October 18, 2023

Before the
U.S. Copyright Office
Washington, D.C.

Copyright and Artificial Intelligence (AI)

Thank you for the opportunity to provide comments to The US Copyright Office via the notice of inquiry on "on copyright and artificial intelligence (AI)." Ownership around intellectual production is critical to generative activities in academia, industry, and creative endeavors. As LLMs and other applications of AI transform these processes, we must resolve open questions around ownership and associated rights in ways that reflect diverse stakeholder interests and expertise.

We are academic researchers in the School of Information Sciences at the University of Illinois at Urbana-Champaign. We write to encourage clarity of interpretation and application of existing copyright laws to intellectual products both used as inputs to AI and outputs co-produced by human users and AI collaborators. We look forward to further opportunities to engage with the Agency's staff to provide additional analysis as the rulemaking process evolves.

I. Introduction

Human-AI Collaboration and Copyright[1]

Generative AI systems raise many questions around copyright and intellectual property, both with respect to the use of copyrighted materials in training these systems and relative to output from these systems. While large language models (LLMs) and other forms of AI raise many benefits and risks to society, the primary issues around copyright are a) the

[1] Addresses Q1: As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

need for clarity regarding how existing rights apply or might be translated to this context, and b) the need for robust enforcement and resources to support that endeavor.

If we consider the analogies between AI and the history of emergent technologies, we can see the ways in which humans have used those technologies to aid in intellectual endeavors, from cameras producing photographic art that is copyrightable by the user of the camera, not the manufacturer of the camera, to computer aided calculations that pertain to patentable technologies, once again by the human in the loop. AI and human collaboration present a new form of sociotechnical intellectual production that challenges existing norms and rules, but does not necessarily invalidate them. Instead, it demands clarity and expeditious reconciliation of ambiguities or conflicts regarding existing rules and this new reality.

In this sense, another closely associated need is: c) transparency requirements about use of AI in producing materials, as well as—specific to generative AI systems themselves—what materials trained those systems, including copyrighted materials that have been transformed or from which patterns have been derived. This information is critical to understanding knowledge production and informed use and consumption of AI-generated materials.

These comments not only provide an interdisciplinary overview of the challenges and sociotechnical reality of generative AI, but also document the broad social impacts of confusion regarding how copyright applies to various intellectual scenarios and production processes. We recognize that the model of the Gartner Hype Cycle, which relative to variables of time and visibility of new technologies, illustrates how the public perceives technology (Dedehayir & Steinert, 2016); at this moment in time, many are experiencing a peak of inflated expectations, while others have already progressed to the trough of disillusionment. This is precisely the moment to clarify rules and rights to efficiently progress toward a plateau of productivity in which AI can enhance intellectual production, rather than merely present challenges without enforceable rules and rights to set appropriate boundaries for the public good.

Relevant Background Literature[2]

[2] Addresses Q3: Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.

It is clear, both in academic literature and common observation, that humans are no longer the sole source of creative outputs. With this technological advancement, as has occurred with other technological advancements over time, new licensing regimes are needed. With this, too, comes a periodic need to review regimes governing copyright as technology advances. This need for periodic review is argued by Hristov (2017), Kaminski (2017) and others and can mean everything from minor adjustments to a reevaluation of underlying legal theories. We add that this is especially true given the volume of works that are both taken in and can be produced by AI models; intake and production at this scale is, as yet, unheard of in human history and, as such, necessitates a timely response.

Current literature suggests various options for these adjustments and reevaluations. These vary in their guidance and suggested direction. This section provides a broad overview of some of this guidance. Throughout our comment, additional sources are also provided where most relevant.

Overarching this guidance are two quandaries. Firstly, the question of what is considered possible to copyright in the first place. We offer the following as an analogy: a professor can create a reading list of research papers; none of these papers were by the individual and all of the bibliographic details of the included papers (such as author, paper title, journal title, etc.) are copied from existing databases. The professor's contribution, in this instance, is the considered selection of what to include in this reading list. As determined in the Supreme Court case Feist vs. Rural (1991) such a list may be considered creative enough to be placed under individual copyright; "these choices as to selection and arrangement, so long as they are made independently by the compiler and entail a minimal degree of creativity, are sufficiently original that Congress may protect such compilations through the copyright laws."

What, then, occurs when facts are derived from a large bank of previously copyrighted data?

This brings up an important set of questions that we do not have the answers to here but, instead, offer as food for thought:

- Can the owners of the copyrighted works in the bank of data veto or prevent the generation of the facts derived from those works in the data?
- Should or can a set of copyrighted works also be considered as "data" at all, especially if what is stored and processed is not really the sequence of words in the book as would be read by a human, but a set of vectors indicating which words and partial words co-occur more frequently?
- Can the generation of these vectors be considered fair use?

- Can the generation of these vectors be considered non-consumptive research?

While we do not have the answers to these questions, we believe these are important to consider and that any resulting answer should be clear in its resultant reasoning.

Secondly, the quandary of the definition of fairness in AI as discussed by Feuerriegel et. al. (2020). They argue that there is no one definition for what true fairness is or not in the context of AI as true fairness hinges on context– including transparency in algorithmic development and AI decision-making.

