https://doi.org/10.51408/issi2025_047

Science and Artificial Intelligence: A Copyright Perspective

Dmitry Kochetkov

d.kochetkov@cwts.leidenuniv.nl, dmitry.kochetkov@urfu.ru
Centre for Science and Technology Studies, Leiden University, Kolffpad 1, 2333 BN Leiden (The Netherlands)
Ural Federal University, 19 Mira Street, 620062 Ekaterinburg (Russian Federation)

Abstract

Artificial intelligence (AI), particularly generative AI (GenAI) and large language models (LLMs), is transforming scientific research and higher education, offering new opportunities while raising significant ethical, legal, and regulatory challenges. This opinion piece explores the intersection of AI and science, focusing on the implications for copyright, peer review, and open science. AI systems, such as LLMs, are increasingly used in research applications, including text generation, data analysis, and peer review, with recent studies suggesting that AI-assisted reviews may improve efficiency and address reviewer shortages. However, concerns about bias, confidentiality, and the lack of guidelines for AI use in peer review persist. The rise of AI also poses challenges to copyright, as LLMs often rely on vast datasets of scientific works, raising questions about fair use, attribution, and licensing. Current regulatory frameworks in the United States, China, the European Union, and the United Kingdom focus on promoting innovation and responsible AI development, but gaps remain, particularly in addressing the use of copyrighted works for AI training. Creative Commons licenses, widely used for open-access outputs, do not fully address the complexities of AI training, and the absence of proper attribution in AI systems challenges the concept of originality. This paper calls for action to ensure that AI training is not considered a fair use exception to copyright law, advocating for authors' rights to refuse the use of their works for AI training and for universities to take a leading role in regulating AI. Governments and international organizations must develop harmonized legislative measures to protect authors' rights and ensure transparency in AI training datasets. The paper concludes that while AI offers transformative potential for science, a careful and responsible approach is needed to balance innovation with ethical and legal considerations, preventing the emergence of an oligopolistic market that prioritizes profit over scientific integrity.

Introduction

While there is no single, universally accepted definition of artificial intelligence (AI), it can be broadly defined as the ability of machines to learn, make decisions, and solve problems in a way that resembles human cognition (Sonone & Dharme, 2019). AI systems are designed to go beyond simple calculations, aiming to solve complex problems autonomously (Fogel, 2005). Generative AI (GenAI), a branch of AI, utilizes deep learning techniques – specifically generative models – to produce creative outputs such as music, images, and text (Ramdurai & Adhithya, 2023). In this opinion piece, I will focus primarily on Large Language Models (LLMs), which are intelligent systems capable of natural language processing (Gao et al., 2023; Hadi et al., 2023). These systems can process and generate human-like language, including tasks like machine translation. However, the true nature of their intelligence remains a subject of debate. Some researchers argue that the apparent intelligence of LLMs may reflect the interviewer's own intelligence rather than the model's, suggesting a "reverse Turing test" (Sejnowski, 2023).

The term "artificial intelligence" (AI) was first coined by John McCarthy at the Dartmouth Conference in 1956, marking the official beginning of AI's history (Strickland, 2021). However, the evolution of AI in the 20th century was marked by significant scientific and technical challenges that have hindered its rapid development. These challenges mainly include computational power limitations and algorithmic constraints (Puttgen & Jansen, 1987). Two periods, usually referred to as “AI winter,” represent the situation of reduction of interest and funding for AI research due to unmet expectations and the failure to deliver the promised breakthroughs. The first AI winter occurred in the 1970s and 1980s, primarily due to the overpromising of AI capabilities by researchers and the subsequent failure to achieve these goals. Similarly, the second AI winter in the late 1980s and early 1990s was caused by the failure of expert systems to deliver on their potential, despite significant investments by corporations (Lloyd, 1995). Algorithmic advances have played a crucial role in overcoming computational limitations (Selman, 2000).

