

Enjeux éthiques de l'Intelligence Artificielle générative pour la recherche scientifique

Raja Chatila
Institut des Systèmes Intelligents et de Robotique (ISIR)
Sorbonne Université, Paris

Raja.Chatila@sorbonne-universite.fr

Pourquoi l'IA générative pour la recherche?

- **Optimisation** du processus de recherche scientifique
 - Analyse et synthèse de documents (état de l'art, ...)
 - Production de **données synthétiques**
 - **Rédaction** de publications, de propositions de projets, de rapports
 - **Programmation**
 - **Evaluation** d'articles ou de projets

Gain de temps et multiplication de la production scientifique



AAAI Launches AI-Powered Peer Review Assessment System

May 16, 2025

Washington, DC — The Association for the Advancement of Artificial Intelligence (AAAI), a leading nonprofit dedicated to advancing scientific research and collaboration, today announced a pilot program that strategically incorporates Large Language Models (LLMs) to enhance the academic paper review process for the AAAI-26 conference. This initiative aims to improve efficiency while maintaining the highest standards of scientific rigor and human oversight.

Enhancing Scientific Review, Not Replacing Human Expertise

The pilot program will thoughtfully integrate LLM technology at two specific points in the established review process:

1. **Supplementary First-Stage Reviews:** LLM-generated reviews will be included as one component of the initial review stage, providing an additional perspective alongside traditional human expert evaluations.
2. **Discussion Summary Assistance:** LLMs will assist the Senior Program Committee (SPC) members by summarizing reviewer discussions, helping to highlight key points of consensus and disagreement among human reviewers.

“This pilot represents a careful, measured approach to incorporating new technology into the scientific review process,” said Stephen Smith, AAAI President. “We’re exploring how LLMs can complement—not replace—the irreplaceable expertise and judgment of our human reviewers.”

Limitations

Liam wants to buy some school supplies. He buys 24 erasers that now cost \$6.75 each, 10 notebooks that now cost \$11.0 each, and a ream of bond paper that now costs \$19. How much should Liam pay now, assuming that due to inflation, prices were 10% cheaper last year?

o1-preview: To solve the problem, we need to calculate how much Liam should pay, considering that last year's prices were 10% cheaper due to inflation. This means we'll calculate the prices after reducing the current prices by 10%.

Raisonnement?

Step 1: Calculate last year's prices by reducing the current prices by 10%.

- Eraser price last year: $\$6.75 \times 0.90 = \6.075
- Notebook price last year: $\$11.00 \times 0.90 = \9.90
- Bond paper price last year: $\$19.00 \times 0.90 = \17.10

Step 2: Calculate the total cost using last year's prices.

- Total cost for erasers: $24 \text{ erasers} \times \$6.075 = \$145.80$
- Total cost for notebooks: $10 \text{ notebooks} \times \$9.90 = \$99.00$
- Total cost for bond paper: $1 \text{ ream} \times \$17.10 = \17.10

Step 3: Sum up all the costs to get the total amount Liam should pay.

$$\$145.80 + \$99.00 + \$17.10 = \$261.90$$

Answer: \$261.90

Un an après (ChatGPT 5.1)

John wants to buy some items. He buys 24 erasers that now cost \$6.75 each, 10 notebooks that now cost \$11.0 each, and a ream of bond paper that now costs \$19. How much should John pay now, assuming that due to inflation, prices were 10% cheaper last year?

Let's compute John's total **current** cost.

He buys:

- 24 erasers @ \$6.75 each
 $24 \times 6.75 = 162$
- 10 notebooks @ \$11.00 each
 $10 \times 11 = 110$
- 1 ream of bond paper @ \$19.00
19

Now add them:

$$162 + 110 + 19 = \boxed{291}$$

Final Answer: John should pay \$291.

