

Generative AI & IP Rights

GLORIA LEE, Pomona College, USA

This paper examines the recent emergence of generative artificial intelligence (AI) and its relation to intellectual property (IP) rights, specifically focusing on the training, output, and applications of these models. The lack of specific IP laws and regulations for generative AI has led to the recent rise of court cases, such as *Getty Images v. Stability AI*, that questions the lawfulness of current training processes for generative AI models. Along with surveying the legal landscape, this paper also explores the technical advancements in algorithms and attacks for the detection of misuse and protection of IP rights. In relation to generative AI outputs, the paper addresses the idea of ownership and what it means for a party to be granted intellectual property rights. Current applications of generative AI are also examined in the context of journalism and academia. Through these investigations, this paper seeks to highlight the urgent need for legal frameworks specifically addressing generative AI and IP and brings to light complex ethical considerations for the future of generative AI.

CCS Concepts: • **Social and professional topics** → **Intellectual property**.

Additional Key Words and Phrases: Generative AI, Copyright, Academic Integrity, OpenAI

ACM Reference Format:

Gloria Lee. 2023. Generative AI & IP Rights. 1, 1 (December 2023), 7 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Artificial intelligence (AI) has become a prominent tool in various sectors of our society, including technology, healthcare, and entertainment fields. With the ability to easily synthesize a large volume of information, artificial intelligence provides companies and individuals quick automated services, ranging from fraud detection to medical diagnoses, that offer a new level of convenience and reasoning capabilities. Generative AI (GenAI), a subset of artificial intelligence, is especially unique because of its ability to generate new original content, and with its recent rise in popularity, generative AI models have seen increased use by the general public. Outside of commonly used chatbots like OpenAI's ChatGPT, generative AI has provided a variety of functionalities, such as image generation (e.g., DALL-E) and code generation (e.g., Github Pilot). The potential of these models as tools that generate original and intelligent outputs seems to be astronomical, possibly changing the game for many fields.

Yet, many ethical considerations arise with the rapid development of generative AI models. By creating content based on existing works, these models pose critical social and legal questions regarding intellectual property (IP). These models have sparked active dispute in several communities of artists and writers who worry about potential copyright infringement of their work in the models' training data. Discussions about what makes an output truly original and how to systematically check for its authenticity and guard against plagiarism continue to revolve within the field. Legal action against generative AI companies (e.g., *Getty v. Stability AI*) are also on the rise, pressuring court systems to start reevaluating current laws within the context of generative AI.

Author's address: Gloria Lee, ggla2020@mymail.pomona.edu, Pomona College, 333 N College Way, Claremont, California, USA, 91711.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

1

Along with ethical concerns about training data, there are various discussions and perspectives on how to properly attribute intellectual property rights for generated outputs. Questions regarding who should be held responsible and reliable for the outputs of generative AI models persist, and authorship for automated articles is debated, along with no clear guidelines about how collaborations with generative AI should look like in the field of journalism. In schools and universities, generative AI chatbots like ChatGPT have recently presented new challenges regarding academic honesty and plagiarism as well. Institutions and educators struggle to deal with the increased use of generative AI by students, and academic integrity policies are currently unprepared to effectively address the situation.

1.1 Structure of This Paper

Reviewing the recent literature, this paper seeks to explore the current status of copyright and patent laws regarding generative AI, highlights relevant current events, and discusses general recommendations for the future of generative AI regarding intellectual property rights. It explores questions like “How should we ethically go about using preexisting works for training generative AI models?” and “Should generative AI models be given IP rights, and if so, which party should be held reliable for these models’ output?”. The review is organized in three large sections focusing on (1) IP rights for generative AI training data, (2) IP rights for generative AI outputs, and (3) current applications of generative AI, along with the ethical considerations behind its usage in a couple of fields.

