Western University
Scholarship@Western

Final Projects Summer 2023

LIS 9704: Librarianship and Evolving Technologies

Summer 2023

“Source?” “I Made It Up”: The Ethics of Citing ChatGPT in Academia

Carling DeKay

Follow this and additional works at: https://ir.lib.uwo.ca/fims_evolvingtech_finalproj_summer2023

Carling DeKay

GRAD LIS 9704

Library and Information Science, Western University

Alex Mayhew

6 August 2023

"Source?" "I Made It Up": The Ethics of Citing ChatGPT in Academia

Introduction

If 2023 has been shaped by any technological phenomenon, it has been the year of Artificial Intelligence (AI). And no AI has quite had the overnight success that the large language model (LLM) Chat Generative Pre-Trained Transformer (ChatGPT) has. In the two-month period from its initial launch, ChatGPT managed to obtain 100 million unique users, a statistic that took social media sites TikTok and Instagram nine months and two years from initial launch to complete respectively (Milmo, 2023a). This type of growth is not just unexpected, but also unprecedented. As a result, the general population has been forced to reckon with not only with how to use a new technological tool but the ways in which this tool had the potential to rapidly reshape the ways in which we perform daily tasks.

Academia is of course no exception to this rule. Almost immediately after its debut, the question of how academia might incorporate ChatGPT into its arsenal of tools emerged as an issue. Given the tools ability to generate text-based content off of simple prompts for free, it was inevitable that the discourse around ChatGPT would quickly turn to discussions on academic integrity. Plagiarism, or the act using someone's else's ideas and words without proper credit or attribution (Purdue OWL, n.d.), has been an ongoing discussion in academic communities for decades but the arrival of AI creates new problems and complications to this issue. While there are many issues posed by AI in academia, the use of ChatGPT as a citation source pushed the accepted notions of academic integrity into a more ethically grey area.

What is ChatGPT?

In order to understand why the use of AI and ChatGPT creates an ethical problem to academic integrity, it is vital to understand what ChatGPT is in the first place. At its most basic level, ChatGPT is a free large language model, created by OpenAI, which can generate human-like language in response to text-based prompts given to it by users who sign up to the site for free (Khan, 2023). It is able to do this thanks to its training where it was fed massive amounts of pre-existing written material (Khan, 2023). While Hollywood might have primed the general public to see the coming of AI as the introduction of a computer's ability to become sentient and to be able to think and act like a human, the AI modelled by ChatGPT is not quite at Skynet levels of unlimited power. There are limitations to ChatGPT's abilities with Bogost analyzing the writing of ChatGPT as formulaic in the "structure, style, and content" in its responses which areonly as convincing as context allows it to be (2023). These limitations are echoed by Sam Altman, one of the cofounders of OpenAI, who tweeted within a month of ChatGPT's lauch that it "was incredibly limited, but good enough at some things to create a misleading impression of greatness" and that it was "a mistake to be relying on it for anything important right now" (Altman, 2022). Though Altman's statement is biased and a means of defence against the software's accuracy, it does reveal OpenAI's awareness of the limitations of ChatGPT as a tool. These limitations can be seen in the accuracy of the responses given by ChatGPT. Khan found that it made up a study it claimed existed when asked to share links about studies on sexual health and linked instead to a study on another topic altogether (2023). This echoes Bogost's assertion that ChatGPT "possesses both knowledge and the means to express it" but admits to "just making things up" when pressed on the source of this intelligence (2023). The reason for this is clear: unlike the AI in Hollywood movies, ChatGPT is not yet sentient and is only able "to simulate human-like responses" rather than form its own unique thoughts from nothing (Khan, 2023). ChatGPT

and other AI chatbots like it can speak like a human but they cannot yet think like a human. Though LLMs mimic human speech thanks to the large amounts of text that they have been trained on, they do not "actually understand" what they say (Schaul et al., 2023). In other words, AI systems like ChatGPT are only ever as good as the datasets it has been trained on.

