Investigating Privacy and Security Risks of Using ChatGPT for Schoolwork

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Ganesh Nanduru

Spring, 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Bryn E. Seabrook, Department of Engineering and Society

STS Research Paper

Introducing ChatGPT and its Data Privacy Concerns

ChatGPT, one of the world's most widely used artificial intelligence products, recently suffered from a cyberattack compromising its users' personal data, login information, and conversations with the chatbot (Mudaliar, 2024). This breach affects students directly as a significant portion of students use ChatGPT - in the U.S., a survey of 588 college students revealed 37% use ChatGPT, and 96% of its student users mainly access it for help on schoolwork ("4 in 10 college students are using ChatGPT on assignments," 2024). Now that chatbots are becoming so widespread in academic environments, it is especially important to inform students of the data privacy concerns that chatbots introduce. Furthermore, if student data is compromised by ChatGPT breaches, the academic community must understand the risk posed to the students themselves as well as their educational institutions.

This paper focuses on employing actor-network theory (ANT) to develop an understanding of the complex network of information passing through ChatGPT, to investigate the security and privacy risks of using the chatbot for schoolwork. The paper identifies key technical, organizational, and cultural actors connected to ChatGPT to answer the question, "How safe is it for students to use ChatGPT for school assignments?" In the process, this paper tackles current policies for AI in the classroom, steps taken for privacy with respect to AI from the government and tech companies, and risks posed to students using ChatGPT and their academic institutions in the event of a data breach.

Applying Actor-Networks to Investigate Risks Faced by Students Using ChatGPT

This paper conducts network analysis to answer the question, "How safe is it for students to use ChatGPT for school assignments?" To gain a foundational scholarly understanding of data privacy with respect to ChatGPT, this paper uses documentary research, analyzing technical actors within publications from OpenAI itself as well as statements made by companies and other organizational actors that influence ChatGPT's development to determine its safety for student use. This paper references scholarly articles to build a strong factual foundation to base its claims on data ethics, privacy, and student usage of AI. To supplement the scholarly research, this paper cites news publications in search of exclusive interviews with key human actors in the OpenAI information network. Finally, to connect this network to academia, the paper explores recent interactions between ChatGPT and educational institutions to establish the existing relationship between the two.

A Brief History of ChatGPT in Education

To understand the role of ChatGPT in the academic system, it is crucial to understand how it gained so much popularity amongst students in the first place. ChatGPT is an artificial intelligence (AI) product, specifically a large language model (LLM). LLMs are designed to interpret natural language and process massive corpora of text, which translates into powerful analytical and generative capabilities (Brown et al., 2022, p. 1878). As a result, ChatGPT is strong at reading lengthy assignment specifications and producing corresponding results. For example, a computer science student could ask ChatGPT for an implementation of a graph traversal algorithm in C++ and paste the chatbot's response directly into their assignment, greatly simplifying their schoolwork.

In response to this new technology, the government has scrambled to produce new guidelines outlining what constitutes its acceptable use. A local example is Glenn Youngkin's

Executive Order 30, which went as far as to establish an "Artificial Intelligence Task Force" to shape school AI policies. Clearly, use of AI in the education system has become so concerning that upper levels of government feel the need to step in and regulate. The order goes on to state, "K-12 schools and postsecondary institutions must embrace innovation… as well as ensure appropriate guardrails and necessary constraints exist to safeguard individual data privacy" (2024, p.3). The excerpt outlines the cautious relationship between the government and AI in education, permitting AI to the extent that it increases productivity whilst not infringing on the users' privacy.

OpenAI, the developer of ChatGPT, is a non-profit AI research organization based in San Francisco whose mission is to ensure artificial intelligence benefits all of humanity. However, in 2019, the organization created a profitable subsidiary named OpenAI Global, LLC. The introduction of a for-profit arm brought in new investors with corporate affiliations and private interests - Microsoft, for example, invested $10 billion into OpenAI and is now selling ChatGPT services as part of its Azure platform (Williams, 2023). Microsoft has a long and ongoing history of data breaches as it is a major target of hackers internationally. Most recently, a Russian hacker group gained unauthorized access into Microsoft senior leadership accounts until January 13, 2024 (Heiligenstein, 2024). Microsoft is one of many firms that OpenAI does business with, and to understand the risks of using ChatGPT for schoolwork, it is necessary to gauge the network OpenAI collaborates with and their additional imposed risks.