2 GENERATIVE AI TRAINING DATA & IP RIGHTS

Using various machine learning techniques, generative AI models have the ability to scrape and parse through extremely large amounts of information to identify patterns and create new content based on their analysis. Because of their ability to generate original content, many ethical concerns regarding intellectual property rights emerge with the training process, prompting questions of whether or not these models are “stealing” other people’s work without consent or proper citations. Currently, there are no IP laws specific to generative AI and use of training data. Hence, lots of backlash, especially from the art community has evolved, leading to the formation of coalitions like EGAI, the European Guild for AI Regulation, to go against the use of art by artificial intelligence companies without the artist’s consent. There have also been efforts by creators to develop “opt-out” tools, which they expect companies to accommodate, and websites like “haveibeentrained.com” that track which works have been used in popular training sets [10].

2.1 Legal Action and Resistance

Along with the resistance and action from creators, there have also been at least five class action litigation cases against generative AI companies concerning copyright infringement, unfair competition, trademark infringement, and data protection. Stability AI, an open-source generative AI company offering audio, language, and image models, has received a lot of backlash from both creators and companies with a recent lawsuit by Getty Images. Claiming trademark infringement, Getty Images is suing Stability AI for its ability to replicate Getty’s trademarks in the outputs of Stability Diffusion. Along with replicating their trademark, Getty Images is claiming that Stability AI had used over 12 million copyright images for model training without proper consent. A similar argument for the misuse of copyrighted works is found in multiple other cases. Along with large institutions like Getty Images, individual creators are also forming class action lawsuits to fight against generative AI companies. Suspicious of ChatGPT’s ability to summarize and answer specific questions unique to their work, book authors Tremblay and Awad are currently taking legal action against OpenAI with the claim that ChatGPT was trained on their books without proper consent. Their case is simply one of

Table 1. Relevant Court Cases

Relevant Cases	Year	Reason for legal action	Case Ruling
Thomas Reuters v. ROSS	2023 - present	Copyright infringement, fair use: Thomas Reuters is suing ROSS for unlawful use of their copyright works for training ROSS's AI models	Awaiting jury trial
Getty Images v. Stability AI	2023 - present	Copyright infringement, trademark infringement, fair use: Getty Images is suing Stability AI for unlawful use of copyright works and trademark replications in Stability AI's generative AI outputs.	Awaiting ruling
Tremblay et al v. OpenAI	2023 - present	Copyright infringement: Putative class action lawsuit suing OpenAI for unlawful use of copyrighted books for training AI models	Awaiting ruling
Andersen et al v. Stability AI	2023 - present	Copyright infringement, rights of publicity: Class action lawsuit formed by artists suing Stability AI for its unlawful use of copyrighted works	Awaiting ruling
Doe et al v. GitHub, Inc.	2022 - present	Copyright infringement, breach of license agreement: Class action lawsuit against GitHub Copilot from a group of anonymous programmers suing for unlawful use of code in developing AI program	Awaiting ruling
Google LLC v. Oracle America, Inc.	2010 - 2021	Copyright infringement: Oracle America sued Google for unlawful use of their Java API for the development of Google's Android platform	Ruled in favor of Google
Authors Guild, Inc. v. Google, Inc.	2005 - 2015	Copyright infringement: Authors sued Google for unlawful use of copyrighted books for its Library Project	Ruled in favor of Google

the many that has recently emerged challenging large tech companies on their practices in training their AI models. See Table 1 for more relevant cases.

Without clear regulations regarding generative AI, most of the legal decisions and directions will likely take place in the court setting until policies are adjusted to fit the current context, and past cases will also set the precedent for future ones. In 2015, the Authors Guild, Inc. v. Google, Inc. ruled in favor of Google after authors filed suit against the company for copying their copyrighted works to create a searchable database for their Library Project. Concluding that the practice was non-infringing, the courts decided that the use was transformative enough and not exposing significant aspects of the original works. With current cases on generative AI, precedents like the Google Books case may play a big role and advantage for AI companies regarding copyright. Looking towards the future, it may also be challenging for parties to provide sufficient evidence against generative AI (i.e. How do artists prove an artistic style?), and open license practices may make it hard to prove that AI systems are indeed 'stealing' people's work. [10]

