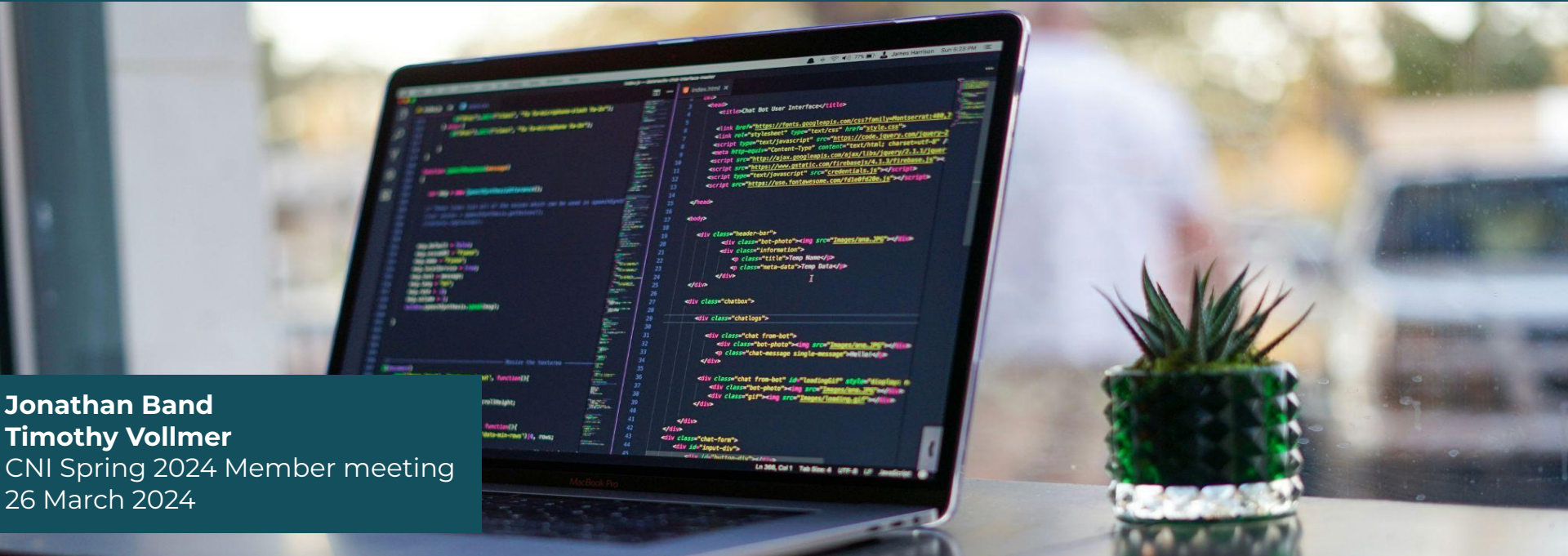


Handling Academic Copyright and AI Research Questions as the Law Develops



Jonathan Band
Timothy Vollmer
CNI Spring 2024 Member meeting
26 March 2024

Overview of the session



1. How's AI being used?
2. Copyright basics
3. 3 © issues implicated by generative AI
4. Challenge of contractual override
5. Publisher licensing banning AI
6. Q&A



Leveling the pitch: How is AI being used?

- Non-generative AI (in use for years)
- Generative AI (new)

What is copyright?

Exclusive rights
to make certain uses
of original expression
for a limited period of
time

Exclusive rights

- Reproduction
- Distribution
- Display
- Performance
- Derivative works

“The **fair use** of a copyrighted work...for purposes such as criticism, comment, news reporting, teaching..., scholarship, or research, is **not an infringement of copyright.**”

17 U.S.C. § 107



"Fair use cat mural, Electronic Frontier Foundation, California, USA" by gruntzooki is licensed under CC BY-SA 2.0

Three Copyright Issues Implicated by Generative AI

1. Does ingestion for training AI constitute infringement?
2. Does the output infringe?
3. Is the output copyrightable?

The answer is different in different jurisdictions

Does Ingestion Infringe?

- Training AI requires ingestion of millions of works
- Software analyzes these works to discover patterns, relationships, and trends, which are reflected in the AI “model.”
- Most of the countries that have examined this issue have determined that ingestion does not infringe.

Does Ingestion Infringe?

- US: Several courts have found that assembly of large databases for AI functions such as search and plagiarism detection are non-infringing.
- But these AI functions were more basic (and less likely to have competitive impact) than generative AI
- Several pending cases, including *New York Times v. OpenAI*
- Israel Ministry of Justice opinion letter finds that copying involved with machine learning is a fair use.

Does Ingestion Infringe?