Yet when it comes time to try and find out what datasets LLMs like ChatGPT have been trained on, it is very difficult to get a clear understanding of this information. One reason why tracking down these datasets are difficult is because researchers who create the tools do not really understand how these tools work in the first place. Sam Bowman, an AI scientist, says of AI systems like ChatGPT: "We don't really know what they are doing in any deep sense … We just don't understand what's going on here. We built it, we trained it, but we don't know what it's doing" (Hassenfield, 2023). When it comes to ChatGPT specifically, OpenAI has been elusive about the specifics of the datasets it has used, however, Schaul et al. were able to discover that the huge dataset includes things such as "all of English language Wikipedia, a collection of free novels by unpublished authors frequently used by Big Tech companies and a compilation of text from links highly rated by Reddit users" (2023). Therefore, although a system like ChatGPT has the ability to answer questions and mimic human language to craft a response, we don't know what exactly what sources it is drawing on to create its answer. This becomes even more problematic when companies like OpenAI allegedly purposefully "do not document the contents of their training data — even internally — for fear of finding personal information about identifiable individuals, copyrighted material and other data grabbed without consent" (Schaul et. al, 2023). These legal fears appear to be well founded with American comedian Sarah Silverman teaming up with fellow authors Christopher Golden and Richard Kadrey to sue OpenAI and Meta, alleging their copyrighted work was used to train their respective AIs without the proper permissions (Milmo, 2023b). This suit was echoed by an open letter signed by more than 10,000 authors

in the Authors Guild calling on AI companies to protect writers and ensure their work is used, in the words of president of the Authors Guild Maya Shanbhag Lang, with proper "consent, credit, and compensation" (2023).

ChatGPT, the CommonCrawl Dataset, and Fanfiction

However, despite this obfuscation surrounding what specific information was in the datasets which were used to train LLMs like ChatGPT, sometimes the truth of the information has been revealed inadvertently, thanks to certain niche subsections of the Internet. Fanficition is a subsection of the Internet composed of people who write transformative fiction of pre-existing works of art to post online for free on dedicated forums and websites like Archive of Our Own (AO3). Unlike the Silverman case, the fanfiction writers do not own open the copyright of their work, even when they legally could, as fanfiction has historically operated in a more grey zone when it comes to copyright law (Eveleth, 2023). However, that does not mean that fanfiction is without original and unique characteristics and tropes. One such trope is the subgenre of the Omegaverse. Originally created by the fans of the CW series *Supernatural*, Omegaverse has become a widespread trope across the fanfiction world (Eveleth, 2023). The stories themselves vary but are generally composed of a "specific sexual hierarchy made up of Alphas, Betas, and Omegas in which Alphas and Omegas can smell one another in particular ways, experience "heats," and (usually) mate for life" (Eveleth, 2023). Many of these stories contain specific sexual acts and terms unique to the Omegaverse which appear exclusively within these fanfiction communities and nowhere else, therefore making it "an ideal way to test how generative AI systems are scraping the web" and what might be a part of datasets used to train LLMs like ChatGPT (Eveleth, 2023). A Reddit user who goes by the username kafetheresu did just that by testing the AI tool Sudowrite (which operates on OpenAI's GPT-3 software) by presenting it with terms unique to the Omegaverse and, unsurprisingly, Sudowrite was able to respond to

these prompts, meaning that part of its datasets used for its learning included fanfiction forums like AO3 (Eveleth, 2023).

This should not be necessarily surprising as we do know that most LLMs are trained using the CommonCrawl dataset, a collection of twelve years' worth of publicly available Internet pages, where sites like AO3 with over 11 million freely available works would be an irresistible to training tool for LLMs like ChatGPT (Eveleth, 2023). However, the community from which these works were pulled has more mixed reviews on its use to train AI. Some of the concern is similar to the aforementioned Silverman lawsuit where the fruits of the labour of these fanfiction writers provided to their communities for free is now be used by large AI companies to profit from and whose AI systems themselves will have copyright protection whereas fanfiction writers often do not (Eveleth, 2023). Additionally, fanfiction is an art form that is inherently community-based where the non-commerciality of the endeavour is a huge draw and where, unlike the AI tools, attribution of own's inspiration and influences is considered a part of the culture and good practice (Eveleth, 2023), something which is impossible for a tool like ChatGPT to do. AO3 and its parent company Organization for Transformative Works are aware of the use of their website in the CommonCrawl, announcing in December 2022 that they have put forth code in requesting that the site not be scraped for data for training AI again (Organization for Transformative Works, 2023). However, this action only protects AO3 for future scraping and does not remove the content already taken from the site that is used in current datasets (Organization for Transformative Works, 2023), meaning that it is very possible that AI will continue to be trained using the data collected from that initial crawl and use the labour of the fanfiction writers without proper attribution and for profit.

