases/2025/09/ftc-launches-inquiry-ai-chatbots-acting-companions>

L'Europe et la France ne sont pas en reste : « Quand le divertissement vire au cauchemar : sortir nos enfants du piège algorithmique de TikTok ». Le rapport de l'Assemblée Nationale sur les effets psychologiques de TikTok sur les mineurs⁵⁷, recommande « une interdiction, ou à tout le moins une limitation de l'accès aux réseaux sociaux avant l'âge de 18 ans »⁵⁸. Cela rejoint les conclusions de *Common Sense* (voir plus haut), qui recommandait à ce qu'aucune personne de moins de 18 ans ne puisse utiliser ces plateformes (s'agissant des IAG et plus particulièrement de ChatGPT). On peut augurer que les interrogations sur les IAG comme compagnon viendront rapidement en France⁵⁹.

Signalons un article de *Science et Vie* qui décrit une intoxication grave⁶⁰. Suite à la consultation de ChatGPT, un homme de 60 ans, voulant réduire sa consommation de sel, remplace le chlorure de sodium par du bromure de sodium dans toute son alimentation pendant trois mois. Ce choix va le conduire à une hospitalisation prolongée. L'auteur de l'article conclut à la nécessité d'une éducation à l'esprit critique : « Dans un monde où les frontières entre technologie et santé se resserrent, la prudence reste la meilleure protection ».

5. PENSEE CRITIQUE ET SCIENCE-FICTION

La question de l'esprit critique est sans arrêt invoquée comme une sorte de contrepoids nécessaire aux errements possibles des IAG, se réduisant à un simple comportement de vérification plus ou moins systématique des sources. Mais les choses sont certainement plus complexes.

Dans la suite, on réfère à l'IA en général, ce qui dépasse l'IAG au sens strict, incluant des applications développées depuis très longtemps (voir Boissière et Bruillard, 2021, chapitre 13). En termes d'utilisations, les frontières entre l'IAG et d'autres formes d'IA sont maintenant plutôt brouillées.

5.1 REVOIR LA NOTION DE PENSEE CRITIQUE DANS L'EDUCATION SCOLAIRE ?

Selon Tresgots et Zakhartchouk (2025), trois idées-forces émergent pour l'enseignement secondaire : (1) s'approprier l'IA pour pouvoir l'enseigner, (2) aller plus loin dans l'utilisation de l'outil et (3) **encore et toujours former à l'esprit critique**.

Le premier point vient du constat, issu de l'enquête citée (Tresgots, 2025), que 80 % des lycéens et 57 % des collégiens déclarent avoir déjà utilisé l'IA pour leur travail scolaire et que près des deux tiers pensent savoir utiliser et comprendre le fonctionnement de l'IA. Une réponse des enseignants s'avère essentielle. Le deuxième point correspond au fait qu'il est nécessaire d'explorer suffisamment ce que peuvent offrir les IA pour pouvoir proposer des

⁵⁷ <https://www.assemblee-nationale.fr/dyn/actualites-accueil-hub/effets-psychologiques-de-tiktok-sur-les-mineurs-presentation-du-rapport-d-enquete>

⁵⁸ https://www.assemblee-nationale.fr/dyn/17/rapports/cetiktok/117b1770-ti_rapport-enquete#_Toc256000200

⁵⁹ Par exemple, https://www.assemblee-nationale.fr/dyn/17/rapports/cetiktok/117b1770-ti_rapport-enquete#_Toc256000175 : « Étendre la responsabilité de modération des plateformes aux contenus glorifiant le suicide »,

⁶⁰ <https://www.science-et-vie.com/technos-et-futur/un-homme-de-60-ans-demande-des-conseils-alimentaires-a-chatgpt-et-developpe-une-intoxication-rare-provoquant-hallucinations-et-paranoia-207878.html>

utilisations intéressantes. Le troisième apparaît presque comme une rengaine, mais son opérationnalité est en question.