Since then, the 21st century has witnessed significant advancements in AI, driven by increased computational power and the availability of vast amounts of data (Hwang, 2018; Liu et al., 2018). These advancements have transformed various sectors, including healthcare, finance, and manufacturing. AI's impact on society and the global order is profound, with implications extending far beyond technology (Rama Padmaja & Lakshminarayana, 2024). The rise of AI has reshaped power dynamics among nations, with countries like the USA, China, and Russia leading the global race for AI dominance (Vijayakumar, 2023).

The development of AI technology presents both challenges and opportunities across various fields (Rama Padmaja & Lakshminarayana, 2024; Wolff et al., 2018). While AI offers immense potential, its advancement raises ethical concerns, including biases, privacy issues, and broader social implications (Rama Padmaja & Lakshminarayana, 2024). Li (2023) identifies 12 key ethical concerns and related strategies for applying AI in healthcare: justice and fairness, freedom and autonomy, privacy, transparency, patient safety and cybersecurity, trust, beneficence, responsibility, solidarity, sustainability, dignity, and conflicts. AI's influence spans all five dimensions of sustainability, with both positive and negative consequences (Khakurel et al., 2018). For instance, an analysis of a Google Scholar sample of questionable scientific papers suspected to be generated by GPT revealed that many address applied, often controversial issues prone to misinformation, such as environment, health, and computing (Haider et al., 2024). Additionally, LLMs may pose a threat to copyright, as they can generate content that potentially violates intellectual property rights (German, 2024). Currently, neither copyright nor “open” licenses can protect scholarly content from unauthorized reuse in AI training (Decker, 2025).

AI is transforming research jobs, and science, that in turn provides LLMs with a vast amount of data for training. The goal of this opinion piece is to analyze the potential consequences of the further development of AI on science, highlighting its positive effects while also mitigating risks. In the next section, I will provide a brief overview of how AI is being used in research applications. I will then analyze the current state of AI regulation, particularly regarding science, identifying any gaps in the current

regulations. Finally, I will outline several suggestions for filling these gaps to ensure the safe and effective use of AI in academic research.

Applications of AI and LLMs in Research and Higher Education

Artificial intelligence (AI), particularly large language models (LLMs), is transforming higher education and research in much the same way it is revolutionizing other industries. AI has the potential to enhance personalized learning experiences, provide feedback to students, identify at-risk learners, and accelerate the research process (Tarisayi, 2024). Applications of AI in these fields include text generation, data analysis, literature review assistance, and peer review (Alqahtani et al., 2023). For instance, AI can automate many tasks involved in conducting systematic literature reviews (De La Torre-López et al., 2023). Another promising use case is the proofreading and editing of scientific texts. While these applications have the potential to revolutionize education and research, challenges remain, including ethical concerns, algorithmic bias, and the need for human oversight (Alqahtani et al., 2023; Peláez-Sánchez et al., 2024). Algorithmic bias refers to systematic errors in AI systems that can lead to unfair and unequal outcomes (Shin & Shin, 2023). Furthermore, Andersen et al. (2024) identified three clusters of AI perception among academics: "GenAI as a workhorse," "GenAI as a language assistant only," and "GenAI as a research accelerator." The authors argue that these variations reflect differences across disciplines and knowledge production models.

Automatic or AI-assisted peer review has been proposed as a potential solution to issues of quality and reproducibility in scientific research. Software tools for automatically evaluating scientific papers using AI, StatReviewer[1] and UNSILO[2], have emerged in recent years[3]. Additionally, tools like the *statcheck* package for verifying statistical analyses have gained traction[4]. Until recently, these tools were considered auxiliary and incapable of replacing human labor (Baker, 2015; Heaven, 2018). However, recent advances in AI are challenging this notion.

