Brain, Digital, & Learning

pISSN : 2384-2474
eISSN : 2586-7490

REVIEW ARTICLE

Utilizing Open Access Publications and Educational Content in LLM-Based Research: Global Trends and Ethical–Policy Considerations

Kuyng Sik Yi[1,2], Chi-Hoon Choi[1,2*]

[1]Department of Radiology, College of Medicine, Chungbuk National University

[2]Department of Radiology, Chungbuk National University Hospital

*Corresponding Author: Chi-Hoon Choi, chihoonc@chungbuk.ac.kr

ABSTRACT

Large language models (LLMs) have become increasingly ingrained in academic research and education. As a result, leveraging open-access publications and publicly available educational materials through LLMs is emerging as a promising strategy to broaden knowledge discovery and enhance teaching and learning. This review surveys global and domestic trends in the application of LLMs within science and education, highlighting notable implementations including Meta's Galactica, KISTI's KONI model, and the open-source CleverBee system. These examples showcase the substantial potential of generative AI when applied to openly available knowledge resources, while also revealing important limitations. In this context, we identify six key ethical challenges associated with LLM use—covering issues of copyright and licensing, personal data privacy, hallucinated content, algorithmic bias, plagiarism, and accountability—and examine evolving legal and regulatory responses in the United States, European Union, and South Korea. Additionally, we propose seven practical recommendations for researchers (such as checking data licenses, removing personal identifiers, verifying model outputs, and transparently reporting AI assistance) to foster responsible use of LLMs. Drawing on up-to-date guidelines from organizations like the Committee on Publication Ethics (COPE) and the International Committee of Medical Journal Editors (ICMJE) as well as recent national policy directives, this review offers comprehensive guidance for scholars across disciplines including neuroscience, linguistics, learning sciences, and artificial intelligence. Our aim is to support the responsible and effective integration of LLM technologies into academic practice—advancing innovation in research and education while maintaining high ethical and legal standards.

Keywords: Large language models (LLM), open access data, public educational content, ai ethics, privacy protection

Brain, Digital, & Learning
2025, Vol. 15, No. 2, 205-216
https://doi.org/10.31216/BDL.2025.15.2.6

Received: May 26, 2025
Revised: June 14, 2025
Accepted: June 17, 2025

Introduction

In recent years, large language models (LLMs) have quickly become embedded in scientific research and educational practice. This integration has been accelerated by the abundance of

publicly available scholarly papers, datasets, and instructional content that can be utilized as sources. LLMs excel at processing natural language and synthesizing information, enabling them to assist with complex tasks such as literature reviews, knowledge retrieval, and automated summarization (Luo et al., 2022). As generative AI tools become more accessible, researchers and institutions are exploring their potential to speed up discovery, personalize learning experiences, and support decision-making across disciplines.

Despite these opportunities, the growing use of LLMs also presents numerous practical and ethical challenges. Applying LLMs to open-access content can lead to concerns about copyright compliance, data privacy, academic integrity, algorithmic bias, and the overall reliability of AI-generated outputs (Gervais et al., 2024; Yazadzhiyan, 2023). Moreover, the rapid deployment of generative AI has outpaced the creation of clear regulatory frameworks. This gap has prompted international organizations and national governments to begin formalizing standards and guidelines for AI use (European Commission, 2023; Personal Information Protection Commission, 2024a). For example, in South Korea, initiatives such as the AI Ethics Standards (Ministry of Science and ICT, 2020) and the proposed AI Basic Act (Republic of Korea, 2024) signal a heightened focus on developing trustworthy and transparent AI systems.

In this evolving landscape, it is crucial that researchers in fields like neuroscience, linguistics, the learning sciences, and artificial intelligence recognize not only the new opportunities but also the responsibilities that accompany LLM integration. While several recent studies have explored technical or domain-specific aspects of LLMs, this review is distinctive in that it integrates ethical, legal, and policy considerations within a cross-national framework, using three representative models (Galactica, KONI, and CleverBee) to concretely illustrate the stakes. By situating practical recommendations within current international guidelines and regulations, it provides a holistic resource that bridges fragmented prior research.

This review addresses that gap by synthesizing the latest global and domestic developments in using LLMs with open academic and educational resources. Specifically, we (1) examine several prominent implementation cases (including Galactica, KONI, and CleverBee), (2) discuss critical ethical issues associated with LLM use (such as hallucination, bias, plagiarism, and data governance), (3) compare major policy frameworks and regulatory initiatives in the United States, European Union, and South Korea, and (4) offer concrete recommendations for responsible LLM use in research, drawing on international guidelines (e.g., ICMJE and COPE). Our overall aim is to help researchers make informed, ethical, and effective use of generative AI in their scholarly work.

