

JOURNALISM IN THE FACE OF CHATGPT

Katharine Chi

Issue At Hand

- Regurgitation of content, specifically published articles.
 - Copyright claims and Fair Use Laws.
 - Detrimental effects on the journalism field and the dissemination of crucial information.
-



How Does ChatGPT Work?

- A language model that is finely tuned from GPT - 3.5, and it is trained on large sets of data from the internet in order to produce human-like conversational responses.
- Paired with Reinforced Learning with Human Feedback(RLHF). This methodology utilizes human demonstrations and comparison preferences in order to produce a desired interaction.



Benefits



- Assisting with time-consuming tasks
 - analyzing large sets of data and records
 - translating stories
- A new platform for news organizations to connect with readers
- Evolving and supporting the journalism field with the growth potential of AI

Drawbacks



- Spreading Misinformation and Lack of Credibility
 - “Regurgitation” and “Hallucination”
 - Bugs
 - Lack of credible sourcing
- Machine Training Bias
- Intellectual Property Violation

New York Times VS. OpenAI

- First major American media organization to sue OpenAI and Microsoft
- Fair Use Laws and Copyright Claims
 - Generative AI, and specifically ChatGPT, has been using millions of the Times's articles to create their own product with neither permission nor payment from the publishing company



"We look forward to continued collaboration with news organizations, helping elevate their ability to produce quality journalism by realizing the transformative potential of AI."

- OpenAI Blog Post

"OpenAI is acknowledging that they have trained their models on The Times' copyrighted works in the past and admitting that they will continue to copy those works when they scrape the Internet to train models in the future,"

- Ian Crosby

Attorney for the New York Times



OpenAI
ChatGPT **4.0**

- Fair Use Laws
 - Justifies training large language models on copyrighted work
- Legal claim over content on the basis of:
 - Material is open to public
 - Not copying the article in its entirety

The New York Times

- Copyright Claims
 - Using copyrighted articles and creating products
 - Absence of permission and payment
- Undermining publishing organization revenue source
 - membership-only content access
- Lack of accreditation

Key Areas of Discussion and Concern

1

Intellectual Property & Innovation

This case study encompasses how fair use laws and copyright laws demand structured legal policies that take into account intellectual property, but also the innovation of knowledge and creativity.

2

Information Accuracy & Misinformation

Albeit the good intentions towards addressing information accuracy, “regurgitation” and “hallucination” bugs create concern towards the credibility and reliability of the AI-based information, while undermining the journalism field.

3

Bias & Liability

The matter of bias is especially volatile as the reporter position is pinpointed to a learning system rather than a human being. There is a lack of liability for a specific person to assume responsibility and allocate care to critical and sensitive information.

THANK YOU

For Listening

Sources:

- <https://help.openai.com/en/articles/6783457-what-is-chatgpt>
- <https://openai.com/blog/openai-and-journalism>
- <https://arxiv.org/pdf/2310.05189.pdf>
- <https://www.sciencedirect.com/science/article/pii/S156625352300177X>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10310434/>
- <https://www.nytimes.com/2024/01/08/technology/openai-new-york-times-lawsuit.html?login=smartlock&auth=login-smartlock>
- <https://www.cnbc.com/2024/01/18/openai-ceo-on-nyt-lawsuit-ai-models-dont-need-publishers-data-.html>