We strongly emphasize the importance of fairness in copyright considerations governing AI inputs and outputs. This viewpoint is emphasized by Lucchi (2023) in a comprehensive article detailing the importance of shifting how society views creativity in response to increasing technological development in generative AI models. Lucchi offers suggestions on how to approach this shift, emphasizing the urgency and complexity of doing so.

We divide our answer into two parts. Firstly, research considering the implications of using copyrighted material to train AI models (the “input” question). Secondly, research considering the implications of managing copyright protection for material generated by AI models (the “output” question).

Levendowski (2018) argues that the use of copyrighted material as training data for AI models falls under fair use, as it serves as a contribution to public knowledge. Levendowski also argues that doing so would help reduce bias in AI systems by extending access to a broader range of training materials and allowing for border bias-mitigation techniques such as algorithmic accountability processes.

Margoni (2018) advocates for a similar allowance of the usage of copyrighted material in training AI models, advocating that the framework of Article 5(1) of the European Union’s (EU) Copyright Directive can potentially exempt the acts of temporary reproduction needed for AI model training processes. Margoni goes on to argue that this exemption is crucial for furthering innovation while simultaneously ensuring copyright protection.

Now we turn to research considering the implications of how to manage copyright protection for material generated by AI models (the “output” question).

Guadamuz (2017) argues for a need for new copyright frameworks to ensure that AI-derived or assisted works receive protection without relying on copyright’s traditional criteria of human authorship. This is because the line between creative works generated by humans and creative works generated by machines is becoming increasingly blurred.

Guadamuz suggests the adoption of a computer-generated work clause similar to the one found in the UK's Copyright, Designs and Patents Act. This clause grants ownership to the person who arranges for the making of the computer-generated work, thus addressing the questions of both its provenance and its economic significance.

Hristov (2017) also suggests that AI-generated works should be attributed to human employers or creators rather than an AI itself, suggesting that the terms “employer” and “employee” be redefined in the U.S. Copyright Act so that the authorship of AI-generated works can be granted to the human employer or programmer. Hristov also advises against redefining authorship to include non-human entities or non-legal persons as “this would open a Pandora's Box of complications and future legal challenges” (Hristov 2017, p. 452), a view shared by Yanisky-Ravid (2017).

One counterpoint is evident in earlier work considering copyright and systems; Wu (1997) argued that determining copyright ownership should be decided on a case-by-case approach where the programmer, the user, or the AI could own an output’s copyright singly or jointly. For example, a programmer could own copyright for repeatable outputs, a user could own copyright for creative inputs, and the AI could own copyright for material that it generated without a large amount of human input.

Ramalho (2017) provides a more recent counterpoint and suggests that the traditional arguments for copyright protection– namely, labor and personality theories– do not align with AI-generated outputs and, thus, that these outputs should remain in the public domain. Ramalho also proposes the “disseminator’s right” similar to what is in Article 4 of the EU Term of Protection Directive, which allows for “a 25-year protection equivalent to the economic rights of the author for the first lawful publication or communication of a previously unpublished work after the expiry of copyright protection” (Ramalho 2017, p. 19). This combination of outputs in the public domain and disseminator’s right, Ralmalho argues, is the best path forward. This would ensure incentives and economic rewards to those to disseminate these AI outputs, much in the same way that publishers disseminate and receive economic rewards for publishing books in the public domain.

At this stage we offer another analogy, this one in terms of a possible process for enacting new laws around AI and copyright. In the development of privacy and security software, it is common to engage what is termed “White Hat Hackers” or a “Red Team”. These security experts try to break a secure system in order to expose its vulnerabilities and thereby enable the redesign of a more secure system. This approach, when applied to new copyright laws, may be useful in identifying cases where the new law as currently designed would fail to provide the protections that the legislation was intended to provide or, additionally, would cause perverse consequences – situations where the law, if enacted, would render an activity illegal that was not intended to be illegal.

We close this answer by circling back to the core of the issue at hand– the necessity for a redesign of existing copyright laws to respond to the current, and ever-developing, presence of AI models. As noted by Murray (2022), "we like to tell a story that protecting intellectual property produces a net social good, whether it is economic, social, or political in nature. … Scholarship, produced over the last twenty-five years, however, has challenged the relatively straightforward account of the incentive function of intellectual property law" (p. 548). We underscore the importance of this moment as an opportunity to generate new possibilities in copyright legislation, one that truly provides a net social good. This could be a step towards righting longstanding issues in copyright law, such as those delineated by Morris (2022) and Whiteleather (2000) in terms of copyright struggles faced by musicians in the process of wrestling with their record labels.

Existing Law and New Legislative Needs[3]

Researchers engaged in large data related research and data scraping have been struggling with End User Licensing Agreements (EULAs) for quite some time. This issue with data scraping for research is not new, but it is very relevant to the discussion of generative AI, as generative AI learning models continue to rely on existing data (which is often scraped from the internet). Not long ago, the question of whether researchers could be criminally prosecuted for scraping data under the Computer Fraud Abuse Act was an issue, but that question has thankfully been put to rest due to Van Buren v. US, 141 S. Ct. 1648 (2021). However, there is still the issue of researchers being prevented from conducting their work due to restrictive terms of service on various websites. This kind of restriction for non-commercial, educational research does nothing to enhance our nation's copyright system—as these uses would otherwise constitute fair uses but for restrictive contracts that website owners put in place.

