A SOCIO-TECHNOLOGICAL ANALYSIS OF CHATGPT's EFFECTS ON INFORMATION SPREAD AND PERCEPTION

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

Eric Xie

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

Joshua Earle, Department of Engineering and Society

Introduction

On February 9th, 2023, ChatGPT, a language-based AI, received a passing score on the United States Medical Licensing Exam without any prior reinforcement or training (Kung, et al., 2023). Since then, the American Medical Association has discussed ChatGPT with Dr. Triola, the founding director of the Institute for Innovations in Medical Education, citing ChatGPT as a powerful tool for supporting medical students through their education ("ChatGPT in Medical Education", 2024). This innovation demonstrates the power of AI to transform critical sectors while raising questions about the implications of AI-generated content in broader contexts, such as information dissemination and public perception in social media.

Large language models like ChatGPT are advanced AI systems capable of generating human-like text based on vast amounts of data. These models have shown remarkable abilities in various tasks, including content creation, language translation, and question-answering. While LLMs offer significant advantages in terms of efficiency and scalability, their impact extends beyond technological innovation. The ability of LLMs to produce convincing and coherent text raises important questions about the authenticity of information, the potential for misinformation, and the ethical implications of AI-generated content. These issues are explored by studying how LLMs like ChatGPT influence the spread and perception of information in society.

While ChatGPT and similar powerful LLMs offer significant benefits to its users, such as the enhancement of accessibility to information and aiding in combating misinformation, they also present challenges, such as the spread of misinformation and ethical concerns. Using the technopolitics framework, the effects of ChatGPT is comprehensively analyzed by evaluating its context, purpose, scale, actors, and synchronization (Kurban, 2017).

This paper begins with an initial outline of the research approach, focusing on the synthesis of academic information and the technopolitics framework. Next, this paper discusses the dual impact of ChatGPT and similar LLMs on misinformation and their contributions to combating false information. The analysis then delves deeper into the technopolitics framework, examining the specific effects of ChatGPT on societal dynamics and power structures. Lastly, this paper highlights the wider implications of AI-generated content.

Methods

This analysis combines information using sources accessed through Google Scholar and ProQuest as provided by the University of Virginia Library, to effectively answer the question: "How has AI-generated content influenced human behavior, perception, and communication on social media platforms and news websites?" This research is primarily guided by previous studies on the evolution and advancement of ChatGPT, the issues associated with its development, interactions with ChatGPT at various political scales, and the interplay between ChatGPT and social media. By employing this methodology, this research aims to understand the multifaceted ways in which AI-generated content alters interactions with digital information.

I analyze the impact of ChatGPT through the technopolitics framework, focusing on the following key components: context, purpose, scale, actors, and synchronization (Kurban, 2017). By utilizing this framework, this research aims to comprehensively analyze ChatGPT's influence by evaluating the context surrounding its development and implementation, its intended and actual purposes described by its creators, the various political scales of its impact, the key actors involved in its ecosystem, and ChatGPT's synchronization and interactions with other organizations. The technopolitics framework provides a detailed assessment of ChatGPT's role

in the sociopolitical landscape, which is shaped by information and communication technologies. First, analyzing the context surrounding ChatGPT reveals how this technology fits into the broader societal background. Understanding ChatGPT's purpose helps uncover the intentions behind its development and utilization, revealing whether it aligns with important values such as democratization or empowerment. Evaluating the impact of ChatGPT at various political scales enables us to identify the extent to which ChatGPT may impact power dynamics and the reconfiguration of political structures at various political levels. Assessing the key actors reveals how the empowerment of individuals and dynamics between traditional and emerging institutions may shift. Finally, analyzing ChatGPT's synchronization demonstrates how this tool influences the spread of information alongside other societal systems. Overall, the technopolitics framework is fitting to the task of analyzing ChatGPT as it allows for a nuanced examination of the societal and political implications that this technology can have.

