AI IN SCHOLARLY COMMUNICATIONS: RISKS, REALITIES, AND POTENTIALS

Presented by Rachel Sweeney

Library Services and the Dominguez Center for Data Science

Bucknell University

AI and Scholarly Communications

WHY ARE WE HERE TODAY?

GOALS

- To examine ethical and legal concerns around AI use, and how it pertains to Bucknell faculty as authors
- Cautions, positives, strategies, and tips for you as authors in a rapidly changing publishing environment
- Information that you can bring to our students– the next generation of scholarly, artistic, and innovation creators

WHAT IS THE SCHOLARLY RECORD?

- No one wants to define it. (Lavoie et al., 2014)
- A popular and often quoted definition:
  - "that which has already been written in all disciplines…that stable body of graphic information upon which each discipline bases its discussions, and against which each discipline measures bases its discussions, and against which each discipline measures its progress." – Roy Atkinson
- For the purposes of this presentation, we'll use this term carefully

WHAT WE TALK ABOUT WHEN WE TALK ABOUT GENERATIVE AI

"Generative AI is a type of artificial intelligence that can learn from and mimic large amounts of data to create content such as text, images, music, videos, code, and more, based on inputs or prompts."

(Harvard University Information Technology, 2023)

THE NEW CHALLENGES IN SCHOLARLY COMMUNICATIONS DUE TO AI

1) Fake Citations, Hallucinations, etc.

2) The Paper Mill Surge

3) Copyright Infringement, DCMA, and Fair Use

CHALLENGE ONE: FAKE CITATIONS/HALLUCINATIONS

"The latest OpenAI systems hallucinate at a higher rate than the company's previous system, according to the company's own tests." (Metz & Weis, 2025)

- Hallucinations in research and publication which can take form as false, misleading, or incorrect citations
- Submissions are increasing because of recursive training (Shumailov et al, 2024)
- In a recent NIH report, it was estimated that LLMs produced in 2025 had increased false citations almost 50% of the time when prompted to perform scholarly research. (Bhattachayya, et al, 2025)

CHALLENGE TWO: PAPER MILLS AND PREDATORS

- End-to-end manuscript development being reported on a a large-scale by major publishers, particularly in STEM publishers
- Surge in "paper mills" and predatory journals since 2023. (Shumailov et al, 2024)
- Falsification of sources, data, and findings that are increasingly hard for editors to detect

CHALLENGE THREE: DISREGARD FOR COPYRIGHTED AND FIREWALLED MATERIALS

"Meta employees turned their attention to Library Genesis, or LibGen, one of the largest of the pirated libraries that circulate online. It currently contains more than 7.5 million books and 81 million research papers." (Reis, 2025)

- Nature, Science, and The Lancet have been scraped. As well as top academic-journal publishers such as Elsevier and Sage (Reis, 2025)
- Much is this is done through "shadow libraries" that contain fully copyrighted work. (Knibbs, 2025)

QUESTIONS? COMMENTS? FOOD OR FOOD/DRINK BREAK?

HOW GEN AI RELATES TO COPYRIGHT

The very nature of Generative AI is to use large amounts of written text, and that licensing has been seen as cost-prohibitive and time-prohibited.

Copyright holders have these rights:

- Make copies e.g., photocopy, scan, import, export
- Distribute and/or publish copies
- Make versions based on the original work
- Perform or display the work publicly
- License it or part of it for monetary gain, or publicity purposes

EXACTLY HOW IS ALL OF THIS HAPPENING?

- GenAI Companies are scraping large amounts of information from behind firewalls, "shadow libraries" and even scans from print journal and books, as well as copyrighted works freely available on the internet
- •Transformative Fair Use is their legal response to copyright infringement.

