Centre for International Governance Innovation

Digital Policy Hub

Digital Policy Hub – Working Paper

Generative AI's Copyright Challenges in Agricultural Extension

Mahatab Uddin

Summer 2024 cohort

About the Hub

The Digital Policy Hub at CIGI is a collaborative space for emerging scholars and innovative thinkers from the social, natural and applied sciences. It provides opportunities for undergraduate and graduate students and post-doctoral and visiting fellows to share and develop research on the rapid evolution and governance of transformative technologies. The Hub is founded on transdisciplinary approaches that seek to increase understanding of the socio-economic and technological impacts of digitalization and improve the quality and relevance of related research. Core research areas include data, economy and society; artificial intelligence; outer space; digitalization, security and democracy; and the environment and natural resources.

The Digital Policy Hub working papers are the product of research related to the Hub's identified themes prepared by participants during their fellowship.

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

Partners

Thank you to Mitacs for its partnership and support of Digital Policy Hub fellows through the Accelerate program. We would also like to acknowledge the many universities, governments and private sector partners for their involvement allowing CIGI to offer this holistic research environment.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Key Points

- Generative artificial intelligence (AI), powered by large language models (LLMs) such as GPT-4 and Llama 2, is applied in various fields such as coding, medicine, law, agriculture and psychology. In some specific applications, its performance is approaching human expertise, suggesting that these applications may be precursors of artificial general intelligence (AGI) systems.
- Generative AI in climate-smart agriculture is still evolving, although AI adoption in this field is already evident. Generative AI can be applied in various agricultural extension and advisory services, including farm mechanization, food processing, water management, crop monitoring and livestock management.
- Most training data sets contain copyrighted works, raising legal questions about their use, especially in agricultural extension services, where training data includes copyrighted images of soil, climate and plant conditions. This issue could be resolved if using these materials for training AI is considered "fair use" as currently defined by jurisprudence, avoiding copyright infringement.
- In case the fair use defence fails, this working paper argues that the adoption of three key recommendations — creating clear data-sharing agreements, implementing remuneration programs such as revenue sharing or royalty payments and using a royalty-based compensation model — could help resolve copyright-related legal disputes and enable the wide-scale application of LLMs, and generative AI in climate-smart agricultural extension and advisory services.

Introduction

Generative AI, run primarily by LLMs, such as GPT-4 and Llama 2, has made significant advancements, demonstrating remarkable capabilities across diverse domains and tasks (Touvron et al. 2023). These models showcase machine intelligence levels surpassing earlier AI models, excelling in fields such as coding, medicine, law, agriculture and psychology, and often operating effectively without specific prompts. In certain domains, their performance approaches human expertise, suggesting they may represent precursors of AGI systems (Bubeck et al. 2023). This paper will first explore potential applications of generative AI and LLMs in agricultural extension and advisory services that require assurance of climate-resilient agricultural practices, enhanced food production and reduced greenhouse gas emissions in the atmosphere (cf. Uddin, Chowdhury and Kabir 2024).

Generative AI raises legal issues related to training data and data output (Susarla 2024). In terms of training data, some generative AI tools are trained using materials scraped from the internet, including copyrighted works, personal information, biometric data and harmful or illegal content. There is ongoing litigation over whether the scraping, downloading and processing of these materials, as well as the trained AI models and their outputs, breach intellectual property (IP) rights, privacy and contract terms. Debates continue regarding the balance of interests between IP owners and AI developers. This working paper aims to address the copyright implications of deploying generative AI in agricultural advisory services.

Agricultural Extension and Advisory Services

Agricultural extension and advisory services refer to any organization in the public or private sectors, including non-governmental organizations (NGOs), farmer groups and private companies, that helps farmers and rural stakeholders access knowledge, information and technology. These organizations also facilitate interactions among different actors and assist in developing technical, organizational and management skills to enhance livelihoods and well-being (Food and Agriculture Organization of the United Nations 2019).