Trends, Review of Literature, and Key Issues

Global and Domestic Trends

Around the world, efforts are underway to harness open-access scholarly resources—such as research papers, textbooks, and open courseware—to power LLMs for research and education. Both major technology companies and academic institutions have begun deploying LLMs trained on large public datasets as research aids and are also developing domain-specific LLMs tailored to particular fields. One prominent example was Meta's Galactica model, introduced in 2022, which was trained on 106 billion tokens of scientific text (including journal articles, textbooks, and

academic websites) and designed to assist with tasks like scientific Q&A and summarization (Turton, 2022). Galactica was discontinued just two days after its public release due to problems with factual inaccuracies and false citations (Turton, 2022), yet it highlighted both the potential and the pitfalls of integrating open scientific data into LLMs.

In addition to general-purpose models, specialized LLMs have emerged in domains such as biomedicine and law. For instance, BioGPT and PubMedGPT were developed by training on vast collections of biomedical literature to support question-answering in medicine and life sciences (Luo et al., 2022; Stanford Center for Research on Foundation Models, 2022). Likewise, companies like Google and Microsoft have incorporated LLM-driven academic search and content synthesis features into their products to help researchers find and summarize scholarly information (Google LLC, n.d.). Within the education sector, universities are beginning to pilot AI-based tutoring systems that leverage open lecture notes and massive open online course (MOOC) content to provide personalized learning support (Gabbyet al., 2024; Hu et al., 2024).

New AI research assistant tools are also being developed to help scholars navigate the literature. For example, *Elicit* (by Ought) uses open-access scholarly databases to retrieve and summarize relevant papers in response to user queries (Ought, 2022). Another example is *CleverBee*, an open-source "deep research" agent introduced in late 2023, which combines an automated web browser with multiple LLMs to autonomously gather information from the internet, synthesize findings, and generate draft research reports (SureScaleAI, 2023). Together, these innovations reflect a global trend of integrating open knowledge resources with LLM technologies to transform many aspects of the research workflow—ranging from literature search and summarization to hypothesis generation and even code writing.

In South Korea, there has been a concerted effort to build data-sharing infrastructure and foster academia–industry collaboration in order to drive large-scale AI research. A growing number of projects focus on developing Korean-language LLMs using specialized domestic datasets. One leading example is KISTI's KONI model, a domain-specific LLM trained on Korean scientific literature, patents, and government reports (J. Kim, 2023). The initial version of KONI was released in 2023 with 800 million parameters, and an expanded version appeared in 2024 after training on more than twice the original data. The updated model showed significant gains in logical reasoning and writing performance, achieving top-tier results on Korean-language LLM benchmarks for scientific argumentation (S. Lee, 2024). KISTI made KONI openly available for research use and implemented retrieval-augmented generation techniques to curb hallucinations and improve the factual reliability of its outputs (S. Lee, 2024). This case demonstrates that even very large LLMs can benefit from being coupled with traditional curated databases to enhance accuracy.

Major Korean technology companies are also adapting their AI systems to leverage public datasets. Naver and Kakao, for instance, have enhanced their large-scale models (such as HyperCLOVA and KoGPT) by incorporating Korean web data as well as open-access academic publications and government documents to improve domain-specific performance (B. Kim et al., 2021; J. W. Kim, 2021). In the education sector, the national Educational Broadcasting System (EBS) has announced plans to create AI tutoring services based on its extensive library of educational materials (Kang, 2024). Additionally, some universities have begun integrating chatbot-based question-answering systems with their course content to support students (Chung-Ang University, n.d.). These developments reflect the growing interest within Korea's research and education communities in combining open knowledge resources with LLMs, bolstered by

government policies that encourage open data usage and AI infrastructure development.

