Human-AI Interaction, Lecture 18

AI and Society

BUSINESS

Anthropic to pay authors $1.5B to settle lawsuit over pirated chatbot training material

UPDATED SEPTEMBER 5, 2025 · 4:50 PM ET

By The Associated Press



Thriller novelist Andrea Bartz is photographed in her home Thursday in the Brooklyn borough of New York City.
Richard Drew/AP

NPR, September 5, 2025, by The Associated Press; photo Richard Drew/AP

June 23, 2025: the order on fair use

An artificial intelligence firm downloaded for free millions of copyrighted books in digital form from pirate sites on the internet. The firm also purchased copyrighted books (some overlapping with those acquired from the pirate sites), tore off the bindings, scanned every page, and stored them in digitized, searchable files.

Order on Fair Use, Bartz v. Anthropic PBC, No. C 24-05417 WHA (N.D. Cal.), Dkt. 231, page 1 (Alsup, J.)

Pirated: Books3, Library Genesis, Pirate Library Mirror

Dkt. 231, pages 2-3

Training vs. pirated copies

The copies used to train specific LLMs were justified as a fair use.

Dkt. 231, page 30

The downloaded pirated copies used to build a central library were not justified by a fair use.

Dkt. 231, page 31

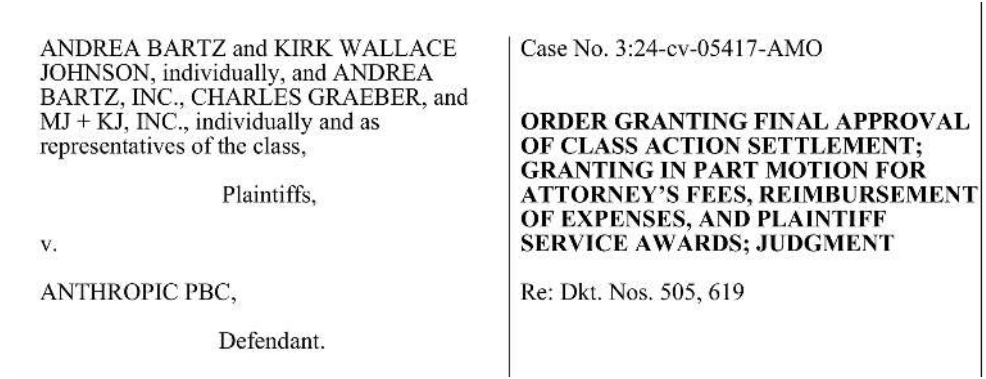
A trial was set on the pirated copies.

July 20, 2026: final approval

$1.5 billion

for **482,460** books on the list

about **$3,000** per book,
before costs and fees



ANDREA BARTZ and KIRK WALLACE JOHNSON, individually, and ANDREA BARTZ, INC., CHARLES GRAEBER, and MJ + KJ, INC., individually and as representatives of the class,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

Case No. 3:24-cv-05417-AMO

ORDER GRANTING FINAL APPROVAL OF CLASS ACTION SETTLEMENT; GRANTING IN PART MOTION FOR ATTORNEY'S FEES, REIMBURSEMENT OF EXPENSES, AND PLAINTIFF SERVICE AWARDS; JUDGMENT

Re: Dkt. Nos. 505, 619

Order Granting Final Approval, Dkt. 680, pages 1 and 4 (Martínez-Olguín, J., July 20, 2026)