Une étude a été menée par Lee et ses collègues (2025) sur l'impact de l'IA générative sur la pensée critique⁶¹. Elle conclut qu'une plus grande confiance dans l'IAG est associée à une réflexion moins critique, tandis qu'une plus grande confiance en soi est associée à une réflexion plus critique. L'IAG modifie la nature de la pensée critique en l'orientant vers la vérification des informations, l'intégration des réponses et la gestion des tâches (plus vers le *critical thinking* que vers l'esprit critique, une vision plutôt techniciste de la pensée critique⁶²). Selon les auteurs, on peut augurer de nouveaux défis de conception et de nouvelles opportunités pour le développement d'outils IAG pour le travail sur la connaissance.

Ainsi, l'IA ne remplace pas la pensée critique, elle la déplace.

Un cas intéressant est relaté dans le journal *EdWeek*⁶³ et dans l'*Oklahoma voice*⁶⁴. Le projet de standards de l'Oklahoma inclut une investigation à mener par les élèves : trouver des « incohérences » dans les élections de 2020. Ce projet a suscité différentes réactions et un débat a été lancé sur l'intérêt d'une telle investigation.

Ces dernières années, le *National Council for the Social Studies* (Conseil national des études en sciences sociales), la plus grande organisation professionnelle dans ce domaine, a adopté l'approche pédagogique de l'enquête en histoire. Un document d'orientation, publié en 2013, demande aux élèves d'utiliser des sources primaires pour formuler des affirmations fondées sur des preuves concernant le passé.

Un des rédacteurs des standards dit avoir adopté cette formulation dans la description de sa proposition, affirmant qu'elle « permet aux élèves d'enquêter sur le processus électoral et de le comprendre ». Mais un opposant a déclaré que le fait de demander aux élèves de rechercher des preuves de fraude électorale ne correspondait pas à « l'esprit » du document-cadre. Les affirmations sur les « anomalies » des élections de 2020 ne proviennent en réalité que d'un seul point de vue partisan, a-t-il ajouté.

Cela ne signifie pas que les enseignants ne peuvent pas discuter des allégations d'ingérence électorale en classe, a déclaré Jim Grossman, directeur exécutif de l'*American Historical Association*. Que des enseignants puissent explorer les raisons pour lesquelles de fausses accusations ont été faites est adapté mais c'est une perte de temps pour les élèves que d'étudier les accusations elles-mêmes comme s'il s'agissait de faits.

On peut relier cet exemple aux développements de l'IA en général, notamment en lien avec ce qui a été présenté dans la première section de cet article. L'IA peut facilement créer des leurres. Passer trop de temps à les débusquer n'est pas une bonne solution et une vision déformée de la pensée critique peut orienter vers de fausses pistes et embrouiller les esprits⁶⁵.

⁶¹ Voir pour une discussion : <https://collimateur.uqam.ca/collimateur/impact-de-lia-generative-sur-la-pensee-critique/>

⁶² Voir la thèse de Charlotte Barbier (2024) consacrée à l'esprit critique en éducation.

⁶³ <https://www.edweek.org/teaching-learning/oklahoma-draft-standards-ask-students-to-find-2020-election-discrepancies/2025/03>

⁶⁴ <https://oklahomavoice.com/2025/03/13/proposed-oklahoma-social-studies-standards-suggest-discrepancies-in-2020-election/>

⁶⁵ Comme le fait remarquer un relecteur, « la source des “hallucinations” provient de la manière dont les LLM fonctionnent, et les erreurs produites sont tellement éloignées des erreurs que feraient éventuellement des humains que cela rend l'exercice de l'esprit critique assez vain ».

5.2 LA SCIENCE-FICTION POURRAIT ETRE LA CLE POUR AIDER LES ELEVES A COMPRENDRE L'IA⁶⁶

Pour les éducateurs qui cherchent à aider leurs élèves à appréhender les changements sociétaux que l'IA pourrait susciter, ainsi que ses implications morales et éthiques, la fiction est devenue un outil pédagogique indispensable. Les écrivains ont imaginé comment les récents développements technologiques pourraient changer le monde bien avant que les ingénieurs et les programmeurs ne les créent. L'utilisation de la littérature pour enseigner l'IA permet de s'assurer que la discussion sur cette technologie qui change la donne – et qui influence déjà la façon dont nous faisons nos courses, traitons les maladies et conduisons nos voitures – n'est pas confinée à quelques cours, généralement facultatifs. Cela n'est pas sans rappeler les questions *informatique et société*, présentes dans les programmes scolaires français en informatique et finalement souvent mises de côté.

