

Peut-on utiliser des IA génératives dans un contexte recherche ?

Florence Maraninchi

www-verimag.imag.fr/~maraninx

Verimag / Grenoble INP - UGA / Ensimag

26 mars 2026 – AG du CERAG



Contenu garanti sans usage d'IA génératives

*CC-BY-NC-SA : Attribution - Utilisation non commerciale - Partage dans les mêmes conditions*¹

¹<https://creativecommons.org/licenses/by-nc-sa/4.0/>

D'où je parle : parcours personnel

+25 ans sur la sûreté des systèmes
embarqués critiques, c-à-d le numérique
dans :



≈ 8 ans sur impacts
socio-environnementaux du numérique,
et même trajectoires décroissantes du
numérique

Intérêt pour les LLM depuis 2020 ^{2, 3, 4}

Le danger des **perroquets stochastiques** ;
les modèles de langue peuvent-ils être trop
grands?

L'arnaque de l'IA

THE AI CON

[Home](#) [Authors](#) [Events](#) [News](#) [Praise](#) [Contact](#)

[Submit Fresh AI Hell!](#)

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these pretrained models and the methodology of fine-tuning them for specific tasks, researchers have extended the state of the art on a wide array of tasks as measured by leaderboards on specific benchmarks for English. In this paper, we take a step back and ask: How big is too big? What are the possible risks associated with this

alone, we have seen the emergence of BERT and its variants [39, 76, 74, 113, 146], GPT-2 [106], T-NLG [112], GPT-3 [25], and most recently Switch-C [45], with institutions seemingly competing to produce ever larger LMs. While investigating properties of LMs and how they change with size holds scientific interest, and large LMs have shown improvements on various tasks [82], we ask whether **enough thought has been put into the potential risks associated with developing them and strategies to mitigate these risks.**

We first consider environmental risks. Echoing a line of recent work outlining the environmental and financial costs of deep learning systems [128], we encourage the research community to priori-



How to Fight Big Tech's Hype and Create the Future We Want

A smart, incisive look at the technologies sold as artificial intelligence, the drawbacks and pitfalls of technology sold under this banner, and why it's crucial to recognize the many ways in which AI hype covers for a small set of power-hungry actors at work and in the world.

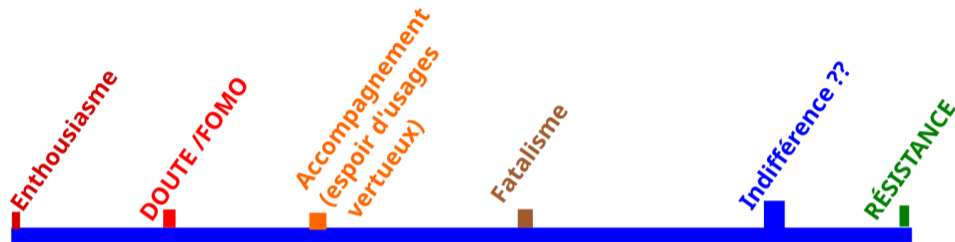
► Read more...

² <https://dl.acm.org/doi/10.1145/3442188.3445922>

³ <https://thecon.ai/>

⁴ <https://www.youtube.com/watch?v=LJmkPILudYY>

Fenêtre d'Overton de l'IAg



Voir “Pourquoi je n'utilise pas ChatGPT ?” (02/25) ⁵
et “Face à l'IA générative, l'objection de conscience” (11/25) ^{6, 7}.