Recent studies have explored the impact of AI and LLMs on peer review, with research indicating that AI-assisted reviews are becoming more prevalent. At ICLR 2024, it is estimated that at least 15.8% of reviews will be AI-assisted (Latona et al., 2024). These AI-assisted reviews tend to assign higher scores to papers and increase acceptance rates (Latona et al., 2024), potentially improving review quality and addressing reviewer shortages (Hosseini & Horbach, 2023). However, such studies are often based on limited samples. For example, Biswas et al. (2023) compared ChatGPT's performance as an AI reviewer to human reviews for a single published article. The authors found that ChatGPT demonstrated commendable ability in identifying methodological flaws, providing insightful feedback on theoretical

[1] StatReviewer. URL: http://statreviewer.com/ (date of access: 22.01.2024).
[2] UNSILO. URL: https://site.unsilo.com/site/ (date of access: 22.01.2024).
[3] At the same time, plagiarism detection systems have existed for much longer. For example, "Antiplagiat," a well-known system in Russia, was established in 2005.
[4] statcheck. URL: https://michelenuijten.shinyapps.io/statcheck-web/ (date of access: 22.01.2024), also R package.

frameworks, and assessing the overall contribution of articles to their respective fields.

Despite these advancements, concerns about bias amplification, confidentiality, and the lack of guidelines for LLM use in peer review persist (Hosseini & Horbach, 2023). Some researchers advocate for AI to assist with manuscript triaging (Bauchner & Rivara, 2024), suggesting that human-AI collaboration could democratize academic culture (Sarker et al., 2024). Nevertheless, researchers recommend disclosing the use of LLMs and maintaining human responsibility for review accuracy and integrity (Hosseini & Horbach, 2023).

The impact of AI on the publishing industry can be described as revolutionary. It is expected that AI will bring about a third digital transformation in the industry (Bergstrom & Ruediger, 2024). Two possible scenarios for the future development of AI in scholarly publishing have been proposed. In the first scenario, AI would make the publishing process more efficient, expanding the range of services offered by publishers. In a more radical scenario, AI could fundamentally change the way scientific communication occurs, transforming the channels used for communication.

The interaction between generative AI (GenAI) and the open access movement is complex (Hosseini et al., 2024). GenAI can make scholarly publications more comprehensible to the public or researchers from other fields. It can also help mitigate the negative consequences of information overload and assist researchers in fully benefiting from open access. However, significant risks are associated with using GenAI to enhance access to scholarly literature. One concern is the potential for systems to provide inaccurate or biased summaries, syntheses, or advice. Another risk is the possibility of facilitating the proliferation of paper mills. Finally, the absence of proper attribution of training data challenges the concept of originality and may discourage the sharing of data and papers.

Open science has led to the generation of vast amounts of data, presenting both opportunities and challenges for the scientific community. AI research can also be part of open science, particularly through the development of open-source LLMs such as Game 2, Nemo Tron-4, and Llama 3.1. Open datasets are crucial to the success of these open-source projects. However, developers face numerous challenges, including language bias and safety issues.

Several community initiatives aim to address these challenges. One such initiative is the Aya project, which seeks to bridge the language barrier by providing a human-curated instruction-following dataset in 65 different languages (Singh et al., 2024). The dataset contains 513 million examples across 114 languages. As a result of this initiative, three key resources have been developed and made freely available: the Aya Dataset, the Aya Collection, and the Aya Evaluation Suite. This initiative serves as a platform for future research collaboration to continue bridging the gap in language resources.

Another issue with open-source LLMs is their susceptibility to malicious exploitation. Yi et al. (2024) identified vulnerabilities in the safety alignment of open-access LLMs, which can significantly increase the success rate and

harmfulness of jailbreak attacks[5]. The study proposes two types of techniques that can make LLMs adeptly reverse-aligned to output harmful content, even in the absence of manually curated malicious datasets.

AI-Related Regulations

In this section, I provide a brief analysis of the regulations related to artificial intelligence (AI) in the United States, China, the United Kingdom, and the European Union.

Interestingly, there is currently no comprehensive regulation governing AI in the UK and the US. The Sunak government issued a framework document in 2023 titled *A Pro-Innovation Approach to AI Regulation* (Department for Science, Innovation & Technology, 2023), which establishes basic principles for AI. The document promotes flexible regulation and aims to foster innovation through the development and use of AI technologies. The British government has also expressed its ambition to make the UK the best place to invest in AI.

In the United States, a framework document was published in October 2023, titled Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (2023). Notably, this document includes actions related to copyright law, stating: "…consult with the Director of the United States Copyright Office and issue recommendations to the President on potential executive actions relating to copyright and AI. The recommendations shall address any copyright and related issues discussed in the United States Copyright Office's study, including the scope of protection for works produced using AI and the treatment of copyrighted works in AI training."