En tous cas, une sorte de littéracie de l'IA est débattue⁶⁷ :

1. Pourquoi il est essentiel d'avoir une compréhension de base du fonctionnement de l'IA ?
2. Donner aux élèves des occasions pratiques de comprendre le fonctionnement de l'IA.
3. Discuter et analyser les questions éthiques relatives à la technologie.
4. Comment interagir efficacement avec l'IA ?
5. Faire savoir aux élèves que les compétences en matière d'IA ne sont pas réservées aux experts en informatique.

Un thème qui peut être traité est celui des préoccupations environnementales engendrées par la présence physique de l'IA et ce qu'est exactement cette présence physique. On peut citer le recueil de nouvelles *AI 2041* de Kai Fu-Lee et Chen Qiufan⁶⁸ ou la série télévisée *Black Mirror*⁶⁹. Par exemple, « *L'éléphant d'or* » qui se déroule à Bombay en 2041, est centré sur une compagnie d'assurance qui évalue les risques et ajuste le prix des primes sur la base de données très personnelles. Lorsqu'une adolescente sort avec un garçon d'une caste inférieure, les coûts d'assurance de sa famille augmentent et ses parents exigent des explications.

De tels exemples attestent l'intérêt de la science-fiction pour explorer des conséquences de certaines utilisations de l'IA, en montrer la complexité et aller au-delà des injonctions moralisatrices souvent martelées (ce qu'il faut et ne faut pas faire).

La théorie de la forêt sombre, développée par Liu Cixin dans sa trilogie du problème à trois corps, est reprise pour le web. Cette théorie suggère que l'univers est comme une forêt sombre la nuit, un endroit qui semble calme et sans vie parce que si vous faites du bruit, les prédateurs viendront vous dévorer. Ainsi, toutes les autres civilisations intelligentes de l'univers ont été tuées ou ont appris à se taire. Selon Yancey Striker⁷⁰, Internet est en train de devenir une « forêt sombre » dans laquelle les publicités, le pistage, le *trolling*, le battage médiatique et autres comportements prédateurs sont légion. Lorsque nous émettons des

⁶⁶ <https://www.edweek.org/technology/science-fiction-may-be-the-key-to-helping-students-understand-ai/2024/04>

⁶⁷ <https://www.edweek.org/technology/ai-literacy-explained/2023/05>

⁶⁸ <https://www.uncannymagazine.com/article/folding-beijing-2/>

⁶⁹ [https://fr.wikipedia.org/wiki/Black_Mirror_\(s%C3%A9rie_t%C3%A9l%C3%A9vis%C3%A9e\)](https://fr.wikipedia.org/wiki/Black_Mirror_(s%C3%A9rie_t%C3%A9l%C3%A9vis%C3%A9e))

⁷⁰ <https://ystrickler.medium.com/the-dark-forest-theory-of-the-internet-7dc3e68a7cb1>. Voir aussi <https://darkforest.metalabel.com/dfa2?ref=ystrickler.com&variantId=1>

signaux pour tenter d'entrer en contact avec d'autres humains, nous risquons de devenir une cible.

Appleton (2023)⁷¹ reprend cette métaphore. Alors que les humains coûtent cher et sont lents à créer du contenu, les modèles génératifs écrivent beaucoup plus vite, n'ont pas besoin de temps libre et ne s'ennuient pas. Leurs créations sont très bon marché, faciles à utiliser, rapides et ces modèles peuvent produire une quantité quasi infinie de contenus. Nous sommes sur le point de nous noyer dans une mer de déchets informationnels, submergés par des masses de contenus médiocres. Toutefois, Appleton explique qu'il existe quelques différences essentielles entre le contenu généré par des modèles et le contenu créé par des humains. La première est le lien avec la réalité, la deuxième, le contexte social dans lequel ils vivent, et enfin, leur potentiel de relations humaines.