The successful sharing of information, provision of resources for modernizing agriculture, promoting rural development and shaping individuals' perspectives on innovative practices all heavily depend on extension training programs (Kassem et al. 2021). Agricultural extension services include all institutions and initiatives that help those involved in agriculture tackle challenges and gain knowledge, skills and technologies to improve their quality of life and prosperity (Davis, Babu and Ragasa 2020). These services stem from three main sectors (Jayasingh, Ashish and Das 2024, 121):

- Public sector: This includes governmental organizations such as ministries and departments of agriculture, as well as agricultural research centres.
- Private non-profit sector: This comprises local and international NGOs, foundations, community boards, associations, bilateral and multilateral aid projects and other non-profit organizations.
- Private for-profit sector: This includes commercial entities such as input manufacturers and distributors, farmer-operated businesses that both use and share agricultural information, agro-marketing and processing companies, trade associations and private consulting and media firms.

Application of LLMs in Climate-Smart Agricultural Extension and Advisory Services

While the adoption and implementation of AI has already been observable in climate-smart agricultural practice (Uddin, Chowdhury and Kabir 2024), the application of generative AI in agricultural extension and advisory services encompassing climate-smart agricultural practices is still evolving. It is possible to apply generative AI in various sectors of agricultural extension and advisory services: farm mechanization, the food processing industry, agricultural water management, crop monitoring and improvement and livestock production management (Gaddikeri, Jatav and Rajput 2023).

Farm Mechanization

LLMs and generative AI can analyze data from various farm management systems, including inventory, financial and labour management. They also aid in the maintenance of farm machinery and equipment by leveraging sensor data and historical records, which helps save time and improve efficiency. In addition, they assist farmers in planning their operations, for example, by identifying the best times for planting, harvesting and applying fertilizers (ibid., 22).

The Food Processing Industry

LLMs and generative AI can analyze data from sensors and cameras to detect defects in food products and estimate their shelf life. They can generate new recipes and meal ideas based on available ingredients and dietary restrictions. Additionally, they aid in food safety testing by assessing contamination risks and supports supply chain management in areas such as inventory, logistics and finance (ibid.).

Furthermore, LLMs and generative AI can provide nutritional information about food products, answer customer questions regarding ingredients and nutrition, create recipes and menus tailored to specific dietary needs and preferences and generate production reports. By automating these tasks, the time and costs associated with manual data entry are reduced (ibid.).

Agricultural Water Management

By leveraging historical weather and soil moisture data, LLMs and generative AI can predict the water needs of crops. These technologies are also capable of analyzing water samples to detect contaminants, including heavy metals, pesticides and bacteria. Furthermore, they can provide text-based recommendations for effective water management practices (ibid.).

Crop Monitoring and Improvement

Data from crop monitoring systems can be analyzed by LLMs and generative AI to forecast yields, supporting farmers in harvest planning and operational optimization. These technologies also provide guidance on pest and disease management and contribute to crop breeding experiments by identifying traits that boost yields. Additionally, by integrating with weather forecasting systems, LLMs and generative AI assess climate change impacts on crop yields and help devise adaptation strategies (ibid., 23).

Application in Livestock Production Management

LLMs and generative AI can analyze data from feed management systems to optimize feed usage and reduce waste. They can also assist farmers in identifying issues with their animals so that they can take timely corrective actions. Additionally, they support livestock breeding experiments to identify traits that boost productivity and predicts when equipment, such as feeders and water troughs, needs maintenance based on sensor data. Moreover, they can generate automated reports, including animal health assessments, to help farmers track their operations and identify areas for improvement (ibid.).

Copyright Challenges Associated with Generative AI

Output-Related Challenges

Texts generated by an AI language model are generally not protected by copyright, as copyright law typically attributes ownership to the human creator of an original work. However, if the generated text involves significant human input or intervention, it may be considered original and eligible for copyright protection. For instance, if a person contributes substantial creative elements based on the AI's responses — by editing, commenting, analyzing or incorporating it into a larger project — the individual who added such content would typically hold the copyright for the final work (Lucchi 2023, 7).