A significant step forward was taken with the development of the *Generative AI Copyright Disclosure Act of 2024* (H.R.7913 - 118th Congress, 2023-2024). This act aims to ensure transparency in the use of copyrighted works for AI training and is currently under consideration in the House of Representatives. If passed, the act would require companies to notify the U.S. Copyright Office about any copyrighted works used in their AI systems. These notifications must be submitted 30 days before or after the public release of the AI system, ensuring transparency and accountability. The act is intended to help copyright holders make informed decisions about licensing and compensation. However, the wording of the document remains vague, raising questions for both AI developers and copyright owners. Additionally, I have concerns about the inability of copyright holders to prohibit the use of their works for AI training, which creates a bias in favor of AI development.

In China, the *Interim Measures for the Management of Generative Artificial Intelligence Services* (Cyberspace Administration of China et al., 2023) were implemented on August 15, 2023. These regulations, comprising 24 articles, aim to strike a balance between fostering innovation and ensuring the security and governance of AI. Article 3 emphasizes the importance of maintaining a harmonious

[5] User prompt injection attacks occur when users deliberately exploit system vulnerabilities to elicit unauthorized behavior from an LLM (see, for example, https://learn.microsoft.com/en-au/azure/ai-services/content-safety/concepts/jailbreak-detection).

relationship between development and innovation while prioritizing security and governance in the field of AI. Articles 5 and 6 highlight the need for collaboration in developing basic technologies, such as chips and software platforms, as well as the creation of shared data resources. Article 16 states that all regulatory measures must be compatible with innovation, and Article 2 clarifies that the regulations apply only to publicly available generative AI services. Service providers are held responsible for the content generated using their services. Chinese regulations are among the most stringent in the world. For example, Article 12 mandates that users must be informed when content is generated using AI as a blanket rule.

On August 1, 2024, the European *Artificial Intelligence Act* (AI Act) entered into force (Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024). This law primarily aims to reduce the risks associated with the use of AI. It focuses mainly on high-risk AI systems, while low-risk systems require transparency. For instance, chatbots must clearly inform users that they are interacting with a machine, and certain AI-generated content must be labeled as such.

In summary, the legislative frameworks in major AI-developing countries primarily focus on either the responsible development and use of AI[6], including content labeling, or on fostering innovation and attracting investment in the AI industry. Notably, only the US addresses copyright issues in connection with AI development, but its regulatory framework remains incomplete and appears biased toward AI developers rather than copyright holders. It is also worth noting that, at present, there are no specific legislative regulations governing AI in the Russian Federation. However, Russia has introduced the concept of "experimental legal regimes for digital innovations[7]," which allows for the testing of technologies that are not yet legally regulated.

Copyright and Licensing

Most scientific works are protected by copyright laws. Copying and retaining these works in AI systems, as well as reproducing them in outputs, involves copyright, making appropriate licensing essential for compliance (Johnson, 2024). The generated output can be considered a derivative work, although this is not explicitly stated in any legal documents.

Creative Commons (CC) licenses are the most widely used for open-access outputs. Approximately 28% of global research output is licensed under the Creative Commons Attribution license (CC BY), while another 22% uses more restrictive Creative Commons licenses (Pollock & Michael, 2024). However, Creative Commons acknowledges that existing CC licenses do not fully address the specific challenges related to using creative works for AI training (Walsh, 2023). Using CC-licensed content raises several questions, such as whether the attribution requirement is fulfilled when training LLM models. In my opinion, this is not the case. For example, the training dataset for ChatGPT contains millions of scientific articles, but

[6] Living Guidelines on the Responsible Use of Generative AI in Research | Research and Innovation (2024) also focuses on responsible use of AI and related issues of research integrity.

[7] Regulated by Federal Law No. 258-FZ, dated July 31, 2020, "On Experimental Legal Regimes in the Field of Digital Innovations in the Russian Federation".

it is unclear exactly which ones were used ("AI Firms Must Play Fair When They Use Academic Data in Training," 2024).