2.2 Technical Advancements

Along with rising legal discourse on intellectual property rights with generative AI, there have been technical advancements in the field investigating various algorithmic ways to possibly detect misuse of works from the training data. With generative AI models becoming increasingly more complex, it is especially important to develop technological tools to help keep the outputs of these models in check. Interested in the ownership of generated AI outputs, Avrahami and Tamir looked into possible mathematical ways to determine if an image output from a generative AI model is copied from a source in the training data. Their proposed algorithm utilizes various classifiers and discrimination models, such as the nearest neighbor algorithm, to objectively gauge how similar an output is to a source. It also uses preexisting algorithms like the LPIPS (Learned Perceptual Image Patch Similarity) to gauge how similar the content of two images are. The algorithm then compares the LPIPS score of the output image and its 1-nearest neighbor with a predefined threshold to determine whether the output is original or copied [1].

While the algorithm by Avrahami and Tamir presents a new algorithmic way to detect possible misuse of training data, it is significantly limited and highly dependent on factors that are currently undetermined, such as the predefined threshold. Therefore, it is critical to first establish these parameters and explicitly define what is considered as misuse. Updates to our legal system and the policies concerning copyright and intellectual property are of the essence for the field to effectively determine these parameters and utilize these algorithms in the future.

Along with detection algorithms, researchers have also focused on examining and developing ways to protect intellectual property rights (IPR) in generative AI models. Using a variety of datasets and generative AI models, Zhong et al. performed a series of three experiments testing the effectiveness of adversarial attacks, watermarking, and attribution in protecting IPR and copyrights. Known as a method of generating incorrect outputs, adversarial attacks are explored in this context through the use of adversarial watermarks. Modifying the inputs of their model, Zhong et al. examined the effectiveness of placing watermarks on copyrighted images to protect against unlawful use in generating output images. They found that watermarks used as adversarial attacks were effective in causing significant disturbances and defending against the unauthorized use of images in generative AI models. Along with adversarial watermarks, Zhong et al. also found that classifiers using spectral domain fingerprints—identifiers that contain spectroscopic information distinct to each image—were highly effective for the attribution of output images from generative AI models. With attribution, people are able to distinguish between real and fake images, and this property allows for more protection of copyrighted works [11].

With the advancement of these detection algorithms and technical tools, copyright protection can be more easily attained, especially in the context of training generative AI. The past literature informs us of the effectiveness of these technological advancements and aids us in understanding where we are at in protecting IPR and copyrights in input images and generative AI models. Examining current practices also prompts more discussion on where the thresholds exactly lie and what exactly constitutes copyright infringement. Looking towards the future, the hope is that these inquiries and discussions will place more pressure on institutions and our legal system to address issues regarding generative AI training more directly.

3 GENERATIVE AI OUTPUTS & IP RIGHTS

Focusing on the outputs of generative AI, there are a lot of considerations regarding whether or not generative AI outputs should be given intellectual property rights, and if so, to whom these rights should be awarded. With the release and growing use of complex models like GPT-3, our assumptions and conception of plagiarism may also need to

be reformed as the development of advanced language models have seen tremendous increases in parameter size (i.e., 175 billion) and evolved to produce unique, nuanced outputs [2]. These developments emphasize the need to address important ethical questions that should act as the precursors to future regulations. Research focusing on generative AI outputs mainly addresses what it means to award intellectual property rights to a party, critiques current practices and policies, and provides recommendations on how generative AI outputs should be treated.

3.1 Ownership

Over the past few decades, artificial intelligence programs have developed to become computationally creative systems, sparking the question of whether intellectual property protections should be afforded for creations that are made entirely or partly with these systems [9]. Considering all the potential parties, researchers Omri Avrahami and Bar Tamir propose six potential candidates for ownership of generative AI outputs: (1) owner of training data, (2) collector of training data, (3) developer of program, (4) user generating outputs, (5) the program itself, and (6) public domain [1]. In doing so, they acknowledge that ownership of these creative models is often ambiguous and overlapping, and there may be more than one party reliable for the outputs of generative AI models with its tremendous parameter size and large creative potential.