[3] Jointly Addresses Q4 and Q5: Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? How important a factor is international consistency in this area across borders?

Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

In the European Union, for instance, data scraping for text and data mining for the purposes of scientific research, teaching, or the preservation of cultural heritage is exempt from restrictive contractual clauses. Art. 7(1) Single Digital Market Directive provides: “Any contractual provision contrary to exceptions provided for in Articles 3, 5 and 6 shall be unenforceable.”

A similar provision in the laws of the United States would be greatly beneficial in the context of text and data mining and AI, as well as in other contexts such as implementation of the Marrakesh Treaty and in Section 108 for libraries and archives.

II. Training

Inferences and the Feasibility of “Unlearning”[4]

In the realm of machine learning, "unlearning" denotes the strategy or method employed to induce a trained model to forget or dismiss particular insights or knowledge it acquired from its training dataset. There are primarily two methodologies: exact unlearning and approximate unlearning (Nguyen et al., 2022).

Exact unlearning ensures that post the unlearning phase, the model behaves precisely as though it had never encountered the unwanted data during its training. Common strategies involve retraining the model from the ground up, excluding the data meant to be forgotten, or employing Sample-wise Inference and Sample-wise Adjustment (SISA). SISA involves tweaking the model parameters for each training sample to mitigate the influence of specific data points. However, exact unlearning procedures are computationally taxing, particularly for expansive models.

On the other hand, approximate unlearning offers a more economically efficient alternative. While it diminishes the impact of undesired data on the model, it doesn't promise total erasure. There are three predominant techniques for approximate unlearning: fine-tuning the model using a new dataset devoid of the undesired data, incorporating differential privacy techniques to intersperse noise within the learning process or dataset, and deploying data augmentation (Bourtoule et al., 2021).

[4] Addresses Q7.3 Is it possible for an AI model to ‘‘unlearn’’ inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to ‘‘unlearn’’ inferences from training?

The application of these techniques varies based on the model's complexity. For straightforward prediction models, such as Linear Regression, Simple Logistic Regression, and elementary Decision Trees, strategies like data regularization prove effective in desensitizing models to individual data entries (Sekhari et al., 2021). In the case of intricate prediction models like convolutional neural networks (CNNs) or recurrent neural networks (RNNs), fine-tuning using a dataset excluding the unwanted data is a practical solution (Mitchell et al., 2021). Conversely, for vast generative models, the inherent complexity and data-generation capability render unlearning exceptionally challenging (Eldan & Russinovich, 2023). While differential privacy and selective retraining emerge as viable options, achieving complete unlearning remains a computationally heavy task.

For complex models, unlearning specific data points in an economically viable manner is still a challenge (Thudi et al., 2022). Given the current state of technology, there might be merit in focusing more on data source control, ensuring only appropriate and desired data is used during initial training. This could reduce the need for unlearning later on. Machine learning researchers, being at the forefront of innovation, should pioneer techniques for improved data handling. Data scientists and ML engineers, responsible for bringing theoretical models into practice, must ensure that the data underpinning their models is pristine and relevant. Furthermore, tech companies leveraging AI solutions need to prioritize data integrity within their operations. Regulatory bodies and AI auditors, on the other hand, can provide the necessary oversight and frameworks, ensuring data practices adhere to ethical and privacy standards. Collectively, this multi-tiered approach ensures that data is treated with the respect and scrutiny it deserves.

Identification of Training Material[5]

Without direct access to the training dataset, identifying whether an AI model was trained on a particular piece of training material–whether that is a document or photograph or other art object–is challenging but not entirely impossible. Techniques known as Membership Inference Attacks aim to determine if a specific data point was in the training set of a machine learning model (Carlini et al., 2023). In these attacks, adversaries with knowledge of the model's architecture and outputs can, to some extent, deduce the likelihood of particular data points being in the training set. However, it's crucial to note that while these techniques can provide some indication or probability, they do not guarantee definitive identification (Carlini et al., 2022). Success rates can vary based on

[5] Addresses Q7.4 Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?

the complexity of the model, the nature of the training data, and the sophistication of the attack. Thus, while potential avenues for inferring membership exist, they come with inherent uncertainties (Carlini et al., 2021). Recognizing the inherent challenges and the veiled processes of LLM training, it would be astute to advocate for tech companies, who assuredly store datasets of their disseminated models, to provide open APIs. Such APIs would enable users to verify copyright details within the training dataset. Exploring approaches like differential privacy can help strike a balance, ensuring the data's confidentiality, integrity, and accessibility.

Copyrighted Works as Training Data[6]

Questions about fair use are complex, since it has been established by 17 U.S.C. § 107 (1976) that several different aspects of use can contribute to a judgment. These include the purpose and character of use, the nature of the copyrighted work, the amount of use, and the effect on the market, if any.