⁵ <https://academia.hypotheses.org/58766>

⁶ <https://atecopol.hypotheses.org/13082>

⁷ https://www.lemonde.fr/campus/article/2026/03/01/face-a-l-ia-generative-des-enseignants-optent-pour-l-objection-de-conscience_6668765_4401467.html

Résumé de l'exposé

- Les impacts environnementaux du numérique étaient déjà sur une trajectoire intenable avant les IAg, c'est pire depuis
- Les autres aspects des IAg sont rédhibitoires : idéologie des vendeurs, besoin de “petites mains”, impacts sur le travail et les apprentissages, accélération, pollution informationnelle, privatisation des communs de la connaissance...
- Peut-on utiliser des IA génératives dans un contexte recherche ? **Non.**
- Est-ce contraignant / pénalisant ? **A moyen terme, non, au contraire**

- 1 IA génératives (de textes) : de quoi parle-t-on ?
- 2 Quelques-unes des motivations d'une position de résistance
- 3 Conclusion

LLM (grands modèles de langue)

- un LLM est un **modèle** (donc une abstraction) de langue naturelle, construit pour prédire le mot suivant (*“perroquet stochastique”*)
- un chatbot est la **mise en forme de dialogue** (attention à l’anthropomorphisation, ce n’est pas un dialogue) de l’usage d’un LLM
- tout ça construit par **apprentissage machine** avec des réseaux de neurones, sur de très grandes masses de textes **pillées sur le web et dans des livres physiques**
- dans des **datacentres** remplis de circuits spécialisés (les **GPU** de Nvidia) qui consomment beaucoup d’électricité et chauffent beaucoup, donc refroidissement (eau), d’où **conflits d’accès aux ressources et luttes locales (Marseille, Eybens, Aragón, Chili, USA, ...)**
- il faut des armées de petites mains pour trier les corpus de textes (**travail du clic abrutissant, voire traumatisant**), améliorer les résultats (forme, contenu, impression de dialogue, ...)

Opacité des outils et évolutions récentes :

Ce ne sont plus seulement des machines à faire des phrases

- Couplage avec un outil de recherche d'informations ciblées et récentes (Retrieval Augmented Generation, **RAG** ⁸)
- **"Routeurs"** qui orientent vers différents types d'outils en interne.
- **"Agents IA"** (attention, style pub enthousiaste ici ⁹)

⁸ https://fr.wikipedia.org/wiki/G%C3%A9n%C3%A9ration_%C3%A0_enrichissement_contextuel

⁹ https://fr.wikipedia.org/wiki/Agent_intelligent

Mes réflexes “systèmes critiques” : comment est-ce évalué et par qui ?

- Comment **spécifier** ce qu'on attend ?
- Comment vérifier que le système informatique **fait ce qu'on attend de lui** ?
L'obligation de preuve est du côté des vendeurs.
- Qui vérifie ? ou au moins audite la démarche de développement ?

- Qui a encore les moyens de faire de la recherche sur les LLM ?

“A few companies account for most of the publications and provide funding to academic researchers through grants and internships. (...) This work calls for increased transparency of industry influence in the field.” ¹⁰

¹⁰<https://aclanthology.org/2023.acl-long.734/>

1 IA génératives (de textes) : de quoi parle-t-on ?

2 Quelques-unes des motivations d'une position de résistance

- Les impacts environnementaux additionnels sont rédhibitoires
- Les IAg cachent une armée de travailleurs du clic traumatisés
- Les IAg dégradent et privatisent l'espace informationnel
- L'appât du gain de temps produit une accélération absurde
- A propos des LLM fiables/sobres/éthiques/souverains/... ?
- Les dégâts sont déjà observables en recherche

3 Conclusion

- 2 Quelques-unes des motivations d'une position de résistance
- Les impacts environnementaux additionnels sont rédhibitoires
 - Les IA_g cachent une armée de travailleurs du clic traumatisés
 - Les IA_g dégradent et privatisent l'espace informationnel
 - L'appât du gain de temps produit une accélération absurde
 - A propos des LLM fiables/sobres/éthiques/souverains/... ?
 - Les dégâts sont déjà observables en recherche

Mythe de la dématérialisation ou matérialité du numérique (mines)

☰ Impact environnemental de l'industrie minière

🌐 14 langues ▾

Article Discussion Lire Modifier Modifier le code Voir l'historique Outils ▾

Les **impacts environnementaux de l'industrie minière** se font ressentir à l'échelle locale, régionale et mondiale. L'extraction minière peut provoquer [l'érosion des sols](#), la [formation de dolines](#), la [perte de biodiversité](#) ou la [contamination des sols](#), des [eaux souterraines](#) et des [eaux de surface](#) par les produits chimiques utilisés lors