Applying ANT to ChatGPT

Researchers have been using actor-network theory to investigate sociotechnical systems since the 1980s. Dr. Bruno Latour, one of the founders of the theory, published a book on ANT introducing it as a method of describing sociotechnical phenomena, namely, "to show why the

social cannot be construed as a kind of material or domain and to dispute the project of providing a 'social explanation' of some other state of affairs" (Latour, 2005, p. 1). One strength of ANT as described by Latour is the inclusion of non-human actors and human actors within a network, as the two strongly impact each other and are equally as significant in a sociotechnical system. This paper draws on Latour's ANT to identify non-human actors, especially cultural values regarding the current state of data ethics, and integrates them with human actors (students, teachers, executives) to identify key data privacy concerns of ChatGPT. Actor-network theory saw widespread use in engineering problem definition, and it has even been used to explore the ethics of ChatGPT already in a 2024 study.

Students from the Communication University of Zhejiang used actor-network theory to conduct an ethical study of ChatGPT (Li & Jhu, 2024). In the study, the authors identify nine actors, four human and five non-human, to construct a network to illuminate potential ethical concerns. When investigating the network for algorithmic bias, Li & Jhu find, "data collection companies frequently prioritize the quantity of data sets and their semantic accuracy, disregarding the inherent value and social consequences of the information" (p. 72). This analysis superbly ties in cultural non-human actors of the moral values surrounding data assembly with organizational actors such as the corporations leading the development of ChatGPT. This paper borrows from the researchers' approach of scanning the current state of data ethics within organizations to identify data privacy risks of ChatGPT. While inspection into the data ethics of companies surrounding a product is a big step in advancing ANT, its capabilities are expanded with the new quantitative abilities unlocked by modern computing.

Modern data analysis techniques are now capable of inferring actor-networks. Students from the Amsterdam School of Communication Research assembled a matrix of connected

tweets to analyze a public discourse that occurred during the The United Nations Conference on Sustainable Development, also known as Rio + 20. Upon constructing a network of words, hashtags, and usernames of X threads related to the conference, authors found, “maps also show a strong activist cluster around the #end- fossilfuelsubsidies linked to the actors @Avaaz and @dilmabr, the latter being the username of the former President of Brazil” (Hellsten & Leydesdorff, 2019, p. 8). By interpreting individual X users and their tweets as actors in a network, the researchers found key ideas in the public discourse of sustainability and identified people such as the former president of Brazil who are pivotal to its development. Introducing quantitative analysis to construct an actor-network is a unique approach that is especially relevant to ChatGPT data privacy in education, as the public opinion of ChatGPT strongly influences whether or not students will use it and school boards will permit it.

One critique of actor-network theory is that it is descriptive and not explanatory: it simply draws up connections between parts of a system without actually explaining what each part does (Amsterdamska, 1990). Accepting this critique of actor-network theory, this paper will take additional steps to explain how data flows within the actor-network encompassing ChatGPT, supplementing its actor-network connections with news articles offering explanations on why and how they were formed. This paper aims not to rely on actor-network theory as the sole instrument to investigate the data safety of using ChatGPT, but rather to use actor-network theory to construct a network to gain a thorough understanding of the data flow surrounding ChatGPT. This paper will investigate this network for vulnerabilities, referencing opinions of experts in the field of data science and cybersecurity to bolster its insights.

To answer the question, “How safe is it for students to use ChatGPT for school assignments?,” this paper adapts the classical Latour actor-network theory with the nuances of

modern discourse analysis such as news media analysis and dataset ethical investigations to construct an actor-network describing the flow of data surrounding ChatGPT. The paper examines this network to determine the safety and privacy of data entered by students using ChatGPT.