But where the process of training itself is concerned, we believe that the first and fourth factors — purpose and character of use, and lack of market effect — are sufficient to decide the question. Recent precedent implies that the task of indexing content (and the closely-related task of modeling it mathematically) are in themselves transformative fair use.

For instance, the courts have held that the creation of a full-text searchable database is sufficiently transformative to count as fair use (*Authors Guild v. HathiTrust 755 F.3d 87 (2d Cir. 2014)).* "Snippet views" of the original texts have also been allowed (*Authors Guild v. Google*, 804 F.3d 202 (2d Cir. 2015)). Finally, search engines have been allowed to model word-usage patterns on websites and to transmit reduced thumbnail images of copyrighted content to assist their users (*Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir., 2007)).

Training a predictive model on a corpus is a more transformative operation than the production of a full-text database or search engine. It extracts — not snippets or thumbnail images from the original source — but extremely high-level generalizations about the way words are related to other words (or pixels to other pixels) in the corpus as a whole. This activity clearly has a transformative character, and many applications of artificial intelligence in research and education stop right here. The whole purpose of such

[6] Addresses Q8 Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

predictive models is to produce generalizations about, for instance, the history of genre patterns or character types in fiction, to help researchers understand cultural history.

Generative applications of a predictive model pose a slightly more complex legal question, since the patterns extracted from a corpus are in this case used to produce new texts or images. However, as Sag (2023) notes, the outputs of models are generally non-infringing. “Generally, pseudo-expression generated by large language models does not infringe copyright because these models ‘learn’ latent features and associations within the training data, they do not memorize snippets of original expression from individual works” (Sag, 2023, p.107).

This level of generality is fundamental to the character of the use: the point of machine learning is to generalize. Memorizing specific examples will make a model less effective and less efficient for fundamental mathematical reasons (Vapnik, 1995). Such failures can happen, as Sag goes on to acknowledge. A famous work may be present so many times in the training data that a model becomes capable of producing a relatively close imitation (Sag, 2023, pp.119-40). This is not an intended effect: models are not created or marketed to reproduce specific works, and do not excel at the task. But if a user did create a copy close enough to have an impact on the market for a work, such an output could infringe copyright and be subject to legal sanction (Henderson et al., 2023).

However, this misapplication must be considered separately from the question of training. It will become impossible to protect the public interests that fair use is designed to serve if we conflate the (uncommon) use of a model to produce infringing *output* with the act of training itself. Quantitative modeling is a transformative activity that generally serves purposes quite distinct from the purpose of reading or viewing a specific work. This is true even when models are predictive. Scholars have trained predictive models, for instance, to understand the transformation of gender stereotypes by predicting the typical attributes and behavior of fictional characters in different decades (Underwood, Bamman, & Lee, 2018). There is no clear line separating these scholarly applications of predictive modeling from generative applications that could also, for instance, produce dialogue appropriate for a 1930s-era detective. Moreover, attempts to exclude copyrighted works from training corpora are likely to exacerbate the problem of implicit bias (Levendowski, 2018). To protect the public’s legitimate interest in the advancement of science and scholarship, the Copyright Office should continue to honor the precedent that training a mathematical model of a corpus is in itself fair use.

Infringement of copyright can only happen if an infringing copy of a specific work is produced. To avoid this legal exposure, the researchers and firms who train models will be well advised to deduplicate their training data, so passages from frequently-quoted works do not occur multiple times. The risk of infringement can also be reduced—and models made more efficient—by limiting the size of a model relative to the size of the dataset (Sag, 2023, pp.142-43). But since this technology is still evolving, it would be unwise for the courts or the Copyright Office to prescribe a particular method of risk reduction. The risk of legal sanction for an infringing copy provides sufficient incentive for researchers and firms to develop strategies that prevent or discourage the misapplication of their models.

The protection of fair use is particularly strong, of course, for models trained in nonprofit and educational contexts, and as researchers we are particularly concerned with those contexts. But the precedents that support the transformative character of indexing and modeling have not been limited to nonprofit contexts; they have also covered organizations like Google and Amazon. We infer that the models produced by machine learning are fair use simply because abstracting high-level patterns from a corpus is in itself a transformative activity.

III. Transparency and Recordkeeping

Transparency and Training Data[7]

It is a common practice to feed vast amounts of copyrighted text, images, video and other data to train AI models under the pretext of 'fair use' doctrine. However, the downloading and storage of copyrighted data to train machine learning models may violate copyright law and impose undue liability on AI developers (Quang, 2021). In recent case law, copyright owners such as authors, artists, programmers, others, argue that companies training AI models, use their copyright material without consent and ultimately benefit commercially from the use of such copyrighted materials (Davis, 2023). Therefore there is a growing demand for developers to disclose records regarding the materials used to train their models.

We recommend that developers must uphold transparency by disclosing records about training data, ML approach used along with the degree of supervision, and the practices related to data management, aggregation, and transformation. Transparency involves

[7] Addresses Q15: In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

developers disclosing the hidden algorithmic decision making and data collection practices (Brevini & Pasquale, 2020) (Kläs et al., 2019).