11

¹¹ https://fr.wikipedia.org/wiki/Impact_environmental_de_l'industrie_minire

Mythe de la dématérialisation ou matérialité du numérique (fabrication des circuits)



12

¹²<https://www.lemondeinformatique.fr/actualites/lire-la-mega-usine-stmicro-et-globalfoundries-a-crolles-se-precise-90619.html>

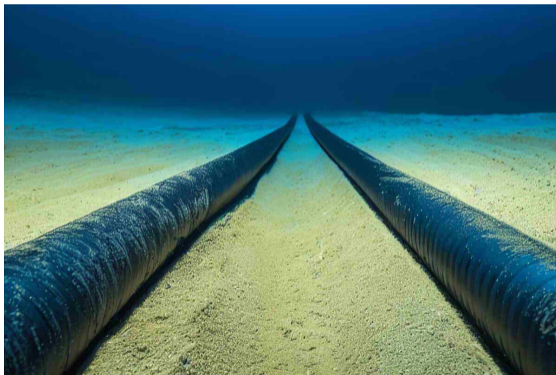
Mythe de la dématérialisation ou matérialité du numérique (transport)



13

¹³ <https://www.rtbf.be/article/le-port-d-anvers-accueille-le-plus-grand-porte-conteneurs-du-monde-10520932>

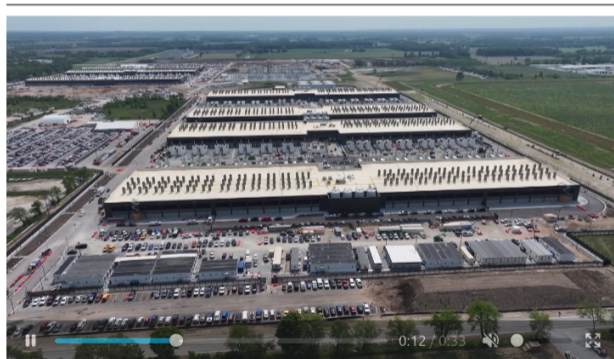
Mythe de la dématérialisation ou matérialité du numérique (câbles sous-marins, satellites) ¹⁴ ¹⁵



¹⁴ <https://zmscable.es/fr/medusa-sistema-cable-submarino/>

¹⁵ <https://en.wikipedia.org/wiki/Starlink>

Mythe de la dématérialisation ou matérialité du numérique (datacenters)



16

16 <https://www.nytimes.com/2025/06/24/technology/amazon-ai-data-centers.html>

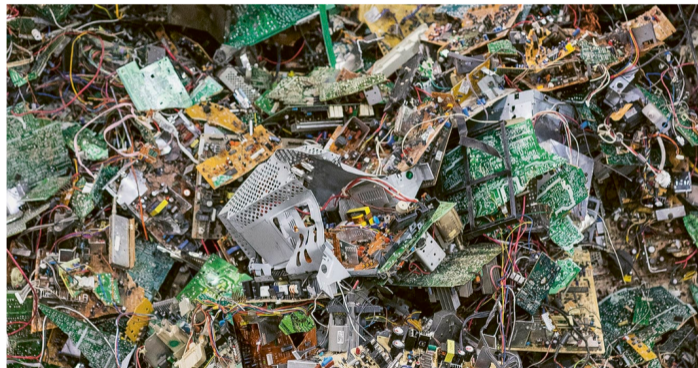
Mythe de la dématérialisation ou matérialité du numérique (production d'électricité)



17

17 <https://www.edf.fr/groupe-edf/comprendre/production/nucleaire>

Mythe de la dématérialisation ou matérialité du numérique (déchets)



18

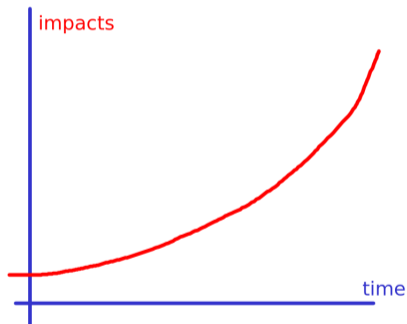
¹⁸ <https://revuehemispheres.ch/repenser-notre-rapport-aux-dechets-numeriques/>

Impacts environnementaux globaux du numérique ¹⁹ ²⁰ : quand et comment cela va-t-il s'arrêter ?