While LLM like ChatGPT may only be as good as the datasets they are trained on in terms of the content they are able to produce, another new consideration provoked by their

accessibility is what to do with that content provided by AI in the first place. From the perspective of AO3, posting AI-generated fanfiction is not in violation of their Terms of Service as they aim to "include maximum inclusivity of fanworks" in what they choose to preserve (Organization for Transformative Works, 2023). While this decision was not without controversy in the fanfiction community (Adarlo, 2023), it brings forth a new dilemma within the fanfiction community about how to treat art created by AI which cannot be properly attributed. Although fanfiction already exists in a morally grey area of copyright law and relies on the idea of the transformative nature of its works, it serves as an interesting case study when it comes to the dilemma of citing ChatGPT and tracing its sources. Despite the legal grey zone in which fanfiction operates, the culture itself believes very much in proper crediting of its inspirations. ChatGPT threatens this aspect as the use of the tool creates a situation where there is no ability to credit those sources. While its use in the fanfiction community is not a legal issue in terms of copyright law, like with the Silverman case, there is an ethical issue in using fanfiction training datasets for these AI systems' for-profit models. This information was taken without the consent of the original creators and it allows people to use AI to create their own works of art, whether for their own individual profit or not, without proper attribution. The proper accreditation of the sources of inspiration is now impossible. The fatal flaw of LLMs like ChatGPT is that they are not whole sentient beings but rather systems capable of parroting ideas but incapable of providing any clues as to where these ideas came from.

ChatGPT, Citations, and Academia

Though they may seem different, the issues with fanfiction reveal the larger problem with using ChatGPT, especially as a tool to be cited within an academic context. While its freely available system is an extremely powerful one that is capable of producing human-like written communication in a matter of minutes, it cannot ever tell you the sources of that

communication. Therefore, unlike traditional sources of knowledge in an academic paper where you know exactly whose idea is being used to build off of for the author's own ideas, there is no way of knowing where ChatGPT got its information from and whose work it was that the author may be drawing on as a source when they cite ChatGPT. The source provided by ChatGPT could be made up entirely. Within an academic context, using someone else's idea or words without being able to provide a proper attribution is considered plagiarism (Purdue OWL, n.d.). Given the explosion of AI in the last year alone, there is not yet definitive consensus on whether or not citing There has, however, been some discussion already about the proper stylistic technique for citing ChatGPT. All three of the major style guides for academic papers (American Psychological Association ((APA)), Modern Language Association ((MLA)), and Chicago style) have released rough guidelines for how to cite ChatGPT (Caulfield, 2023). While the system in which accreditation varies among them, the fact that each style is already thinking of how to approach acknowledging the use of AI in an academic paper shows that the use of AI is only becoming more and more widespread in academia. While the stylistic elements of the citations differ in accordance with each specific style, they all involve crediting ChatGPT and OpenAI specifically by name (Caufield, 2023). They also contain some interesting variations like APA advising using descriptions of how the tool was used in a methodology or introduction section while MLA advising that if you only used AI to locate sources, you do not need to cite which AI tool you specifically used (Caulfield, 2023). Interestingly even with all of these considerations for hot to cite ChatGPT, ChatGPT is still not considered a "credible source" by academic standards for factual information and using it to write an assignment for you is still considered plagiarism, yet using it as a source is not despite its issues in sourcing its ideas (Caulfield, 2023).