Case Studies: Galactica, KONI, and CleverBee

Galactica (Meta, 2022): Galactica was a large-scale LLM developed by Meta AI intended to tackle the problem of information overload in scientific research. It was trained on an unusually broad range of scientific data—including journal articles, textbooks, protein sequences, and chemical formulas—and was built not only to retrieve documents but also to store and reason over scientific knowledge (Taylor et al., 2022). Galactica initially demonstrated strong performance in domains like mathematics, medicine, and chemistry, likely due to the high quality and diversity of its training data (Taylor et al., 2022). A domain-specific variant called "GeoGalactica" showed promise on geoscience tasks (Deng et al., 2023). However, the Galactica public demo was shut down just three days after launch because the model frequently produced hallucinations (fabricated information) and biased content. Major issues associated with the Galactica model—including hallucinated citations, social bias, overstated capabilities, and implications for industry standards—are summarized in Table 1. Subsequent critiques pointed out that the model's capabilities had been overstated and highlighted its limitations in scientific reasoning, as well as the potential negative impact of such AI systems on human creativity in research (Nature Editorial, 2024). In the wake of the Galactica episode, Meta reportedly postponed the release of its next major LLM, reflecting a broader industry realization that simply making models larger is not a guaranteed path to better performance (Browning, 2025).

Table 1. Major issues associated with Galactica

Issue	Case Summary	Citation
I . Fabrication of Information and Compromised Scientific Credibility	Galactica generated and cited non-existent academic papers and included inaccurate content in scientific summaries.	AI Business. Wodecki, B. (2022, November 17). Meta's Galactica AI criticized as "dangerous" for science by renowned experts. AI Business. Retrieved from https://aibusiness.com/nlp/meta-s-galactica-ai-criticized-as-dangerous-for-science
II. Bias and Generation of Harmful Content	The model produced racially discriminatory, antisemitic, and misogynistic responses, revealing significant social bias.	AIAAIC (AI, Algorithmic, and Automation Incidents Consortium). Galactica generates inaccurate, racist, homophobic, and offensive responses [Internet]. Published 2022 Nov 20 [cited 2025 May 18]. Retrieved from: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/galactica-generates-inaccurate-racist-homophobic-and-offensive-responses
III. Limitations as a Scientific Research Tool	Despite offering features such as scientific summarization, explanation generation, and equation completion, its practical utility fell short of expectations.	LinkedIn News. (2022, November 21). Meta's science-focused AI goes awry. *LinkedIn News*. Retrieved fromhttps://www.linkedin.com/news/story/metas-science-focused-ai-goes-awry-5501756/
IV. Technical Constraints and Industry-Wide Implications	Galactica's failure led to industry-wide reflection that bigger LLMs are not inherently better. Meta subsequently postponed the release of its next major model (e.g., LLaMA 4).	Browning K. Meta is delaying the rollout of its flagship AI model [Internet]. The Wall Street Journal. Published 2025 May 15 [cited 2025 May 18]. Retrieved from: https://www.wsj.com/tech/ai/meta-is-delaying-the-rollout-of-its-flagship-ai-model-f4b105f7

KONI (KISTI, 2023–2024): Korea's KONI model, developed by the Korea Institute of Science and Technology Information (KISTI), represents a contrasting example of a domain-focused LLM. KISTI assembled a large repository of scientific and technical documents—including both Korean and international research papers, patents, and expert reports—to train this model of approximately 8 billion parameters. KONI achieved leading performance on Korean-language science and technology benchmarks (for tasks such as complex question answering and logical reasoning) and successfully integrated retrieval-augmented generation methods to minimize hallucinations and enhance factual accuracy (S. Lee, 2024). This case illustrates how a language- and field-specific LLM, grounded in open research data, can be effectively constructed to address local research needs.

CleverBee (Open Source, 2023): CleverBee is another notable case—a fully open-source research assistant platform introduced in 2025. Written in Python, CleverBee is designed to leverage a wide range of online resources (from public websites to academic databases) by using multiple LLMs in tandem. Operating through an automated browser environment, it can autonomously search the internet, aggregate and summarize information, and generate draft research reports, thereby significantly streamlining tasks like literature reviews and data analysis for investigators. This system demonstrates the potential of combining open web resources with LLM-driven agents to automate parts of the research process and assist human researchers.

Taken together, these cases underscore both the opportunities and the challenges of integrating LLMs with openly available knowledge bases. Galactica demonstrated the feasibility of using vast scientific corpora with an LLM but also exposed the risks of generating incorrect or misleading content. By contrast, the experiences with KONI and CleverBee highlight the practical benefits of tailoring LLMs to specific domains or integrating them with retrieval systems, while at the same time emphasizing the need for ongoing strategies to prevent hallucinations and to verify AI-generated information.