In the application of LLMs or generative AI to climate-smart agricultural extension and advisory services, copyright protection for the final work (for instance, advice regarding farming or any farming-related activities) will not necessarily be an impediment for end-user farmers. This is because farmers are meant to receive the advice and suggestions directly from the copyright owner (if the final work qualifies for copyright protection) and use the same by acquiring the prior informed consent of the relevant copyright owner. However, in adopting and applying LLMs or generative AI to climate-smart agricultural extension and advisory services, actual copyright-related challenges will arise if copyright-protected materials are used for training relevant AI systems (namely input-related challenges), as described below.

Input-Related Challenges

A big gap exists regarding the IP management, especially copyright management, of data inputs, meaning the management of IP issues pertaining to the data used for training relevant AI systems (Strowel 2023, 491; Franceschelli and Musolesi 2022; Bonadio, Dinev and McDonagh 2022; Quang 2021, 1407; Sobel 2021; Abbamonte 2021). The main question in this regard is to determine whether using any copyright-protected material in training an AI system amounts to infringement of copyright (Lucchi 2023, 11). Machine learning relies on extensive training data to deliver accurate results, particularly in areas such as natural language processing, facial recognition, translation and stop sign detection (Brown et al. 2020; Radford et al. 2019, Devlin et al. 2018). This is especially crucial for Chat GPT, which depends on vast amounts of data for effective performance. To generate an interactive and reliable output, Chat GPT needs to gather information from publicly accessible websites on the internet, including text, images and other content (Lucchi 2023, 11). Supporting the training of AI algorithms requires the adoption and application of various techniques such as text and data mining (TDM) and generative deep learning approaches (Franceschelli and Musolesi 2022). TDM processes involve extracting and analyzing extensive data sets to reveal valuable insights and patterns (Alpaydın 2004, 2). These insights and patterns can subsequently be used to improve the performance of AI models (ibid.).

TDM is an essential tool in AI, helping researchers and data scientists analyze large amounts of unstructured data to extract useful information that would be difficult to gather manually. By examining these data sets, AI algorithms can identify patterns and make accurate predictions, leading to new content, discoveries and innovations. Without access to large volumes of data, AI would find it challenging to learn and improve. Therefore, the future of AI relies on TDM's ability to process data on a large scale. However, a significant challenge is that AI systems require exact copies of artwork in their training data (Lemley and Casey 2021, 743); this means creating a training set with millions of examples by replicating copyrighted images, videos, audio or text-based works (Lucchi 2023, 11). Therefore, using Chat GPT or similar models for agricultural extension and advisory services will necessitate creating a training set with millions of examples by replicating copyrighted images, videos, audio and text related to agricultural science and education. As a result, AI training necessitates the question as to whether machine copying should be considered fair use or covered by other copyright exceptions.

While it is true that some large data sets are purely informational and not protected by copyright, most training data sets contain copyrighted works. For example, the collections used to develop AI algorithms for text, facial recognition and image recognition include copyrighted materials. This raises the question as to whether using these works is legal and what conditions or restrictions could or should be entailed (Lucchi 2023, 12). This issue is particularly relevant in the adoption and application of LLMs and generative AI in agricultural extension services because input or training data for this kind of generative AI necessarily requires using copyright-protected images of soil conditions, climate conditions, plant conditions and various images connected with the agricultural sector. It is possible to resolve this challenge if uses of copyright-protected materials for training AI systems are considered as fair use, meaning that they do not amount to infringement of the relevant right holders' copyright. The following two sections of this paper discuss whether the fair use argument is applicable to uses of copyright-protected materials for training AI systems. The discussion will cover two major jurisdictions in the Western Hemisphere — the United States and the European Union.

Fair Use or Not: The Example of the United States

At present, collecting data for TDM is considered as fair use in the United States (Carroll 2019, 894; Lemley and Casey 2021, 746). In the case of *Authors Guild v Google, Inc* (2015),[1] Google Books obtained permission for searching entire libraries in order to provide search functions as well as excerpts from books. However, it is not yet clear if these conclusions are applicable to data collection and input for machine learning. Accordingly, there is no guarantee that the courts will extend this precedent to similar technologies, including those involved in agricultural extension or advisory services (Lemley and Casey 2021, 763). In the United States, data collection for TDM may be permitted if it is seen as a transformative use, but it is not always clear when one copyrighted work is being transformed into another different and distinctive copyrighted work.[2] Furthermore, in the *Authors Guild v Google, Inc* case, the court determined that Google's digitization of copyrighted books to create a comprehensive index and improve search functionality constituted fair use.[3]

1 *Authors Guild v Google, Inc*, 804 F Supp (3d) 202 (2015) at 214–15.

