**COPYRIGHT + TECHNOLOGY CONFERENCE 2023**
**SEPTEMBER 14, 2023**
**9:30 AM - 5:00 PM**
**FORDHAM UNIVERSITY SCHOOL OF LAW**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**10:00 AM TO 10:15 AM – WELCOME AND INTRODUCTIONS**

***Speaker*:** Bill Rosenblatt, President, GiantSteps Media Technology Strategies

**10:15 AM TO 11:15 AM – THE MODEL TRAIN SET: AI TRAINING MODELS AND THEIR IMPACT ON COPYRIGHT LIABILITY**

There are various techniques for training machine learning systems that use preexisting works in different ways and therefore have different implications for copyright. Assessing the potential liability of those techniques requires understanding of how these techniques work. For example, the techniques used to train large language models (such as ChatGPT) are materially different than those for diffusion or image classification models, and they can change again at the fine-tuning level. Throughout these different processes, notions of reproduction, distribution, and display may or may not be implicated, and indeed traditional notions of what these terms mean may be subject to strain and challenge in the world of AI. In this session, we’ll explain how content becomes data for AI purposes and identify where potential reproduction, distribution, and display of content may occur.

***Speakers*:**

• Moderator: Aleksander Goranin, Partner, Duane Morris
• Matthew Sag, Professor of Law, Artificial Intelligence, and Data Science, Emory University School of Law
• Yacine Jernite, Machine Learning and Society Lead, Hugging Face
• Rebecca Blake, Advocacy Liaison, Graphic Artists Guild

**Introduction**

5 minutes

Aleksander Goranin (moderator, Duane Morris) introduces the speakers and the topic.

To set the stage and provide a launch pad for discussion, Alex provides a short introduction to key generative AI terminology and concepts, i.e., training data, embeddings, weights/parameters, neural networks, pretraining, and prompting.

Alex then offers a roadmap to the four copyright liability questions the panel will discuss after their opening remarks, inspired by the wave of generative AI lawsuits filed since ChatGPT’s launch in Nov. 2022. Specifically:

Fair use: Should the reproduction of copyrighted books, images, and other material during the embedding process be treated as a fair use under 17 U.S.C. § 107?

Derivative works*:* Should the AI model itself be considered a “derivative work”? How about any output synthesized by the model?

Fine-tuning and prompting: If an individual model user fine tunes an AI on a particular artist’s work, or prompts the model to output content similar to an individual artist’s work, should the infringement and fair use analysis differ?

Vicarious copyright liability: Should AI model developers and deployers be considered to have the ability to “control” unauthorized reproduction because they can apply technological measures such as filtering or blocking certain problematic prompts and outputs.

*Supporting written material*:

Rishi Bommasani et al. (Stanford Center for Research on Foundation Models), On the Opportunities and Risks of Foundation Models, 2021, available at <https://crfm.stanford.edu/assets/report.pdf>

Timothy B. Lee and Sean Trott, “Large Language Models, Explained With a Minimum of Math and Jargon,” July 27, 2023, available at <https://www.understandingai.org/p/large-language-models-explained-with>

**Opening statements from the panelists**

5 minutes each

Yacine Jernite (Hugging Face) summarizes three principal points at which a generative AI model can be trained, i.e., (1) pretraining a model, (2) subsequent fine-tuning of a pretrained model, and (3) prompting/instructing a deployed model immediately before inference time. Yacine then introduces his views on the fair use and derivative works questions.

*Supporting written material*:

Giadi Pistilli, Carlos Munoz Ferrandis, Yacine Jernite, & Margaret Mitchell (Hugging Face), “Stronger Together: On the Articulation of Ethical Charters, Legal Tools, and Technical Documentation in ML,” June 2023, available at <https://dl.acm.org/doi/fullHtml/10.1145/3593013.3594002>

Rebecca Blake (Graphic Artists Guild) discusses the creator perspective on copyright liability for AI model training. She explains the Guild’s position against the fair use defense for use generative AI model training, as well as the rationales and motivations behind creators’ message of “consent, credit, and compensation” for use of copyrighted works in AI training.

*Supporting written material*:

“Graphic Artists Guild Position on AI Image Generative Technologies,” June 17, 2023, available at <https://graphicartistsguild.org/graphic-artists-guild-position-on-ai-image-generative-technologies/>

Matthew Sag (Emory Univ. School of Law) presents the core concepts behind (a) his past scholarship on non-expressive fair use, from the Authors Guild/Google Books litigations and (b) his recent scholarship on technological measures to mitigate unauthorized reproduction in generative AI.