Key Ethical Issues in LLM Use

Copyright and Licensing: Open-access materials are still subject to licensing terms (for instance, Creative Commons licenses like CC BY or CC BY-NC), and using such content to train or prompt LLMs requires careful compliance with those conditions. If an LLM's output closely mirrors a source text, proper attribution or even explicit permission may be required to avoid copyright infringement. Researchers should proactively verify that any text or data they use is legally permissible for reuse, and they must refrain from feeding proprietary or protected content into models without authorization (Creative Commons, 2023; Gervais et al., 2024; Liu et al., 2024; Longpre et al., 2023; U.S. Copyright Office, 2024).

Personal Data and Privacy: Publicly available datasets can include personally identifiable information (PII) for example, individuals' names or identification numbers. Using such data in model training without proper safeguards can violate privacy regulations. In jurisdictions like South Korea, regulators permit only limited use of public personal data under a "legitimate interest" clause, and even then, robust protections such as anonymization or filtering are required (Personal Information Protection Commission, 2024a, 2024b; KISA, 2023). Researchers have a responsibility to screen datasets for any PII and to remove or mask sensitive details before incorporating data into LLM training or analysis.

Hallucination and Factual Integrity: LLMs have a well-documented tendency to sometimes produce false or

fabricated information—a phenomenon commonly referred to as "hallucination." For example, the Galactica model produced citations to non-existent papers and misrepresented scientific facts, calling into question the reliability of its outputs (Ayala & Béchard, 2024; Main, 2022; Wodecki, 2022). To counteract this issue, it is critical to cross-check AI-generated assertions against authoritative primary sources. Techniques like retrieval-augmented generation—which ground an LLM's responses in a vetted knowledge base or dataset—can also help ensure that a model's outputs remain anchored to verifiable information (Wired Staff, 2024; Yazadzhiyan, 2023).

Bias and Discrimination: Large language models can inadvertently perpetuate or even amplify biases present in their training data, reflecting societal stereotypes related to attributes such as race, gender, or culture. For instance, Galactica produced some responses that were offensive and culturally biased, which prompted significant criticism (AIAAIC, 2022; Ars Technica, 2022). It is imperative for researchers to scrutinize LLM outputs for signs of bias. When problematic patterns are detected, steps such as applying bias-mitigation algorithms or having human reviewers in the loop should be taken to prevent discriminatory or harmful content (Lee et al., 2024; Shrawgi et al., 2024).

Academic Integrity and Plagiarism: The use of LLMs to write or translate scholarly text raises the risk of unintentional plagiarism, particularly if a model's output closely mirrors existing sources. Most academic publishers now insist that AI tools cannot be credited as authors and that human authors retain full responsibility for the content of their work (COPE, 2023a; Lee & Tang, 2023; MIT Communication Lab, 2023; Nature Editorial, 2024). To uphold academic integrity, researchers should significantly rewrite or edit any AI-generated material and ensure that all content meets originality standards with proper citation of sources.

Accountability and Transparency: If an LLM's use leads to errors or biased outcomes, the researcher—not the tool—must be accountable for these issues. LLMs should be treated strictly as assistive instruments rather than independent experts, with humans making all final interpretations and decisions. Equally important is transparency about AI involvement in research. Scholars should clearly disclose if and how an LLM was used in their work, which model was employed, what content it generated or influenced, and how those outputs were verified or edited before publication. This level of transparency enhances reproducibility and builds trust with peer reviewers and readers (American Chemical Society, 2024; COPE, 2023b; Elsevier, 2023; ICMJE, 2023; Resnik & Elliott, 2023;).

Emerging Legal and Policy Frameworks

As generative AI systems become more embedded in society, governments and international bodies have been racing to create legal and policy frameworks to ensure these technologies are deployed responsibly. A broad global consensus is emerging that LLM deployment must be bounded by principles that protect human rights, promote fairness and transparency, and prevent harmful outcomes such as discrimination (European Commission, 2023; Republic of Korea, 2024; U.S. Congress, 2024). At the same time, regulatory approaches differ from one jurisdiction to another, especially regarding how to enforce copyright laws, safeguard personal data, ensure algorithmic accountability, and categorize the risks of AI applications. The use of LLMs in educational and research settings often falls into legal gray areas, which is motivating policymakers to translate high-level ethical principles into concrete regulations and guidelines. Examining the initiatives in the United States, the European Union, and South Korea reveals different models of AI governance that

are now taking shape. These differences and priorities across jurisdictions are summarized in Table 2, which compares key AI-related policy instruments from each region and their implications for higher education and research.