2 A transformative use is a kind of use that "alter[s] the first [work] with new expression, meaning, or message." See *Campbell v Acuff-Rose Music, Inc*, 510 US 569, 579 (1994).

3 *Authors Guild v Google, Inc*, 770 F Supp. (2d) at 207–8.

The digitization process was designed to help users find and access copyrighted books, serving as a valuable resource for researchers and members of the public without competing with original works. In contrast, generative AI allows users to create content that directly competes with original material, producing new works based on existing content. While Google's indexing complies with fair use, the accessibility of generative AI raises complex copyright challenges as it enables the creation of derivative works that may impact the market for original content (Lucchi 2023, 13). For this reason, numerous court cases are currently under way in the United States to clarify the definitions of "derivative work" and "transformative use" under IP law, especially regarding copyrighted material used to train AI systems.[4]

Various famous generative AI platforms such as OpenAI are currently facing lawsuits for alleged copyright infringement in connection with training their AI systems using data sets that their creators have not legally acquired.[5] In *Tremblay et al v OpenAI, Inc*,[6] the plaintiffs argue that OpenAI used their copyrighted material without prior authorization in order to train the AI system for Chat GPT. It has been claimed that Chat GPT can effectively summarize a range of books, indicating that the chatbot has fully engaged with and grasped the content of these works.[7]

In the case of *Silverman v Open AI Inc*, the plaintiffs similarly allege that OpenAI improperly used copyrighted material, specifically the book *The Bedwetter*, in order to train Chat GPT.[8] The claimants of this class action suit are a group of authors who assert that Chat GPT can generate summaries of their novels when given an appropriate prompt, thereby demonstrating its familiarity with their content, because Chat GPT's AI system was trained using their copyrighted materials.[9]

In the case of *Getty Images (US), Inc v Stability AI, Inc*, the prominent photo agency alleges that the developer of the AI art tool Stable Diffusion illegally scraped a significant number of its images.[10] This was reportedly done to train the system without obtaining the necessary permissions and/or compensation to Getty Images. Additionally, Stable Diffusion generated a modified version of Getty's watermark, aimed at concealing and facilitating copyright infringement. According to the plaintiffs, this modified watermark also violates the Digital Millennium Copyright Act concerning copyright management information.[11]

Since these cases are still pending, the legal confusion regarding LLMs' and generative AI's ongoing training methods remain unresolved. However, the above-stated cases signify the first major legal challenges related to AI and copyright (Lucchi 2023, 14). Until these cases are resolved, possibilities for the adoption and application of LLMs or generative AI in agricultural extension and advisory services will remain uncertain.

4 See, for example, *Getty Images (US), Inc v Stability AI, Inc*, No. 1:23-cv-00135-GBW (D Del Mar. 29, 2023) [*Getty Images*]; *Silverman v OpenAI, Inc*, No. 4:23-cv-03416 (ND Cal 7 July 2023); *Tremblay v OpenAI, Inc*, No. 4:2023-cv-03223 (ND Cal 7 July 2023).

5 *Ibid*. It is expected that the resolution of these litigations will depend on the interpretation of the fair use doctrine.

6 *Tremblay v OpenAI, Inc*, No. 4:2023-cv-03223 (ND Cal 7 July 2023).

7 *Ibid*.

8 *Silverman, supra* note 4.

9 *Ibid*.

10 *Getty, supra* note 4 *Images (US), Inc. v. Stability AI, Inc*., No. 1:23-cv-00135-GBW (D. Del. Mar. 29, 2023).

11 See *Digital Millennium Copyright Act*, 17 USC, s 1202(b). Specifically, the plaintiffs argued that the defendant's actions violated the provisions of the Digital Millennium Copyright Act because it altered or removed copyright management information (CMI) embedded in the plaintiffs' images and instructed the AI system to exclude any CMI from its generated output.

