Talkin' 'Bout AI Generation:
Copyright and the Generative-AI Supply Chain

Katherine Lee[*]
A. Feder Cooper[†]
James Grimmelmann[‡]

"Does generative AI infringe copyright?" is an urgent question. It is also a difficult question, for two reasons. First, "generative AI" is not just one product from one company. It is a catch-all name for a massive ecosystem of loosely related technologies, including conversational text chatbots like ChatGPT, image generators like Midjourney and DALL·E, coding assistants like GitHub Copilot, and systems that compose music and create videos. Generative-AI models have different technical architectures and are trained on different kinds and sources of data using different algorithms. Some take months and cost millions of dollars to train; others can be spun up in a weekend. These models are made accessible to users in very different ways. Some are offered through paid online services; others are distributed on an open-source model that lets anyone download and modify them. These systems behave differently and raise different legal issues.

The second problem is that copyright law is notoriously complicated, and generative-AI systems manage to touch on a great many corners of it. They raise issues of authorship, similarity, direct and indirect liability, fair use, and licensing, among much else. These issues cannot be analyzed in isolation, because there are connections everywhere. Whether the output of a generative-AI system is fair use can depend on how its training datasets were assembled.

Whether the creator of a generative-AI system is secondarily liable can depend on the prompts that its users supply.

*In this Article, we aim to bring order to the chaos. To do so, we introduce the **generative-AI supply chain**: an interconnected set of stages that transform training data (millions of pictures of cats) into generations (a new, potentially never-seen-before picture of a cat that has never existed). Breaking down generative AI into these constituent stages reveals all of the places at which companies and users make choices that have copyright consequences. It enables us to trace the effects of upstream technical designs on downstream uses, and to assess who in these complicated sociotechnical systems bears responsibility for infringement when it happens. Because we engage so closely with the technology of generative AI, we are able to shed more light on the copyright questions. We do not give definitive answers as to who should and should not be held liable. Instead, we identify the key decisions that courts will need to make as they grapple with these issues, and point out the consequences that would likely flow from different liability regimes.*

Introduction

Generative artificial-intelligence (i.e., "generative-AI") systems like ChatGPT, Claude, Bard, DALL·E, and Ideogram are capable of turning a user-supplied prompt like `"give three arguments why marbury v. madison was wrongly decided"` into a persuasive essay, or `"a robot cowboy riding a rocket ship"` into a work of digital art. Their unpredictability and complexity means that they break out of existing legal categories. In particular, the fact that generative-AI systems involve training on millions of examples of human creativity means that they raise serious copyright issues. These copyright issues have not gone unnoticed. Numerous groups of plaintiffs have sued leading generative-AI companies for copyright infringement, with potential damages reaching into the billions of dollars.

This Article is an attempt to think carefully and systematically about how copyright applies to generative-AI systems. Our first contribution is to be precise about what "generative AI" is. It is not just one product from one company. Instead, "generative AI" is a catch-all name for a massive ecosystem of loosely related technologies, including conversational text chatbots like ChatGPT, image generators like Midjourney and DALL·E, coding assistants

like GitHub Copilot, and systems that compose music, create videos, and suggest molecules for new medical drugs. Generative-AI models have different technical architectures and are trained on different kinds and sources of data using different algorithms. Some take months and cost millions of dollars to train; others can be spun up in a weekend. These models are also made accessible to users in very different ways. Some are offered through paid online services; others are distributed open-source, such that anyone could download and modify them.

This Article takes the complexity and diversity of generative-AI systems seriously. To provide a clear framework for thinking about the different kinds of generative-AI systems and the different ways they are created and used, the Article introduces what we call the **generative-AI supply chain**: an interconnected set of stages that transform training data (millions of pictures of cats) into generations (a new and hopefully never-seen-before picture of a cat that may or may not ever have existed).

1. The supply chain starts with **creative works**: all of the books, artwork, software, and other products of human creativity that generative AI seeks to learn from and emulate.

2. Next, works and other information must be converted into **data**: digitally encoded files in standard, known formats.

3. Individual items of data are useless for AI training by themselves. Instead they must be compiled into **training datasets**: vast and carefully structured collections of related data. The process requires both extensive automation and thoughtful human decision-making.

4. To create a generative-AI **model**, its creator picks a technical architecture, assembles training datasets, and then runs a training algorithm to encode features of the training data in the model. Model training is both a science and an art, and it involves massive investments of time, money, and computing resources.

5. The model that results from this initial training process is called a "base" or "pre-trained model," because it is often just a starting point. A model can also be **fine-tuned** to improve its performance or adapt it to a specific problem domain. This process, too, involves extensive choices — and it need not be carried out by the same entity that did the initial training.

6. A model by itself is an inert artifact. It can be used only by technical experts with substantial computing resources. To make a model usable by a wider userbase, it must be **deployed**: embedded in some larger software system that provides a convenient interface. ChatGPT has a conversational text-box interface that allows users to interact with a GPT model

hosted on OpenAI's servers. Midjourney is deployed as a Discord bot; users request images by sending messages to it. Other systems are provided as downloadable apps, or released publicly for other developers to modify and deploy themselves.

7. A deployed system can be used to **generate** outputs: new creative works that are based on statistical patterns in the training dataset but combine them in new ways. An output — or "generation" — is based on a prompt supplied by the user: an input that describes the particular features they want the output to have. This is typically the only part of the supply chain that users see.

8. The supply chain does not end with generation. The developers of a generative-AI system can perform **alignment** by rating prompts and generations: further adjusting the model and the system it is embedded in to better achieve users' (and their own) needs. Those needs can include safety, helpfulness, and legal compliance. In this way — as in many others — the supply chain feeds back into itself. It is not a simple cascade from data to generations. Instead, each stage is regularly adjusted to better meet the needs of the others.

Breaking down generative AI into these constituent stages reveals all of the places at which companies and users make choices that have copyright consequences.