When considering ownership, it is also important to recognize what it means for intellectual rights to be awarded to parties. In the context of copyright, creators are credited for their works and are protected from potential misuse or profit of their original works by others. Yet, along with these protections, copyrights also hold creators reliable and responsible for their creations. With generative AI, it is not possible for a non-human program to be held responsible for its outputs [4]. This is also true in the context of patent law, which historically has been tied to a human inventor. Yet, when attributing intellectual property rights to individual users, creators of training sets, or companies of generative AI, attribution can become complicated as outputs become more creative and far removed from the individual components themselves (e.g., algorithm, training data, or user prompt).

3.2 Current Legal Practices

Similar to regulations regarding the training of generative AI models, there are no legal policies in the U.S. specific to the outputs of generative AI models. Current copyright and patent laws have not been amended to address the developments in generative AI, and most of the legal action has taken place in the courts. Because of the lack of specific regulations in the U.S, generative AI outputs are generally being considered as part of the public domain, not being strictly tied to a particular individual or party.

Looking outside of the U.S, similar patterns are found in other countries like Japan and Indonesia, where AI systems are not specifically regulated outside of normal copyright and patent laws [8]. Although there are differences in the legal structures of these two countries, the common lack of regulations directly focusing on generative AI emphasizes the need for governments to amend current regulations to effectively address the growing use and possible future issues regarding generative AI. With artificial intelligence systems making choices that greatly impact human lives, it is only natural that our legal system also changes to account for these developments and keep these systems accountable.

4 CURRENT APPLICATIONS OF GENERATIVE AI

Generative AI has been used across multiple fields, becoming an increasingly important tool in our everyday lives. Ranging from creating video game animations to processing natural languages, generative AI applications have been broad and instrumental in the development of various industries. Looking particularly in the journalism and academic

fields, this paper hopes to showcase some of the many ways these tools have infiltrated our lives and address some of the ethical considerations arising in these specific fields.

4.1 Generative AI Ethics in Journalism

With the need to gather and analyze large amounts of information from a variety of sources, journalists have largely benefited from the use of generative AI. Tools like Quakebot, a seismic activity tracker, and GPT-3, a text generator, have been used in prominent journals like the Los Angeles Times to help automate more of the writing process [7]. Ever since the first publications by the Associated Press in 2014, there has been a large increase of AI-generated or AI-assisted articles, and companies like Bloomberg and Google now have dedicated generative AI tools specifically designed to assist their writers in producing news. With these developments specific to journalism, generative AI systems will only grow to play a larger role in the field.

Yet, with these developments, there is more need to critically reflect on the impact of these changes and the ethical ramifications regarding the use of generative AI in journalism. With the rise of automated journalism, Montal and Reich investigated how authorship of automated articles is attributed, along with the level of disclosure and algorithmic transparency of such pieces. Interviewing workers at organizations using automatic tools, the researchers explored how organizations view the need for appropriate bylines and disclosures in automated media and what they believe the readers need when engaging with such content. Following their survey, they found that none of the workers viewed the algorithm as the author, and instead attributed authorship to a specific human or organization who was involved in the formation of the program (e.g., the developer) [6].

Similar to the U.S. legal system, research has shown that most news organizations lack any policies regarding collaborations with generative AI tools [6]. Hence, a lot of the choices are left to individual journalists themselves, who often struggle with the lack of information and insight on how collaboration with generative AI should look. With the importance of transparency and clear communication, clear policies regarding algorithmically generated content may assist journalists and help establish clear ethical grounds for the field, improving the experience of both journalists and readers alike.

4.2 Generative AI Ethics in Academia

With the rise of intelligent language models, there has been increasing concerns about the disruptive nature of generative AI tools like Chat-GPT, especially in the context of academia. Using GPT-3, the latest ChatGPT model offers users an even wider range of natural language processing (NLP) cases and logical malleability, allowing users to interact with the program and calibrate its responses to their needs. This has brought concerns regarding academic integrity in primary schools and higher education.