ChatGPT is Vulnerable Through Insecure Organizational Connections and Harmful Cultural Actors

After investigating the actor-network encompassing ChatGPT, this paper finds that the chatbot has access to sensitive personal information that poses a risk to both the people described by the personal information and users of the chatbot. Larger actor-networks pose more data security risks as they have a greater number of connections that can be compromised, and ChatGPT is encompassed in a massive actor-network due to the enormous amount of data required to train it. Students using the chatbot face the risk of encountering private information that is unsafe to include in their schoolwork, and students must be careful not to provide personal details while prompting ChatGPT, as OpenAI saves their chat history. Within its actor-network, ChatGPT is connected to organizational actors including Microsoft and CommonCrawl that introduce new vulnerabilities due to risks of data breaches. An important cultural actor linking ChatGPT to students is their trust in the chatbot to produce accurate and helpful information for their work. However, to accomplish this, ChatGPT's information network directly involves the students themselves - using their account details to produce its responses therefore putting the students' personal data at risk. Due to risks posed by vulnerable actors surrounding ChatGPT, this paper advises educational institutions to adopt the current professional approach taken by companies integrating ChatGPT in their work - to avoid strict bans and instead offer training to students, advising them on the risks associated with using it for school assignments.

Training Data

Large language models like ChatGPT respond to the user's questions based on patterns learned during pre-training, a process where an AI model learns features and representations of a dataset. In order to train a model to understand entire languages, AI developers must pre-train models on enormous amounts of data. However, by using larger datasets, it is harder to verify that the data used is safe and ethical. OpenAI, in their publication for the GPT-3 model, stated they trained it on the CommonCrawl dataset, prepared by Common Crawl, a non-profit organization (Brown et al., 2022, p. 1885). CommonCrawl is a massive dataset generated by "crawling" the internet, or downloading information by systematically browsing public websites. New websites are discovered by clicking links on known websites, eventually generating a vast network of website connections (Gillis, 2022). Envisioning OpenAI and CommonCrawl as two actors in the ChatGPT data flow, one observes that CommonCrawl is an outlet to millions of websites and data sources. With a larger actor-network, there are more data security risks posed to ChatGPT, such as one of the data sources containing harmful or sensitive private information.

Due to the sheer size of CommonCrawl and the autonomous nature of its data collection, it is difficult to verify its safety for training models. CommonCrawl, in their terms of use, state, "CC cannot guarantee the truthfulness, authenticity, quality, lawfulness or accuracy of the Crawled Content" ("*Terms of use,*" 2024). CommonCrawl does respect requests to not track data such as those included in a website's robots.txt or nofollow, and the crawler is run with the support of Amazon web services, as stated on CommonCrawl's website ("*Common crawl - FAQ,"* n.d.). While this approach protects those with the information technology background necessary to block their data from CommonCrawl, it assumes the consent of any other website it visits without the flags necessary to ward off the web crawler. The process of assuming consent

of online sources is potentially dangerous to those who do not want their data to be used by ChatGPT, but are unaware that their website was crawled by CommonCrawl. Connecting CommonCrawl to Amazon, another organizational actor in the data industry, introduces a new level of technical capability to the crawler, as Amazon has significant computational resources to expand CommonCrawl's scope of data collection. However, as more information and more actors enter ChatGPT's data network, the risk and impact of breaches increases.

Breaches and Leaks

In addition to storing information gathered during pre-training, ChatGPT also stores data during its interactions with users. This includes user IP address, location, chat history, and contact information (Arnott, 2024). Data collection of this nature links ChatGPT's information bank directly to the human actors using it, introducing a new hazard of students unknowingly leaking information about themself and their school to organizations connected to ChatGPT. When discussing how ChatGPT collects user data, Arnott identifies two critical risks: "ChatGPT training from your data and sharing sensitive information… with other users outside of your organization," and, "OpenAI itself becoming a victim of a data breach, exposing the data your users have submitted" (2024). Not only does OpenAI encounter the risk of compromising personal information in the event of a data breach, but other actors partnered with OpenAI such as Salesforce and Microsoft can leak sensitive data from the chatbot if they are hacked.