Recent works recommend some good practices for developers to reduce concerns about potential copyright infringement and improve transparency in the creation of AI models. They should disclose the origins of the training data that model employs, i.e. reference their sources appropriately to maintain accuracy, credibility, and avoid plagiarism. However, unlike humans these models do not inherently “support” or “reject” particular claims to provide validity to their work, instead they rely on using patterns and structures from training data. Therefore, an effective way for AI models to reference their work is by specifying the sources utilized or outlining the methodologies employed to gather and evaluate the data (Lucchi, 2023). Additionally, they could also disclose “uncertainty estimates” of the model to reflect data quality and appropriateness of data used (Kläs et al., 2019).

However, Pasquale (Pasquale, 2015) points out that requiring companies to be transparent is not the end-solution. Historically, companies have parried transparency rules with more complex systems. Even if companies are transparent, their systems remain complex and unintelligible to the general audience. Therefore it is essential to combine transparency with explainability to render transparency intelligibly through explainability.

Specific Transparency Requirements[8]

In so far as transparency is critical both for those who create generative AI applications and those who co-produce creative outputs with such AI systems, transparency must be meaningful and consistent.

It must be clear for informed use of such systems what sources trained the model, how those sources were aggregated, and what learning approaches and/or modeling techniques were employed to gain insights and train a given model based upon these materials. Data-centric explanations provide considerable transparency about machine learning systems, even when models themselves are opaque, black-boxed, or otherwise uninterpretable to humans (Anik & Bunt, 2021).

It must also be clear to those who read, view, and use creative works whether or not AI was employed to produce the content. Not only would transparency about the mode of

[8] Addresses 15.1: What level of specificity should be required?

production support information literacy and aims regarding an informed public, but comparisons regarding this transparency need and those of other AI use-cases emergent along with specialized uses of LLMs, such as in healthcare and law make clear the need to understand AI systems and their impact regarding issues like bias (e.g., de Hond et al., 2022).

As such, two alternative models for clear, consistent, and sufficient transparency are available to structure such transparency requirements. First, the emergence of best practices and standards around related issues such as data quality management (DQM). Data quality reporting (Birigazzi et al., 2019) is required in specific critical domains; this approach to standardization for data quality by domain and targeted intentionality in management of said data quality is rooted in a longer literature, demonstrating how empirical research on these issues can analogously aid in standard setting for this domain, as well (e.g., Shankaranarayanan & Cai, 2006). Second, the broader literature on explainability offers many possible approaches to make clear and concrete relevant details to relevant audiences, while also recognizing feasibility limitations (Ehsan et al., 2021).

Scoping Transparency Disclosures[9]

Innovation in AI and applications of said systems in specific contexts have been subject to much secrecy, even when appropriate use and subsequent innovations would benefit from more transparency (Alì & Yu, 2021). Beyond the considerations mentioned in response to questions 15 and 15.1, we emphasize that there is interest in transparency with respect to the public in order to engender trust (Anhalt-Depies et al., 2019), in addition to the rights to disclosure that individual rights holders may have.

Disclosures should be made regarding the use of copyrighted materials to both the public and to copyright holders; efforts toward explainability can embed such transparency mechanisms in the systems by design, including automated notices (Felzmann et al., 2020). While there are barriers, including with respect to low quality data leading to provenance ambiguity (Soylu et al., 2022), the benefits of transparency outweigh these challenges and efforts should be made (Wischmeyer, 2020). Further, public disclosures should be made regarding the use of public records in training AI models and systems. There is a compelling public interest with respect to this information, as we may consider the potential for data breaches and/or inadvertent disclosures (Agelidis, 2016) with respect to AI systems.

[9] Addresses 15.2 To whom should disclosures be made?

IV. Generative AI Outputs

A. Copyrightability

Authorship and Human-AI Collaboration[10]

Yes. Creativity is the most important factor here. The Supreme Court has long considered that creative works (Feist Publ'ns, Inc., v. Rural Tel. Serv. Co., 499 U.S. 340, 1991), even when a machine aids the author, deserves copyright protection (Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 1884). That means that it is completely feasible that an author who used a machine to create a work be considered a copyright owner as long as that work has not been created exclusively by the machine or that human participation has not been irrelevant in the creative creation process.

It is important to highlight that each scenario should require fact-specific, case-by-case consideration since Courts must go deeper to evaluate the level of creativity ("*Modicum of creativity*". *Feist, 499 U.S. at 362*)[11] that is directly related to the type of author's intervention in the creation process. There are some instances in which a human author could be involved throughout the creation process of derivative work, using a copy-reliant technology, such as the training dataset, the models, and the prompt.

Datasets can be copyrightable as compilations when the compilation itself features a sufficiently original selection or arrangement. Thus, it is possible that the strategies adopted to compose a given training dataset, as well as how it is organized, are considered by the Courts as a relevant human intervention in the process of generating a derivative work. This is not a general rule, considering that a dataset can often be the result of an indiscriminate data collection process on the internet, without the intention of creating a dataset with particular attributes.