In the United States, for example, lawmakers have proposed new legislation such as the *American Privacy Rights Act of 2024* (draft), which emphasizes comprehensive data privacy protections and includes provisions for algorithmic transparency and accountability (U.S. Congress, 2024). The European Union is finalizing its *Artificial Intelligence Act* (proposed 2023), which takes a risk-based approach to regulating AI with the aim of ensuring transparency, safety, and the protection of fundamental rights in high-impact AI systems (European Commission, 2023). South Korea recently enacted the *Framework Act on Artificial Intelligence* (2023), which establishes national principles for AI ethics and governance while promoting transparency, safety, and the preservation of human dignity in AI deployments (Republic of Korea, 2024). Despite differences in legal mechanisms, these efforts share common priorities such as privacy protection, transparency requirements, and safety measures. Notably, the emerging policies are intended not only to regulate AI developers and companies but also to guide researchers and institutions in their responsibilities when integrating LLMs into academic work. By delineating clear boundaries and expectations, the legal frameworks in these regions are helping to shape a more ethical and accountable use of AI in research and education. Table 2 summarizes key legislative and policy instruments from each region, with a focus on their implications for higher education and research.

Table 2. Summary of international and domestic legal regulations and policy guidelines

Region	Major AI-Related Policy/Regulation	Key Focus
United States	American Privacy Rights Act of 2024 (draft)	Emphasizes comprehensive data privacy and includes provisions for algorithmic transparency and accountability.
European Union	EU Artificial Intelligence Act (2023)	Implements a risk-based approach to AI regulation, aiming to ensure transparency, safety, and protection of fundamental rights.
Korea	Framework Act on Artificial Intelligence (2023)	Establishes national AI ethics principles and governance, promoting transparency, safety, and the protection of human dignity in AI deployment.

Practical Considerations for Researchers

To complement broad ethical principles and regulations, researchers need concrete strategies for day-to-day use of LLMs in academia. Below, we outline seven practical recommendations:

Verification of Data Sources and Licenses: Before using any > dataset or text as training material or input for an LLM, confirm > that it is legal to do so. Open-access content often comes with > specific licensing conditions (for example, some Creative Commons > licenses restrict commercial use or require attribution), and > violating these terms can lead to copyright problems. If an > AI-generated output closely resembles any source material, proper > citation or explicit permission is essential (Creative Commons, > 2024; U.S. Copyright Office, 2024). Thorough due diligence on data > licensing helps prevent the unauthorized use of protected content.

Removal of Sensitive Information: Filter all training and input > data to remove personally identifiable information (e.g., real > names, national identification numbers, contact details). Many > regulations—for instance, in South Korea—explicitly prohibit > using certain personal identifiers and mandate anonymization or > pseudonymization

of personal data before reuse. Any dataset > containing sensitive personal information should either be > excluded entirely or processed with robust de-identification > techniques to protect privacy and ensure the model does not learn > inappropriate details (Personal Information Protection Commission, > 2023; Republic of Korea, 2020).

Validation and Cross-Checking of Outputs: Independently verify > any output produced by an LLM against reliable sources. Important > facts, figures, and references provided by an AI should be > manually checked for accuracy. Where possible, use additional > tools or search engines to cross-verify information (or even query > multiple AI models to see if they concur) – this can help > identify errors, particularly in sensitive or high-stakes research > (European Commission, 2023; Korea Institute of Science and > Technology Information [KISTI], 2024; Wired Staff, 2024).

Context-Appropriate Use of Models: Do not treat LLMs as > infallible experts. Use them as supportive tools for drafting, > brainstorming, or summarizing information – not as replacements > for human analysis or decision-making. Retain full responsibility > for interpreting results and drawing conclusions. For example, an > LLM might help summarize related work, but it should never be > allowed to determine the outcome of an experiment or the > interpretation of data (ICMJE, 2023; MIT Communication Lab, 2023).

Transparency and Disclosure: Be transparent about any use of AI > in the research process. Disclose in manuscripts (or other > communications) if an LLM was used, which model was employed, what > it was used for, and how its output was modified by the human > author. This practice is increasingly required by journals and > aligns with publishing ethics, helping others understand and > reproduce the work (COPE, 2023a; Elsevier, 2023). Clear disclosure > builds trust among readers and reviewers regarding the > authenticity of the research.