*(The note about prices being 10% cheaper last year is just background—since we're asked how much he pays **now**, we use the current prices given.)*

Chute de performance devant la complexité

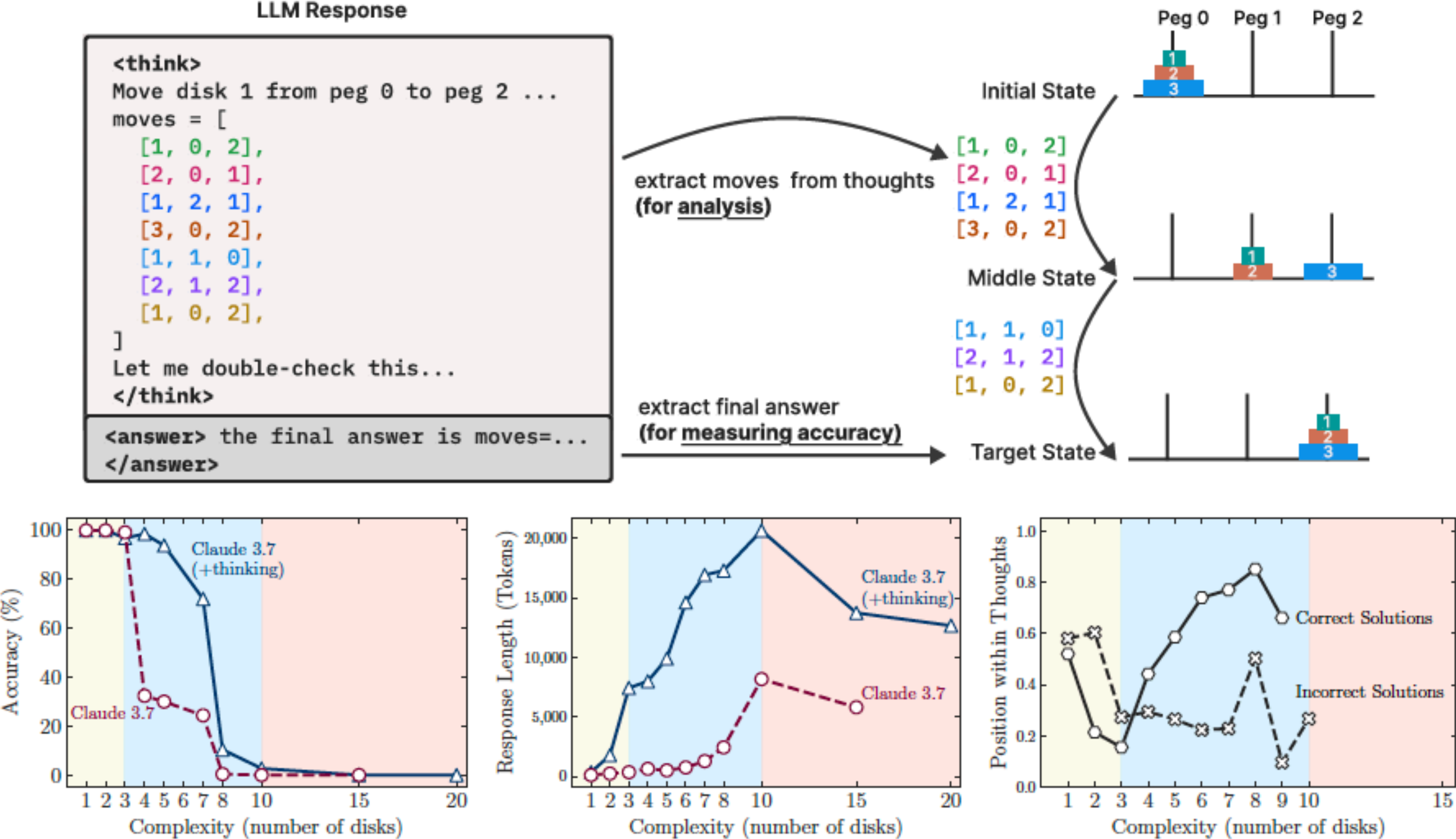


Figure 1: Top: Our setup enables verification of both final answers and intermediate reasoning traces, allowing detailed analysis of model thinking behavior. Bottom left & middle: At low complexity, non-thinking models are more accurate and token-efficient. As complexity increases, reasoning models outperform but require more tokens—until both collapse beyond a critical threshold, with shorter traces. Bottom right: For correctly solved cases, Claude 3.7 Thinking tends to find answers early at low complexity and later at higher complexity. In failed cases, it often fixates on an early wrong answer, wasting the remaining token budget. Both cases reveal inefficiencies in the reasoning process.