*Supporting written material*:

Matthew Sag, “The New Legal Landscape for Text Mining and Machine Learning,” Feb. 27, 2020, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331606>

Matthew Sag, “Copyright Safety for Generative AI,” May 4, 2023, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593>

**Panel discussion of the four copyright liability topics implicated by AI model training**

30 minutes

In this program segment, the panel will interactively debate four substantive open questions regarding copyright liability (infringement and defenses) implicated by large language models (LLMs) and image generators. Many of these questions are prompted—but not yet answered—by the new round of generative AI lawsuits currently pending in the Northern District of California and the District of Delaware.

1. Fair use: The question of whether the fair use defense can and should cover intermediate reproduction of copyrighted works during AI model training—i.e., during the embedding process—has received considerable commentary and attention. Rather than repeat the debate, the panel will focus instead on different frames for the issue, such as:

In the *Andersen v. Stability AI* litigation, plaintiffs have called image generator StabilityAI a “21st century collage tool” implying that it merely remixes copyrighted images in a non-transformative way. Is that a fair characterization of the technical operation of the tool? Should the specifics of the technical operation of the tool dictate the result of the fair use analysis?

Assume technological tools are developed that can determine attribution, i.e., the extent to which one image or text influences the final output of a generative AI model. Should that empirical evidence bear on either a *de minimis* defense or fair use defense?

*Supporting written material*:

Complaint, Jan. 13, 2023, *Andersen et al. v. StabilityAI et al*, 3:23-cv-00201 (N.D. Cal.), available at <https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.1.0.pdf>

Mark Lemley & Bryan Casey, “Fair Learning,” Jan. 20, 2020, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3528447>

2. Derivative works: A copyright liability theory appearing in many of the recent generative AI complaints asserts that if a large language model or image generator is trained on copyrighted works both the AI model itself and the generated output of the model infringe the § 106(2) derivative works exclusive right because they are “based on” the copyrighted material. Is this legal theory viable?

*Supporting written material*:

Complaint, Jun. 28, 2023, *Tremblay et al. v. OpenAI et al*, 3:23-cv-03223 (N.D. Cal.), available at <https://storage.courtlistener.com/recap/gov.uscourts.cand.414822/gov.uscourts.cand.414822.1.0_1.pdf>

Daniel Gervais, “AI Derivatives: the Application to the Derivative Work Right to Literary and Artistic Productions of AI Machines,” Feb. 8, 2022, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4022665>

3. Fine-tuning and prompting: Copyrighted works need not be used only in the pretraining stage of the AI training process. Consider the below scenarios; our panel will discuss whether this use of copyrighted works should affect the copyright liability or fair use analysis:

A user would like to take a pretrained model but then perform additional rounds of training—“fine tuning”—specifically on the works of a select artist, e.g. the novels of Stephen King, to update the model weights to better capture the artist’s style and voice. Does or should this undercut a fair use defense?

One of the interesting emergent capabilities of LLMs is to take a document the user provides to it and to “read” the document and answer questions about it. Moreover, many of the latest LLMs have large “context” windows for prompts, permitting a user to pass in at least 100-300 pages of text as part of a query or instruction. Assume a user feeds a Stephen King novella or a physics textbook and asks the LLM questions about information contained in the text. Does or should this undercut a fair use defense?

4. Vicarious copyright liability: Under copyright law, a defendant is liable for vicarious copyright infringement if he or she has (1) the right and ability to supervise the directly infringing activity and (2) a direct financial interest in the infringing activity. Generative AI model developers and deployers have available to them various “guardrails” or “copyright safety” measures that could mitigate direct reproduction of copyrighted works as AI model output—filtering and blocking mechanisms for instance. Should the availability of these measures affect AI model developers’ and deployers’ secondary liability for infringement?