What the settlement covers

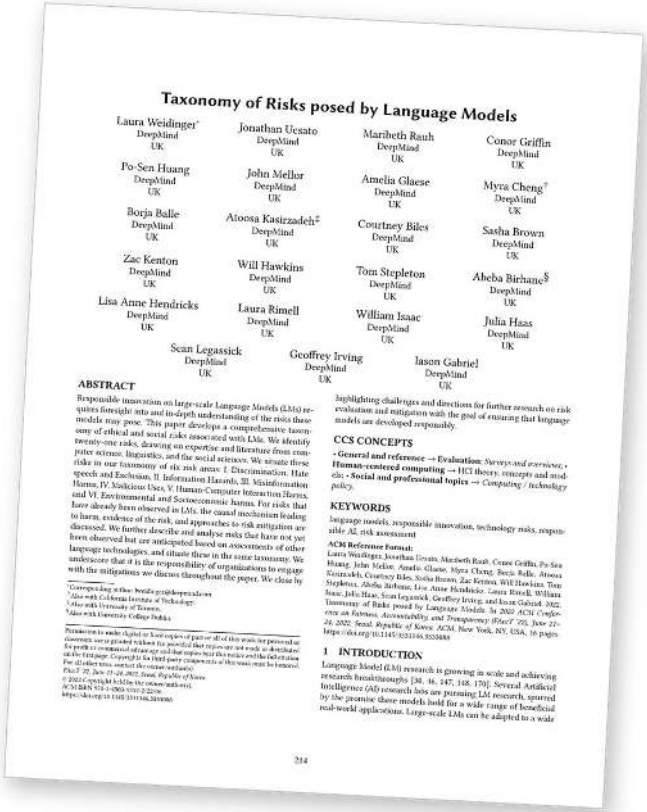
This case concerned AI inputs, not AI outputs.

Memorandum Opinion on Preliminary Approval, Dkt. 437, page 5 (Alsup, J., October 17, 2025)

Authors keep any claims about what the model outputs.

Besides the people who use it, who else does an AI tool **affect**?

Today's readings



Taxonomy of Risks posed by Language Models

Laura Weidinger*
DeepMind
UK

Jonathan Uesato
DeepMind
UK

Maribeth Rauh
DeepMind
UK

Conor Griffin
DeepMind
UK

Po-Sen Huang
DeepMind
UK

John Mellor
DeepMind
UK

Amelia Glaese
DeepMind
UK

Myra Cheng†
DeepMind
UK

Borja Balle
DeepMind
UK

Atoosa Kasirzadeh‡
DeepMind
UK

Courtney Biles
DeepMind
UK

Sasha Brown
DeepMind
UK

Zac Kenton
DeepMind
UK

Will Hawkins
DeepMind
UK

Tom Stepleton
DeepMind
UK

Abeba Birhane§
DeepMind
UK

Lisa Anne Hendricks
DeepMind
UK

Laura Rimell
DeepMind
UK

William Isaac
DeepMind
UK

Julia Haas
DeepMind
UK

Sean Legassick
DeepMind
UK

Geoffrey Irving
DeepMind
UK

Iason Gabriel
DeepMind
UK

ABSTRACT

Responsible innovation on large-scale Language Models (LMs) requires foresight into and in-depth understanding of the risks these models may pose. This paper develops a comprehensive taxonomy of ethical and social risks associated with LMs. We identify twenty-one risks, drawing on expertise and literature from computer science, linguistics, and the social sciences. We situate these risks in our taxonomy of six risk areas: I. Discrimination, Hate speech and Exclusion, II. Information Hazards, III. Misinformation Harms, IV. Malicious Uses, V. Human-Computer Interaction Harms, and VI. Environmental and Socioeconomic harms. For risks that have already been observed in LMs, the causal mechanism leading to harm, evidence of the risk, and approaches to risk mitigation are discussed. We further describe and analyse risks that have not yet been observed but are anticipated based on assessments of other language technologies, and situate these in the same taxonomy. We underscore that it is the responsibility of organizations to engage with the mitigations we discuss throughout the paper. We close by highlighting challenges and directions for further research on risk evaluation and mitigation with the goal of ensuring that language models are developed responsibly.

CCS CONCEPTS

• **General and reference** → **Evaluation**; *Surveys and overviews*; • **Human-centered computing** → HCI theory, concepts and models; • **Social and professional topics** → *Computing / technology policy*.

KEYWORDS

language models, responsible innovation, technology risks, responsible AI, risk assessment

ACM Reference Format:
Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks posed by Language Models. In *2022 ACM Conference on Fairness, Accountability and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3531146.3533088

*Corresponding author. Iwoulds.gorg@deepmind.com
†Also with California Institute of Technology.
‡Also with University of Toronto.
§Also with University College Dublin.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
FAccT '22, June 21–24, 2022, Seoul, Republic of Korea
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9352-2/22/06
https://doi.org/10.1145/3531146.3533088

1 INTRODUCTION

Language Model (LM) research is growing in scale and achieving research breakthroughs [30, 36, 247, 248, 170]. Several Artificial Intelligence (AI) research labs are pursuing LM research, spurred by the promise these models hold for a wide range of beneficial real-world applications. Large-scale LMs can be adopted to a wide



opinion

DOI:10.1145/3625251

Pamela Samuelson

Legally Speaking
Legal Challenges to Generative AI, Part II

Deliberating on inconclusive AI-generated policy questions.