However, if the use of content is subject to copyright exclusions, the licensee's abilities are limited. In fact, such an exclusion is currently being considered for legislation in the US. Moreover, the US fair use doctrine allows for the unlicensed use of copyrighted works under certain circumstances. AI training is often considered a case of fair use (Johnson, 2024; Walsh, 2023). For instance, OpenAI argues that this position is "supported by long-standing and widely accepted precedents" (*OpenAI and Journalism*, 2024).

Publishers are also responding to market changes by developing licensing agreements for the use of content in LLM training (Schonfeld, 2024). Currently, the number of such deals is relatively low[8], and they primarily cover content distributed through subscription services. If a publishing contract includes the full transfer of rights to the publisher, the publisher can license the content for AI training without seeking the authors' consent (Hansen, 2024). This underscores the importance of the rights retention strategy. Major publishers, along with Clarivate, are rapidly developing new AI-based businesses, which are evolving into data cartels (Pooley, 2024). This could lead to a situation where the academic AI market adopts the same oligopolistic structure as the current academic publishing market.

A Call for Action

Science and artificial intelligence (AI) are closely linked. Research provides data, which is crucial for training large language models (LLMs) and advancing data science more broadly. At the same time, generative AI (GenAI) is revolutionizing research. Open-source LLMs are an essential part of open science. While AI presents significant opportunities for scientific advancement, it also poses substantial risks. Legislation in this field is still evolving, and regulatory and policy documents often focus on attracting investment in AI or promoting its responsible development and use. The use of publicly available research outputs for training LLMs falls into a "grey area." At the moment, the community lacks any meaningful discussion on the reuse of academic content for LLMs' training. Attempts to raise this issue are made, but their impact is rather limited (Decker, 2025). Below, I offer some thoughts on actions that can be taken in the near future.

First and foremost, AI training should not be considered an exception to copyright law (i.e., under the fair use doctrine). Recognizing LLM training as a case of fair use undermines efforts to reform copyright regulation. In my opinion, LLM training should not qualify as fair use for at least two main reasons:

1. *Non-commercial use is not guaranteed:* Many AI systems already operate on paid subscription models. Even if no fees are currently charged, there are no legal restrictions preventing these models from becoming commercialized in the future.

[8] Generative AI Licensing Agreement Tracker. URL: https://sr.ithaka.org/our-work/generative-ai-licensing-agreement-tracker/.

2. *Content can be reproduced with high accuracy:* AI-generated content often closely resembles the original, making it subject to copyright and attribution requirements.

This issue is particularly relevant in the US context, but given that most AI developers are based in the US, it is critical for the global development of the industry. Some researchers argue that it will take years for US courts to address the issue of licensing content for LLM training (Bergstrom & Ruediger, 2024). This is a major concern for the academic community, as the market will continue to evolve, researchers will increasingly rely on AI for interacting with scholarly output, and it will become more difficult to implement changes (see below for further discussion of limitations and challenges).

Authors should have the option to refuse the use of their work for training GenAI models or specific groups of such models. This should be explicitly stated in the licensing terms. There are two possible strategies to achieve this:

1. *Examine existing licenses:* The Creative Commons BY-ND (Attribution-NoDerivatives) license could be considered restrictive for AI training, but only if regulatory frameworks recognize AI-generated content as derivative works. However, determining whether AI-generated content qualifies as a derivative work is complicated by the fact that LLMs can produce different responses for each query, making it difficult to assess similarity to the original. The BY-NC (Attribution-NonCommercial) license may also be restrictive for training models intended for commercial use[9].
2. *Introduce a new "NT" (no train) extension:* This would explicitly prohibit the use of licensed works for AI training. However, since the original datasets used for LLM training are not publicly accessible, the prospects for enforcing such licensing terms remain uncertain. Additionally, publishing contracts should specify that publishers cannot use articles to train their LLMs or other AI models without author consent.