A FOCUS ON ONE IMPACTFUL CASE

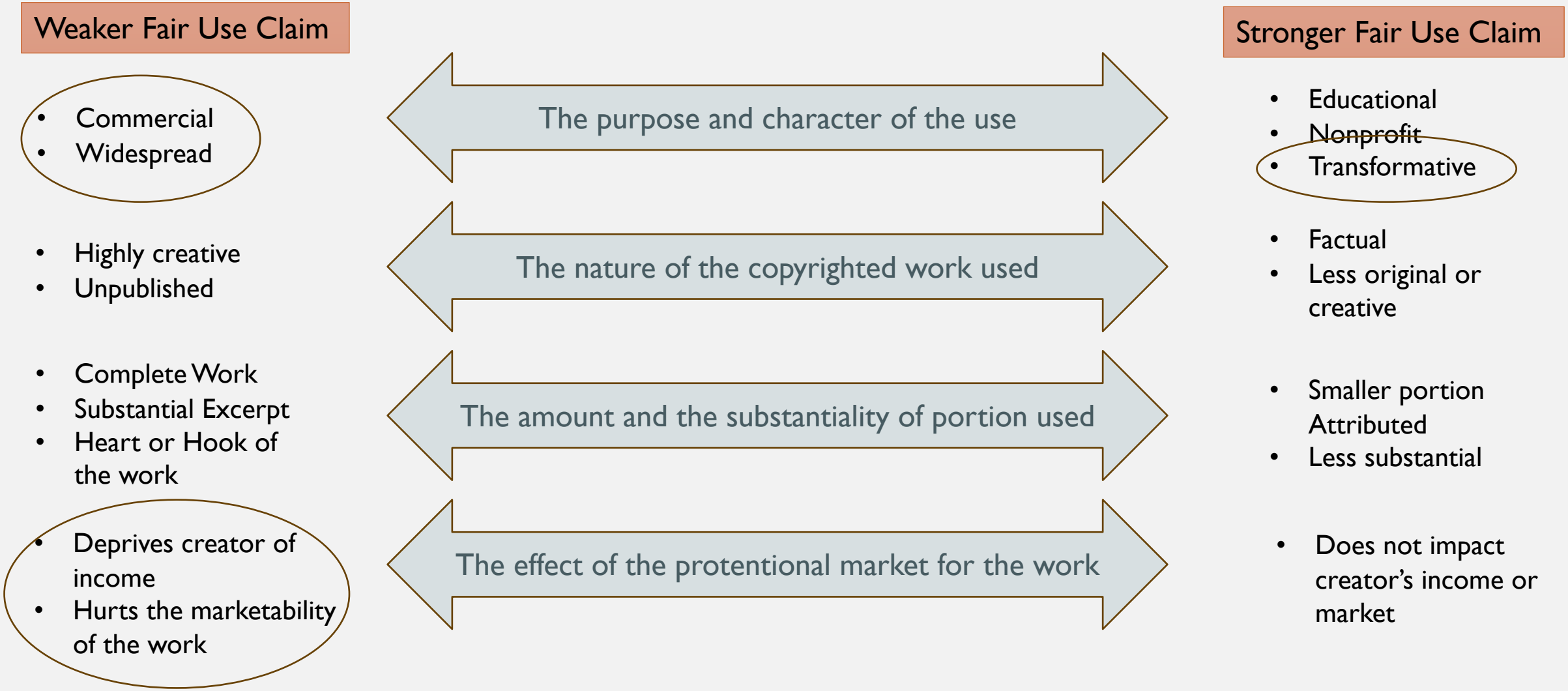
- Kadrey et al v. Meta
  - Large amounts of copyrighted materials scraped and imported into LLMs
  - This case involves commercial publications but the same things are happening to scholarly publishing and output
  - Judges are in high disagreement over this case, which it is something to really watch

FAIR USE

Fair use is a part of U.S. Copyright Law (Section 107) that allows for certain uses of copyright protected materials.

Basic definition of Fair Use: "Fair use permits a party to use a copyrighted work without the copyright owner's permission for purposes such as critism, comment, news reporting, teaching, scholarship, or research." (Copyright Alliance, 2024)

FAIR USE DETERMINATION FOR USING A COPYRIGHTED WORK (OR AN EXCERPT OF)



BASICS OF A FAIR USE (TRANSFORMATIVE) DEFENSE

- Transformative Fair Use is the main legal defense that GenAI companies use
- Transformative use is born out of other parts – weights—of Fair Use provisions
- Its roots are in the late 20th Century and directly related to the advancement of technology

- Roy Orbison and "Pretty Woman" being sampled – music sampling in general
- Andy Warhol, Prince and the Conde Nast saga
- **At the root of the problem is that one person's transformation is another person's piracy, and it is highly subjective to the courts and individual decisions.**

[REAL] ELECTRONIC MESSAGES BETWEEN META EMPLOYEES

"the problem is that people don't realize that if we license one single book, we won't be able to lean into fair use strategy."

"I feel that using pirated material should be beyond our ethical threshold."

"torrenting from a [Meta-owned] corporate laptop doesn't feel right 😃."

(Knibbs, pg. 136, 2025)

TIME FOR QUESTIONS SO FAR

(Before we move into positive!!)