DALL-E, Midjourney, and Stable Diffusion are among the generative AI technologies widely used to produce images in response to user prompts. The output images are, for the most part, indistinguishable from images humans might have created.

Generative AI systems are capable of producing human creator-like images because of the extremely large quantities of images, paired with textual descriptions of the images' contents, on which the systems' image models were trained. A text prompt to compose a picture of a dog playing with a ball on a beach at sunset will generate a responsive image drawing upon embedded representations of how dogs, balls, beaches, and sunsets are typically depicted and arranged in images of this sort.

The intellectual achievement underlying the development of these generative AI models is unquestionably impressive. Some recent lawsuits against Stability AI have, however, raised serious questions about whether the use of copyrighted images as training data constitutes copyright infringement and whether generative AI outputs are infringing derivative works of images on which these models were trained.

My July 2023 Legally Speaking column discussed a different generative AI lawsuit, *Doe v. GitHub*—a proposed class action lawsuit proposed initiated by several programmers. They claim GitHub, OpenAI, and Microsoft violated various laws (but not copyright) because the programming assistance tool Copilot draws upon OpenAI's Codex large language model (LLM), which was trained on billions of lines of open source code available on public Internet sites, to suggest code sequences for particular tasks in response to users' text prompts.

Stability, similar to GitHub et al., has moved to dismiss the lawsuits on various grounds. As this column goes to press, courts have not ruled on those motions. Because *Communications* readers may be more interested in the substantive issues posed in the litigations than in the motions, this column focuses on the substantive claims and likely defenses. Because Stability has not yet publicly defended itself on the merits, this column discusses what are likely to be its principal defenses.

Complaints against Stability

Getty has brought two lawsuits against Stability, one in the U.S. and one in the U.K. Sarah Andersen, a graphic artist, is the lead plaintiff in a similar proposed class action lawsuit against Stability. Getty claims Stability copied more than 12 million photographs, along with textual descriptions of the photographs' contents, from Getty websites. It says that Stability used these images and captions as training data for Stable Diffusion's image model.

The pairing of high-quality photographs with concise descriptions is said to make Getty images much more valuable to Stability as training data than ordinary uncaptioned images available on the Internet. Getty owns copyright on many of these photographs and holds nonexclusive licenses on many others. (Under U.S. law, Getty can only sue Stability for infringement of copyrights and get remedies against it for images in which Getty owns rights.) The complaint states the outputs produced by Stable Diffusion are often substantially similar to, and therefore infringing derivatives of, im-