The technopolitics framework stresses the importance of the perspectives of key actors, or social groups that are profoundly affected by AI-generated content. We define the key actors to be general social media users, content creators, policymakers, and AI developers (Caramancion, 2023). General social media users, including young adults, political activists, and members of minority communities, are particularly impacted by AI-generated content. These groups rely on social media for information and communication, making them vulnerable to misinformation and manipulation by AI algorithms. (Abdullah et al., 2022). Content creators, including independent journalists, educational content providers, and social media influencers, play a crucial role in the AI-generated content ecosystem. They use AI tools for content creation and curation, but also face challenges in maintaining the authenticity and credibility of their work. Policymakers, including legislators focused on technology policy, data protection

authorities, and international organizations, are tasked with regulating AI technologies. Their decisions shape the legal and ethical framework for AI-generated content, balancing innovation with the protection of public interests. Lastly, AI developers, encompassing researchers in academic institutions and engineers in tech companies, are responsible for creating AI-generated content tools. Their design choices and ethical considerations directly influence the societal impact of these technologies, including issues of bias, transparency, and accountability.

Because of the vast audience that is affected by such a powerful technology, there exist several auxiliary groups that are also impacted by ChatGPT whose perspective is also relevant to this topic. Included in these groups are educators, academic researchers, and various advocacy groups (Ray, 2023; Baidoo-Anu & Ansah, 2023; Bozkurt et al., 2023). The perspectives of these groups are important, but this project focuses on the identified primary groups that are directly impacted by ChatGPT to maintain a feasible scope.

Through the selection of these groups, the information survey focuses on understanding how each of these groups are uniquely affected by ChatGPT and AI-generated content. This approach promotes an in-depth understanding of the distinct perspectives and experiences unique to these groups to shed light on the nuances of how these technologies reshape societal dynamics (Stahl, 2023). Furthermore, by concentrating on a variety of stakeholders, this research crafts more nuanced analyses and recommendations for the future. Policies that take the insights of these groups into consideration will be far more effective and inclusive (Stahl, 2023). Overall, by centering this research around the viewpoints of these social groups, the findings significantly contribute to the informed development of future policy related to AI.

Analysis

The technopolitics framework acts as the primary analytical lens to better understand the multifaceted effects of ChatGPT on information dissemination and public perception in this research (Kurban, 2017). This framework offers a comprehensive approach to scrutinize the societal and political ramifications of AI-generated content. By focusing on the key components of the framework, context, purpose, scale, actors, and synchronization, the complex dynamics that shape the impact of ChatGPT in our current society can be unraveled.

The purpose of this analysis has two major components. First, ChatGPT must be interpreted within the broader technopolitical landscape to fully identify how ChatGPT influences and is influenced by sociopolitical factors. Second, the deeper implications of these effects in terms of power dynamics, social structures, and the distribution of information, must be uncovered. Through this lens, a nuanced understanding of the role of AI-generated content in shaping contemporary discourse and its potential consequences for society can be obtained.

Context

The launch of ChatGPT for public use in November 2022 marked the start of a new era where Google, Meta, Microsoft, and other leading technology companies around the world would begin developing and iterating their own LLMs (Wu et al., 2023). However, this seemingly sudden breakthrough was the result of several years of global research and development in AI and machine learning. Focusing on OpenAI specifically, their vision for ChatGPT was made clear in 2016 in an article focusing on the potential for generative models to change the future (Karpathy, 2016). OpenAI continued to make significant advancements in natural language processing, including the ability to refine a language model based on human preferences and training models to adhere to human instructions (Ouyang et al., 2022; Ziegler et

al., 2019). The impact of these advancements is clearly displayed in the capabilities of the current version of ChatGPT.

As the capabilities of LLMs and AI grow, so do the concerns of potential repercussions of these models. Initially, these concerns focused on broad issues such as mitigating AI's unintended behaviors or ensuring adaptability to various environments (Amodei et al., 2016). These concerns became far more targeted and specific as the developing early iterations of ChatGPT made the capabilities of this technology clear. In testing the security of ChatGPT, researchers were able to extract personally identifiable information, such as names, phone numbers, and email addresses, as well as online communication logs from GPT-2, ChatGPT's predecessor (Carlini et al., 2021). This raised significant privacy concerns given that these LLMs are trained using data containing private information (Carlini et al., 2021). This trend of growing repercussions continues with the most powerful version of ChatGPT available today, GPT-4. In fact, GPT-4 was found to be more vulnerable when given malicious prompts, potentially because it follows the attacker's instructions more precisely (Wang et al., 2024).