A Recent Case of Concern

A recent US Supreme Court ruling on a non-technological case, *Andy Warhol Foundation for the Visual Arts, Inc v Goldsmith*, has raised concerns about negative impacts on the IP rights of AI-generated works.[12]

The petitioner, the Andy Warhol Foundation for the Visual Arts, Inc. (AWF), manages the works of artist Andy Warhol. Respondent Lynn Goldsmith, a photographer, licensed a photo of the musician Prince to the magazine *Vanity Fair* in 1984. Warhol then used this photo to create an illustration and a series of prints known as the "Prince Series." After Prince's death in 2016, Condé Nast approached AWF to republish the illustration for a special edition but ended up licensing a different piece from the series. Goldsmith later claimed copyright infringement, leading AWF to file a declaratory judgment action, with Goldsmith countersuing. The district court ruled in favour of AWF, stating Warhol's use was fair use, but the Second Circuit reversed this decision in 2021. AWF then petitioned for certiorari on the fair use factor, which the US Supreme Court granted.[13] The Supreme Court determined that the licensing of Andy Warhol's "Orange Prince" portrait to Condé Nast did not constitute a transformative use of Lynn Goldsmith's reference photo because "Goldsmith's original photograph of Prince, and the Andy Warhol Foundation's (AWF) copying use of that photograph…share substantially the same purpose, and the use is of a commercial nature."[14]

While the AWF invoked fair use to justify creating derivative works, the Supreme Court ruled that the foundation had no fair use defence for licensing a derivative version of the photograph for commercial purposes.[15] This recent decision could lead to significant limitations on the transformative use doctrine, as the Supreme Court seems to have narrowed its scope (Patry 2023).

Hence, it will be intriguing to observe how US courts apply the rules from this case to AI training input licensing. If a court determines that data ingestion — namely collecting and modifying raw data for AI training — constitutes infringement, it could create significant legal challenges for AI systems. Most data used by generative AI, including text and images, has been obtained without explicit permission from rights holders. The issue at hand is whether using copyrighted works for training data constitutes infringement or if the training purpose is distinct enough to qualify for a fair use defence (Lucchi 2023, 14–15).

Fair Use or Not: The Example of the European Union

The European Union has adopted a protectionist stance, ensuring accountability for the use of training data. Article 4(1) of the Directive on Copyright in the Digital Single Market[16] offers a broad exception for TDM. This enables individuals, such as commercial AI developers and educators, to create copies of works or databases for information extraction and retain them as necessary for AI training.[17] However, rights holders can exclude TDM exemptions in their contracts with miners (entities involved in TDM) to protect their commercial interests.[18]

12 *Andy Warhol Foundation for the Visual Arts, Inc v Goldsmith*, 143 US 1258 (2023).

13 *Ibid*.

14 *Ibid*. at 38.

15 *Ibid*. at 2.

16 EC, *Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*, [2019], OJ L 130/92, art 4(1), online: <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

17 *Ibid*. at arts 4(1) and 4(2).

18 *Ibid*. at art 4(3).

This provision has garnered considerable criticism for providing a copyright exception that is viewed as being too restrictive (Lucchi 2023, 15). Unlike the traditional understanding of copyright as primarily protecting original expression, this provision appears to cover factual information and data, which has resulted in significant backlash (Margoni and Kretschmer 2022). Nonetheless, the implementation of this option to opt out, and the extent to which AI developers will adhere to it, are still uncertain (Lucchi 2023, 14–15).

Another issue related to data aggregation involves the implementation of EU data protection laws (Hacker, Engel and Mauer 2023, 1112–23). Data aggregation is crucial for training and improving generative AI models (Lucchi 2023, 15). It involves collecting and combining large amounts of data from various sources to enhance a model's performance and capabilities (ibid.). However, processing personal data in the European Union must follow strict rules set by the General Data Protection Regulation. These challenges have not been thoroughly examined in both legal and policy contexts and require further investigation and resolution (ibid.).

