

LLM 101

Opportunities and Challenges



David J. Gunkel

Northern Illinois University (USA)
dgunkel@niu.edu | @David_Gunkel

Large Language Model

A large language model (LLM) is a type of artificial intelligence program that processes and generates human-like text based on patterns and examples it has learned from a vast amount of data. These models are designed to understand and respond to queries, complete sentences, or even generate content in a conversational, informative, or creative manner.



ChatGPT
OpenAI



OpenAI
Generative Pre-trained
Transformer



Google
Bidirectional Encoder
Representations
from Transformers



ANTHROPIC
Claude
“Constitutional AI”

Opportunities & Challenges

Write original content
Write computer code
Standardized tests
Converse with users
Machine Translation
Answer Questions

Education
Copyright
Disinformation
Technological Unemployment
Bias & Prejudice
Platform Capitalism

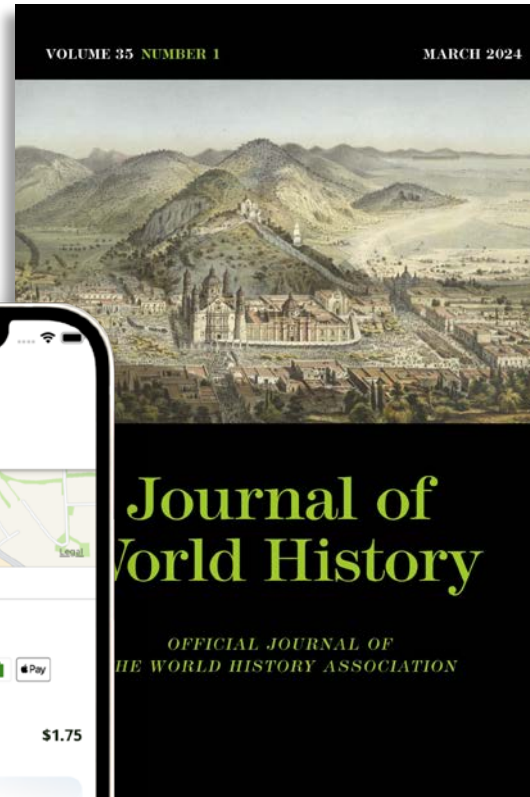
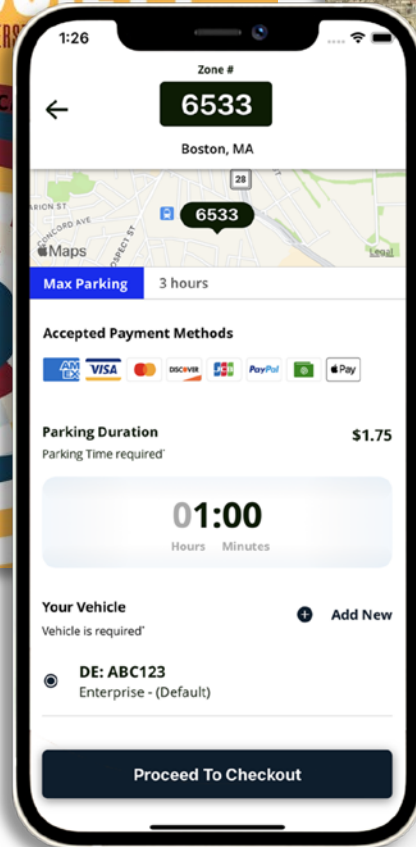
Education

What is the role of writing in higher education?

2 different roles

Writing as an End | Writing as a Means

Education



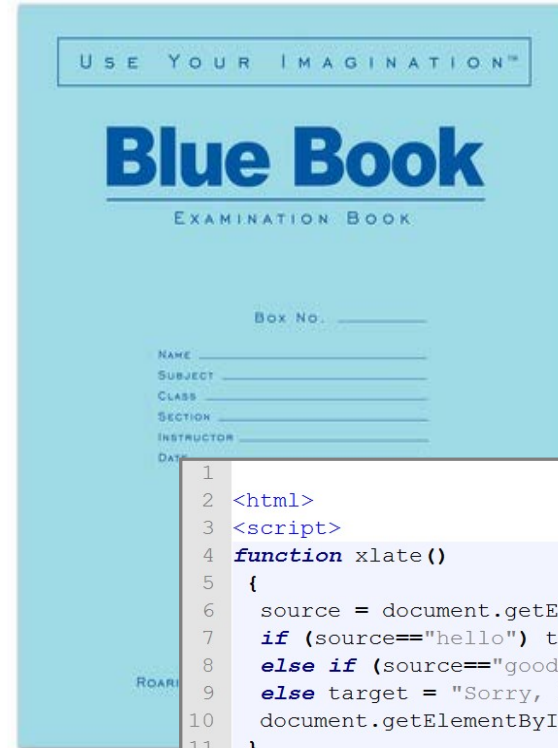
Writing as an End

- Develop innovative content
- Contribution to knowledge
- Product instead of process