In addition to the danger of a data breach, data used to train ChatGPT is also vulnerable to extraction by malicious actors querying the chatbot to gain unauthorized access to its data. For example, technology news outlets reported a hack discovered by Google DeepMind researchers that tricks ChatGPT into outputting its training data, containing sensitive private information including people's names, phone numbers, and email addresses (Ray, 2023). The hack was

simple as asking the chatbot to repeatedly say a single word, such as poem, company, or make. Not only does this new attack compromise the data security of the chatbot, but it introduces Google as an adversarial actor connected to ChatGPT, performing its own private research to target the data security of the bot — Google has its own chatbot, Google Bard, that competes with ChatGPT. While OpenAI implements safety measures such as reinforcement learning from human feedback to strengthen the safety of their models' responses, their developers state, "Perhaps the greatest limitation of our models is that, in most cases, they follow the user's instruction, even if that could lead to harm in the real world" (Ouyang et al., 2022, p. 27749). Malicious actors, whether individual hackers or organizational researchers such as Google DeepMind, have the ability to publicize attacks using the chatbot to obtain its own sensitive information. These actors pose a significant threat to students who trust OpenAI with personal data or details about their assignments.

Response to Data Privacy Concerns

In response to these vulnerabilities, companies such as Amazon and Apple have restricted their employees' use of ChatGPT to protect confidential information (Mok, 2023). Academic institutions have also started to restrict usage of ChatGPT, although the reasoning behind these restrictions is more to promote original work and protect against misinformation. However, the rejection of ChatGPT by large organizational actors establishes a cultural actor of mistrust in ChatGPT that has spread from the business world to school policy. For example, a Seattle news outlet reported that Seattle Public Schools had outright blocked ChatGPT from school internet and school devices, with the justification, "Original thought and original work is required of students, and the concern here is that sites like this can produce content that is not original" (Clarridge, 2023). The Seattle educational board's response implies a new cultural actor linking

school administration to ChatGPT: that chatbots must be avoided for aid with schoolwork because they do not produce original content. While this cultural actor protects students' data by discouraging ChatGPT usage, schools must actually acknowledge and publish the data privacy risks posed to students when they create an account and share personal information with OpenAI to adequately prepare their students to face these risks.

While companies like Amazon and Apple have rejected ChatGPT, others embrace it to boost productivity. PwC, a major professional services firm in the U.S., invested significant money into integrating ChatGPT into their workflow, intending to, "power up its U.S. workforce of over 65,000 employees with basic to advanced knowledge of ChatGPT technology" (Kawamoto, 2023). The company made a $1 billion investment into boosting its AI capabilities, and it has also entered a partnership with OpenAI and Microsoft to scale PwC into its industry applications. Joe Atkinson, the company's chief products and technology officer, stated, "the training will help them understand how to interact with the chatbot to get the best first draft and recognize AI's strength is in the crafting of words and not always in the crafting of the facts" (Kawamoto, 2023). Atkinson poses an alternative cultural approach to ChatGPT: that organizations can reap its benefits by preparing their human actors in advance for its data integrity risks. However, circumstances are different for students, as they do not currently have financial backing from large organizational actors like PwC to invest in ChatGPT training.

Students' Perspectives of ChatGPT

Students are placed in a unique situation where they generally have access to ChatGPT ("4 in 10 college students are using ChatGPT on assignments," 2024), they are in an academic environment demanding deliverables such as essays and code assignments to be produced regularly, and they are still in a developmental stage in their professional life where they are not

trained professionally to use ChatGPT but have the ability to use it for their schoolwork. In some cases, students are even provided access to ChatGPT at the same level as professional clients. For example, HooHacks, a hackathon open to high school and collegiate participants hosted by the University of Virginia, offered its students, "free Perplexity Pro AI credits for one year to all participants" ("HooHacks 2024," 2024). Perplexity Pro credits can be used to access sophisticated OpenAI models, such as GPT-4 Turbo, which would normally be locked behind a paywall. Therefore, by signing up for HooHacks, which is open to all students and free to register, a student has access to ChatGPT models that are used at the professional level, only without the corporate training that would typically precede professional usage of ChatGPT. This introduces new corporate actors into the educational side of ChatGPT's network, where private sponsors can push the chatbot into the hands of its students, posing it as a trustworthy productivity tool. Not just limited to the U.S., ChatGPT is gaining the trust of students worldwide.