*Supporting written material*:

*A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (9th Cir. 2001), available at <https://scholar.google.com/scholar_case?case=14102696336550697309&q=am+records+napster&hl=en&as_sdt=6,33>

**Questions from the audience**

10 minutes

**11:30 AM TO 12:30 PM – AUTHOR! AUTHOR? WHAT IS AN “AUTHOR” AND HOW CAN AI BE ONE**

The Copyright Office notes that it “will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author.” This has led to the Office issuing several high profile registrations rejections for works created in whole or in part using artificial intelligence tools. To determine whether, and to what extent, a human author is involved in the creation of works made with generative AI tools, it is important to understand just what we mean by the term “author.” Our panelists will discuss what it means to be an “author,” whether that “author” must be a human being, and the various doctrines adopted by courts throughout the years in trying to identify and define an author. The panelists will compare the U.S. approach to AI-generated work authorship with those of jurisdictions such as the UK that recognize ownership rights.

***Speakers:***

• Moderator: Bill Rosenblatt, President, GiantSteps Media Technology Strategies
• Heather Whitney, Associate, Morrison & Foerster
• Gia Jung, Associate, Cotchett, Pitre & McCarthy LLP
• Vejay Lalla, Partner, Fenwick & West

|  |  |
| --- | --- |
| **11:30-11:40 am** | **Introduction of Speakers**  |
| **11:40-11:50 am** | **History of Technologies in Authorship Determinations*** [Jung, G., “Do Androids Dream of Copyright?: Examining AI Copyright Ownership” (Berkeley Technology Law Journal, Vol. 35, No. 4)](https://btlj.org/wp-content/uploads/2022/01/0009-35-4-Jung.pdf)
* [*Naruto v. Slater*, No. 16-15469 (9th Cir. 2018)](https://copyrightsociety.sharepoint.com/%3Ab%3A/g/ETTwHXmrOaxMv83rzrC8rSgBEsOm2IB8PZUFnUNmbiHJ1Q?e=zrOrfj)
 |
| **11:50-11:55 am** | **Older Cases*** *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884)
* [*Urantia Foundation v. Maaherra*, 114 F.3d 955 (1997)](https://copyrightsociety.sharepoint.com/%3Ab%3A/g/EVs1YfCs_BBEkxnqiIaa5g8BycCvUFf-DHbkHT_HzKylrw?e=3s2GOJ)
 |
| **11:55 am -12:00 pm** | **Current Copyright Office Practice** * Compendium of U.S. Copyright Office Practices, January 28, 2021, Chapter 300, § 313.2
 |
| **12:00-12:10 pm** | **Recent Decisions on AI-Assisted Authorship*** [Memorandum Opinion for Summary Judgment in *Thaler v. Perlmutter*, District of Columbia, 22-cv-01564-BAH (2023)](https://copyrightsociety.sharepoint.com/%3Ab%3A/g/EU8dr7PgeHBGskGYH4ype-sB-qilUSTidkVmHbRqAvOFcg?e=2uDEQx)
* [U.S. Copyright Office letter to counsel for Kristina Kashtanova February 21, 2023 indicating cancellation of copyright registration for AI-generated images in Kashtanova’s graphic novel “Zarya of the Dawn”](https://www.copyright.gov/docs/zarya-of-the-dawn.pdf)
 |
| **12:10-12:20 pm** | **Practical Advice for Practitioners and Clients** |
| **12:20-12:30 pm** | **Audience Q&A**  |

### **1:15 PM TO 2:15 PM – KEYNOTE ADDRESS: AI AND COPYRIGHT CONTRACTS CREATORS: THE PERSISTENT PLIGHT OF FREELANCE AUTHORS ACROSS THE CREATIVE INDUSTRIES**

### ***Speaker*:** Pina D’Agostino

### Ms. D'Agostino is the co-director of the new interdisciplinary research center, Centre for AI & Society, at York University in Toronto and is the co-author of Leading Legal Disruption: *Artificial Intelligence and A Toolkit for Lawyers and the Law*.

**2:45 PM TO 3:45 PM – ARE YOU FOR REAL? IDENTIFYING AND DETECTING AI-GENERATED CONTENT**

The amount of content submitted to commercial services that is generated by AI is poised to explode. Determining which works come from humans and which come from AI–and to what degree–will become more and more important as the volume grows. In this session, we look at efforts to identify AI-generated works proactively as well as emerging technologies for telling the difference after the fact. Are there solutions that will work well enough for rights administration purposes? Our panelists will discuss.