Weidinger et al., FAccT 2022; Samuelson, CACM, Nov. 2023

21 risks in **six areas**

I. Discrimination, hate speech and exclusion

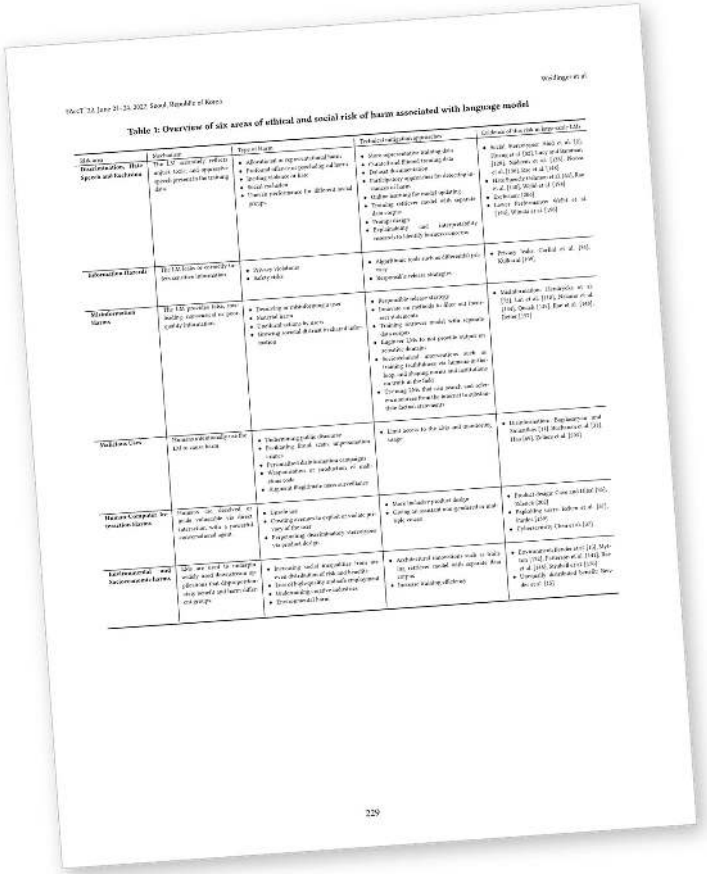
II. Information hazards

III. Misinformation harms

IV. Malicious uses

V. Human-computer interaction harms

VI. Environmental and socioeconomic harms



Weidinger et al., FAccT 2022, Table 1

Areas V and VI

V. Human-computer interaction harms

Anthropomorphising systems can lead to overreliance or unsafe use.

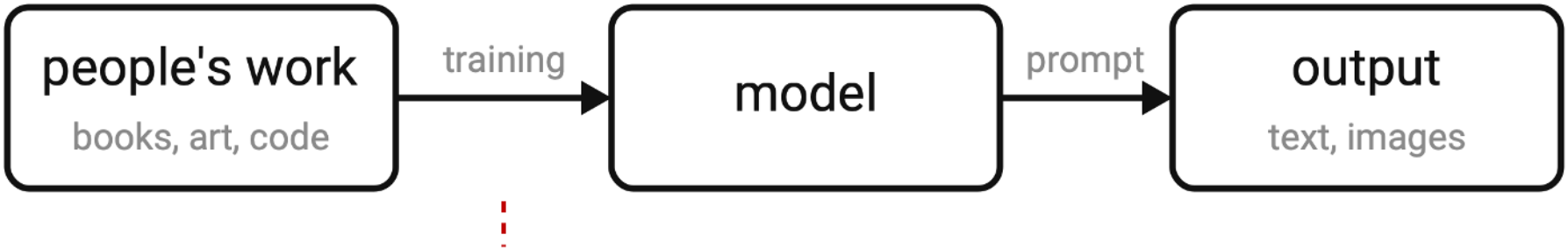
Weidinger et al., FAccT 2022, Section 2.5.2

VI. Environmental and socioeconomic harms

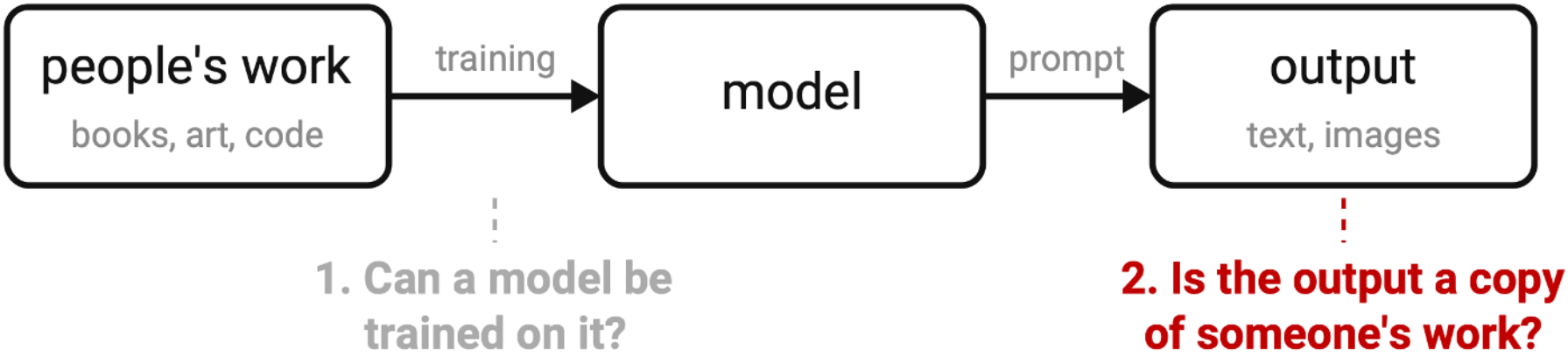
Undermining creative economies. LMs may generate content that is not strictly in violation of copyright but harms artists by capitalising on their ideas, in ways that would be time-intensive or costly to do using human labour.