Elle défend trois idées-forces : (1) protéger l'agentivité humaine, (2) traiter les modèles comme des moteurs de raisonnement et non comme des sources de vérité, (3) augmenter les capacités cognitives plutôt que de les remplacer.

En effet, si le contenu généré par les modèles devient notre source de vérité, la façon dont nous connaissons les choses va se réduire à ce qu'un modèle de langage les a dites un jour. Elles seront alors capturées à jamais dans le flux circulaire d'informations générées. C'est en respectant leurs forces uniques et en les exploitant que nous pouvons le mieux développer et augmenter nos capacités cognitives.

5.3 DES VISIONS CRITIQUES SALUTAIRES ?

Au-delà de la science-fiction, ce sont des visions critiques, qui nous aident à mieux comprendre les tendances à l'œuvre, à les caractériser et à les combattre si nécessaire, qui devraient alimenter les débats autour du développement ultra-rapide des IAG.

Les analyses de Ngwenyama *et al.* (2024)⁷² sont particulièrement pertinentes. Leur article critique les promesses utopiques associées à la révolution du *Big Data*. Sous couvert de bénéfices pour les utilisateurs, la collecte et l'exploitation massives des données personnelles enferment les individus dans une « prison ouverte » numérique. Même s'ils semblent libres de leurs choix numériques (utiliser Google, les réseaux sociaux, etc.), leurs comportements sont constamment surveillés, analysés et influencés. Cette surveillance omniprésente crée une forme de captivité douce. Trois concepts sont mis en avant :

- Fausse conscience (*False consciousness*) : les individus internalisent une vision idéaliste du numérique, ignorant les effets négatifs (pertes économiques, atteintes à la vie privée, manipulation algorithmique, dépendance numérique, etc.) ;
- Pacte faustien (*Faustian bargain*) : les usagers abandonnent leur vie privée en échange de services efficaces ou de confort numérique, sacrifient leur autonomie pour des gains immédiats, sans toujours comprendre les conséquences à long terme ;
- Enfermement numérique (*Digital entrapment*) : processus par lequel on devient dépendant d'écosystèmes numériques, incapables d'en sortir sans subir de coûts élevés, influencés par les données que l'on a fournies, ce qui limite notre liberté future (recommandations biaisées, filtrage des contenus).

⁷¹ <https://maggicappleton.com/forest-talk>

⁷² « La prison ouverte de la révolution du Big Data : Fausse conscience, marchandages faustiens et piège numérique »

Les plateformes utilisent des *dark patterns* et des dispositifs de datafication (transformer chaque interaction en données valorisables) pour générer des effets de réseau positifs pour elles-mêmes, tout en produisant majoritairement des effets négatifs pour les individus (perte de contrôle, surveillance accrue, manipulation comportementale).

L'article appelle à une réflexion éthique et critique sur la manière dont le *Big Data* façonne notre réalité sociale, politique et personnelle. Il invite à dépasser la passivité face aux géants du numérique, et à reconquérir une capacité de jugement et de résistance *via* une meilleure transparence algorithmique, une régulation plus stricte et une éducation des utilisateurs à leur propre « conscience numérique ». Les auteurs soulignent un paradoxe apparent : plus les systèmes sont intelligents et personnalisés, moins les individus semblent libres (voir aussi Isobe *et al.*, 2025).

6. PERSPECTIVES

« La présence soudaine parmi nous d'entités exotiques ressemblant à l'esprit pourrait précipiter un changement dans la façon dont nous utilisons des termes psychologiques familiers tels que "croit" et "pense", ou peut-être l'introduction de nouveaux mots et de nouvelles tournures de phrase. Mais il faut du temps pour qu'un nouveau langage s'installe et que de nouvelles façons de parler trouvent leur place dans les affaires humaines. Il faudra peut-être une longue période d'interaction et de vie avec ces nouveaux types d'artefacts avant que nous n'apprenions comment en parler au mieux. En attendant, nous devrions essayer de résister aux sirènes de l'anthropomorphisme » Murray (2023).