Seulement les gaz à effet de serre :

- Quelques % des émissions totales
- En croissance (dont IA)

+ Autres impacts
(biodiversité, eau, mines...)



¹⁹ <https://www.sciencedirect.com/science/article/pii/S2666389921001884>

²⁰ <https://ia2-2025.imag.fr/ressources/CroissanceIA-Aussois.pdf>

Quels impacts environnementaux additionnels des IAg ?

CO₂ et réchauffement, eau, biodiversité, pollutions, ... : **on redresse l'exponentielle.**

- Fabrication et déchets (très rapides) des circuits spécialisés
- Alimentation en électricité pour la fabrication des LLM et leur utilisation
- Refroidissement des datacentres (eau ²¹ ou électricité)
- Effets d'accélération de secteurs non numériques

²¹https://hal.science/hal-04698568v1/file/De_l_eau_dans_les_nuages.pdf

Attention aux indicateurs d'impacts *par requête* et issus de *moyennes globales*

1 requête ChatGPT = 1/2 ℓ d'eau, 1 requête ChatGPT c'est 10x requête Google en électricité, 1/15 de cuillère à thé d'eau, ... 20s de vidéo, 12.86g de CO₂... ~~1 requête ChatGPT = 1/2 ℓ d'eau, 1 requête ChatGPT c'est 10x requête Google en électricité, 1/15 de cuillère à thé d'eau, ... 20s de vidéo, 12.86g de CO₂...~~

- Raisonner localement sur l'eau (*stress hydrique*)
- Questionner la construction des chiffres
- Penser infrastructure globale plutôt qu'utilisation unitaire ²²
- Attention aux indicateurs choisis par les bigtech (que penserions-nous de l'industrie du tabac disant "*une cigarette n'augmente le risque de maladie que de 0.000001% pour les enfants*" ?)

²² <https://ketanjoshi.co/2025/08/23/big-techs-selective-disclosure-masks-ais-real-climate-impact/>



CANDRIAM
BY NEW YORK LIFE INVESTMENT MANAGEMENT

Le crédit peut être sauvage.

Domptez le risque. Ne le fuyez pas. Notre approche >

TRIBUNE

Explosion de l'IA : « La France devra arbitrer entre alimenter ses data centers et décarboner ses transports et son chauffage » 🇫🇷

La consommation électrique des centres de données explose avec l'essor de l'IA. Face à cette tendance, le projet de Stratégie nationale bas carbone 3 pose des chiffres sur la table, mais peine encore à proposer des réponses concrètes, analysent Hugues Ferreboeuf et Pauline Denis, respectivement chef de projet et ingénieure de recherche numérique au sein du think tank The Shift Project.

Domptez le risque. Ne le fuyez pas. Notre approche >

CANDRIAM

- 2 Quelques-unes des motivations d'une position de résistance
- Les impacts environnementaux additionnels sont rédhibitoires
 - Les IA_g cachent une armée de travailleurs du clic traumatisés
 - Les IA_g dégradent et privatisent l'espace informationnel
 - L'appât du gain de temps produit une accélération absurde
 - A propos des LLM fiables/sobres/éthiques/souverains/... ?
 - Les dégâts sont déjà observables en recherche

Travail du clic ²⁵, ²⁶ pour ChatGPT et autres Chatbots LLM

Eviter que des chatbots produisent des discours racistes/sexistes/haineux/choquants... repose sur la construction de filtres automatiques, par apprentissage **supervisé**, donc grâce à des êtres humains qui sont exposés aux discours racistes/sexistes/haineux/choquants... toute la journée.