Education

Writing as a Means

- Writing for student evaluation
- Evidence of comprehension
- Process instead of product



The Impact of Misinformation on Public Health

A Qualitative Analysis

Nia Norris

Department of Communication Studies, Northern Illinois University

COMS 647: Communication Technology

Dr. David J. Gunkel

December 6, 2021

```
1
2 <html>
3 <script>
4 function xlate()
5 {
6   source = document.getElementById("english").value;
7   if (source=="hello") target = "Guten Tag";
8   else if (source=="goodbye") target = "Auf Wiedersehen";
9   else target = "Sorry, I do not know that one.";
10  document.getElementById("german").value = target;
11 }
12 </script>
13 <body>
14 <form>
15   Basic Translator - English to German
16   <p><input type="text" id="english">
17     <input type="button" value=">>" onClick="xlate()" >
18     <input type="text" id="german"></p>
19 </form>
20 </body>
21 </html>
```

Education

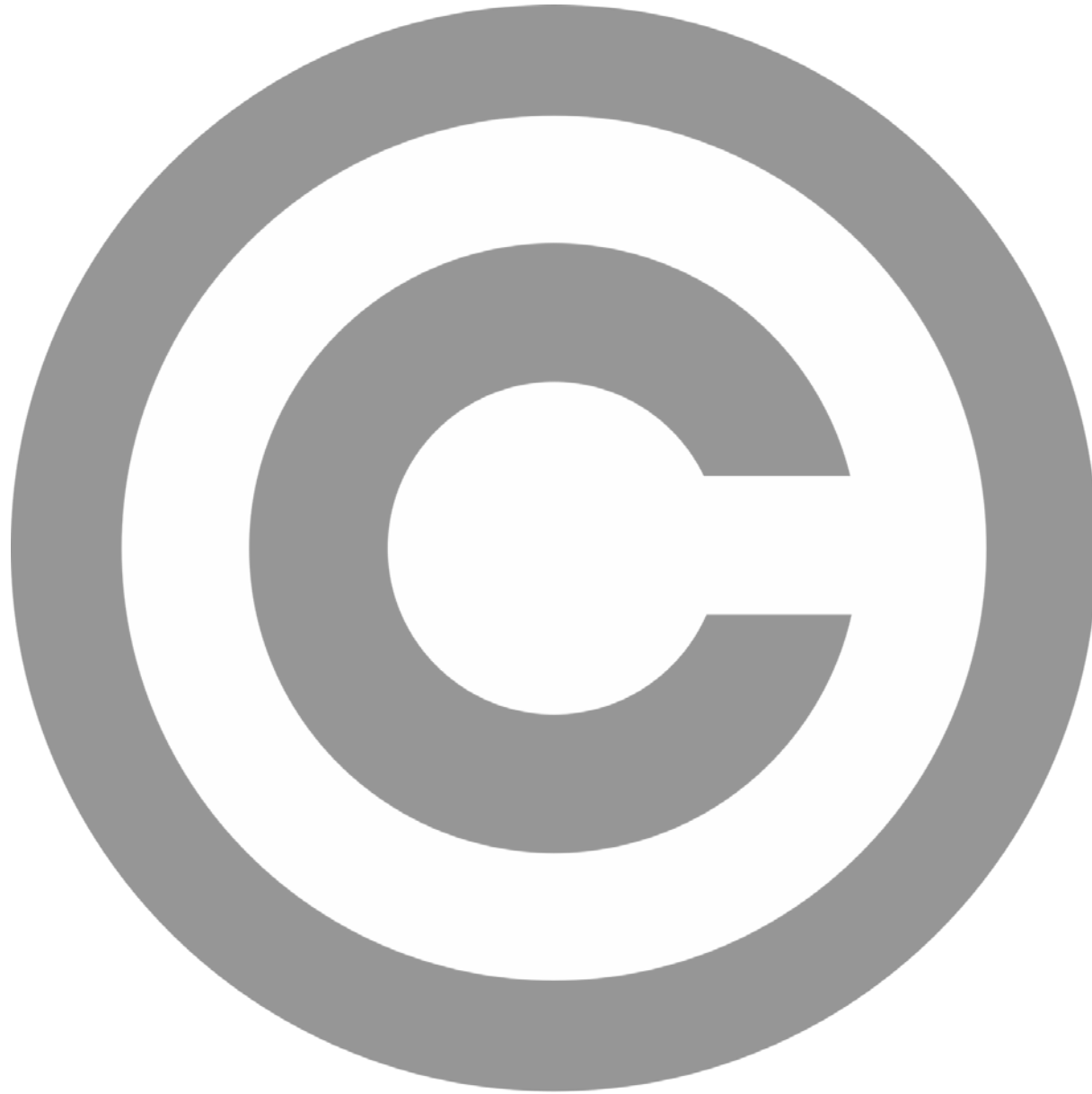
Writing as an End

- Cultivate critical users
- Develop best practices
- Access and engagement
- Faculty model responsible use

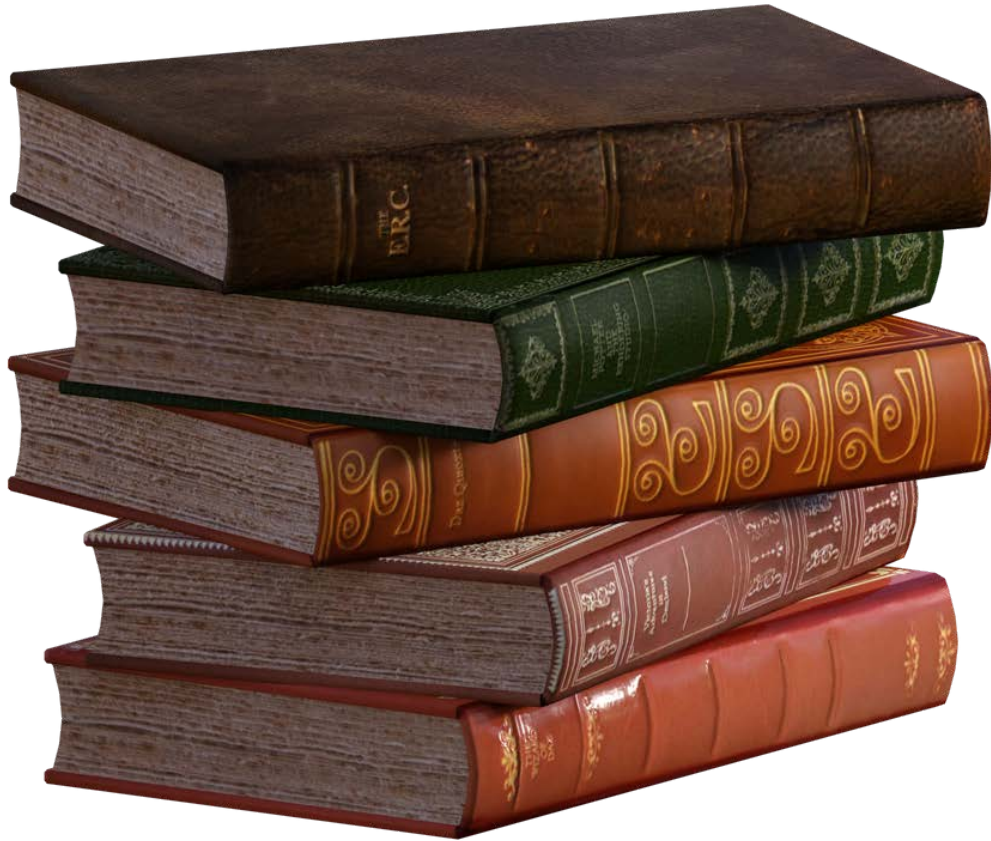
Writing as a Means

- Limit student access and use
- Revise evaluation methods
- Teach the debate
- Faculty confidence with tech