Adherence to Institutional and Disciplinary Ethics: Follow all > AI-related ethics guidelines set by your institution or > professional discipline. Such guidelines often forbid listing an > AI as a co-author and require that human researchers take > responsibility for all content. Some universities and journals > also require prior approval or review for the use of AI in > drafting manuscripts. Complying with these standards preserves the > integrity of the scholarly record and protects your professional > reputation (KISA, 2023; Ministry of Science and ICT, 2020;).

Continuous Learning and Policy Awareness: Stay informed about > the rapid developments in AI technology and related policies. > Changes in data protection laws or journal standards may directly > affect how LLMs can be used in research and publishing. Building > your AI literacy – along with critical thinking about AI outputs > and ongoing ethical reflection – is essential for long-term, > responsible engagement with these tools (Republic of Korea, 2024; > U.S. Congress, 2024). By keeping up to date with both technical > advances and regulatory changes, researchers can adapt their > practices to remain ethical and effective.

Conclusion

The rapid growth of LLM technologies presents a pivotal opportunity to enhance the use of open-access academic and educational resources. As universities, research institutes, and public agencies continue to release high-quality digital content, integrating LLMs offers a scalable way to accelerate knowledge discovery, personalize learning, and improve research productivity. However, with great potential comes the responsibility to deploy these technologies within well-defined ethical and legal boundaries.

This review highlights the importance of integrating LLMs into scholarly practice in ways that are not only innovative

but ethically grounded. The analysis of Galactica, KONI, and CleverBee reveals how technological ambition must be tempered by safeguards for factual reliability, data integrity, and human oversight. Looking forward, cross-sector collaboration—between academia, industry, and regulators—will be essential to align LLM development with educational and scientific values. Researchers are encouraged to remain critically engaged with the evolving AI landscape and to contribute to shaping responsible norms of use.

We also compared evolving legal and policy frameworks from the United States, European Union, and South Korea. While jurisdictional differences exist, all three emphasize principles of transparency, safety, and the protection of individual rights. Finally, we outlined seven practical recommendations for researchers, ranging from license checks to responsible disclosure, aimed at embedding ethical reasoning into everyday AI use.

Integrating LLMs responsibly into scholarly work demands more than technical competence; it requires awareness of changing legal norms, adherence to disciplinary ethical standards, and a commitment to human-centered research values. Researchers must recognize the limitations of generative AI and apply these tools judiciously, always treating LLM outputs as advisory rather than definitive.

This review is aimed especially at scholars and practitioners in fields such as neuroscience, linguistics, the learning sciences, and artificial intelligence—areas that are poised both to benefit from LLM innovations and to shape how these tools are adopted in academia. In closing, we encourage researchers to:

Proactively evaluate the ethical and legal suitability of any > dataset or material before using it with an AI system.

Thoroughly verify AI-generated content by cross-referencing it > with trusted sources and applying human expertise.

Be transparent about the use of AI tools in all scholarly > communications, including publications and presentations.

Stay updated on national and international AI policies and > guidelines that are relevant to your discipline.

Advocate for institutional policies on AI usage that uphold > research integrity and ethical standards.

By grounding their use of AI in strong ethical practices, researchers can leverage LLMs not only to advance their own projects but also to help build a more trustworthy and responsible digital knowledge ecosystem.

Conflicts of Interest

The author(s) declared no conflicts of interest.

References

AIAAIC (AI, Algorithmic, and Automation Incidents Consortium). (2022, November 20). Galactica generates inaccurate, racist, homophobic and offensive responses. AI Algorithmic and Automation Incidents Repository. Retrieved from https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/galactica-generates-inaccurate-racist-homophobic-and-offensive-responses

American Chemical Society (ACS). (2024). Artificial Intelligence (AI) Best Practices and Policies at ACS Publications. https://researcher-resources.acs.org/publish/aipolicy

Ars Technica. (2022, November 18). New Meta AI demo writes racist and inaccurate scientific literature, gets pulled. Ars Technica. Retrieved from https://arstechnica.com/information-technology/2022/11/after-controversy-meta-pulls-demo-of-ai-model-that-writes-scientific-papers/

Ayala, O. M., & Béchard, P. (2024). Reducing hallucination in structured outputs via retrieval-augmented generation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Industry Track) (pp. 228–238). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-industry.19

Browning, K. (2025, May 15). Meta is delaying the rollout of its flagship AI model. The Wall Street Journal. Retrieved from https://www.wsj.com/tech/ai/meta-is-delaying-the-rollout-of-its-flagship-ai-model-f4b105f7