- Enquête sur les travailleurs du clic au Kenya, pour ChatGPT ²³
- Le documentaire “Les sacrifiés de l’IA” ²⁴

²³ <https://theconversation.com/>

[forcats-du-numerique-comment-une-decision-de-justice-au-kenya-fragilise-la-sous-traitance-des-multinationales-du-web-206896](https://www.france.tv/documentaires/documentaires-societe/6888928-les-sacrifies-de-l-ia.html)

²⁴ <https://www.france.tv/documentaires/documentaires-societe/6888928-les-sacrifies-de-l-ia.html>

²⁵ <https://www.casilli.fr/tag/diplab/>

²⁶ <https://cacm.acm.org/news/the-invisible-labor-force-powering-ai/>

LISTEN TO THE 404 MEDIA PODCAST



404

SIGN OUT

ACCOUNT

ABOUT RBS SUPPORT/FAQ PODCAST FOIA FORUM ARCHIVE MERCH ADVERTISE THANKS PRIVACY

AI

'AI Is African Intelligence': The Workers Who Train AI Are Fighting Back

Kenyan workers are still the underpaid labor behind AI training, moderation, and sex chatbots. The Data Labelers Association is fighting back.

JASON KOEBLER · MAR 12, 2025 AT 11:08 AM



27

27 <https://www.404media.co/ai-is-african-intelligence-the-workers-who-train-ai-are-fighting-back/>

- 2 Quelques-unes des motivations d'une position de résistance
- Les impacts environnementaux additionnels sont rédhibitoires
 - Les IA_g cachent une armée de travailleurs du clic traumatisés
 - Les IA_g dégradent et privatisent l'espace informationnel
 - L'appât du gain de temps produit une accélération absurde
 - A propos des LLM fiables/sobres/éthiques/souverains/... ?
 - Les dégâts sont déjà observables en recherche

Pollution informationnelle ²⁸

NEXT S3 1.5 To

[Récap] Nous avons découvert des milliers de sites d'info générés par IA : tous nos articles

« Mais tu écris n'importe quoi, en fait »



Entamée début 2024, notre enquête sur les « sites d'information » générés par IA (mais qui ne le précisent pas) nous avait permis d'en identifier 250 en octobre 2024, 500 à Noël, 1 000 fin janvier 2025, près de 3 000 début mars, 6 000 en août, 9 000 en décembre... et plus de 10 000 depuis février 2026. Cette page résume la quarantaine d'articles que nous y avons consacrés depuis février 2025.

²⁸ <https://next.ink/173214/recap-nous-avons-decouvert-des-milliers-de-sites-dinfo-generes-par-ia-tous-nos-articles/>

Cas de la traduction non humaine

Pollution de Wikipedia

AI Translations Are Adding ‘Hallucinations’ to Wikipedia Articles ²⁹

*“AI translated articles swapped sources or added unsourced sentences with no explanation, while others **added paragraphs sourced from completely unrelated material.**”*

²⁹ <https://www.404media.co/ai-translations-are-adding-hallucinations-to-wikipedia-articles/>

Reproduction et amplification des biais

“... As people in positions of privilege with respect to a society’s *racism, misogyny, ableism, etc.*, tend to be *overrepresented in training data* for LMs (...), this training data thus includes encoded biases, many already recognized as harmful” – Extrait de l’article “Stochastic Parrots” ³⁰.

Voir aussi Karën Fort ³¹

— conférence au colloque Néala’25 sur *l’amplification* des biais ³²

— You’ll be a nurse, my son! Automatically Assessing Gender Biases in Autoregressive Language Models in French and Italian

³⁰ <https://dl.acm.org/doi/10.1145/3442188.3445922>

³¹ <https://members.loria.fr/KFort/publications/hal/>

³² <https://www.youtube.com/watch?v=981Qmx0z0G0>

Non représentation de certaines langues et cultures ³³

Le contenu du web n'est qu'une toute petite portion de la production humaine

"the imbalance between how many people speak a language in the physical world and how much that language is represented in online data."