The growing negative impact of issues within ChatGPT that come with the increasing capabilities reveals a clear and worrying pattern. Furthermore, issues identified earlier on in ChatGPT's development, such as training data leakage observed in GPT-2, were not addressed before the creation of subsequent iterations, and continue to occur in GPT-3.5 and GPT-4. As AI researchers continue to develop more powerful models, the continuation of this trend could lead to potentially disastrous results.

Purpose

ChatGPT is designed to be a step in OpenAI's plan towards iteratively developing an artificial general intelligence capable of solving human-level problems (Achiam et al., 2023; Brundage et al., 2022). The current iteration, GPT-4, is meant to be used in a wide range of applications, such as dialogue systems (such as chatbots or virtual assistants), text summarization, and machine translation, converting text from one language to another (Achiam et al., 2023). However, the current issues that ChatGPT faces, especially as a model that is open for public use, could significantly impact the spread and perception of information. As ChatGPT becomes integrated into various applications, the impact of the unresolved privacy and security challenges ingrained within ChatGPT will spread across digital spaces. The repercussions that grow alongside ChatGPT's capabilities may potentially be magnified across the digital spaces, especially given that the model is intended to see widespread use as a general intelligence tool.

Scale

ChatGPT's role can be analyzed on two major political scales: the communal and national scales (Cox, 2017). At both scale, ChatGPT's impact on information dissemination can depend significantly on the unique social, cultural, and political dynamics within each group. On the communal scale, ChatGPT can enhance customer service roles and facilitate cross-cultural dialogues within local communities, leading to more effective communication and information exchange within these groups (Baldassarre et al., 2023). However, because ChatGPT's responses are preferentially aligned with certain demographic groups, this bias lowers the quality and reliability of information that is disseminated (Santurkar et al., 2023). On a national scale, the way ChatGPT affects information is heavily influenced by factors such as government regulations, language preferences, and cultural norms. One major example of this is the temporary ban of ChatGPT in Italy due to OpenAI's unlawful collection of user's personal data

and lack of safeguards preventing underage users from accessing inappropriate material (Satariano, 2023). Further investigations into the legality and ethics of ChatGPT were conducted by the Office of the Privacy Commissioner of Canada, and other countries such as Germany, France, Ireland, and Spain (Lindrea, 2023). Alternatively, OpenAI themselves have decided to make their service unavailable in China, North Korea, Russia, and Iran (Satariano, 2023). This disparity in the interactions between different national policies and ChatGPT changes the way information is handled on a national scale.

Actors and Synchronization

The spread of information in our society from ChatGPT is heavily guided by how a few key groups and systems interact with ChatGPT. The key groups identified consists of general social media users, content creators, policymakers, and AI developers. Analyzing the interaction between ChatGPT and various social media platforms is crucial to understanding ChatGPT's influence on information dissemination. A combination of three major factors contributes to this dynamic.

First, almost 62% of adults use social media as their primary source of news, a proportion that continues to grow over time (Kim & Dennis, 2019). Second, the average person is unable to distinguish between human and AI-generated content, however, most Americans don't trust information regarding current events that comes from ChatGPT (Partadiredja et al., 2020; McClain, 2024). 38% of Americans state they either have "not too much" trust (18%) or "no trust at all" (20%) in the information that comes from ChatGPT about the 2024 U.S. presidential election. In contrast, only 2% of Americans have "a great deal" or "quite a bit" of trust, 10% have "some" trust, and the remaining percentage being unsure or have not heard of ChatGPT (McClain, 2024). Lastly, many resources exist to assist with the use of ChatGPT to

create social media posts. These resources come in many forms, such as blogs detailing specific prompts for users to generate specific types of content, or custom versions of ChatGPT made specifically for social media content creation (Macready, 2023; Orbix AI). Combining these factors clarifies the significant impact that ChatGPT has on information dissemination. Social media content creators are given resources to spread content generated by ChatGPT across the primary news source for 62% of adults - content that is largely considered untrustworthy with current events and is often indistinguishable from human-generated content by the average person.