Copyright



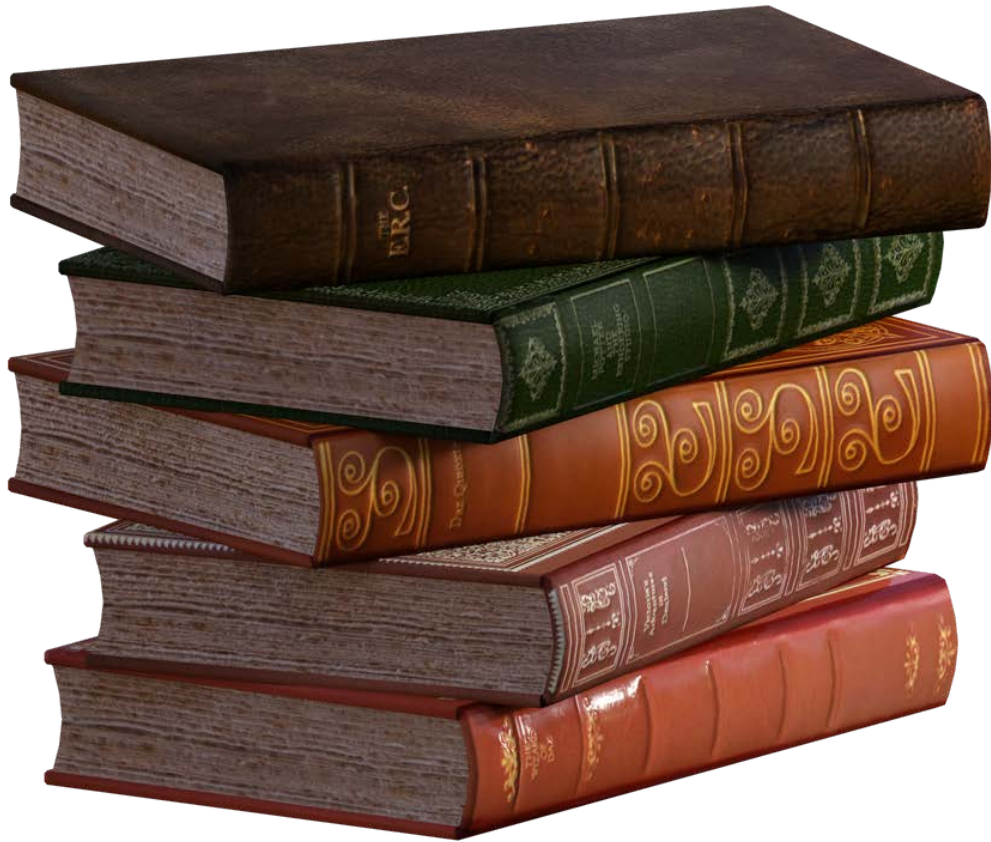
Copyright



Training Data

Training LLMs requires massive amounts of textual data. Basically every piece of writing that resides on the Internet and all books and documents that are accessible in digital form.

Copyright



Public domain

Government documents, literary classics, publications with expired copyright, etc.

Copyright

Publications that can only be used with permission and licensing agreements.

Copyright

Plaintiffs allege that OpenAI and Microsoft used their proprietary content to train the LLM and that this not only violates copyright protections but facilitates the production of new content that unfairly competes with the news organizations.

OpenAI and Microsoft have sought protection under the fair use exception and have argued that compensating publishers for use of their content would render the training of LLMs prohibitively expensive and hinder technological progress.