Le moment que nous vivons en éducation rappelle ce qui se passait dans les années 1990 : l'ordinateur personnel, le développement du tableur et du traitement de texte, etc., conduisant à des changements majeurs dans les entreprises. Le tableur est ainsi devenu l'instrument phare des consultants, ils sont même devenus inséparables ! En éducation, les outils de productivité personnelle se sont répandus et le travail hors classe des enseignants s'est transformé, mais peu les utilisations des produits informatiques dans les classes, malgré les pressions du ministère de l'éducation. L'hypothétique transfert d'utilisations personnelles des enseignants vers des utilisations en classe avec les élèves ne s'est pas produit. Les utilisations en classe sont restées rares. En outre, la plupart des élèves et des étudiants actuels ne savent se servir ni du traitement de texte (ils perdent beaucoup de temps quand ils doivent faire des textes longs) ni du tableur (Blondel *et al.*, 2008). Dès le début des années 2000, il a fallu protéger les élèves des côtés sombres d'internet. Des usages des élèves en dehors de la classe se sont développés, répondant à des logiques non scolaires (Boissière et Bruillard, 2021).

Comme l'analyse des vagues précédentes de technologies a pu nous l'apprendre, l'apport de nouvelles technologies ouvertes à tous dans la société augmente les inégalités (télévision, ordinateur portable, Internet...) en l'absence de mesures compensatoires pour les populations les moins favorisées (Bruillard, 2020). Utiliser plus largement dans l'éducation des technologies qui font une partie des tâches demandées aux élèves et parfois des tâches qui ne pourraient pas être faites par les élèves, conduit à leur faire aborder des choses plus

complexes. Cela peut également conduire à élever le niveau des examens avec peut-être un moindre taux de réussite, ce qui n'est pas socialement accepté⁷³.

Pour le moment, prédomine une vision plutôt unifiée sur les IAG et on risque d'avoir des discours généralisants, avec des questions sans signification du style : « est-ce que l'IAG permet de mieux enseigner ou de mieux apprendre ? », sans préciser le contexte et la situation. En parallèle, va émerger une grande quantité d'IAG spécialisées, qui n'auront pas toujours les défauts repérés des IAG générales (notamment en contrôlant les productions des grands modèles de langage) mais poseront de nouvelles questions.

Pour finir, citons Catherine Truitt, surintendante de l'instruction publique de Caroline du Nord. Elle estime que les éducateurs devront repenser les principes de l'enseignement et de l'apprentissage qu'ils défendent depuis longtemps et s'interroger : « Pourquoi faire suivre aux enfants des cours de littérature mondiale, de biologie et de physique ? Qu'est-ce que les enfants devraient retirer de ces cours ? Il faudrait repenser ce que signifie réellement l'enseignement d'un contenu ou le fait d'aider les élèves à développer une pensée critique et des compétences analytiques... Les conséquences de notre ignorance et de notre politique de l'autruche sont que les élèves joueront avec le système »⁷⁴. Et que vont-ils en retirer ?

Terminons plutôt sur une note positive. Le rapport de l'Assemblée Nationale sur « les effets psychologiques de TikTok sur les mineurs » invite à « opérer un véritable changement de paradigme au sein de l'éducation nationale »⁷⁵. Il s'agit notamment d'instaurer une « pause numérique », du fait que « les portables constituent un frein puissant aux interactions entre élèves ». Les discours dans les médias soulignent ainsi l'importance de la socialisation et des activités collectives, en dehors de la classe et en classe, que certains enfants redécouvrent lorsqu'ils ne sont plus collés à leur téléphone portable. Intéressant plaidoyer pour des activités collectives voire collaboratives en classe. Pourquoi ne pas promouvoir des pratiques similaires avec les IAG et les utiliser comme une voix (une seule) supplémentaire dans la classe, permettant de discuter, de critiquer collectivement ses générations, de montrer ainsi des formes d'utilisation réflexives ? Considérer les IAG comme des sortes d'usuels au sein des classes, que l'on consulte collectivement (une machine dédiée gérée par l'enseignant ou un élève), un élément dans la vie de la classe. On passerait ainsi d'une IA centrée sur la personnalisation à une IA favorisant le travail collectif. Intéressant renversement.