Section 2.6.2

Three questions in Samuelson's column

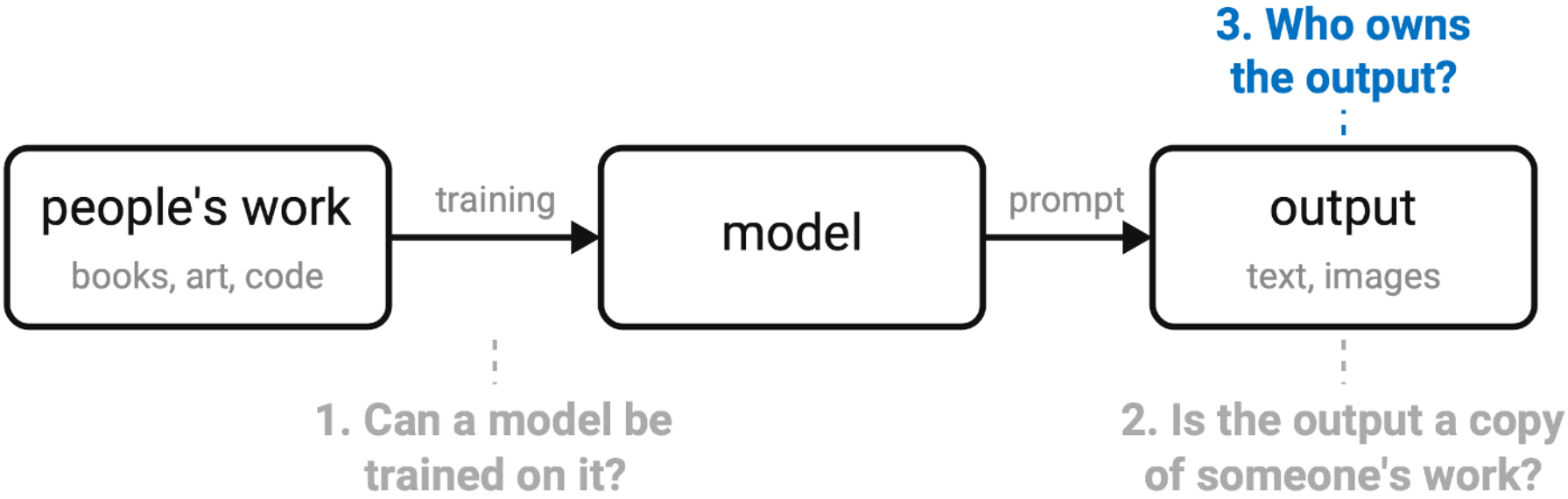


Three questions in Samuelson's column



Samuelson: the lawsuits “raised serious questions about whether the use of copyrighted images as training data constitutes copyright infringement and whether generative AI outputs are infringing derivative works of images on which these models were trained” (CACM 2023, p. 16)

Three questions in Samuelson's column



“the next question is whether the outputs themselves are copyrightable” (CACM 2023, p. 19)

What copyright protects

“copyright law does not protect **styles** as such”

Samuelson, Science 2023, p. 161; Samuelson, Communications of the ACM 2023, p. 18

Fair use: four factors

1. **the purpose and character of the use**
2. the nature of the copyrighted work
3. the amount and substantiality of the portion used
4. **the effect of the use upon the potential market**

17 U.S.C. § 107; Samuelson, Science 2023, p. 160

Discussion: Weidinger and Samuelson

1. Which risk on Weidinger's list
do you now see in tools you use?

2. Does training a model look like the earlier
computational uses Samuelson describes?

3. Which of the six areas could your own project
touch, and which would a user test never show you?

Factor 1: transformative use

In short, the purpose and character of using copyrighted works to train LLMs to generate new text was quintessentially transformative. Like any reader aspiring to be a writer, Anthropic's LLMs trained upon works not to race ahead and replicate or supplant them — but to turn a hard corner and create something different.