Concerning the AI model, the material produced by the system may be considered a work of authorship to be claimed by the developer. On the one hand, when the application of an existing algorithm or an already known model architecture is applied to a training

---

[10] Addresses Q18: Under copyright law, are there circumstances when a human using a generative AI system should be considered the ''author'' of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

[11] *"The standard of originality is low, but it does exist."*

dataset, there will probably be no standard of creativity necessary for copyright protection of that model. On the other hand, in terms of the copyright of the model output, it is necessary to consider whether the model architecture was created in a way that mitigates the relationship between the training dataset and the AI model output. In other words, it is important to evaluate whether the model is capable of learning from existing works, but without reproducing them in its generations. In this scenario, if the outputs are the result of a detailed designed model, often involving discoveries of new model architectures, it is feasible to confer authorship of the output on the model developer.

Finally, it is possible for the user of a given generative AI system to adopt creative choices when guiding the system through prompts. Taking the case of artists as an example, it appears that the system follows the instructions contained in the prompt and generates a series of images, which will be evaluated by the user who, in turn, will determine which elements should predominate, instructing the system to discard some elements, while highlighting others. This situation differs from those cases in which the user's prompt is too short to configure the necessary modicum of creativity, which should not justify a claim for authorship.

Clarity and the Copyright Act[12]

Copyright law in the United States already includes the requirement of human involvement as a criterion for copyright protection. Therefore, there is no need to make changes to the current legislation. According to the law, the validity of a work's copyright is conditioned on the ability to demonstrate the presence of creative expression, which must come from someone considered the legitimate and evident author of the final work.

In terms of standards, as previously mentioned, the case of Burrow-Giles Lithographic Co. v. Sarony (1884), helps to elucidate the issue. On that occasion, the US Supreme Court recognized the creative authorship of a photographer in the way he conducted the session and chose/organized the costumes and lighting. However, the Court also admitted that in some situations, photographs may not meet the criteria for copyright protection if the photographer does not demonstrate the required level of creativity. Therefore, considering current case law, the presence of a degree of authorial creativity is crucial to ensuring copyright protection and current copyright laws already reflect this requirement. As a result, there is no need to change copyright law in this regard.

[12] Addresses Q19: Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?

Legal Protections for AI-Generated Material[13]

The scenario of legal uncertainty regarding copyright protection about works created with the help of systems equipped with generative artificial intelligence ends up discouraging developers of such technologies from continuing to create and improve the capabilities of those systems. Thus, the legal protection of AI-generated material presents itself as a policy matter, since, given how such emerging technology has spread on a large scale in the most diverse sectors, its innovation potential promotes positive impacts.

It is noteworthy that the objective of granting copyrightability to a given work is precisely to serve as an incentive for creativity while increasing the number and variety of works available in the public domain. Therefore, if we were to withhold legal protection for works created by or with systems equipped with generative AI, we would need to also consider that this would disincentivize production. This would specifically adversely impact academic, research, and teaching applications of AI, which are supported by the fair use doctrine, allowing the use of existing works in those cases. As such, we should recognize the ways these applications already align with existing rights and law and extend those legal protections analogously.

Protecting AI-Generated Material via Copyright[14]

The legal protection of works created using systems equipped with generative artificial intelligence must take the form of copyright, as highlighted in the previous comments. Certainly, not only the factor of creativity in the preparation of new work must be duly observed in the analysis of the specific case by the competent authorities, but also the other requirements for copyright protection under the terms of Copyright Law.

Copyright and the Promotion of AI-Generated Intellectual Production[15]

[13] Addresses Q20: Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

[14] Addresses Q20.1: If you believe protection is desirable, should it be a form of copyright or a separate sui generis right? If the latter, in what respects should protection for AI-generated material differ from copyright?

[15] Addresses Q21: Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection ''promote the progress of science and useful arts''? If so, how?

The ultimate goal of copyright law is to promote the progress of science and the useful arts. Granting legal protection to works created through generative AI systems is in line with the idea of promoting such progress, given that recent advances in the field of artificial intelligence have demonstrated that intelligent systems can also be sources of creativity and innovation. Despite the evolution in the interpretation and application of the copyright clause, it cannot be denied that its purpose remains the same in the sense of seeking a balance between the rights of those created with the public interest, to promote progress and innovation. In this sense, the extension of existing legal protection for works created through generative AI systems via case law is capable of encouraging innovation, giving AI developers confidence that innovations in that field (such as the creation of new model architectures) will be legally protected, as long as it does not infringe the copyright of other existing works.

B. Labeling or Identification

Labeling AI-generated Material[16]

Requiring labeling of generative AI materials will allow public identification of how that material was procured which allows openness, labeling, and tracking of information used and gathered. For example, let’s look at OpenAI evolution. It all started with ChatGPT 3, then ChatGPT 3.5, and now ChatGPT 4 with subscription. Each one of these databases carries more information and processing capabilities than the next. Knowing this from the point of view of what system and or version the information originated from will help to track what’s being created in synergy with this technology.