Anti-trust/Competition Issues

Large, established companies such as Google, Facebook, Amazon and OpenAI have extensive collections of language and image data for AI applications (Foster 2019, 1). This access gives them a significant competitive edge in the AI industry. Consequently, they can leverage their data sets to train and develop more advanced AI models, improving their products and services. For new entrants into the industry, this poses legal challenges, as the ownership and licensing of data sets can be intricate and are subject to IP rights, privacy regulations and other legal factors.

Moreover, the expense of creating or licensing a data set "from scratch" can be substantial, making it challenging for smaller companies to compete with established players (Lucchi 2023, 12).

Furthermore, anti-trust concerns may arise if dominant market players control access to the data sets essential for developing AI models, potentially stifling innovation and competition. Therefore, ensuring fair and open access to training data is a vital legal issue in the progress and deployment of AI technology.

Recommendations

- **Data-sharing agreement:** Creating clear data-sharing agreements with data providers can be an effective step for any LLM- or generative AI-based service provider, including agricultural extension service providers (cf. Leistner and Antoine 2022). These agreements can help navigate the complexities of using protected content for AI training while ensuring compliance with copyright laws and safeguarding content owners' rights (European Commission 2020). The agreements would define the scope of data usage, outline limitations, specify necessary permissions and arrange the required licences for utilizing copyrighted material in AI training (Kop 2021). Data-sharing agreements offer AI developers a legally binding framework for overseeing access to and use of protected content during AI training. These agreements specify authorized data usage and ensure adherence to agreed terms. They also help to identify non-viable data and establish criteria for acceptable content use, including allowed AI algorithms and guidelines

for data timing and duration. In addition, these agreements often include clauses that restrict data usage, preventing developers from retrieving, repurposing or monetizing protected material beyond the initial scope (Lucchi 2013, 18).

- **Adopting remuneration programs:** Another effective strategy for protecting AI training data is to implement remuneration programs, such as revenue sharing or royalty payments, to ensure that creators of copyrighted materials used in AI systems receive fair compensation (Senftleben 2023; Frosio, forthcoming 2025). This strategy is essential for acknowledging the value of copyrighted content and ensuring that creators receive fair compensation for their works used in AI systems. By adopting revenue-sharing or royalty structures like any other generative AI developers, the LLM- or generative AI-based agricultural advisory and extension service providers can directly connect financial gains to the use of copyrighted materials. This creates a strong incentive for content creators to provide their work as training data, allowing them to benefit from the system's success. In a revenue-sharing arrangement, creators would receive a predetermined share of the revenue generated by the AI system based on their contribution to the training data, ensuring that they are rewarded for the value that their work brings to the AI's functionality (Lucchi 2023, 18).

- **Royalty-based compensation model:** LLM- or generative AI-based agricultural extension and advisory service providers can also introduce a royalty-based compensation model, which will allow content creators to receive a set fee for each instance that an AI system uses their copyrighted works. This fee structure may consist of a fixed amount per use or a percentage of the revenue generated by the AI. Such a model ensures that creators are fairly compensated for the entire duration that their works are utilized, directly tying the fee to their usage (ibid.). Implementing revenue sharing or royalty structures necessitates clear agreements between AI developers and content creators that define the terms of compensation. These agreements must specify how revenue sharing or royalties are calculated, as well as their frequency and duration. Transparent remuneration mechanisms safeguard the interests of both parties, fostering a fair and sustainable ecosystem for the use of copyrighted works in AI systems. Overall, compensating content creators through these models recognizes the value of their contributions and ensures they receive fair rewards for their role in the success of generative AI tools, promoting a mutually beneficial relationship and equitable practices (ibid.). In short, the adoption of any or all of these three suggestions would help in removing copyright-related legal disputes and open the scope for a wider-scale application of LLMs or generative AI in climate-smart agricultural extension and advisory services.

Conclusion

The successful deployment of generative AI in agricultural advisory and extension services relies heavily on addressing copyright-related provisions, particularly those concerning TDM for generative AI. However, a comprehensive legal framework in this area has yet to be established. While the European Union has adopted a protectionist stance by ensuring accountability for the use of training data, the situation in the United States remains uncertain, pending the outcomes of several ongoing court cases. Once these cases are resolved, the extent to which investors and stakeholders in the generative AI sector can rely on the fair use defence for using copyrighted materials to train their AI systems will become

clearer. Regardless of the viability of this defence, generative AI developers can mitigate legal risks by considering strategies such as establishing data-sharing agreements, adopting remuneration programs and implementing royalty-based compensation models. These approaches can help safeguard against copyright infringement claims and support the effective deployment of generative AI in agricultural advisory and extension services in the face of evolving copyright laws.