Parshin Shojaei *et al.* (2025) <https://ml-site.cdn-apple.com/papers/the-illusion-of-thinking.pdf>

Questions sur les systèmes d'IA générative

- Opacité; peu d'explicabilité
- Absence d'ancrage dans le monde réel
- Absence de signification, de compréhension des concepts
- Limites de la généralisation
- Absence de causalité et de raisonnement (malgré les apparences)
- Corrélations fallacieuses ("hallucinations")
- Imprévisibilité (sensibilité à la formulation des entrées) et non-reproductibilité
- Mélange indiscernable d'informations vraies et fausses
- Impacts environnementaux

IA générative et recherche

Exemples de pratiques de recherche pour lesquelles l'IA générative peut être utilisée (Ofis)

- Définition de la question de recherche ou de l'hypothèse de départ
- Développement et argumentation de l'idée originale
- Sélection ou création de la méthodologie de recherche
- Assistance à la réalisation d'expérience, traitement d'images
- Résumé d'articles, support à la réalisation d'états de l'art
- Ecriture et optimisation de codes informatiques
- Rédaction, traduction, formatage de textes et d'images
- Réalisation de bibliographie, sélection d'articles
- Evaluation des projets ou des manuscrits, identification de pairs

Recommandations de l'Ofis

- Principes fondamentaux
 - Contrôle humain et responsabilité
 - Transparence
- Points de vigilance pour l'intégrité scientifique
 - Fiabilité
 - Propriété (intellectuelle)
 - Confidentialité

Exigences pour une IA de confiance

Groupe d'Experts de haut niveau sur l'IA (UE) Avril 2019

Fondement éthique du règlement européen sur l'IA



1. **Préservation de l'action et de la supervision humaine**
2. **Robustesse technique et sécurité** : résilience et sûreté de fonctionnement, sécurité, précision, fiabilité, reproductibilité
3. **Respect de la vie privée et gouvernance des données**: qualité et intégrité des données
4. **Transparence**: traçabilité, explicabilité, communication
5. **Diversité non-discrimination et équité**: absence de biais injustes, accessibilité, participation des parties prenantes
6. **Bien-être social et environnemental**: durabilité, respect de l'environnement, impact social, démocratie
7. **Responsabilité** : auditabilité, réduction des incidences négatives, communication, arbitrages, recours.

<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

Recommandations aux chercheurs (Commission UE)

1. Remain ultimately responsible for scientific output.
2. Use generative AI transparently.
3. Pay particular attention to issues related to privacy, confidentiality and intellectual property rights when sharing sensitive or protected information with AI tools.
4. Respect applicable national, EU and international legislation
5. Continuously learn how to use generative AI tools properly to maximise their benefits, including by undertaking training.
6. Refrain from using generative AI tools substantially in sensitive activities that could impact other researchers or organisations (for example peer review, evaluation of research proposals, etc).



[Living Guidelines on the Responsible Use Of Generative AI In Research](https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf)

https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf

Recommendations aux institutions (Commission UE)

1. Promote, guide and support the responsible use of generative AI in research activities.
2. Actively keep track of the evolution and use of generative AI systems within their organisations.
3. Reference or integrate these generative AI guidelines into their general research guidelines for good research practices and ethics.
4. Whenever possible and necessary, implement locally hosted or cloud-based generative AI tools that they govern themselves (confidentiality, data protection, security)



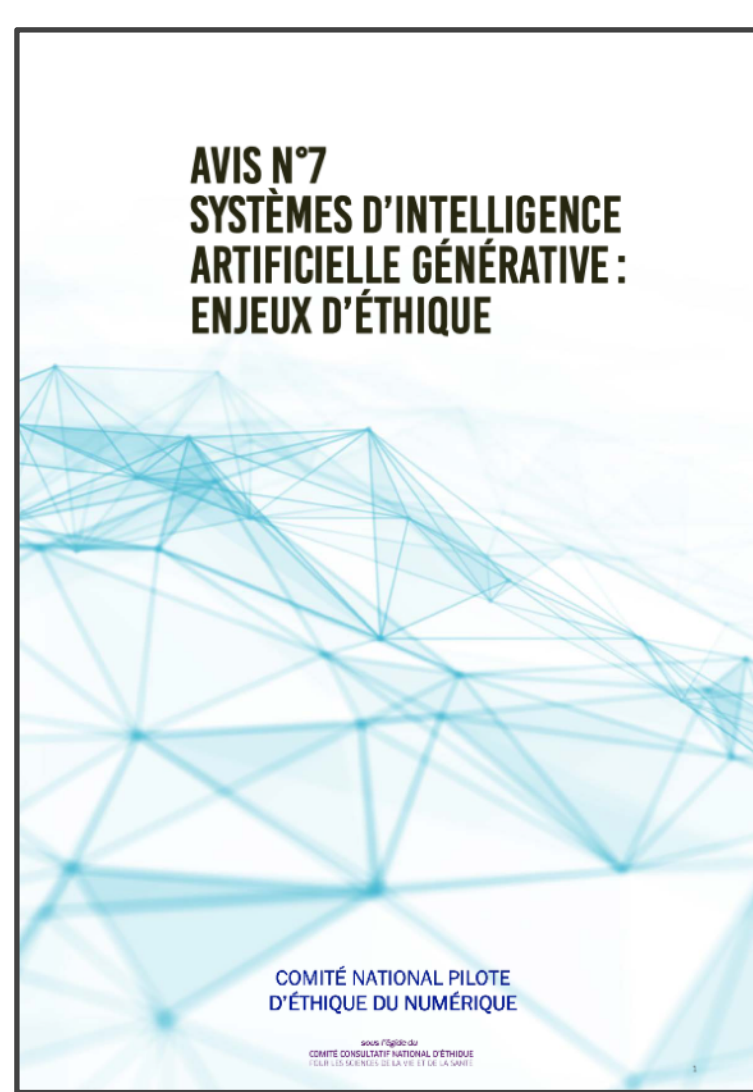
Living Guidelines on the Responsible Use OF Generative AI In Research