There is no precedent for this technology and ownership is an unknown area. These language models’ power is created by the amount of data that has been created by others. The question of ownership should extend beyond the company who created the software. This software would not be without the data of others and now the user of the software has also put in some labor and originality to the output. There has to be considerations of all the labors in the use of generative AI to accredit all appropriate creators.

Labeling can be broken into two fundamental processes (Epstein et al., 2023), 1) focusing on the process through which content is created and 2) seeking to identify content that was generated by AI and labeling that attempt to identify content that has the potential to be misleading.

[16] Addresses Q28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

Labeling or identifying AI-generated materials may be helpful in answering the question of "who owns them." The company that develops AI and AI users are two bodies that potentially want to claim copyright of AI-generated materials. For example, tensions might arise between OpenAI, which owns ChatGPT, and users, who generate music, novels, or poems by prompting ChatGPT.

There has been no law on deciding who owns the copyright of AI-generated materials; once there are laws or legal precedents, techniques to label or identify AI-generated materials can be useful for court rulings of such copyright disputes. For instance, if the law decides companies who own AI systems own the copyright of their outputs and a YouTube video uploaded by a user is decided to be AI-generated, the video might be subject to copyright removal by the platform.

Responsibility and Labeling[17]

Identifying a work as AI-generated is enticing for neither AI companies nor AI users at this point since there are no laws favoring either party as the copyright owner of AI-generated materials. Moreover, developing tools to identify AI-generated materials requires a large amount of human resources and money. Companies may see an incentive to do so if some day, the law decides that they own the outputs of AI they create. If users are favored by the law, the government or academics might need to take the initiative of developing such tools as users do not have the resources.

Building tools is a costly process. However, government investments should be allocated to generative AI labeling to those companies, organizations, and individuals with current knowledge domain and ability to train up proficient labelers in these new areas.

Barriers to Labeling AI-generated Content[18]

Adversarial attacks refer to a class of techniques used to manipulate AI models, particularly neural networks, by introducing carefully crafted, human-imperceptible input data that are designed to deceive the models' predictions (Chakraborty et al., 2018). Fake news detection models have been found vulnerable to adversarial attacks (Zhou et al., 2019). Similarly, users can inject imperceptible noises into AI-generated materials to make them less "AI-like."

[17] Addresses Q28.1 Who should be responsible for identifying a work as AI-generated?
[18] Addresses Q28.2 Are there technical or practical barriers to labeling or identification requirements?

AI detection tools can be accurate even without adversaries. Non-native English writing has been misclassified as AI-generated, raising concerns about fairness and robustness (Liang et al., 2023).

Moreover, the fundamental question of "who owns the data" must also be addressed beforehand. AI creators, AI users, and training data contributors (basically everyone online when it comes to large language models) are all potential copyright owners of AI-generated materials.

Consequences[19]

Chilling effects and content creator frustration are natural consequences of copyright-related content removals (Fiesler et al., 2023). If content creators' products (e.g., videos, and news articles) are misclassified as AI-generated and taken down by platforms, a heightened level of frustration and chilling effects is expected. Biases arising from inaccurate predictions on minority populations are also severe and should be taken into account.

V. Summary of Recommendations

As a summary of our comments, we wish to strongly emphasize that we believe existing copyright law is sufficient to protect interests and appropriately incentivize creative production via the use of generative AI, yet that we also must:

1. Extend this clearly when applied via the courts in case law;
2. Apply and clarify rights via guidance, given that creators and users of these systems are not necessarily experts in Copyright law or its implications;
3. Provide clarity regarding enforcement of the law as applies to uses of AI in generating creative works; and
4. Require transparency and explainability regarding the use of copyrighted materials in training systems and the use of AI in generating creative works.

Thank you for your consideration of our comments and suggestions.

[19] Addresses Q28.3 If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?

Respectfully submitted,

Gowri Saini Balasubramaniam
Sara Rachel Benson
Anita Say Chan
Keith Jacobs
Karen V. Jenkins
Smirity Kaushik
Jiaqi Ma
Madelyn Rose Sanfilippo
Eryclis Rodrigues Bezerra Silva
Emmy Tither
Michael Twidale
Ted E. Underwood
Yaman Yu
Kyrie Zhou

References

Agelidis, Y. (2016). Protecting the Good, the Bad, and the Ugly: “Exposure” Data Breaches and Suggestions For Coping With Them. *Berkeley Technology Law Journal*, *31*(2), 1057-1078.

Alì, G. S., & Yu, R. (2021). Artificial intelligence between transparency and secrecy: from the EC whitepaper to the AIA and beyond. *European Journal of Law and Technology*, *12*(3).

Anhalt-Depies, C., Stenglein, J. L., Zuckerberg, B., Townsend, P. A., & Rissman, A. R. (2019). Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biological Conservation*, *238*, 108195.

Anik, A. I., & Bunt, A. (2021, May). Data-centric explanations: explaining training data of machine learning systems to promote transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Birigazzi, L., Gregoire, T. G., Finegold, Y., Golec, R. D. C., Sandker, M., Donegan, E., & Gamarra, J. G. (2019). Data quality reporting: good practice for transparent estimates from forest and land cover surveys. *Environmental Science & Policy*, *96*, 85-94.