Universities as Key Players in AI Regulation

Universities should take a leading role in regulating AI. On the one hand, universities often act as publishers or maintain their own repositories, making it feasible to implement content licensing approaches in practice. On the other hand, universities conduct research and develop GenAI models, placing them at the forefront of addressing the ethical aspects of these processes. Furthermore, universities can provide evidence to support legislative regulation. Having said that, I must acknowledge that universities lack the regulatory power that governments possess. However, it is concerning that many current community documents in the field of open science, such as the *Barcelona Declaration on Open Research Information* (2024), do not address AI-related issues.

[9] However, can we be certain that today's open models will not be commercialized in the future?

Legislative Measures and International Cooperation

Governments and international organizations must develop and implement legislative measures to protect authors' rights and prevent the unauthorized use of their works for training GenAI models. One of the first steps should be the mandatory disclosure of training datasets by developers.

The challenge lies not only in adopting national AI laws but also in harmonizing these laws globally. Without international coordination, commercial developers could exploit "safe harbors" to serve their own interests. Therefore, it is essential for large intergovernmental organizations, such as UNESCO, to take on this task. Another challenge is that AI models cannot be "untrained." If restrictions are imposed only on new models, existing models would gain a non-market advantage. Conversely, applying restrictions retroactively to existing models could destabilize the industry. A responsible dialogue is needed to find a balanced solution. One possible approach is retrieval-augmented generation, which allows models to reference relevant papers in their outputs ("AI Firms Must Play Fair When They Use Academic Data in Training," 2024).

Conclusion

The author of this article does not oppose AI. In fact, while writing this manuscript, the Yandex. Translate service was used to assist with reading Chinese text and proofreading the English version. The development of AI brings numerous opportunities for research, but it requires a careful and responsible approach that considers the interests of all stakeholders. Otherwise, there is a risk of fostering an oligopolistic market driven by profit maximization, resembling the current dynamics of the academic publishing sector. As an author, I would like the option to refuse the use of my work for training GenAI models, especially for commercial purposes.

Acknowledgments

The author would like to express his gratitude to Ludo Waltman and Erna Sattler for their valuable comments and suggestions, which have greatly improved this paper.

Contribution

Dmitry Kochetkov: Conceptualization, Writing – original draft

References

AI firms must play fair when they use academic data in training. (2024). *Nature*, 632(8027), 953–953. https://doi.org/10.1038/d41586-024-02757-z

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy*, 19(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Andersen, J. P., Degn, L., Fishberg, R., Graversen, E. K., Horbach, S. P. J. M., Schmidt, E. K., Schneider, J. W., & Sørensen, M. P. (2024). *Generative Artificial Intelligence*

(GenAI) in the research process – a survey of researchers' practices and perceptions. SocArXiv. https://doi.org/10.31235/osf.io/83whe

Baker, M. (2015). Smart software spots statistical errors in psychology papers. *Nature*. https://doi.org/10.1038/nature.2015.18657

Kramer, B., Neylon, C., & Waltman, L. (2024). *Barcelona Declaration on Open Research Information*. https://doi.org/10.5281/ZENODO.10958522

Bauchner, H., & Rivara, F. P. (2024). Use of artificial intelligence and the future of peer review. *Health Affairs Scholar*, 2(5), qxae058. https://doi.org/10.1093/haschl/qxae058

Bergstrom, T., & Ruediger, D. (2024). *A Third Transformation? Generative AI and Scholarly Publishing*. Ithaka S+R. https://doi.org/10.18665/sr.321519

Biswas, S., Dobaria, D., & Cohen, H. L. (2023). ChatGPT and the Future of Journal Reviews: A FeasibilityStudy. *The Yale Journal of Biology and Medicine*, 96(3), 415–420. https://doi.org/10.59249/SKDH9286

De La Torre-López, J., Ramírez, A., & Romero, J. R. (2023). Artificial intelligence to automate the systematic review of scientific literature. *Computing*, 105(10), 2171–2194. https://doi.org/10.1007/s00607-023-01181-x

Decker, S. (2025, April 15). *Guest Post - The Open Access – AI Conundrum: Does Free to Read Mean Free to Train?* The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2025/04/15/guest-post-the-open-access-ai-conundrum-does-free-to-read-mean-free-to-train/