https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf

Exemple: Utilisation de l'IA générative dans des universités danoises

- Etude via un questionnaire (2 534 réponses, 32 cas d'utilisation) aux chercheurs dans cinq phases de recherche : génération d'idées, conception de la recherche, collecte de données, analyse de données, rédaction et rapports.
- Révision linguistique, aide à la rédaction, analyse des données: **opinions positives**
- Conception d'expériences et évaluation par les pairs: **opinion négatives**
- Création/modification d'images et données synthétiques: **controversés**
- **Nécessité de lignes directrices spécifiques à chaque discipline**

Andersen, J. P *et al* (2025). Generative Artificial Intelligence (GenAI) in the research process – A survey of researchers' practices and perceptions. *Technology in Society*, 81, 102813



Préconisation G3 : Mutualisation des pratiques des systèmes d'IA générative

- Il est nécessaire de bâtir un écosystème capable de recenser les bonnes et mauvaises pratiques en matière d'utilisation des systèmes d'IA générative dans différents types d'applications. Notamment, il est nécessaire de créer une plateforme de mutualisation et une agence de contrôle. Les résultats doivent être mis à disposition de tous les membres de la communauté d'IA générative.

Juin 2023

Risques pour la recherche

- Mise en cause de la méthode scientifique
- Risque de recherche incrémentale
- Risque d'utilisation de concepts erronés et de production de résultats erronés
- Risque de plagiat et de violation de la propriété intellectuelle
- Risque de divulgation de données personnelles
- Responsabilité légale et morale concernant les résultats
- Obligations de transparence ; mécanisme de détection/marquage des résultats générés (règlement européen sur l'IA)

R. Chatila, C. Kirchner, S. Petitjean, C. Tessier

Intelligence artificielle : quels enjeux éthiques pour la recherche ?

In Une recherche responsable: Ethique de la recherche scientifique. Edition Quae, A paraître (2026)

Production et utilisation de résultats

- Attribution d'auteur à un document principalement produit par un système d'IA ?
 - Qualité d'un résumé ou d'une évaluation par IA d'un texte scientifique?
 - Qualité et réalité des références?
 - Production et utilisation de données synthétiques?
- Biais selon choix des données
 - Perte de représentativité de la distribution des données: incomplétude, simplification
 - Perte de caractéristiques temporelles
 - Validation
 - ...

R. Chatila, C. Kirchner, S. Petitjean, C. Tessier

Intelligence artificielle : quels enjeux éthiques pour la recherche ?