Chung-Ang University. (n.d.). CAU e-Advisor system for student-centered AI-assisted university life. Retrieved from https://eadvisor.cau.ac.kr/info.html

Committee on Publication Ethics (COPE). (2023a). Artificial intelligence and authorship. https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Committee on Publication Ethics (COPE). (2023b). Authorship and AI tools. https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

Creative Commons. (2023, August 18). Understanding CC licenses and generative AI. Creative Commons. Retrieved from https://creativecommons.org/2023/08/18/understanding-cc-licenses-and-generative-ai/

Creative Commons. (2024). Understanding CC licenses and AI training. Retrieved from https://creativecommons.org/

Elsevier. (2023). The use of generative AI and AI-assisted technologies in writing. Retrieved from https://www.elsevier.com/

European Commission. (2023). EU Artificial Intelligence Act: Overview and final text. Retrieved from https://artificialintelligenceact.eu/

Deng, C., Wang, S., Zhu, Y., & Wang, Y. (2024). K2: A foundation language model for geoscience knowledge understanding and utilization. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM ’24) (pp. 161–170). Association for Computing Machinery. https://doi.org/10.1145/3616855.3635772

Gabbay, H., & Cohen, A. (2024). Combining LLM-generated and test-based feedback in a MOOC for programming. In Proceedings of the 11th ACM Conference on Learning at Scale (L@S ’24) (pp. 177–187). Association for Computing Machinery. https://doi.org/10.1145/3657604.3662040

Gervais, D., Shemtov, N., Marmanis, H., & Rowland, C. Z. (2024). Copyright liability for LLM outputs. Journal of the Copyright Society of the USA, 71. Retrieved from https://www.copyright.com/blog/copyright-liability-for-llm-outputs/

Google LLC. (n.d.). Google Search Academic Enhancements. (Retrieved 2025)

Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. IEEE Transactions on Learning Technologies, 17, 1445–1459. https://doi.org/10.1109/TLT.2023.3234002

International Committee of Medical Journal Editors (ICMJE). (2023). Defining the role of authors and contributors. https://www.icmje.org/

Kang, Y. (2024, February 18). EBS munje pulmyŏn AI-ga sujun chindan… kyosa-ga kongbubŏp onlain mentŏring [EBS diagnoses student levels using AI after solving questions; Teachers provide online mentoring for study methods]. Korea Economic Daily. https://www.hankyung.com/article/2024021817181

Kim, B., Kim, H., & Lee, S.-W. (2021). What changes can large-scale language models bring? Intensive study on HyperCLOVA. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 3405–3424). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.274

Kim, J. (2023, December 20). KISTI develops specialized generative language model “KONI”. Yonhap News. Retrieved from https://www.yna.co.kr/view/AKR20231220091200063

Kim, J. W. (2021, November 16). Kakao, Han'gugŏ t'ŭkhwa AI konggae… Neibeowa kuknae taehyŏng AI kyŏngjaeng [Kakao unveils Korean-specialized AI, sparking major domestic AI competition with Naver]. Korea Economic Daily. https://www.hankyung.com/article/2021111630991

Korea Institute of Science and Technology Information (KISTI). (2024). KONI: Korean scientific large language model. https://kisti.re.kr/

Korea Internet & Security Agency (KISA). (2023). Guidebook on de-identified information usage and AI self-assessment. https://www.kisa.or.kr/

Lee, J., Tang, L., Xu, W., & Wang, J. (2023). Do language models plagiarize? In Proceedings of the ACM Web Conference 2023 (WWW ’23) (pp. 3637–3647). Association for Computing Machinery. https://doi.org/10.1145/3543507.3583199

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2024). Large language models portray socially subordinate groups as more homogeneous, consistent with a bias observed in humans. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’24) (pp. 1321–1340). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658975

Lee, S. (2024, July 31). KISTI releases updated “KONI” model for science and technology. Chungcheong News. Retrieved from https://www.ccnnews.co.kr/news/articleView.html?idxno=342851

LinkedIn News. (2022, November 21). Meta’s science-focused AI goes awry. LinkedIn News. Retrieved from https://www.linkedin.com/news/story/metas-science-focused-ai-goes-awry-5501756/

Liu, X., Sun, T., Xu, T., Wu, F., Wang, C., Wang, X., & Gao, J. (2024). SHIELD: Evaluation and defense strategies for copyright compliance in LLM text generation. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 1640–1670). Association for Computational Linguistics.