³³<https://aeon.co/essays/generative-ai-has-access-to-a-small-slice-of-human-knowledge>

Attaques par déni de service sur les architectures de gestion de communs

Les robots des IAq qui parcourent le web écroulent les serveurs (Humanum, OpenEditions, la BNF, ..)

À rebours de l'ouverture (science ouverte, open data, ...)
on réinstalle des barrières pour se défendre.

Privatisation des communs de l'information

- Le web se “referme” par un mécanisme d'*enclosure* ³⁴
- Risques spécifiques sur l'écosystème du logiciel libre ³⁵, exemple du forum StackOverflow
- ARCEP janvier 2026 *IA générative : des défis pour l'avenir de l'internet ouvert* ³⁶

“... préserver la richesse de l'internet, où chaque utilisateur conserve sa *liberté de choix* (...) où la *diversité* des contenus et des expériences numériques est préservée. L'ouverture de l'internet ne se limite pas à la couche des réseaux télécoms : elle doit s'étendre à toutes les portes d'entrée à internet, y compris celles que *l'IA pourrait un jour fermer à clé.*”

³⁴ <https://michiël.buddingh.eu/enclosure-feedback-loop>

³⁵ <https://next.ink/222770/le-vibe-coding-met-en-danger-lopen-source-selon-des-economistes/>

³⁶ https://www.arcep.fr/uploads/tx_gspublication/rapport-IA-generative-defis-internet-ouvert-janvier2026.pdf

- 2 Quelques-unes des motivations d'une position de résistance
- Les impacts environnementaux additionnels sont rédhibitoires
 - Les IA_g cachent une armée de travailleurs du clic traumatisés
 - Les IA_g dégradent et privatisent l'espace informationnel
 - **L'appât du gain de temps produit une accélération absurde**
 - A propos des LLM fiables/sobres/éthiques/souverains/... ?
 - Les dégâts sont déjà observables en recherche

Gain de temps ou accélération jusqu'à l'absurdité ?



2 Quelques-unes des motivations d'une position de résistance

- Les impacts environnementaux additionnels sont rédhibitoires
- Les IA_g cachent une armée de travailleurs du clic traumatisés
- Les IA_g dégradent et privatisent l'espace informationnel
- L'appât du gain de temps produit une accélération absurde
- **A propos des LLM fiables/sobres/éthiques/souverains/... ?**
- Les dégâts sont déjà observables en recherche

Est-ce que c'est seulement possible ?

- L'indépendance ("souveraineté") impossible du numérique (conception et fabrication de circuits, logiciel, ...) ³⁷
- Outils construits sans pillage de données ? ³⁸
- Peut-on se passer des travailleurs du clic ?
- Modèle économique et rentabilité ?

37

<https://theconversation.com/semi-conducteurs-lindependance-technologique-ne-se-limite-pas-a-la-fabrication-209637>

38

<https://www.mediapart.fr/journal/economie-et-social/230226/>

[harry-potter-le-petit-prince-elton-john-ou-amel-bent-les-multiples-infractions-au-droit-d-aute](#)

Cadres d'usages responsables/éthiques/... pour l'ESR

Une affirmation gratuite

Exemple du guide de l'INRAE ³⁹ :

- *“Les assistants IAG sont capables de rechercher et synthétiser des informations, à partir de requêtes formulées en langage naturel, sans que l'utilisateur ait besoin de maîtriser le vocabulaire du domaine. (...) ces outils innovants peuvent générer des réponses synthétiques et personnalisées, plutôt que de simples listes de liens. ”*

³⁹ <https://hal.inrae.fr/hal-05382841v1>

Cadres d'usages responsables/éthiques/... pour l'ESR

Une contradiction de fond

Exemple du guide de l'INRAE ⁴⁰ :

- *“Qualité et pertinence des résultats : les résultats générés dépendent de la performance de l'IAG et de ses données d'entrée (quantité, qualité, fraîcheur, etc.). Des résultats **biaisés, incomplets ou même erronés** peuvent être obtenus.”*
- *“Notez bien que l'utilisateur est **seul responsable devant la loi** des contenus qu'il produit à l'aide de l'IAG et du respect de la réglementation relative à ses usages.”*
- *“il est indispensable de **croiser les sources et vérifier** les informations obtenues”*

⁴⁰ <https://hal.inrae.fr/hal-05382841v1>

Comment croiser les sources si perte de redondance ?