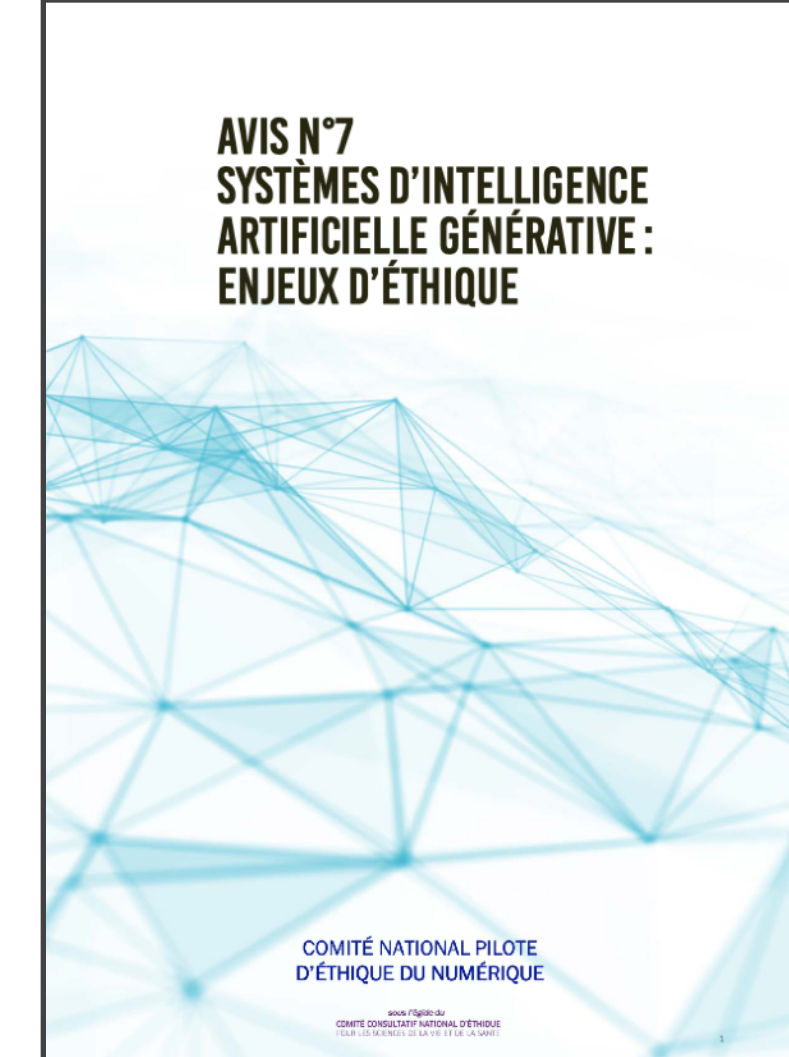
In Une recherche responsable: Ethique de la recherche scientifique. Edition Quae, A paraître (2026)

CNPEN Avis N°7

Préconisation C3 : Utilisation de sources de qualité pour l'apprentissage.

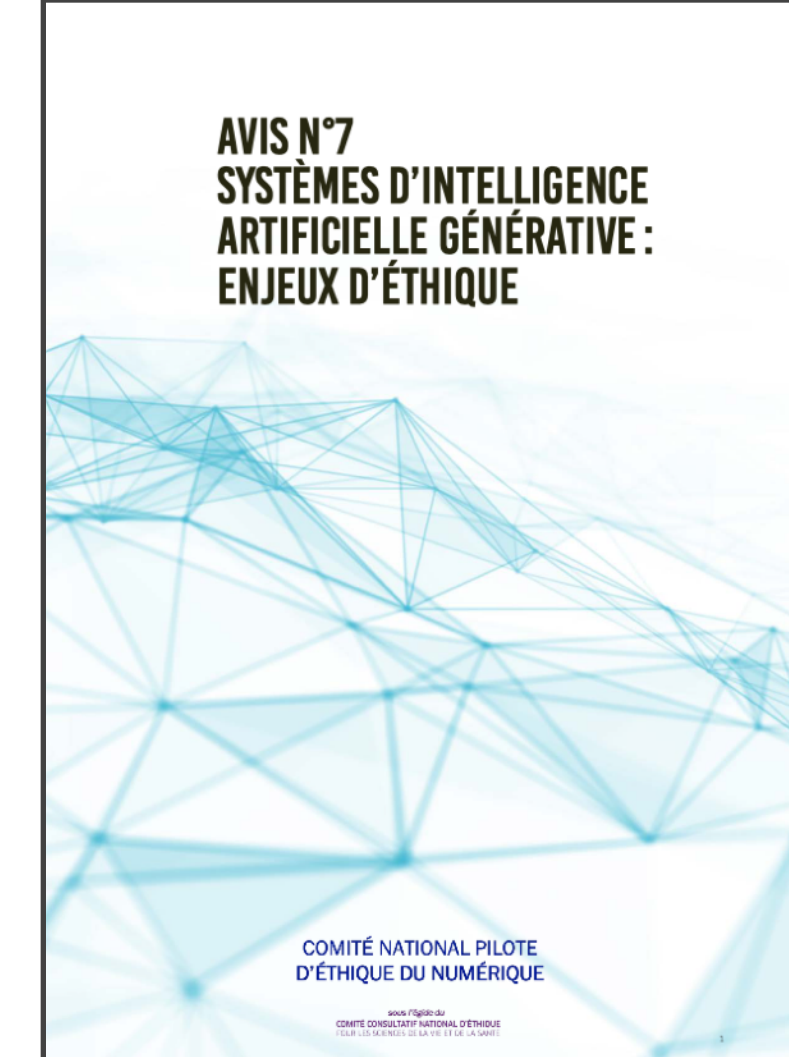
- Les concepteurs doivent privilégier l'usage de sources de qualité, jugées selon des critères rendus explicites, pour la constitution et l'usage des corpus d'apprentissage des modèles d'IA générative (pré-entraînement), ainsi que pour leur optimisation, et ce quelle que soit la méthode d'apprentissage.
- Notamment, il est nécessaire de s'interroger sur la transparence et la finalité des **contenus artificiels ou synthétiques** dans les corpus d'apprentissage.