WHEN AI IS MOST APPROPRIATE TO INCORPORATE IN SCHOLARLY RESEARCH

- Data is open
- Success and failure are easy to assess
- Rules are explicit
- Tasks are repetitive

(Mousalason-Sandy quoting Kolluri, 2025, pg. 30)

EMBRACING (OR ACCEPTING?) AI WITH INTEGRITY (UKRIO)

- Follow any governance or policy mandated by your institution, your department
- Appropriate transparency and explainability of usage
- Accountability for your work
- Smaller tasks
- Use as an assistant that needs to be carefully vetted

(*Embracing AI with Integrity - UK Research Integrity Office*, 2025)

QUESTIONS TO YOU

- What constitutes appropriate use of AI?
- If your answer is never, why?
- When does language support veer into substantive content generation? Where is that line? Does it move?
- How should journals or institutions balance transparency, accountability, and inclusiveness when setting policy around AI?

TOWARDS A BETTER FUTURE

- Bring this to your students!
- Liberal arts colleges and universities are in a unique position and should have a bigger seat at the national table on AI, authorship and potential harms
- Be careful, but remember you are the brain trust here. Gen AI cannot replace you.
- Progress is being made
- Possible alliances are being discussed across creative industries

CONCLUSION

Higher education stands at a crossroads: unchecked AI threatens the scholarly record, authorship, and the value of faculty work. With leadership and literacy, Bucknell can model how AI enriches—rather than replaces—scholarship. This seminar invites faculty to take that first step.

MY AI DISCLOSURE

- Chat GPT, Claude, and Gemini to help me brainstorm for an opener. I eventually went a different route, but I took bits and pieces and crafted out of them for slides 2 and 14.
- Chat GTP to make slide 3 (Goals) more succinct and less "Death by Powerpoint." Success? Debatable. One prompt and then tweaked their response into my own version.
- Gemini successfully identified discover the UKROI document. Score one for AI.

QUESTIONS & DISCUSSION

Ways to reach me!

Rachel Sweeney, Scholarly Communications & Copyright Librarian

rs080@Bucknell.edu

REFERENCES

- Atkinson, R. (1990). Text mutability and collection administration. *Library Acquisitions: Practice & Theory*, *14*(4), 355–358. https://doi.org/10.1016/0364-6408(90)90006-G
- Bellos, D., & Montagu, A. (2024). *Who owns this sentence? A history of copyrights and wrongs*. W.W. Norton & Company.
- Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus*, *15*(5), e39238. https://doi.org/10.7759/cureus.39238
- *Embracing AI with integrity—UK Research Integrity Office*. (n.d.). Retrieved September 26, 2025, from https://ukrio.org/ukrio-resources/embracing-ai-with-integrity/
- Federal Courts Find Fair Use in AI Training: Key Takeaways from Kadrey v. Meta and Bartz v. Anthropic. (n.d.). *Jackson Walker*. Retrieved September 19, 2025, from https://www.jw.com/news/insights-kadrey-meta-bartz-anthropic-ai-copyright/
- Hao, K. (2025). *Empire of AI: Dreams and nightmares in Sam Altman's OpenAI*. Penguin Press

REFERENCES

- Kadrey, R. (n.d.). *UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA SAN FRANCISCO DIVISION. 3*.
- Knibbs, K. (n.d.). Meta Secretly Trained Its AI on a Notorious Piracy Database, Newly Unredacted Court Docs Reveal. *Wired*. Retrieved September 28, 2025, from https://www.wired.com/story/new-documents-unredacted-meta-copyright-ai-lawsuit/
- Lavoie, B., Childress, E., Erway, R., Ixchel Faniel, Malpas, C., Schaffner, J., & Werf, T.V. D. (2014). *The Evolving Scholarly Record*. https://doi.org/10.25333/C3763V
- Metz, C., Francisco, K.W. M. reported from S., & Seattle, K.W. from. (2025, May 5). A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse. *The New York Times*. https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html
- Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631* (8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y
- Suchak, T., Aliu, A. E., Harrison, C., Zwiggelaar, R., Geifman, N., & Spick, M. (2025). Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US national health database. *PLOS Biology*, *23*(5), e3003152. https://doi.org/10.1371/journal.pbio.3003152
- UK Research Integrity Office (UKRIO), Woodhams, J., Dally, K., Neave, S., Parry, J., & Scott, J. (2025). *Embracing AI with integrity*. UK Research Integrity Office (UKRIO). https://doi.org/10.37672/UKRIO.2025.06.embracingAIwithintegrity