Dkt. 231, pages 13-14

Factor 4: the market for the books

But Authors' complaint is no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works. This is not the kind of competitive or creative displacement that concerns the Copyright Act.

Dkt. 231, page 28

Andersen v. Stability AI

SARAH ANDERSEN, et al.,

Plaintiffs,

v.

STABILITY AI LTD., et al.,

Defendants.

Case No. 23-cv-00201-WHO

ORDER GRANTING IN PART AND DENYING IN PART MOTIONS TO DISMISS FIRST AMENDED COMPLAINT

Dkt. Nos. 160, 162, 163, 164, 169

They allege that Stable Diffusion used plaintiffs' artistic works as "training images" and as a result Stable Diffusion can produce output images "in the style" of those images.

Andersen v. Stability AI Ltd., No. 23-cv-00201-WHO (N.D. Cal.), Dkt. 223, page 1 (Orrick, J., Aug. 12, 2024)

Aug. 12, 2024: motions to dismiss

Defendants'

motions to dismiss the Copyright Act claims are DENIED.

Dkt. 223, page 33

Whether true and whether the result of a glitch (as Stability contends) or by design (plaintiffs' contention) will be tested at a later date.

Dkt. 223, page 9

At this stage, the judge decides only whether the claims are plausible.

The Copyright Office's own example

Prompt

professional photo, bespectacled cat in a robe reading the Sunday newspaper and smoking a pipe, foggy, wet, stormy, 70mm, cinematic, highly detailed wood, cinematic lighting, intricate, sharp focus, medium shot, (centered image composition), (professionally color graded), ((bright soft diffused light)), volumetric fog, hdr 4k, 8k, realistic

Output



U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2 (January 2025), p. 19

“prompts do not alone provide sufficient control”

A drawing as the input

Prompt

"a young cyborg woman (((roses))) flowers coming out of her head, photorealism, cinematic lighting, hyper realism, 8k, hyper detailed."

Input



Output



Registered, but only for the drawing

U.S. Copyright Office, Part 2, p. 23 (Kris Kashtanova, Rose Enigma, 2023)

Selecting and editing outputs



(1) Generate Candidate Images with Prompt: *meadow trail lithograph*



(2) Select and Upscale Image



(3) Use Freehand Editing Tool to Select Region



(4) Generate Candidate Images with Prompt: *meadow stream lithograph*



(5) Select and Upscale Image

These are the controls from Lecture 13.

U.S. Copyright Office, Part 2, p. 26

Four harms artists describe

1. Economic loss
2. Digital artwork forgery
3. Hegemonic views and stereotyping
4. Chilling effects on cultural production and consumption

Jiang et al., AIES 2023, Section 4

Which tasks could an LLM speed up?



POLICY FORUM | ARTIFICIAL INTELLIGENCE

GPTs are GPTs: Labor market impact potential of LLMs

Research is needed to estimate how jobs may be affected

TYNA ELOUNDOU, SAM MANNING, PAMELA MISHKIN, AND DANIEL ROCK Authors Info & Affiliations

SCIENCE • 20 Jun 2024 • Vol 384, Issue 6702 • pp. 1306-1308 • DOI: 10.1126/science.adj0998

Eloundou et al., Science 2024

19,265 work tasks, rated by people and GPT-4

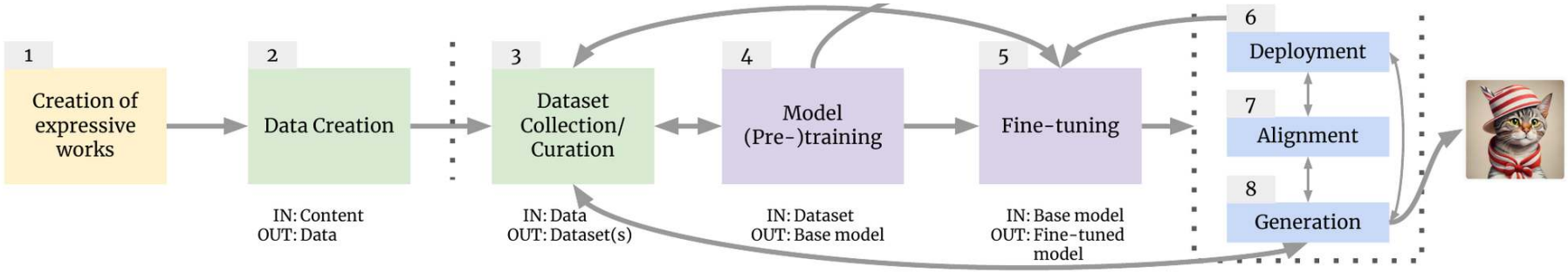
Eloundou et al., arXiv:2303.10130v5, pp. 6-7