Researchers from Beijing conducted a study on Taiwanese students, polling them on their information technology preferences for schoolwork. Of the 916 students surveyed, the researchers report, "In the entire sample, 442 students (48.3%) preferred using Google for academic help-seeking, while 474 students (51.7%) preferred using ChatGPT for the same purpose. Overall, the usage rates of both types of tools were relatively high" (Xiang & Yang, 2024, p. 16). An interesting takeaway from this article is that Google and OpenAI are framed as competitors in the service of providing information; earlier, we noticed Google DeepMind employees were the ones who discovered the "poem" data leak hack against ChatGPT. This cements Google as an adversarial, competitive actor to ChatGPT. As ChatGPT gains powerful

adversarial attention from its expanding actor-network, its data security is increasingly threatened.

Nevertheless, the researchers found that students who overestimate the capabilities of ChatGPT trust it more often, explaining that the chatbot is much more intelligent and knowledgeable than them, so its answers should be treated as authoritative. Younger students are likely to prefer ChatGPT to Google, with the older students stating Google is more factually reliable and better for accessing recent literature. These views are concerning, as the trends indicated in the results suggest newer generations of students are starting to use ChatGPT more frequently with less regard for its data ethics and safety. Students blindly trusting ChatGPT with their information is a new cultural actor that may harm them in the event of ChatGPT compromising their data.

Preparing Students for the Professional World

As demonstrated in the case of PwC, ChatGPT is being used widely amongst large organizational actors in the tech industry to boost productivity. Along with a paid subscription to sophisticated OpenAI models, companies are investing in AI training to teach employees the proper usage and risks associated with ChatGPT. To prepare students for the professional world where chatbots are used in everyday tasks, schools should adopt the practice of educating students on the dangers of ChatGPT, as well as teaching them safe prompting techniques. Even for the companies who reject ChatGPT and ban its usage in the workplace, employees have the ability to seek help from the chatbot via personal accounts, which is no different from schools banning it. To address this paper's research question - student usage of ChatGPT brings their school's private information and their personal information into the data network of ChatGPT, which branches out to many vulnerable organizations and malicious actors. However, ChatGPT

is highly useful for generative tasks and will see widespread use amongst students due to a cultural actor instilling students with trust in the chatbot's capabilities. Schools should accept that students will feel encouraged to use ChatGPT without proper knowledge of its data privacy and focus their efforts on ensuring their students understand the tool fully before trusting it with their private information and assignments.

Limitations & Future Work

While this paper addresses the research question holistically by analyzing diverse sources including news articles, journal publications, and product information directly from the developers' websites, it is still limited by the scope of its research. An actor-network completely encapsulating ChatGPT would take years of research and investigation to construct, so this paper opts to narrow the scope of the network to concisely address its research question.

Additionally, a known criticism of actor-network theory is its inability to explain why certain connections are made instead of just describing what the connections are. While this paper aims to supplement its actor-network connections surrounding ChatGPT with news articles reporting on why those connections were made, in doing so it relies on media outlets with their own political and commercial agendas, which complicates the actor-network with an added layer of uncertainty behind each explanation. For further work on understanding the data privacy of ChatGPT, researchers should find opportunities to obtain information directly from employees of relevant organizations such as OpenAI, CommonCrawl, and Microsoft. Interviewing developers of the GPT models and training datasets will reveal authentic, expert perspectives on the data privacy of students using ChatGPT. Finally, the information presented in this paper is subject to change, as data privacy standards continue to evolve with new cybersecurity tools and technologies and institutions respond by changing their policies on ChatGPT usage.

Takeaways for Educational Institutions Designing a ChatGPT Policy

ChatGPT is a widely applicable and popular tool among students seeking to elevate their productivity on their schoolwork. However, using ChatGPT introduces new risks to the students' data privacy. Not only does ChatGPT connect to CommonCrawl, an organizational actor connecting to personal information all over the internet, it also accesses the user's personal information during chat interactions. This sensitive data can be leaked in data breaches of actors with access to ChatGPT's data - its owner OpenAI or any of OpenAI's partnered firms including Microsoft and Salesforce are vulnerable. In response, many companies like Amazon and Apple have established a new opposing cultural actor by banning ChatGPT in the workplace, and organizational actors within education such as Seattle Public Schools have followed suit. However, despite these risks, other actors like PwC and the aforementioned partners have embraced ChatGPT, adopting it into their workforce at scale. While educational institutions should not follow the profit-seeking intentions of companies rapidly integrating ChatGPT into their toolkit, they stand to learn from companies training their employees how to safely use it. Students face a significant risk of leaking personal and institutional information when interacting with ChatGPT, but it is better for their professional growth to educate students on responsible ChatGPT usage, as opposed to banning and ignoring the issue. By focusing on preparing the human actors facing the risks of vulnerable data security, schools can find a balance between embracing ChatGPT's technical capabilities while protecting their students' data.