This is the paradox of ChatGPT. It cannot be used as a source of factual information or to write an assignment for you, but it can be used as a tool in the research process and a

source itself (Caulfield, 2023), thereby allowing ideas from ChatGPT to be used in academia without proper attribution. Additionally, ChatGPT's ability to promote academic honestly cannot be relied on as ChatGPT's ability to properly source information is questionable at best, and our ability to rely on it to self-regulate is impossible as the sources it uses to build its answers are a mystery even to the people who developed the software. Therefore, while we might be able to release rough guidelines on how to approach citing ChatGPT in accordance with varying academic styles, none of these solutions really address the core elephant in the room when it comes to using ChatGPT in academic writing. That elephant is that even when when the use of AI in an academic paper is befitting the contents of the paper (for instance a study on how ChatGPT responds to certain keywords), it is impossible to use these tools without using the millions and millions of uncredited and unacknowledged voices that make up the dataset in which it was trained on.

Conclusion

There has not been a single software tool that has exploded in widespread use and popularity faster than ChatGPT. The LLM has brought public widespread access to AI and with it, new ethical concerns and considerations. Unlike Hollywood's treatment of AI, ChatGPT is not yet sentient and is only capable of mimicking ideas it has been trained on using unknown datasets. These datasets are anonymous by design, meaning that all ideas posited by ChatGPT are ones without proper accreditation. Although academic style systems have released rough guides on how to cite ChatGPT, its generative nature means that to use it will always requiring citing ideas without credit, something which could be fits the academic definition of plagiarism. As we move forward into this brave new world of AI technology, the ethics of who gets to profit over the ideas expressed by AI whose access was never freely given by the original creators will only become more and more pressing. Unlike the parrot in a cartoon, there is no way of knowing whose mannerisms ChatGPT is picking up on. Like

the fanfiction community, academia prides itself on its ability to properly credit and source ideas. ChatGPT threatens this ability and the question of how we should cite the tool in academia is not just a question of how to fit it within our style guides but an ethical one of should we be allowed to do so in the first place.

Works Cited

Adarlo, S. (2023 May 19). *Writers Furious When Fanfiction Site Won't Ban AI-Generated Work.* The Byte. https://futurism.com/the-byte/fanfiction-ai-generated-work

Altman, S. [@sama]. (2022 December 10). *ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness. it's a mistake* [Tweet]. Twitter. https://twitter.com/sama/status/1601731295792414720?lang=en

Caulfield, J. (2023 May 15). *ChatGPT Citations: Formats & Examples. S*cribbr. Retrieved August 5, 2023 from https://www.scribbr.com/ai-tools/chatgpt-citations/

Eveleth, R. (2023 May 15). *The Fanfic Sex Trope That Caught a Plundering AI Red-Handed*. Wired. https://www.wired.com/story/fanfiction-omegaverse-sex-trope-artificial-intelligence-knotting/

Hassenfield, N. (2023 July 15). *Even the scientists who build AI can't tell you how it works.* Vox. https://www.vox.com/unexplainable/2023/7/15/23793840/chat-gpt-ai-science-mystery-unexplainable-podcast

Khan, A. (2023 May 9). *What is ChatGPT? For People Who Still Don't Get It.* Vice. https://www.vice.com/en/article/z3mn55/what-is-chatgpt-openai-ai-tech

Milmo, D. (2023a, Feb 2). *ChatGPT reaches 100 million users two months after launch*. The Guardian. https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

Milmo, D. (2023b, July 10). *Sarah Silverman sues OpenAI and Meta claiming AI training infringed copyright*. The Guardian. https://www.theguardian.com/technology/2023/jul/10/sarah-silverman-sues-openai-meta-copyright-infringement

Organization for Transformative Works. (2023 May 13). *AI and Data Scraping on the Archive.* https://www.transformativeworks.org/ai-and-data-scraping-on-the-archive/

Purdue OWL. (n.d.). *Plagiarism Overview.* Retrieved August 6 2023 from https://owl.purdue.edu/owl/avoiding_plagiarism/index.html

Schaul, K., Chen, S. Y., & Tiku, N. (2023 April 19). *Inside the secrete list of websites that make AI like ChatGPT sound smart.* The Washington Post. https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/

The Authors Guild. (2023 July 18). *More than 10,000 Authors Sign Authors Guild Letter Calling on AI Industry Leaders to Protect Writers.* https://authorsguild.org/news/thousands-sign-authors-guild-letter-calling-on-ai-industry-leaders-to-protect-writers/