- EU Copyright in the Digital Single Market Directive Art. 3: permits text and data mining for noncommercial scientific research of lawfully accessed works (is TDM the same thing as AI?)
- EU CDSMD Art. 4: permits text and data mining for other purposes subject to opt-out.
- EU proposed AI Act: would require disclosure of categories of inputs.
- Most companies that scrape works from the Internet respect robot exclusion protocols.
- Singapore and Japan also have exceptions for ingestion.

Does the Output Infringe?

- This can be assessed under traditional copyright principles
- In the US: 1) was there access to the original work?; 2) is the output substantially similar in protected expression?
- Courts have rejected the claim that outputs are per se derivative works

Is the Output Copyrightable?

- US: Only works reflecting human originality can receive copyright protection. So work created completely by GAI is not protectable. But user directs the AI with prompts: at what point is there sufficient human originality?
- U.S. Copyright Office: Must disclose what is created by GAI. This might be more difficult than Office thinks.
- Most other countries have similar human originality requirement, except UK: Copyright Designs and Patents Act 1988 expressly provides for copyright protection of computer-generated works without a human creator.

Why is training AI in research contexts fair use?

1. Purpose & character of use
2. Nature of copyrighted work
3. Amount and substantiality
4. Effect on market for original



CONTRACTUAL OVERRIDE

Even if a use is fair, or if the content is not protected by copyright at all, there may be a contract that restricts scraping, TDM, AI, etc.

Current Model License: Fair Use Savings

Nothing in this Agreement, including but not limited to Authorized Uses, shall be interpreted to diminish the rights and privileges of the Licensee or Authorized Users with respect to any of the Licensed Materials, including exceptions or limitations to the exclusive rights of copyright owners, such as fair use, under Section 107 of the U.S.

Current Model License: Text Data Mining

Authorized Users may use the Licensed Materials to perform and engage in text and/or data mining activities for **academic research, scholarship, and other educational purposes...** and may **utilize and share the results of text and/or data mining in their scholarly work and make the results available for use by others,** so long as the purpose is not to create a product for use by third parties **that would substitute for the Licensed Materials.**

New agreements are banning AI

Regardless of whether
generative or ***non-generative***

Publisher Language

Licensee and Authorized Users may not:

1. use the Subscribed Products in combination with an artificial intelligence tool (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool);

Known publisher concerns

1. Security
2. Competing or commercial product
3. Charging separately to make more money

Fixing Publisher Language

Licensee and Authorized Users may not:

1. use the Subscribed Products in combination with an artificial intelligence tool ~~(including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool)~~
to the extent doing so would: create a competing or commercial product or service for use by third parties;

Fixing Publisher Language

Licensee and Authorized Users may not:

1. use the Subscribed Products in combination with an artificial intelligence tool ~~(including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool)~~ **to the extent doing so would: create a competing or commercial product or service for use by third parties; unreasonably disrupt the functionality of the Subscribed Products;**

Fixing Publisher Language

Licensee and Authorized Users may not:

1. use the Subscribed Products in combination with an artificial intelligence tool ~~(including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool)~~ **to the extent doing so would: create a competing or commercial product or service for use by third parties; unreasonably disrupt the functionality of the Subscribed Products; or reproduce or redistribute the original Subscribed Products to third parties.**

Fixing Publisher Language

Licensee and Authorized Users may not:

1. use the Subscribed Products in combination with an artificial intelligence tool (~~including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool~~) **to the extent doing so would: create a competing or commercial product or service for use by third parties; unreasonably disrupt the functionality of the Subscribed Products; or reproduce or redistribute the original Subscribed Products to third parties. Further, artificial intelligence tools may not be used without commercially reasonable information security standards to undertake, mount, load, or integrate the Subscribed Products on Licensee's or Authorized Users' servers or equipment.**

Addressing 3P Generative AI

We also needed to separately address conditions under which **third-party generative AI** could be used:

Only pursuant to an agreement (e.g. enterprise or API) entered into by Subscriber that imposes commercially reasonable security measures; no data or Subscribed Products are shared with third parties; and all Subscribed Products are removed from the secure computing environment at the termination of this Agreement.

We need a coordinated front

Without consistency or a directive,
our negotiation position is constrained

See also:

UCB Library Copyright Office Comments:

<https://www.regulations.gov/comment/COLC-2023-0006-8194>

Office of Scholarly Communications blog post:

<https://update.lib.berkeley.edu/2023/10/24/uc-berkeley-library-to-copyright-office-protect-fair-uses-in-ai-training-for-research-and-education/>

CNI Fall 2023 Closing Plenary:

<https://www.youtube.com/watch?v=9ABziVkJAm0>

Authors Alliance Post (licensing focused):

<https://www.authorsalliance.org/2024/01/10/licensing-research-content-via-agreements-that-authorize-uses-of-artificial-intelligence/>

Fair use rights to conduct text and data mining and use artificial intelligence tools are essential for UC research and teaching

<https://osc.universityofcalifornia.edu/2024/03/fair-use-tdm-ai-restrictive-agreements/>