Préconisations C6 et G4 : transparence et maintien des *distinctions* (données produites/non produites par IA) et filigrane réglementaire



Juin 2023

CNPEN Avis N°7



Juin 2023

Préconisation G8: Chaîne de responsabilité.

- La responsabilité légale sur les systèmes d'IA générative et les modèles de fondation doit être attribuée aux fournisseurs des modèles de fondation et aux déployeurs d'applications spécifiques d'IA générative à partir de tels modèles. De plus, la responsabilité morale s'étend aux concepteurs des modèles de fondation et aux développeurs des systèmes d'IA générative utilisant de tels modèles

La responsabilité des utilisateurs reste engagée dans la manière dont ils utilisent les résultats

Challenge pour les chercheurs: s'adapter à un outil opaque

Rôle de l'utilisateur humain

- Formulation des requêtes et itérations: biais cognitifs
- Obséquiosité des LLMs: biais de confirmation, risques de manipulation
- Vérification des références et des sources: recherche et lecture
- Vérification de la véracité des informations: discernement
- Vérification de programmes produits par un système d'IA: tests ou vérification/validation?

Conclusions

- L'IA générative induit une transformation des processus de recherche, de production et d'évaluation scientifiques
- Motivation de la part des chercheurs, des organisations, des agences de recherche, des éditeurs et des organisateurs de conférences
- Plusieurs réflexions convergentes des institutions sur des recommandations d'éthique et d'intégrité scientifique: transparence, contrôle humain, respect de la propriété, ...
- Mais plutôt un encouragement à l'utilisation

Néanmoins:

“Collectivement nous devons être vigilants, particulièrement dans le cadre des métiers scientifiques, quant à l’incidence de l’utilisation de ces outils numériques sur nos propres capacités de réflexion, d’imagination, d’argumentation, de synthèse et de rédaction.”

R. Chatila, C. Kirchner, S. Petitjean, C. Tessier`

Intelligence artificielle : quels enjeux éthiques pour la recherche ?

In Une recherche responsable: Ethique de la recherche scientifique. Edition Quae, A paraître (2026)

Quelques références

Messeri, L., Crockett, M.J. Artificial intelligence and illusions of understanding in scientific research.
Nature 627, 49–58 (2024). <https://doi.org/10.1038/s41586-024-07146-0>

Claudi L. Bockting, Eva A. M. van Dis, Robert van Rooij, Willem Zuidema & Johan Bollen
Living guidelines for generative AI — why scientists must oversee its use
Establish an independent scientific body to test and certify generative artificial intelligence, before the technology damages science and public trust.
Nature 622, 693-696 (2023). doi: <https://doi.org/10.1038/d41586-023-03266-1>

Université de Bologne
<https://www.unibo.it/en/university/statute-standards-strategies-and-reports/artificial-intelligence/use-cases-of-generative-artificial-intelligence-in-research-staff2019s-scientific-production>

David B. Resnik, Mohammad Hosseini
The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool
AI and Ethics. <https://doi.org/10.1007/s43681-024-00493-8>

R. Chatila, C. Kirchner, S. Petitjean, C. Tessier
Intelligence artificielle : quels enjeux éthiques pour la recherche ?
In Une recherche responsable: Ethique de la recherche scientifique. Edition Quae, A paraître (2026)