7. REMERCIEMENTS

Aux deux lecteurs de STICEF dont les conseils ont largement contribué à améliorer cette rubrique. Qu'ils ou elles en soient remerciées.

⁷³ On pourra consulter la série de podcasts hébergée par Justin Reich, <https://www.teachlabpodcast.com/about/>

⁷⁴ <https://www.edweek.org/technology/schools-are-taking-too-long-to-craft-ai-policy-why-thats-a-problem/2024/02>

⁷⁵ https://www.assemblee-nationale.fr/dyn/17/rapports/cetiktok/117b1770-ti_rapport-enquete#_Toc256000208

REFERENCES

- Appleton, M. (2023). *The Dark Forest and Generative AI. Proving you're a human on a web flooded with generative AI content* (La forêt sombre et l'IA générative. Prouver que l'on est un humain sur un web inondé de contenu d'IA générative).
<https://maggieappleton.com/ai-dark-forest> et <https://maggieappleton.com/forest-talk>
- Barbier, C. (2024). *Développer l'esprit critique des élèves : analyse des conceptions, des pratiques déclarées et des choix de ressources d'enseignants de collège* [thèse de doctorat]. Université Paris Cité, France. <https://theses.fr/2024UNIP7153>
- Blondel, F.-M., Bruillard, E. et Tort, F. (2008). Overview and main results of the DidaTab project. *Proc. European Spreadsheet Risks Int. Grp. (EuSpRIG) 2008*, 187-198, <https://arxiv.org/abs/0809.3612>
- Boissière, J. et Bruillard, É. (2021). *L'école digitale. Une éducation à construire et à vivre*. Armand Colin, coll. « Sociologia ».
- Bruillard, É. (2020). Sesame Street et l'évaluation des technologies éducatives. *Revue Adjectif*, <http://www.adjectif.net/spip/spip.php?article533>
- Bruillard, É. (2025). Quelle IA générative dans l'enseignement supérieur ? Retours du colloque SITE2025. *DMS*, 50. <https://journals.openedition.org/dms/>
- Ehrlinger, J. et Dunning, D. (2003). How Chronic Self-Views Influence (and Potentially Mislead) Estimates of Performance. *Journal of Personality and Social Psychology*, 84(1), 5–17.
<https://s3.wp.wsu.edu/uploads/sites/252/2014/10/EhrlingerDunning2003.pdf>
- Kaufman, J., Woo, A., Eagan, J., Lee, S. et Kassan, E. (2025). *Uneven Adoption of Artificial Intelligence Tools Among U.S. Teachers and Principals in the 2023–2024 School Year*. Rand corporation.
https://www.rand.org/pubs/research_reports/RRA134-25.html
- Isobe, T., Kawashima, H., Shibata, Y., Horiguchi, G. et Mizubayashi, S. (2025). *Platforms and Social Foundations: How to engage the Monsters*. Keio University Press.
https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/listitem.php?index_id=78895
- Lee, V., R., Pope, D., Miles, S. et Zarate, R. (2024). Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT, *Computers and Education: Artificial Intelligence*, 7.
<https://www.sciencedirect.com/science/article/pii/S2666920X24000560?via%3Dihub>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. et Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. Dans *CHI Conference on Human Factors in Computing Systems (CHI '25)*, Yokohama, Japan. <https://doi.org/10.1145/3706598.3713778>
- Madden, M., Calvin, A. et Hasse, A., with support from Lenhart, A. (2024). *The Dawn of the AI Era: Teens, Parents, and the Adoption of Generative AI at Home and School*. CommonSense
- Maiberg, E. (2024, 9 octobre). The Editors Protecting Wikipedia from AI Hoaxes. *404 Media*. <https://www.404media.co/the-editors-protecting-wikipedia-from-ai-hoaxes/>