Acknowledgements

The author would like to thank Ataharul Chowdhury, Paul Samson and Matthew da Mota.

About the Author

Mahatab Uddin is an expert on climate change law, IP law, technology transfer and sustainable development. He is an adjunct professor and post-doctoral researcher at the School of Environmental Design and Rural Development at the University of Guelph, Canada, where he researches possible legal frameworks for AI-run climate-smart agricultural practices — research he expanded on during his fellowship with the Digital Policy Hub. He holds a Ph.D. in public international law on climate change and sustainable development at Aarhus University, Denmark; an M.Sc. in sustainable development from Uppsala University, Sweden; and an LL.M. in IP law from Stockholm University, Sweden.

Works Cited

Alpaydın, Ethem. 2004. *Introduction to Machine Learning*. Cambridge, MA: MIT Press.

Abbamonte, Giuseppe B. 2021. “The Rise of the Artificial Artist: AI Creativity, Copyright and Database Right.” *European Intellectual Property Review* 43 (8): 702–10. https://tind.wipo.int/record/45277?ln=en.

Bonadio, Enrico, Plamen Dinev and Luke McDonagh. 2022. “Can Artificial Intelligence Infringe Copyright? Some Reflections.” In *Research Handbook on Intellectual Property and Artificial Intelligence*, edited by Ryan Abbott, 245–57. Cheltenham, UK: Edward Elgar.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. 2020. “Language Models are Few-Shot Learners.” Preprint, arXiv, May 28. https://doi.org/10.48550/arXiv.2005.14165.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee et al. 2023. “Sparks of Artificial General Intelligence: Early Experiments with GPT-4.” Preprint, arXiv, March 22. https://doi.org/10.48550/arXiv.2303.12712.

Carroll, Michael W. 2019. “Copyright and the Progress of Science: Why Text and Data Mining Is Lawful.” *UC Davis Law Review* 53: 893–964. https://lawreview.law.ucdavis.edu/sites/g/files/dgvnsk15026/files/media/documents/53-2_Carroll.pdf.

Davis, Kristin E., Suresh Chandra Babu and Catherine Ragasa, eds. 2020. *Agricultural Extension: Global Status and Performance in Selected Countries: Synopsis*. Washington, DC: International Food Policy Research Institute.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2018. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” Preprint, arXiv, October 11. https://doi.org/10.48550/arXiv.1810.04805.

European Commission. 2020. *Towards a European strategy on business-to-government data sharing for the public interest*. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing. Luxembourg: Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/d96edc29-70fd-11eb-9ac9-01aa75ed71a1/language-en.

Food and Agriculture Organization of the United Nations. 2019. *The international Code of Conduct for the sustainable use and management of fertilizers*. Rome, Italy: Food and Agriculture Organization of the United Nations.

Foster, David. 2019. *Generative Deep Learning: Teaching Machines to Paint, Write, Compose and Play*. Newton, MA: O’Reilly.

Franceschelli, Giorgio and Mirco Musolesi. 2022. “Copyright in generative deep learning.” *Data & Policy* 4: e17. https://doi.org/10.1017/dap.2022.10.

Frosio, Giancarlo. Forthcoming 2025. “Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?” In *A Research Agenda for EU Copyright Law*, edited by Enrico Bonadio and Caterina Sganga. Cheltenham, UK: Edward Elgar.

Gaddikeri, Venkatesh, Malkhan Singh Jatav and Jitendra Rajput. 2023. “Revolutionizing Agriculture: Unlocking the Potential of ChatGPT in Agriculture.” *Food and Scientific Reports* 4 (3): 20–25. www.researchgate.net/publication/369977080_Revolutionizing_Agriculture_Unlocking_the_Potential_of_ChatGPT_in_Agriculture.