Confronter les productions des IA_g à d'autres sources (aka résultats de recherches "sur le web") ; mais que va devenir l'information sur le web si :

- elle est envahie par des textes générés par IA ?
- elle est privatisée ("enclosure") ?
- elle disparaît par masquage
(les moteurs de recherche donnent des résumés IA_g)⁴¹ ?

⁴¹<https://affordance.framasoft.org/2026/02/de-google-books-a-lia-lhistoire-numerique-repasse-les-plats/>

2 Quelques-unes des motivations d'une position de résistance

- Les impacts environnementaux additionnels sont rédhibitoires
- Les IA_g cachent une armée de travailleurs du clic traumatisés
- Les IA_g dégradent et privatisent l'espace informationnel
- L'appât du gain de temps produit une accélération absurde
- A propos des LLM fiables/sobres/éthiques/souverains/... ?
- Les dégâts sont déjà observables en recherche

Productivité et embouteillages ⁴², ⁴³

How AI is transforming research: More papers, less quality, and a strained review system – Mathijs De Vaan, Toby E. Stuart

Encore une histoire de pollution informationnelle :

*“AI is rapidly reshaping scientific research. Even as AI tools help researchers write more papers faster, many of these studies are of marginal scientific merit. The resulting flood of polished but potentially superficial work is making it harder for reviewers, funders, and policymakers to **separate worthy papers from unimportant and potentially misleading work.**”*

⁴²<https://newsroom.haas.berkeley.edu/how-ai-is-transforming-research-more-papers-less-quality-and-a-strained-review-system/>

⁴³<https://www.science.org/doi/10.1126/science.adw3000>

Agriculture/recherche — intensive/extensive ⁴⁴

AI research assistants, intrinsic values, and the science we want – Curmudgeon Corner – Published: 01 February 2024, Volume 40, pages 235–237, (2025)

*“Do we want a science of texts (partially) automated by machines that other machines then process and summarize and in which **no one reads what we write?** Do we want an **intensive**, instrumental science **optimized for quantity and efficiency** or, in contrast, do we prefer to imagine and pursue an **extensive** science that seeks **value, meaning, and depth**, and which is not just an instrument for something but rather an end in itself? ”*

⁴⁴ <https://link.springer.com/article/10.1007/s00146-023-01861-4>

Usage pour revue de littérature (résumés ?) ⁴⁵

INFORMATION LITERACY ARTICLE

Are AI-generated summaries suitable for studying and research?

FEBRUARY 24, 2026

Despite didactic, ethical, and environmental concerns, the use of GenAI is on the rise in academia. For most applications, the jury is still out on whether and how they will benefit education and research in the long term. But it's already safe to conclude that one popular use case is, in fact, a bad one: AI-generated summaries.

⁴⁵ <https://www.tue.nl/en/our-university/library/library-news/24-02-2026-are-ai-generated-summaries-suitable-for-studying-and-research>

- 1 IA génératives (de textes) : de quoi parle-t-on ?
- 2 Quelques-unes des motivations d'une position de résistance
- 3 Conclusion

Nous avons un problème collectif

Quels que soient les avantages qu'on peut trouver à l'usage de ces objets numériques, à quelles évolutions de nos contextes de travail, quelles dégradations des conditions-mêmes d'exercice de ce travail, participe-t-on ?

- Pour la recherche
 - Avantages individuels à court terme vs dégâts collectifs à moyen terme
- Plus largement
 - Impacts environnementaux intenable et responsabilité de la recherche
 - Risques : dépendance accrue au numérique / à des systèmes privés et privatifs

Fin. Merci de votre attention