Longpré, S., Mahari, R., Chen, A., et al. (2024). A large-scale audit of dataset licensing and attribution in AI. Nature Machine Intelligence, 6, 975–987. https://doi.org/10.1038/s42256-024-00878-8

Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. Y. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. Briefings in Bioinformatics, 23, bbac409. https://doi.org/10.1093/bib/bbac409

Main, N. (2022, November 22). Meta AI bot contributed to fake research and nonsense before being pulled offline. Gizmodo. Retrieved from https://gizmodo.com/meta-ai-bot-galactica-1849813665

Ministry of Science and ICT. (2020). AI ethics standards. Ministry of Science and ICT. https://www.msit.go.kr/

MIT Communication Lab. (2023). Using Generative AI for Your Scientific Writing? Be Aware of Journal Policies. https://mitcommlab.mit.edu/

Nature Editorial Policies. (2024). Guidance on generative AI in scientific publishing. https://www.nature.com/

Nature Editorial. (2024, September 11). AI is complicating plagiarism. How should scientists respond? Nature. Retrieved from https://www.nature.com/articles/d41586-024-02371-z

Ought. (2022, March 22). Elicit: The AI research assistant. Ought. Retrieved from https://ought.org/updates/2022-03-22-elicit

Personal Information Protection Commission (Republic of Korea). (2024a). Guidelines on processing pseudonymized information. https://www.pipc.go.kr/

Personal Information Protection Commission (Republic of Korea). (2024b). Guidelines on processing publicly available personal information for AI development and services. https://www.pipc.go.kr/

Personal Information Protection Commission. (2023). Guidelines on processing publicly available personal information for AI development. https://www.pipc.go.kr/

Republic of Korea. (2020). Personal Information Protection Act (Act No. 16930, March 31, 2020). https://www.law.go.kr/

Republic of Korea. (2024). Framework Act on Artificial Intelligence (Act No. 2024-120, December 27, 2024). https://likms.assembly.go.kr/

Resnik, D. B., & Elliott, K. C. (2023). The ethics of disclosing the use of artificial intelligence tools in scientific manuscripts. Research Ethics, 19(3), 1–9. https://doi.org/10.1177/17470161231169480

Stanford Center for Research on Foundation Models. (2022, December 15). Stanford CRFM introduces PubMedGPT 2.7B. Stanford Institute for Human-Centered AI. https://hai.stanford.edu/news/stanford-crfm-introduces-pubmedgpt-27b

Shrawgi, H., Koci□, A., & Ray, S. (2024). Uncovering stereotypes in large language models: A task complexity-based approach. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024) (pp. 1841–1857). Association for Computational Linguistics. https://aclanthology.org/2024.eacl-long.111

SureScaleAI. (2023, December). CleverBee: The open source Deep Researcher tool. GitHub. Retrieved from https://github.com/SureScaleAI/cleverbee

Turton, W. (2022, November 21). Meta's AI bot Galactica was supposed to help scientists. It spewed nonsense instead. Gizmodo. https://gizmodo.com/meta-ai-bot-galactica-1849813665

U.S. Congress. (2024). American Privacy Rights Act of 2024 (draft bill). https://www.congress.gov/

U.S. Copyright Office. (2024). Copyright and artificial intelligence: Part III – Generative AI training. https://www.copyright.gov/

Wired Staff. (2024, June 14). Reduce AI hallucinations with this neat software trick. Wired. https://www.wired.com/story/reduce-ai-hallucinations-with-rag/

Wodecki, B. (2022, November 17). Meta's Galactica AI criticized as "dangerous" for science by renowned experts. AI Business. https://aibusiness.com/nlp/meta-s-galactica-ai-criticized-as-dangerous-for-science

Yazadzhiyan, H. (2023, January 15). What are LLM hallucinations? Causes, ethical concerns and prevention. ResearchGate. https://www.researchgate.net/publication/376829203_What_are_LLM_hallucinations_Causes_ethical_concerns_and_prevention

Authors Information

Yi, Kuyng Sik: Department of Radiology, College of Medicine, Chungbuk National University and Chungbuk National University Hospital. First Author. https://orcid.org/0000-0002-4274-8610

Choi, Chi-Hoon: Department of Radiology, College of Medicine, Chungbuk National University and Chungbuk National University Hospital. Corresponding Author. https:// orcid.org/0000-0003-4137-7376