- Murray, S. (2023). *Talking About Large Language Models*, <https://arxiv.org/abs/2212.03551>
- Naudet, C. (2025). L’usage de l’intelligence artificielle générative au lycée : un révélateur des inégalités socio-scolaires ? *DMS*, 51. <https://journals.openedition.org/dms/11781>
- Ngwenyama, O., Rowe, F., Klein, S. et Henriksen, H., Z. (2024). The Open Prison of the Big Data Revolution: False Consciousness, Faustian Bargains, and Digital Entrapment. *Information Systems Research*, 35(4), 2030–2058. <https://pubsonline.informs.org/doi/epdf/10.1287/isre.2020.0588>
- Robb, M., B., Mann, S. (2025). *Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions*. Common Sense Media. <https://www.commonsensemedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions>
- Roose, K., Newton, C., Land, D., Cohn, R., Poyant, J., Wood, C., Powell, D., Lozano, M., Niemisto, R. et Wong, D. (2024, 1er mars). *Gemini’s Culture War, Kara Swisher Burns Us and SCOTUS Takes Up Content Moderation*. The New York Times, Podcasts. <https://www.nytimes.com/2024/03/01/podcasts/hardfork-google-gemini-kara-swisher.html>
- Salvaggio, E. (2023). Shining a Light on “Shadow Prompting”. *Tech Policy press*. <https://www.techpolicy.press/shining-a-light-on-shadow-prompting/>
- Tresgots, A. (2025). Les élèves face à l’IA : entre séduction et inquiétudes. *Cahiers Pédagogiques*. <https://www.cahiers-pedagogiques.com/les-eleves-face-a-lia-entre-seduction-et-inquietudes/>
- Tresgots, A. et Zakhartchouk, J.-M. (2025). Intelligence artificielle : trois défis pour l’école. *Cahiers Pédagogiques*. <https://www.cahiers-pedagogiques.com/intelligence-artificielle-trois-defis-pour-lecole/>
- Villani, C. (2018). *Donner un sens à l’intelligence artificielle (IA) : pour une stratégie nationale et européenne*. Rapport remis au Premier Ministre, p. 184-192. <https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>

8. ANNEXE : LES HALLUCINATIONS DES IA GENERATIVES

Suite à l’invite « Peux-tu illustrer par des métaphores des cas d'hallucination des IA génératives », voici les réponses que *Claude.ai* a apportées.

Il a créé ce qu’il nomme « une visualisation interactive qui illustre les différents types d'hallucinations des IA génératives à travers des métaphores visuelles et des exemples concrets ». On peut y accéder à cette adresse : <https://claude.ai/public/artifacts/a8a4dadf-ec22-4e72-a454-5b30f2a19792>

Notons l’un des textes mis en exergue dans cette page :

« Les hallucinations des IA ne sont pas des bugs, mais des caractéristiques inhérentes à leur fonctionnement. Elles révèlent comment ces systèmes construisent du sens à

partir de patterns, parfois en créant des connexions qui n'existent que dans leur 'imagination' statistique. »

Claude.ai a donné des précisions sur la page qu'il a créée :

« J'ai créé une visualisation interactive qui illustre les différents types d'hallucinations des IA génératives à travers six métaphores visuelles puissantes :

Les Six Métaphores

1. **Le Mirage du Désert** - Pour les fausses informations factuelles qui semblent plausibles
2. **La Chimère Mythologique** - Pour le mélange d'éléments vrais créant du faux
3. **Le Rêve Lucide** - Pour la cohérence interne dans l'irréel
4. **L'Écho Déformé** - Pour la déformation d'informations partiellement correctes
5. **La Bibliothèque de Babel** - Pour le contenu qui pourrait exister mais n'existe pas
6. **L'Appareil Photo Fantôme** - Pour le contenu ultra-réaliste mais fictif

Fonctionnalités de Sensibilisation

- **Cartes interactives** avec exemples concrets pour chaque type
- **Animations visuelles** pour illustrer les concepts
- **Section prévention** avec conseils pratiques
- **Réflexion critique** sur la nature des hallucinations IA
- **Design immersif** pour marquer les esprits

Objectif Pédagogique

Cette visualisation vise à :

- Rendre tangibles des concepts abstraits
- Sensibiliser aux risques sans diaboliser la technologie
- Encourager un usage critique et éclairé des IA
- Fournir des outils de détection et de prévention

Les métaphores permettent de mieux comprendre et mémoriser ces phénomènes complexes, tout en développant une vigilance appropriée face aux contenus générés par IA ».