# Copyright, Creativity, and Machine Learning: Bridging Perspectives

# (Virtual Roundtable)

# Workshop on Creativity & Generative AI

Copyright law has long served as a framework for protecting intellectual property, allowing creators to control the reproduction and distribution of their works. However, the advent of generative AI (GenAI) has introduced new challenges to traditional copyright structures. This white paper summarizes a roundtable discussion on the implications of GenAI and potential pathways to ensure compatibility with conventional copyright principles.

The participants at the workshop discussed three main questions:

First, what are we seeking to protect through copyright law. Second, the participants shared common misconceptions folks working in their respective fields have about genAI and/or creative work, and risk factors that policymakers should pay attention to. Third, the participants shared their perspectives on what copyright law should generally protect moving forward.

## I. What are we seeking to protect through copyright law?

At its core, copyright law is designed to:

- Prevent unauthorized reproduction and distribution of protected works.
- Ensure creators receive proper credit and attribution.
- Regulate derivative works and the scope of permissible transformation.
- Maintain accessibility to artistic creation without undue gatekeeping. Balancing measures include “fair use.”

The ability to make a living as an artist, while not explicitly guaranteed by copyright, is part of upholding the constitutional mandate for copyright, which is to promote useful arts by granting creators exclusive rights.

## II. Misconceptions and Policy Considerations

Although there were variations on technical and legal understanding of GenAI and copyright, participants generally acknowledge that GenAI disrupts traditional copyright frameworks in several ways:

- Training Data Concerns: AI models are trained on vast datasets, often including copyrighted works. The legality of using such data without explicit permission remains contested. More specifically, should open datasets like LAION and Common Crawl be regulated? The use of openly available datasets in AI training raises ethical and legal

concerns and policymakers must determine whether restrictions on such datasets align with copyright principles. Further complicating this issue is the question of whether datasets are also protectable creative works.

- Reproduction vs. Inspiration: AI models can generate content that closely resembles copyrighted works, raising questions about what constitutes an unlawful derivative work.
- Attribution and Ownership: AI-generated art blurs the lines of authorship.
- Market Disruption: The ability to generate high-quality art, music, and literature at scale challenges the livelihoods of human artists. Moreover, the effectiveness of "guardrails" remains questionable, as AI can recreate protected content if trained on sufficient data.
- Lack of Legal Precedent: Current laws do not clearly address AI's role in copyright, leading to uncertainty in product design, enforcement, and regulation. There are a number of currently pending copyright infringement lawsuits against AI companies, and this space will continue to evolve.
- The Role of Corporations: Large technology companies leverage copyright law to protect their own intellectual property while advocating for extended copyright durations (e.g., the continued protection of Mickey Mouse). These companies have significant influence over legislative outcomes.
- The Threat of Gatekeeping: If copyright law is structured in a way that raises the barrier to entry for new artists, it could hinder creative expression rather than protect it. Additionally, artists and technical experts debate whether AI "learns" in a way comparable to human learning. This anthropomorphization of AI can influence public perception of fair use and copyright law enforcement.

## III. Potential Solutions

To address these challenges, stakeholders should consider the following:

- Regulation of Training Data: Mandating transparency in dataset composition and ensuring consent from copyright holders. Given the decentralized nature of AI development, technical and non-technical measures are needed to prove that copyrighted materials were not improperly included in training datasets.
- Incentivizing High-Provenance Datasets: Encouraging the creation of ethically sourced and properly licensed training materials.
- Reevaluating Copyright Protections: Expanding copyright law to address AI-generated works. Furthermore, artistic style is not currently protected under copyright law, but AI-generated outputs challenge the boundaries between inspiration and direct reproduction.
- Implementing Royalty-Based Models: Companies utilizing GenAI could establish revenue-sharing agreements with artists whose works contribute to AI training.
- Exploring Technical Solutions: Developing mechanisms to detect and prevent copyrighted content from being used in AI training and output generation.
- Clarifying Legal Liability: Determining whether responsibility for copyright infringement falls on the AI developer, dataset curator, or end user.
- User Accountability: Establishing clear guidelines on whether users should be responsible for the prompts they input into AI models.

As AI-generated content becomes more prevalent, the way we define and protect artistic labor must evolve. Interdisciplinary spaces and discussions help affected communities come together to share their perspectives on the impact on labor as well as the promises and limitations of policy mechanisms. Aside from the rich discussion amongst the participants, one final takeaway from this roundtable was that policymakers, artists, and technology leaders must collaborate to ensure a fair and sustainable creative ecosystem in the digital age.