Identifying ChatGPT as an example of a “disruptive technology,” Eke investigates how the model brings threats to academic integrity and addresses how institutions, publishers, and other stakeholders should respond to best utilize the tool for knowledge, while also mitigating its harms as much as possible. To do so, Eke states that institutions must first embrace ChatGPT as a helpful tool for both students and staff, rather than banning its use in academics. As well as embracing its use, he suggests that there should be more education and training surrounding ChatGPT to ensure that it is responsibly used. Eke also argues that institutions and publishers need to update their academic integrity policies to have clear guidelines on how to acknowledge the use of AI and guard against potential misuse and plagiarism. Lastly, he states that there should be more progress and focus dedicated to creating affordable programs able to identify and flag the misuse of AI tools [3].

Developing technical tools to flag inappropriate generative AI outputs becomes especially significant in academia in order to identify dishonest work and plagiarism. Along with the likelihood of misuse, generative AI language models like ChatGPT also bring other risks to the academic field. Recognizing its potentially disruptive nature, Lund et al. states that the use of ChatGPT in scholarly research may exacerbate bias, create citation problems (i.e., ChatGPT does not provide all the references), and adds onto the “Matthew Effect”, which refers to the cycle that occurs when researchers with higher citation counts are constantly referred to more than those with lower citation counts [5].

Yet, while emphasizing the ethical concerns, the researchers also bring to light all the benefits ChatGPT may bring that could make the researchers’ lives easier (e.g., do tedious tasks, help with the review process, and create useful metadata and summaries) [5]. By automating tasks that require extra labor, generative AI models offer students and educators an extra hand in the learning process. However, the benefits of using these tools unfortunately do not come without the additional risks and consequences, which emphasize the need for more formal guidelines and future training about the use of generative AI in academia [5].

5 CONCLUSION

With the expanding use of generative AI, there is a growing need to address these changes and update the legal frameworks regarding intellectual property. With a lack of formal policies specifically addressing generative AI, the court system now faces numerous cases that seek to challenge the status quo, and over the next few years, the courts will need to make difficult decisions that will greatly shape how generative AI models are utilized across various sectors of our society. Institutional guidelines, such as expectations for journal bylines and policies for academic honesty, will also need to be reformed, and more education and proper training on how to responsibly use these tools will be important for future collaborations with generative AI. These considerations are especially critical during this time of rapid development, and it is imperative that the field proactively addresses these ethical nuances present in the intersection of generative AI and intellectual property.

REFERENCES

- [1] Omri Avrahami and Bar Tamir. 2021. Ownership and Creativity in Generative Models. (2021).
- [2] Nassim Dehouche. 2021. Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). *Ethics in Science and Environmental Politics* 21 (2021).
- [3] Damian Okaibedi Eke. 2023. ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology* 13 (2023).
- [4] Daniel Gervais. 2020. Is Intellectual Property Law Ready for Artificial Intelligence? *GRUR International* 69, 2 (2020).
- [5] Brady D. Lund, Nishith Reddy Mannuru, Ting Wang, Bing Nie, Somipam Shimray, and Ziang Wang. 2023. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology* 74, 5 (2023), 570–581.
- [6] Tai Montal and Zvi Reich. 2017. I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism. *Digital Journalism* 5, 7 (2017).
- [7] John Pavlik. 2023. Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator* 78, 1 (2023).
- [8] Rian Saputra, Tiara Tioline, Iswantoro, and Sanju Kumar Singh. 2020. Artificial Intelligence and Intellectual Property Protection in Indonesia and Japan. *GRUR International* 69, 2 (2020).
- [9] Deepak Somaya and Lav R. Varshney. 2018. Embodiment, Anthropomorphism, and Intellectual Property Rights for AI Creations. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* 1, 1 (2018), 278–283.
- [10] Alain Strowel. 2023. ChatGPT and Generative AI Tools: Theft of Intellectual Labor? *IIC - International Review of Intellectual Property and Competition Law* 54, 4 (2023).
- [11] Haonan Zhong, Jiamin Chang, Ziyue Yang, Tingman Wu, Pathum Chamikara Mahawaga Arachchige, Chehara Pathmabandu, and Minhui Xue. 2023. Copyright Protection and Accountability of Generative AI: Attack, Watermarking and Attribution. *Companion Proceedings of the ACM Web Conference 2023* 1, 1 (2023), 94–98.