References

4 in 10 college students are using ChatGPT on assignments. Intelligent. (2024, March 11). https://www.intelligent.com/4-in-10-college-students-are-using-chatgpt-on-assignments/

Amsterdamska, O. (1990). Surely You Are Joking, Monsieur Latour! [Review of *Science in Action*, by B. Latour]. *Science, Technology, & Human Values*, *15*(4), 495–504. http://www.jstor.org/stable/689826

Arnott, B. (2024, February 15). *Yes, CHATGPT saves your data. here's how to keep it secure.* Forcepoint. https://www.forcepoint.com/blog/insights/does-chatgpt-save-data

Brown, T., Mann, B., Ryder, N., Subbiah, M., D, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2022, July 20). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*. 1877-1901.

Clarridge, C. (2023, January 24). *CHATGPT banned by Seattle schools — for now - axios Seattle*. Axios Seattle. https://www.axios.com/local/seattle/2023/01/24/chatgpt-banned-seattle-schools-artificial-intelligence

Common crawl - FAQ. Common Crawl - Open Repository of Web Crawl Data. (n.d.). https://commoncrawl.org/faq

Gillis, A. S. (2022, September 30). *What is a web crawler? everything you need to know from techtarget.com*. WhatIs. https://www.techtarget.com/whatis/definition/crawler

Heiligenstein, M. X. (2024, February 20). *Microsoft data breaches: Full timeline through 2024*. Firewall Times. https://firewalltimes.com/microsoft-data-breach-timeline/

Hellsten, I., & Leydesdorff, L. (2019). Automated analysis of actor–topic networks on twitter: New approaches to the analysis of socio-semantic networks. *Journal of the Association for Information Science and Technology*, *71*(1), 3–15. https://doi.org/10.1002/asi.24207

HooHacks 2024. (2024). https://hoohacks-2024.devpost.com/

Kawamoto, D. (2023, May 18). *PWC's HR, Tech leaders prepare to train U.S. workforce on CHATGPT technology*. HR Executive. https://hrexecutive.com/pwcs-hr-tech-leaders-prepare-to-train-u-s-workforce-on-chatgpt-technology/

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Li, Y., & Zhu, J. (2024). An ethical study of generative AI from the actor-network theory perspective. *International Journal on Cybernetics & Informatics, 13(1)*, 67–78. https://doi.org/10.5121/ijci.2024.130106

Mok, A. (2023, July 11). *Amazon, Apple, and 12 other major companies that have restricted employees from using chatgpt*. Business Insider. https://www.businessinsider.com/chatgpt-companies-issued-bans-restrictions-openai-ai-amazon-apple-2023-7

Mudaliar, A. (2024, February 1). *ChatGPT Leaks Sensitive User Data, OpenAI Suspects Hack*. Spiceworks Inc. https://www.spiceworks.com/tech/artificial-intelligence/news/chatgpt-leaks-sensitive-user-data-openai-suspects-hack/

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in Neural Information Processing Systems* (Vol. 35, pp. 27730–27744). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf

Ray, T. (2023, December 4). *CHATGPT can leak training data, Violate Privacy, says Google's Deepmind*. ZDNET. https://www.zdnet.com/article/chatgpt-can-leak-source-data-violate-privacy-says-googles-deepmind/

Terms of use. Common Crawl. (2024, March 7). https://commoncrawl.org/terms-of-use

VA Exec. Order No. 30 (2024). https://www.governor.virginia.gov/media/governorvirginiagov/governor-of-virginia/pdf/eo/EO-30.pdf

Williams-Alvarez, J. (2023, December 7). *OpenAI's unusual board: Should it change its structure to govern effectively?* The Wall Street Journal.

https://www.wsj.com/articles/openais-unusual-board-should-it-change-its-structure-to-govern-effectively-e3d5ee76

Xiang, M., & Yang, X. (2024). *Google or ChatGPT: Who Is the Better Helper for University Students*. https://doi.org/10.48550/arXiv.2405.00341