1.8% of jobs “could have over half their tasks affected by LLMs”

46% with “software developments that complement LLM capabilities”

Eloundou et al., Science 2024, abstract; arXiv:2303.10130v5 (2023), p. 2

The generative-AI supply chain



Lee, Cooper & Grimmelmann, CSLAW 2024, Fig. 1

“The supply chain reveals all of the places at which companies and users make choices that have copyright consequences.”

Structured debate

Should training a generative model on people's work need their **consent**?

Structured debate: 30 minutes

1. Left half argues yes, right half no
2. Groups of 3-4, 6 minutes: your two best points
3. Openings, 2 minutes each
4. Rebuttals, 2 minutes each
5. Open floor, 4 minutes
6. What could a design do? 8 minutes

From the readings

Yes

Jiang et al.: a proposed law requiring “the explicit consent of content creators”

Weidinger et al.: harm “not strictly in violation of copyright”

No

Bartz order: “Authors cannot rightly exclude anyone from using their works for training or learning as such.”

Samuelson: courts rejected claims over copies made for “computational uses”

Jiang et al. 2023, Sec. 7; Weidinger et al. 2022, Sec. 2.6.2; Dkt. 231, p. 12; Samuelson, CACM 2023, p. 17

Designs that already exist

Opt out

(A second EU exemption authorizes commercial actors to engage in TDM, although copyright owners can opt out of this exemption, as some have done.)

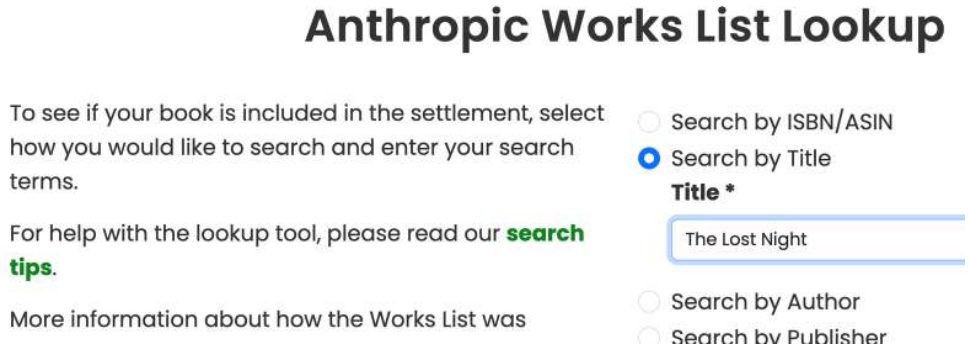
Samuelson, Science 2023, p. 159

Protect your images

For instance, University of Chicago student Shawn Shan and his collaborators, advised by security professor Ben Y. Zhao, created a tool called Glaze that allows artists to add perturbations to their images which would prevent diffusion model based generators from being used to mimic their styles [112].

Jiang et al., AIES 2023, Sec. 6.2

Look up your work and get paid



Works List Lookup, anthropiccopyrightsettlement.com

Where in the supply chain does each one act? Which would the **artists** accept, and which would a **model builder**?

Milestone 4: Decide what it does on its own

Due Tuesday of week 10

Takeaways

1. AI tools also affect people who **never use them**.
2. Training was **fair use** in Bartz, but **pirated** copies were not.
3. Copyright does not protect a **style**, only specific **works**.
4. People who **select** and **edit** an output can **own** more of it.

Next: Lecture 19, Project Studio

No readings and no reading post.

Milestone 4 is due: bring your prototype to test.