Comments of Patrun, Inc.

To the U.S. Copyright Office, Library of Congress, on the Notice of Inquiry and Request for Comments, Issue 1021

October 30, 2023

On behalf of Patrun, Inc., we write in response to the notice of inquiry regarding artificial intelligence (AI) and copyright. Our expertise resides at the intersection of data science and AI policy.

Patrun, Inc. is the first content licensing platform for generative AI models. By introducing a system of fair compensation to owners when their content is used to build an AI model, we seek to promote a transparent, ethical AI ecosystem. Our mission is to champion the rights of content owners by ensuring they receive the true market value of their work while providing the assurance of legal, ethical data sourcing for AI companies. Patrun aims to create a future where the benefits of AI become accessible to everyone. We're working to ensure that content owners profit more than they ever could have in the past thanks to the multiplicative value of their original work as it is used derivatively in AI.

Our comments focus on these questions:

> *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? 40 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.*

> *Under the fourth factor of the fair use analysis, how should the effect on the potential market for of a copyrighted work used to train an AI model be measured? 46 Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?*

> *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.*

The inquiry's stated purpose is to help the Office assess whether further legislative or regulatory steps are warranted. We strongly urge Congress to put forth bipartisan legislation as soon as possible to address ambiguities in copyright law around the use of copyrighted material for AI training. Today's [Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence](#) sets the stage for Congress to establish comprehensive, binding laws on AI and individual rights. In light of the current struggles in copyright law, we recommend that Congress introduce data

ownership protections through pattern rights. **Pattern rights** refer to the right of an individual to decide whether an AI developer can extract patterns from data that represents him/her (whether a created work, image, behavioral patterns, etc) to build an AI model. This new category of rights not only fills the gap seen in copyright law but also addresses a much bigger problem lying outside the scope of copyright.

In the sections that follow, we provide information and analysis regarding the inquiry questions.

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?

We recommend looking to the OECD Principles on Artificial Intelligence, which the U.S. has endorsed, and the European Union Artificial Intelligence Act (EU AI Act) which will soon have an impact worldwide as it goes into effect.

OECD Principles on Artificial Intelligence

The United States has endorsed the OECD Principles on Artificial Intelligence, which promote AI that is innovative, trustworthy, and respects human rights. Adopted in May 2019 by member countries, the framework is the first of its kind to be signed by governments. It has become the most important standard for AI policymaking worldwide and has had a significant influence on the development of the EU AI Act. We call your attention to the AI Principle on Transparency:

> *AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:*
>
> *to foster a general understanding of AI systems,*
> *to make stakeholders aware of their interactions with AI systems, including in the workplace,*
> *to enable those affected by an AI system to understand the outcome, and,*
> *to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.*

While the OECD framework was created before the emergence of generative AI as a major policy issue, the transparency obligation is still very much applicable. Below, we will show how it was incorporated into the EU AI Act and why a similar approach should be considered in the U.S. Second, the Principle on Human-Centered Values and Fairness is also relevant in the context of generative AI:

> *AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and **internationally recognised labour rights**.*

> *To this end, AI actors should implement mechanisms and safeguards, such as **capacity for human determination**, that are appropriate to the context and consistent with the state of art.[1] (emphasis added)*

This principle states that AI systems should not violate existing international labor standards. While there is no single international copyright law, as copyright is administered by country, harmonization on copyright protection exists. Nearly 180 countries have signed the Berne Convention through the World Intellectual Property Organization (WIPO), which establishes minimum standards for the protection of the rights of creators around the world. The U.S. should take this opportunity to lead the international community by example, through legislation that protects the rights of content owners when their work is used to build AI.

The European Union Artificial Intelligence Act (EU AI Act)

Like the OECD Principles, the original EU AI Act was drafted before the widespread emergence of generative AI, leaving this category of AI policy unaddressed. However, as of June 14, 2023, the European Parliament adopted an amendment[2] with requirements for **transparency in the use of training data protected under copyright law**:

> *Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system, shall in addition:*
>
> *a) comply with the transparency obligations outlined in Article 52 (1),*
>
> *b) train, and where applicable, design and develop the foundation model in such a way as to ensure adequate safeguards against the generation of content in breach of Union law in line with the generally-acknowledged state of the art, and without prejudice to fundamental rights, including the freedom of expression,*
>
> *c) without prejudice to Union or national or Union legislation on copyright, **document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law (emphasis added)**.*

The importance of the requirement to document the use of training data protected under copyright cannot be overstated. Without access to the underlying dataset, it will be impossible for governments to enforce copyright protections in AI training at scale. Until a documentation requirement is introduced, the matter will be taken up repeatedly in the courts, as we are currently seeing. A combination of data documentation and independent auditing is the only means of holding AI developers accountable in the long run.