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., ... & Papernot, N. (2021, May). Machine unlearning. In *2021 IEEE Symposium on Security and Privacy (SP)* (pp. 141-159). IEEE.

Brevini, B., & Pasquale, F. (2020). Revisiting the Black Box Society by rethinking the political economy of big data. *Big data & society*, *7*(2), 2053951720935146.

Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramer, F., ... & Wallace, E. (2023). Extracting training data from diffusion models. In *32nd USENIX Security Symposium (USENIX Security 23)* (pp. 5253-5270).

Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramer, F., & Zhang, C. (2022). Quantifying memorization across neural language models. *arXiv preprint arXiv:2202.07646*.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)* (pp. 2633-2650).

Chakraborty, A., Alam, M., Dey, V., Chattopadhyay, A., & Mukhopadhyay, D. (2018). Adversarial attacks and defenses: A survey. arXiv preprint arXiv:1810.00069.

Davis, W. (2023). *Sarah Silverman is suing openai and meta for copyright infringement*. The Verge.https://www.theverge.com/2023/7/9/23788741/sarah-silverman-openai-meta-chatgpt-llama-copyright-infringement-chatbots-artificial-intelligence-ai

de Hond, A. A., Leeuwenberg, A. M., Hooft, L., Kant, I. M., Nijman, S. W., van Os, H. J., ... & Moons, K. G. (2022). Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ digital medicine*, *5*(1), 2.

Dedehayir, O., & Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, *108*, 28-41.

Eldan, R., & Russinovich, M. (2023, October 4). *Who's Harry Potter? making Llms Forget*. Microsoft Research. https://www.microsoft.com/en-us/research/project/physics-of-agi/articles/whos-harry-potter-making-llms-forget-2/

Epstein, Z., Arechar, A. A., & Rand, D. (2023). What label should be applied to content produced by generative AI?. PsyArXiv https://doi.org/10.31234/osf.io/v4mfz

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021, May). Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, *26*(6), 3333-3361.

Fiesler, C., Paup, J., & Zacher, C. (2023). Chilling Tales: Understanding the Impact of Copyright Takedowns on Transformative Content Creators. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), 1-21.

Henderson, P., Li, X., Jurafsky, D., Hashimoto, T., Lemley, M. A., & Liang, P. (2023). Foundation Models and Fair Use. arXiv preprint arXiv:2303.15715.

Kläs, M., Sembach, L. (2019). Uncertainty Wrappers for Data-Driven Models. In: Romanovsky, A., Troubitsyna, E., Gashi, I., Schoitsch, E., Bitsch, F. (eds) Computer Safety, Reliability, and Security. SAFECOMP 2019. Lecture Notes in Computer Science(), vol 11699. Springer, Cham.

Levendowski, A. (2018) How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem, 93 Wash. L. Rev. 579. https://digitalcommons.law.uw.edu/wlr/vol93/iss2/2

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. arXiv preprint arXiv:2304.02819.

Lucchi, N. (2023). ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. European Journal of Risk Regulation, 1-23. doi:10.1017/err.2023.59

Mitchell, E., Lin, C., Bosselut, A., Finn, C., & Manning, C. D. (2021). Fast model editing at scale. *arXiv preprint arXiv:2110.11309*.

Nguyen, T. T., Huynh, T. T., Nguyen, P. L., Liew, A. W. C., Yin, H., & Nguyen, Q. V. H. (2022). A survey of machine unlearning. *arXiv preprint arXiv:2209.02299*.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Quang, J. (2021). Does Training AI Violate Copyright Law?. *Berkeley Tech. LJ*, *36*, 1407.

Sag, M. (2023). Copyright Safety for Generative AI (SSRN Scholarly Paper 4438593). https://doi.org/10.2139/ssrn.4438593

Sekhari, A., Acharya, J., Kamath, G., & Suresh, A. T. (2021). Remember what you want to forget: Algorithms for machine unlearning. *Advances in Neural Information Processing Systems*, *34*, 18075-18086.

Shankaranarayanan, G., & Cai, Y. (2006). Supporting data quality management in decision-making. *Decision support systems*, *42*(1), 302-317.

Soylu, A., Corcho, Ó., Elvesæter, B., Badenes-Olmedo, C., Yedro-Martínez, F., Kovacic, M., ... & Roman, D. (2022). Data quality barriers for transparency in public procurement. *Information*, *13*(2), 99.

Thudi, A., Jia, H., Shumailov, I., & Papernot, N. (2022). On the necessity of auditable algorithmic definitions for machine unlearning. In *31st USENIX Security Symposium (USENIX Security 22)* (pp. 4007-4022).

Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in English-language fiction. Journal of Cultural Analytics. https://doi.org/10.22148/16.019

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. Springer.

Wischmeyer, T. (2020). Artificial intelligence and transparency: opening the black box. *Regulating artificial intelligence*, 75-101.

Zhou, Z., Guan, H., Bhat, M. M., & Hsu, J. (2019). Fake news detection via NLP is vulnerable to adversarial attacks. arXiv preprint arXiv:1901.09657.