Department for Science, Innovation & Technology (2023). *A pro-innovation approach to AI regulation* (No. 815). https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (2023). https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Fogel, D. B. (2005). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence* (1st ed.). Wiley. https://doi.org/10.1002/0471749214

Gao, Y., Baptista-Hon, D. T., & Zhang, K. (2023). The inevitable transformation of medicine and research by large language models: The possibilities and pitfalls. *MedComm – Future Medicine*, 2(2), e49. https://doi.org/10.1002/mef2.49

Generative AI Copyright Disclosure Act of 2024, H.R.7913—118th Congress (2023-2024) (2024). https://www.congress.gov/bill/118th-congress/house-bill/7913

German, D. M. (2024). *Copyright related risks in the creation and use of ML/AI systems* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2405.01560

Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2023). *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects*. https://doi.org/10.36227/techrxiv.23589741.v4

Haider, J., Söderström, K. R., Ekström, B., & Rödl, M. (2024). GPT-fabricated scientific papers on Google Scholar: Key features, spread, and implications for preempting evidence manipulation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-156

Hansen, D. (2024, July 30). *What happens when your publisher licenses your work for AI training?* Authors Alliance. https://www.authorsalliance.org/2024/07/30/what-happens-when-your-publisher-licenses-your-work-for-ai-training/

Heaven, D. (2018). AI peer reviewers unleashed to ease publishing grind. *Nature*, 563(7733), 609–610. https://doi.org/10.1038/d41586-018-07245-9

Hosseini, M., & Horbach, S. P. J. M. (2023). *Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other Large Language Models in scholarly peer review.* https://doi.org/10.21203/rs.3.rs-2587766/v1

Hosseini, M., Horbach, S. P. J. M., Holmes, K., & Ross-Hellauer, T. (2024). Open Science at the generative AI turn: An exploratory analysis of challenges and opportunities. *Quantitative Science Studies*, 1–24. https://doi.org/10.1162/qss_a_00337

Cyberspace Administration of China, National Development and Reform Commission of the People's Republic of China, Ministry of Education of the People's Republic of China, Ministry of Science and Technology of the People's Republic of China, Ministry of Industry and Information Technology of the People's Republic of China, Ministry of Public Security of the People's Republic of China, State Administration of Radio and Television (2023). *Interim Measures for the Management of Generative Artificial Intelligence Services*. http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Johnson, B. (2024, July 31). *True or False? Addressing Common Assumptions About Copyright and AI*. Copyright Clearance Center. https://www.copyright.com/blog/addressing-common-assumptions-copyright-ai/

Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., & Zhang, W. (2018). The Rise of Artificial Intelligence under the Lens of Sustainability. *Technologies*, 6(4), 100. https://doi.org/10.3390/technologies6040100

Latona, G. R., Ribeiro, M. H., Davidson, T. R., Veselovsky, V., & West, R. (2024). *The AI Review Lottery: Widespread AI-Assisted Peer Reviews Boost Paper Scores and Acceptance Rates* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2405.02150

Li, Y. (2023). Specifics of regulatory and legal regulation of Generative Artificial Intelligence in the UK, USA, EU and China. *Law Journal of the Higher School of Economics*, 3, 245–267. https://doi.org/10.17323/2072-8166.2023.3.245.267

Living guidelines on the responsible use of generative AI in research | Research and innovation. (2024). https://research-and-innovation.ec.europa.eu/document/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

Lloyd, J. W. (1995). Surviving the AI Winter. *Logic Programming: The 1995 International Symposium*, 33–47.

OpenAI and journalism. (2024, January 8). https://openai.com/index/openai-and-journalism/

Peláez-Sánchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024). The impact of large language models on higher education: Exploring the connection between AI and Education 4.0. *Frontiers in Education*, 9, 1392091. https://doi.org/10.3389/feduc.2024.1392091

Pollock, D., & Michael, A. (2024, December 10). *News and Views: How much content can AI legally exploit?* https://www.deltathink.com/news-and-views-how-much-content-can-ai-legally-exploit

Pooley, J. (2024). Large Language Publishing: The Scholarly Publishing Oligopoly's Bet on AI. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, 7(1), 1–11. https://doi.org/10.18357/kula.291