The New York Times

8 Daily Newspapers Sue OpenAI and Microsoft Over A.I.

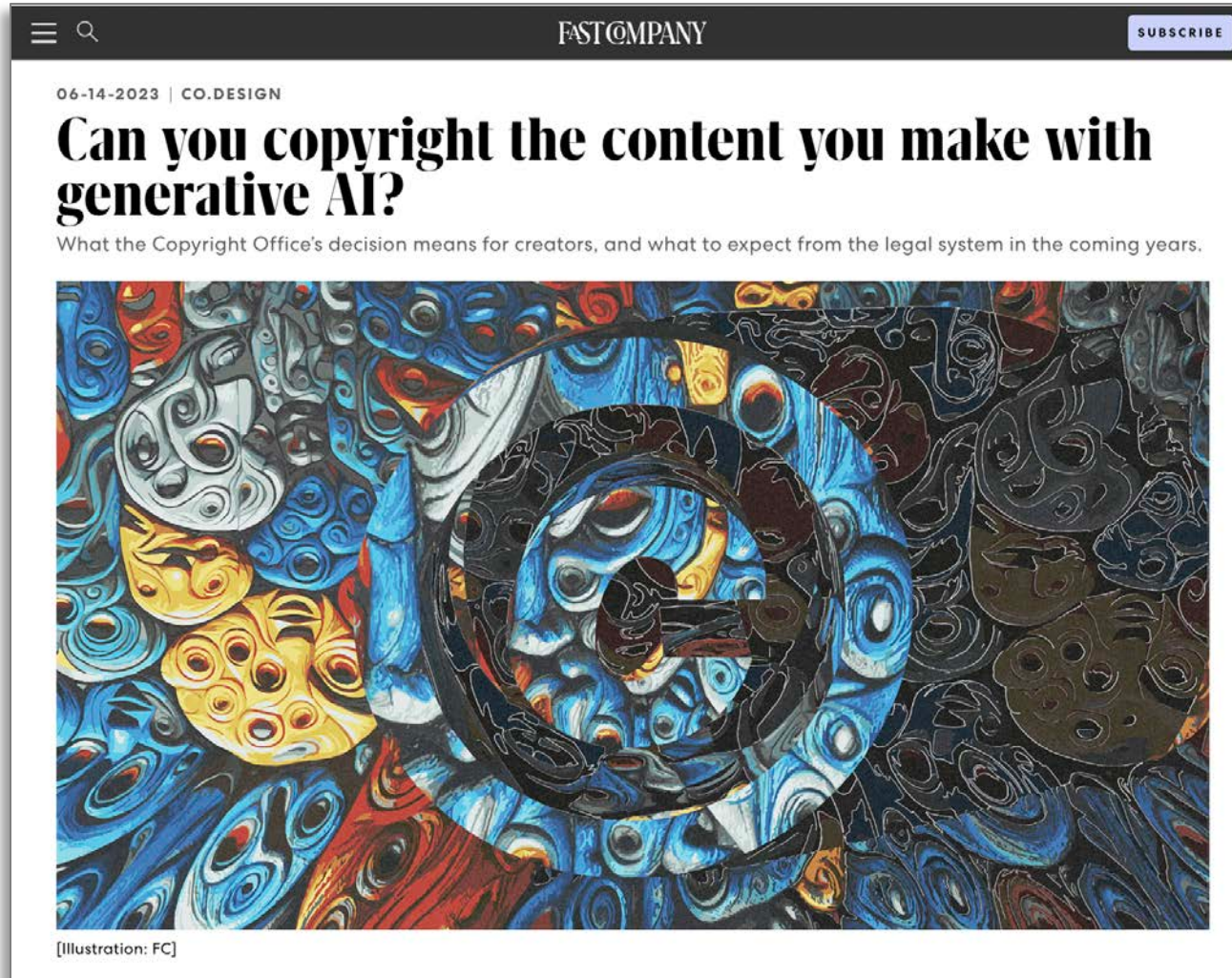
The suit, which accuses the tech companies of copyright infringement, adds to the fight over the online data used to power artificial intelligence.

 Share full article



The firm behind the lawsuit, Alden Global Capital, bought Tribune Publishing, which operated metropolitan dailies including The Chicago Tribune, in 2021. Taylor Glascock for The New York Times

Copyright



NO

AI generated content cannot be registered for copyright.

Problem – IP protections and author identification and rights for AI generated work is uncertain.

Attribution

Artificial Intelligence (AI)

Springer Nature is monitoring ongoing developments in this area closely and will review (and update) these policies as appropriate.

1. [AI authorship](#)
2. [Generative AI images](#)
3. [AI use by peer reviewers](#)

AI authorship

Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our [authorship](#) criteria. Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript.

 APA STYLE

Q MENU

How to cite ChatGPT



By [Timothy McAdoo](#)
Last updated: February 23, 2024
Date created: April 7, 2023
7 min read
 [51 Comments](#)

References



 f x in ✉ 📄

Plagiarism



The attempt of any student to present as his or her own work that which he or she has not produced is regarded by the faculty and administration as a serious offense. Students are considered to have cheated if they copy the work of another during an examination or turn in a paper or an assignment written, in whole or in part, by someone else.

Students are responsible for plagiarism, intentional or not, if they copy material from books, magazines, or other sources without identifying and acknowledging those sources or if they paraphrase ideas from such sources without acknowledging them.

Plagiarism

Is prompting an LLM to generate content for you a form of cheating? Or is it a writing tool?

Is it plagiarism to cut-and-paste or otherwise include LLM generated content in your own work?

Are LLMs a “source” that can be cited. How should they be cited? And how do we verify the citation?

Ideas/Recommendations

1. Faculty development/training in AI
2. AI education across the curriculum
3. Update existing policy statements
4. Develop university guidelines for AI