Here we point your attention to the [Foundation Model Transparency Index](#)released by Stanford, which scored the industry's 10 major foundation model developers using 100 indicators. The report underscores the inseparable relationship between model transparency and accountability, stating that:

---

[1] https://oecd.ai/en/dashboards/ai-principles/P6

[2] https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html#sdocta6

Transparency is an essential precondition for public accountability, scientific innovation, and effective governance of digital technologies. Without adequate transparency, stakeholders cannot understand foundation models, who they affect, and the impact they have on society.[3]

It recommends that policymakers prioritize transparency "with sufficient precision," making it "a top priority in legislative proposals related to foundation models.[4]" Specifically,

Policymakers might also acknowledge that requirements to disclose a summary of any copyrighted training data are too vague and a more specific definition, such as the definition we provide in Appendix B, may be desirable to improve compliance.

Due to the transnational nature of internet commerce, the first country or entity to legislate and enforce a documentation requirement will likely cause a global convergence on AI policy. The EU AI Act is predicted to affect the U.S. in a manner like the Global Data Protection Regulation (GDPR).

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.

The unauthorized use of a copyrighted work to train AI models does not constitute fair use due to its commercial nature, which has significant impacts on the potential market for the work. Even under the "transformative use" argument, which has had a far-reaching impact on copyright case law, fair use does not hold because, by its nature, AI cannot create new expression.

Here we discuss case law relevant to this question. The U.S. Copyright Act of 1976 sets forth that the creation of a derivative work is a right reserved exclusively for the copyright owner. A derivative work is one "based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted."

In 2014, the Acuff-Rose vs Campbell decision[5] gave judicial approval to the concept of "transformative" fair use. In this case, the court ruled that "parody has an obvious claim to transformative value," as "it can provide social benefit, by shedding light on an earlier work, and, in the process, creating a new one."[6] This transformative use argument led to a diminishment in the copyright owners' exclusive right to create derivative works, as many rulings since came down to a determination of whether the work qualifies as "derivative" or "transformative."

---

[3] Page 3 https://crfm.stanford.edu/fmti/fmti.pdf

[4] Page 7 https://crfm.stanford.edu/fmti/fmti.pdf

[5] "Under the first of the four § 107 factors, "the purpose and character of the use, including whether such use is of a commercial nature ... ," the enquiry focuses on whether the new work merely supersedes the objects of the original creation, or whether and to what extent it is "transformative," altering the original with new expression, meaning, or message. The more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use." https://supreme.justia.com/cases/federal/us/510/569/

[6] https://supreme.justia.com/cases/federal/us/510/569/

However, the recent *Andy Warhol Foundation vs. Goldsmith* decision shifted focus away from the derivative-transformative dichotomy and towards other elements, such as how the new work competes in the market. This ruling recognized **the owners' right to create derivative works as the primary intent of the copyright act,** and that "transformative use" often compromises the exclusive right of the owner to create derivative works:

> "But an overbroad concept of transformative use, one that includes any further purpose, or any different character, would narrow the copyright owner's exclusive right to create derivative works. To preserve that right, the degree of transformation required to make "transformative" use of an original must go beyond that required to qualify as a derivative[7]…

> Campbell did describe a transformative use as one that "alter[s] the first [work] with new expression, meaning.[8] **But Campbell cannot be read to mean that §107(1) weighs in favor of any use that adds new expression, meaning, or message.** Otherwise, "transformative use" would swallow the copyright owner's exclusive right to prepare derivative works, as many derivative works that "recast, transfor[m] or adap[t]" the original, §101, add new expression of some kind (emphasis added).