Hacker, Philipp, Andreas Engel and Marco Mauer. 2023. “Regulating ChatGPT and Other Large Generative AI Models.” In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1112–23. New York, NY: Association for Computing Machinery.

Jayasingh, Debi Kalyan, Anand Ashish and Kiran Sourav Das. 2024. "Artificial Intelligence in Agricultural Extension." In *Innovative Agricultural Strategies and Concepts in Extension*, edited by Jyotishree Anshuman and Biswajit Mallick, 115–29. New Delhi, India: AkiNik Publications.

Kassem, Hazem S., Bader Alhafi Alotaibi, Muhammad Muddassir and Ahmed Herab. 2021. "Factors influencing farmers' satisfaction with the quality of agricultural extension services." *Evaluation and Program Planning* 85: 101912. https://doi.org/10.1016/j.evalprogplan.2021.101912.

Kop, Mauritz. 2021. "The Right to Process Data for Machine Learning Purposes in the EU." *Harvard Journal of Law & Technology* 34: 1–40. https://jolt.law.harvard.edu/assets/digestImages/Mauritz-Kop_The-Right-to-Process-Data-for-Machine-Learning-Purposes-in-the-EU.pdf.

Leistner, Matthias and Lucie Antoine. 2022. *IPR and the use of open data and data sharing initiatives by public and private actors*. Brussels, Belgium: European Parliament. www.europarl.europa.eu/RegData/etudes/STUD/2022/732266/IPOL_STU(2022)732266_EN.pdf.

Lemley, Mark A. and Bryan Casey. 2021. "Fair Learning." *Texas Law Review* 99 (4): 743–78. https://texaslawreview.org/fair-learning/.

Lucchi, Nicola. 2023. "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems." *European Journal of Risk Regulation* 1–23. https://doi.org/10.1017/err.2023.59.

Margoni, Thomas and Martin Kretschmer. 2022. "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology." *GRUR International: Journal of European and International IP Law* 71 (8): 685 –701. https://doi.org/10.1093/grurint/ikac054.

Patry, William. 2023. "*Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*: did the U.S. Supreme Court tighten up fair use?" *Journal of Intellectual Property Law & Practice* 18 (9): 628–41. https://doi.org/10.1093/jiplp/jpad060.

Quang, Jenny. 2021. "Does Training AI Violate Copyright Law?" *Berkeley Technology Law Journal* 36 (4): 1407–50. https://doi.org/10.15779/Z38XW47X3K.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei and Ilya Sutskever. 2019. "Language Models are Unsupervised Multitask Learners." OpenAI. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Senftleben, Martin. 2023. "Generative AI and Author Remuneration." *International Review of Intellectual Property and Competition Law* 54: 1535–60. https://doi.org/10.1007/s40319-023-01399-4.

Sobel, Benjamin. 2021. "A Taxonomy of Training Data: Disentangling the Mismatched Rights, Remedies, and Rationales for Restricting Machine Learning." In *Artificial Intelligence and Intellectual Property*, edited by Jyh-An Lee, Reto Hilty and Kung-Chung Liu, 221–42. Oxford, UK: Oxford University Press.

Strowel, Alain. 2023. "ChatGPT and Generative AI Tools: Theft of Intellectual Labor?" *International Review of Intellectual Property and Competition Law* 54 (1): 491–94. https://doi.org/10.1007/s40319-023-01321-y.

Susarla, Anjana. 2024. "Generative AI could leave users holding the bag for copyright violations." *The Conversation*, March 22. https://theconversation.com/generative-ai-could-leave-users-holding-the-bag-for-copyright-violations-225760.

Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov et al. 2023. “Llama 2: Open Foundation and Fine-Tuned Chat Models.” Preprint, *arXiv*, July 18. https://doi.org/10.48550/arXiv.2307.09288.

Uddin, Mahatab, Ataharul Chowdhury and Ashad Kabir. 2024. “Legal and ethical aspects of deploying artificial intelligence in climate-smart agricultural technologies.” *AI & Society* 39 (3): 221–34. https://doi.org/10.1007/s00146-022-01421-2.