Puttgen, H., & Jansen, J. (1987). Knowledge based systems applied to power systems: A passing fad or a useful tool here to stay? *26th IEEE Conference on Decision and Control*, 408–412. https://doi.org/10.1109/CDC.1987.272830

Rama Padmaja, C. V., & Lakshminarayana, S. (2024). The rise of AI: A comprehensive research review. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 13(2), 2226. https://doi.org/10.11591/ijai.v13.i2.pp2226-2235

Ramdurai, B., & Adhithya, P. (2023). The impact, advancements and applications of Generative AI. *International Journal of Computer Science and Engineering*, 10(6), 1–8. https://doi.org/10.14445/23488387/IJCSE-V10I6P101

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (2024). http://data.europa.eu/eli/reg/2024/1689/oj/eng

Sarker, S., University of Virginia, Susarla, A., Michigan State University, Gopal, R., University of Warwick, Thatcher, J. B., & University of Colorado / University of Manchester. (2024). Democratizing Knowledge Creation Through Human-AI Collaboration in Academic Peer Review. *Journal of the Association for Information Systems*, 25(1), 158–171. https://doi.org/10.17705/1jais.00872

Schonfeld, R. C. (2024, October 15). *Tracking the Licensing of Scholarly Content to LLMs.* The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2024/10/15/licensing-scholarly-content-llms/

Sejnowski, T. J. (2023). Large Language Models and the Reverse Turing Test. *Neural Computation*, 35(3), 309–342. https://doi.org/10.1162/neco_a_01563

Selman, B. (2000). Compute-intensive methods in artificial intelligence. *Annals of Mathematics and Artificial Intelligence*, *28*(1/4), 35–38. https://doi.org/10.1023/A:1018943920174

Shin, D., & Shin, E. Y. (2023). Data's Impact on Algorithmic Bias. *Computer*, 56(6), 90–94. https://doi.org/10.1109/MC.2023.3262909

Singh, S., Vargus, F., D'souza, D., Karlsson, B., Mahendiran, A., Ko, W.-Y., Shandilya, H., Patel, J., Mataciunas, D., O'Mahony, L., Zhang, M., Hettiarachchi, R., Wilson, J., Machado, M., Moura, L., Krzemiński, D., Fadaei, H., Ergun, I., Okoh, I., … Hooker, S. (2024). Aya Dataset: An Open-Access Collection for Multilingual Instruction Tuning. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 11521–11567. https://doi.org/10.18653/v1/2024.acl-long.620

Sonone, S., & Dharme, A. (2019). A review paper on simulated intellect. *International Journal of Physics and Mathematics*, 1(1), 31–33. https://doi.org/10.33545/26648636.2019.v1.i1a.6

Strickland, E. (2021). The Turbulent Past and Uncertain Future of AI: Is there a way out of AI's boom-and-bust cycle? *IEEE Spectrum*, 58(10), 26–31. https://doi.org/10.1109/MSPEC.2021.9563956

Tarisayi, K. S. (2024). Strategic leadership for responsible artificial intelligence adoption in higher education. *CTE Workshop Proceedings*, 11, 4–14. https://doi.org/10.55056/cte.616

Vijayakumar, A. (2023). Potential impact of artificial intelligence on the emerging world order. *F1000Research*, 11, 1186. https://doi.org/10.12688/f1000research.124906.2

Walsh, K. (2023, August 18). *Understanding CC Licenses and Generative AI. Creative Commons*. https://creativecommons.org/2023/08/18/understanding-cc-licenses-and-generative-ai/

Wolff, J., Gordon, S., & Guo, D. (2018). The Rise of Artificial Intelligence. *Advances in Social Sciences Research Journal*. https://doi.org/10.14738/assrj.56.4722

Yi, J., Ye, R., Chen, Q., Zhu, B., Chen, S., Lian, D., Sun, G., Xie, X., & Wu, F. (2024). On the Vulnerability of Safety Alignment in Open-Access LLMs. *Findings of the Association for Computational Linguistics ACL 2024*, 9236–9260. https://doi.org/10.18653/v1/2024.findings-acl.549