***Speakers*:**
**•** Moderator: Howie Singer, Adjunct Professor, Music Business, NYU
• Claire Leibowicz, Head of AI and Media Integrity, Partnership on AI
• Judith Germano, Founding Member, Germano Law
• Diana Pfeil, Chief Technology Officer, Pex
• Ben Colman, CEO, Reality Defender

|  |  |
| --- | --- |
| **2:45-2:50 pm** | Introduction of Speakers  |
| **2:50-3:10 pm** | **Discussion:** Is it currently possible to detect on either the input or output side whether content was created in whole or in part by AI? If yes, how do these techniques work, how different is the performance for various media types, and will these methods be effective as the learning models evolve over time?* [Diana Pfeil, Real or fake: Identifying AI-generated music and voices. Pex blog post, July 2023](https://pex.com/blog/real-or-fake-identifying-ai-generated-music-and-voices/)
 |
| **3:10-3:20 pm** | **Discussion: can metadata and watermarking or other robust identification techniques be of assistance?*** [Bill Rosenblatt, Google And OpenAI Plan Technology To Track AI-Generated Content. Forbes.com, July 22, 2023](https://www.forbes.com/sites/billrosenblatt/2023/07/22/google-and-openai-plan-technology-to-track-ai-generated-content/?sh=2bb43090131b)
 |
| **3:20-3:35 pm** | **Discussion: What legal obligations or regulatory changes are likely in terms of the obligations on services and the rights of creators?*** [Complaint in purported class action *Tremblay et al v. OpenAI*](https://llmlitigation.com/pdf/03223/tremblay-openai-complaint.pdf)
* [Complaint in purported class action *Kadrey et al v. Meta Platforms*](https://llmlitigation.com/pdf/03417/kadrey-meta-complaint.pdf)
 |
| **3:35-3:45 pm** | **Audience Q&A** |

**4 PM TO 5 PM – TURN THE PAGE: THE FUTURE OF LIBRARIES IN THE WAKE OF *HACHETTE V INTERNET ARCHIVE***

This past March, a district court in New York awarded summary judgment to a group of book publishers in their litigation against the Internet Archive over certain forms of digital book lending. The ruling could have profound implications for libraries as they work to make content available that they preserve in digital form or that is “born digital” on acquisition. On this panel we’ll discuss potential ways forward for libraries if the ruling is let stand on appeal: is controlled digital lending (CDL) dead, or are there ways to achieve its aims that avoid what is now considered liability? Or is licensing going to be the way to fulfill libraries’ missions in the digital age?

***Speakers*:**
**•** Moderator:Christopher Kenneally, Sr. Director, Content Marketing, Copyright Clearance Center
• Bhamati Viswanathan, Faculty Fellow, New England Law
• Lisa Janicke Hinchliffe, Professor & Coordinator for Research and Teaching Professional Development, University of Illinois Urbana-Champaign
• Maria Bustillos, Founder, Popula

|  |  |
| --- | --- |
| **4:00 – 4:10 pm** | **Introduction of Speakers** * Maria Bustillos, Author
* Lisa Janicke Hinchliffe, Librarian
* Bhamati Viswanathan, Lawyer
 |
| **4:10 – 4:25 pm** | **Basis of Case: Violation of First Sale Doctrine (Viswanathan)*** Opinion & Order, *Hachette Book Group et al v. Internet Archive,* S.D.N.Y, 1:20-cv-04160-JGK-OTW
* Consent Judgment and Permanent Injunction Subject to Reservation of Right of Appeal, *Hachette Book Group et al v. Internet Archive,* S.D.N.Y, 1:20-cv-04160-JGK-OTW
* 17 U.S.C. § 109
 |
| **4:25 – 4:35 pm** | **Fair Use Defense (Hinchliffe)*** 17 U.S.C. §§ 107-108
1. <https://www.copyright.gov/title17/92chap1.html#107>
2. <https://www.copyright.gov/title17/92chap1.html#108>
* [Roger C. Schonfeld, Karin Wulf, Rick Anderson, Lisa Janicke Hinchliffe, Joseph Esposito, Roy Kaufman, The Internet Archive Loses on Controlled Digital Lending. *The Scholarly Kitchen*, March 28, 2023.](https://scholarlykitchen.sspnet.org/2023/03/28/internet-archive-controlled-digital-lending/)
 |
| **4:35 – 4:50 pm** | **Internet Archive and “Public Good” (Bustillos)** * Moving forward, First Sale for eBooks
* Own vs. license and the impact on libraries
* <https://www.thenation.com/article/culture/internet-archive-lawsuit-libraries-books/>
 |
| **4:50 – 5:00 pm** | **Audience Q&A** |