Per this new ruling, we would argue that the most relevant determiner of fair use is the commercial impact criteria. Under this criteria, it is easy to demonstrate that AI-generated works create competition for the original owner. Not only are content creators not paid for the use of their work in the model; but they are also not receiving due compensation for AI-generated works that bring in a profit. The Foundation Model Transparency Index shows that creators are not being compensated for their licenses:

> Within the upstream domain, **no company scores points for indicators about data creators, the copyright and license status of data, and mitigations related to copyright.** The industry-wide lack of transparency on these issues relates directly to pressing societal concerns related to copyright and intellectual property, which are the subject of ongoing litigation.[9]

Under the "transformative" criteria for fair use, the unauthorized use of copyrighted material for AI models does not qualify because it is **categorically different from human transformative use. Transformation requires new expression. But there can be no new expression by AI; therefore, AI cannot satisfy the definition of transformation.** AI developer companies will argue that the AI model transforms the work so substantially that it changes the meaning and therefore meets the fair use standard. This claim fails the creativity test, the central principle for copyright law.

---

[7] Page 22, https://www.supremecourt.gov/opinions/22pdf/21-869_87ad.pdf

[8] 510 U. S., at 579. https://www.supremecourt.gov/opinions/22pdf/21-869_87ad.pdf

[9] Page 4,

The *Feist Publications, Inc. v. Rural Telephone Service Co*. decision established the principle that **human creativity**, not level of effort (referred to as "sweat of the brow"), is the minimum standard for copyright protection. To quote the case commentary[10]:

> This case created the idea/expression dichotomy, in which ideas are not protected, but expression is protected. Mere facts cannot be copyrighted, but the selection and arrangement of them may be copyrighted if there is some originality to it.

Any claim to transformative use necessarily requires human creativity, such that something new is added that changes the meaning of the original. The human creative element is what gives the work meaning because it has *intention.* But when a machine ingests data for machine learning (ML) training, the process is purely mathematical:

1) The machine ingests large amounts of text and converts it into aggregate math.
2) The machine extracts patterns from the aggregate math to create rules or procedures, also known as algorithms.[11]

The text, as the expression of the copyright owner ingested by the model, is the DNA of machine learning. **By its very nature as a mathematical operation, the machine learning model and any outputs it produces are derivative, not transformative.** No human creativity is involved. The model can only extract the expression of the original work, which is why it can produce text that perfectly mimics the style of an author or work.

Under the fourth factor of the fair use analysis, how should the effect on the potential market for a copyrighted work used to train an AI model be measured? 46 Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

Before we can consider the potential market effect, we must center the discussion on the problem of author harm. Because AI models extract patterns from the author's expression, the expression itself is being copied and the authors' copyright is violated. It also creates the additional harm of potentially affecting the reputation of the author and original work, as new versions of the work could be produced.

Second, and separately, the effect on the market value for the copyrighted work includes all three considerations – the model could be used to create a work that competes with the same copyrighted work, the body of works of the same author, and the market for that general class of works. A ghostwritten book by AI affects the value of a copyrighted book by diluting the market and decreasing the demand for it.[12]

---

[10] https://supreme.justia.com/cases/federal/us/499/340/
[11] Algorithm: a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer. As defined by dictionary.com.
[12] Here's an example of what this looks like for authors on a day-to-day basis: https://janefriedman.com/i-would-rather-see-my-books-pirated/

But the market effect goes further beyond, encompassing any other text or that would never have existed in that particular form without the authors' expression that fed the model. Any text that derives from the model could potentially come into competition with the copyrighted work.

5. 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.

While we believe that copyright law clearly extends to the generative AI use case, it does not directly or completely address it. There is a legislative gap that is much broader than copyright only, which is the right of companies to extract patterns from *any* kind of digital asset. Any type of material that can be extracted as a pattern and leveraged in machine learning should be protected under the new legal category of pattern rights. Can a company use your face to build a model? What about your online buying patterns?

The right for any digital asset to be extracted as a pattern must lie with the owner of that material. Until pattern rights are codified by U.S. law, we will continue to see lawsuits as AI developer companies build datasets using personal and copyrighted data. Pattern rights is the only concept that is comprehensive enough to cover the many kinds of AI use cases and legal actions that will come onto the horizon.

Please do not hesitate to contact us with any further questions.

Kind regards,

Kara Doriani O'Shee

AI Ethicist
Patrun, Inc