Docket No. 2023-6

Artificial Intelligence & Copyright

88 Fed. Reg. 59942 (Aug. 30, 2023)

Comments of the Software & Information Industry Association

Introduction

SIIA is an association of roughly 380 companies involved in the business of information. Our members produce databases, educational software, and a variety of other publications used to do everything from perform medical research to evaluate the desirability of a stock trade, to tracking down missing persons. As an association, its mission is to ensure a healthy information ecosystem: one that promotes information's creation, dissemination, and productive use.

Our members have wholeheartedly embraced the promise of AI and predict advances that will revolutionize data management, analysis, and dissemination. Our members actively use AI on many fronts—in the classroom, in fraud detection, in money laundering investigations, and in locating missing children. They have invested billions in its development, acquisition, and use. At the same time, however, SIIA has repeatedly argued that the use of AI must comply with existing statutory requirements and respect for established intellectual property rights. We also recognize that the benefits of this technology also come with risks such as privacy concerns, algorithmic bias, and ethical implications. SIIA supports a risk-based approach to AI regulation, recognizing that in some cases that "AI is different."

When the Founders drafted the first copyright act in 1790, they did not envision that machines would be able to perfectly mimic the

human authorship that forms *Feist*'s "creative spark."[1] The 1976 Act reflects an accrued legislative wisdom that, while technologies may require legislation in particular contexts, the Act as a whole ought to be technology neutral. That technological neutrality–largely due to the resilience of the fair use doctrine, the idea/expression dichotomy, and similar longstanding tenets–has formed the basis for the Act's resiliency even as technologies come along that test its premises. In the AI context, such challenges are working their way through the courts and, while our members may not all agree on how a particular fact pattern ought to be resolved, they do agree that legislation at this time is premature. We submit these comments to make a few general points around which there is member consensus.

First, copyrighted works do not lose their protection by virtue of appearing on the internet. Much of the information business relies on an interdependent relationship between copyright law and established legal and technological protocols that facilitate public access to information. Many of our members use these rules to acquire and use publicly available information for non-generative AI-related activity, such as providing search engines and accessing state government records to performing threat assessments. Others search the internet for publicly available information, such as social medial sentiment, analyze that information and use artificial intelligence models to create an AI system that predicts the direction of a particular security price. Still others develop high quality content and materials, representing valuable company intellectual property and seek the legal safeguards and framework of the Copyright Act to protect their ability to put that content online. The Copyright Office itself has issued access exemptions for text and data mining in its 1201 rulemaking for specific uses and in limited circumstances. We do not believe it is the goal of this proceeding to upset existing justifications or legitimate uses for non-generative AI activity.

Second, if a copyright owner elects to restrict *access* to its work, or limit its use for generative AI purposes, such arrangements ought to be respected. Many of our members already license their works for use as AI training data. Licensing ought to be encouraged: it lowers litigation risk and ensures that developers are the most likely to get

---

[1]   *Feist Pubs. v. Rural Telephone*, 499 U.S. 340, 345 (1991).

access to the highest quality training data. In the long run, those license fees will ensure incentives to create and maintain those data sets. By prioritizing a collaborative approach that fosters cooperative relationships between content owners and AI developers, the industry can cultivate an environment that encourages innovation while safeguarding the integrity of intellectual property rights. We urge caution against adopting stringent regulatory measures, such as the expansive text and data mining exemption adopted by Singapore[2] that will curtail the transformative potential of AI and impede the flow of digital information.

Our answers to individual selected questions are addressed to the context of generative AI.

8.2 How should the [fair use] analysis apply to entities that collect and distribute copyrighted material for training but did not engage in the training?

Without expressing any opinion on liability for ingesting works for AI training, firms that are selling copies of others' copyrighted works for training purposes face a different liability risk. While the law and common practice has to a limited degree acknowledged a personal right to make certain kinds of copies, well-established precedent has prohibited others from profiting from being the agent of another's fair use. Thus, for example, time-shifting does not entitle a consumer to sell the made copy, nor may a copy shop provide scholarly articles for educational purposes, even though an individual may make a fair use copy for their own personal study.

A finding of fair use in such a context would be extremely troubling, as even the most slavish pirate will merely claim that they are providing access to unprotected ideas for the purpose of analysis. Such an argument proves too much. Appropriate licensing practices are a valuable means of reducing litigation risk in this circumstance.

---

[2]   Singapore Copyright Act of 2021, available at https://sso.agc.gov.sg/Acts-Supp/22-2021/Published/20211007?DocDate=20211007 (accessed Oct. 30, 2023).

9. Should copyright owners have to affirmatively consent to the use of their works as training materials, or should they be provided with the means to opt out?

Whether or not copyright owners have to affirmatively consent to the use of their works for generative AI will ultimately be determined by the multiple cases working their way through the courts. However, our members do agree that it would be useful for the adoption of voluntary technical standards that would result in improve choice and control over online content. One of our members has developed such a standard.

15. 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?
24. … Are existing civil discovery rules sufficient to address this situation?

SIIA supports transparency as part of the ethical and responsible deployment of artificial intelligence in a risk-based regime. Transparency can be an important component to enabling evaluation of algorithmic bias, and potential privacy harms. Such transparency would not extend, of course, to the disclosure of trade secrets. That same transparency may in a given case provide a copyright owner with information about the use of their works.

However, SIIA opposes the creation of special disclosure requirements for AI systems as a matter of *copyright* policy. Existing law requires that a defendant prove both access to the work and copying. There is no evidence that existing discovery rules have proven inadequate to prove those elements, and it is simply premature to make any such policy judgments at the very least until the existing batch of cases make their way through the courts.

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

SIIA believes that the Copyright Office guidance got the important questions mostly right. The originality standard is not a high one, but it does require human creativity. However, originality is a continuum: it is entirely possible that a series of iterative, human-drafted queries could result in a particular work meeting the originality standard. These matters will have to be evaluated on a case-by-case basis, and the Office was right to require some disclosure of the use of generative AI in registration applications while it tries to determine what that line is and how to draw it. We note, however, that AI is rapidly evolving. As use of these tools becomes more sophisticated and widespread and the tools themselves evolve, these disclosure requirements may need to be revisited if they prove to be unduly burdensome for authors and copyright owners.

19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when AI-generated material is subject to copyright protection?

No. We believe that both the Copyright Office and the courts have correctly interpreted the law on these points.

20. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI?
Other than in the context of the registration process, no additional disclosure requirements are necessary. While there may be certain specific contexts (such as automated decision making) where it may be useful to disclose AI involvement, we do not see a copyright-related reason for such disclosures.

Thank you for considering our views.

Respectfully submitted,

/s/

Christopher A. Mohr

President