To combat this, it falls to policymakers to establish appropriate regulations and to AI developers to create more reliable and trustworthy solutions. Progress has already been made towards achieving these goals. Many of the largest social media platforms, such as Facebook, Instagram, Tiktok, and Youtube, have begun requiring labels to flag AI-generated content (Battle, 2023). The development of AI made specifically to detect misinformation has also been implemented into these platforms. Community-based fact-checking artificial intelligence systems, such as Birdwatch on the "X" platform, have been found to effectively reduce the virality of misleading posts, underscoring the ability of AI to combat misinformation in real-world situations (Drolsbach & Pröllochs, 2023). In fact, ChatGPT itself has shown its ability to distinguish between mis/disinformation against legitimate news content for knowledge within its data corpus, achieving 100% accuracy on a current standard fake news test designed for human subjects (Caramancion, 2023).

Discussion

ChatGPT's influence on the perception and spread of information is complex and multifaceted but can be examined in full using the Technopolitics framework. Since its release,

ChatGPT has shown its potential to revolutionize various fields within our society. However, the current privacy and security challenges that ChatGPT continues to face with each iteration of the model raises concerns regarding the growing repercussions of a more powerful model. As OpenAI continues to advance ChatGPT and integrate it into various applications, the implications of these unresolved issues could potentially have far-reaching consequences.

As OpenAI continues working towards expanding the capabilities and influence of ChatGPT, other organizations must continue to follow suit in implementing appropriate regulations and limitations to counteract the downsides of this technology. Events such as the temporary ban of ChatGPT in Italy and the varied responses from other countries will assist in ensuring safe and ethical practices are upheld. The development of artificial intelligence tools capable of identifying misinformation within social media also helps to counteract the potential harm that can be caused by ChatGPT. The spread of AI-generated content across digital spaces highlights the need for effective regulation and the development of reliable solutions.

In conclusion, ChatGPT is undeniably a powerful tool with the potential to transform various aspects of our society. Its capabilities in automating tasks and enhancing communication are extremely useful in appropriate contexts. However, as ChatGPT's capabilities increase, the potential downsides of ChatGPT's misuse, such as privacy breaches or the spread of misinformation, need to be addressed with increased urgency as well. As we harness the benefits of ChatGPT, we must also implement measures to mitigate these risks to ensure that ChatGPT's impact on our society is both positive and responsible.

Bibliography

Abdullah, M., Madain, A., & Jararweh, Y. (2022, November). ChatGPT: Fundamentals,

applications and social impacts. In 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1-8). IEEE.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52-62.

Baldassarre, M. T., Caivano, D., Fernandez Nieto, B., Gigante, D., & Ragone, A. (2023, September). The social impact of generative ai: An analysis on chatgpt. In Proceedings of the 2023 ACM Conference on Information Technology for Social Good (pp. 363-373).

Battle, P. (2023, November 15). These social media platforms, except for one, have added AI protections - thestreet. TheStreet. https://www.thestreet.com/technology/these-social-media-platforms-except-for-one-have-added-ai-protections

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., ... & Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. Asian Journal of Distance Education, 18(1).

Brundage, M., Mayer, K., Eloundou, T., Agarwal, S., Adler, S., Krueger, G., Leike, J., & Mishkin, P. (2022e, March 3). Lessons learned on Language Model Safety and misuse. Lessons learned on language model safety and misuse. https://openai.com/research/language-model-safety-and-misuse

Caramancion, K. M. (2023, June). Harnessing the Power of ChatGPT to Decimate Mis/Disinformation: Using ChatGPT for Fake News Detection. In 2023 IEEE World AI IoT Congress (AIIoT) (pp. 0042-0046). IEEE.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting training data from large language models. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 2633-2650).

Chatgpt in medical education: Generative AI and the Future of Artificial Intelligence in Health Care. American Medical Association. (2024, February 23). https://www.ama-assn.org/practice-management/digital/chatgpt-medical-education-generative-ai-and-future-artificial

Cox, K. R. (2017). Spaces of dependence, spaces of engagement and the politics of scale, or: looking for local politics. In Politics (pp. 257-280). Routledge.

Drolsbach, C. P., & Pröllochs, N. (2023). Diffusion of community fact-checked misinformation on twitter. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), 1-22.

Karpathy, A., Abbeel, P., Brockman, G., Chen, P., Cheung, V., Duan, Y., Goodfellow, I., Kingma, D., Ho, J., Houthooft, R., Salimans, T., Schulman, J., Sutskever, I., & Zaremba, W. (2016, June 16). Generative models. https://openai.com/research/generative-models/

Kim, A., & Dennis, A. R. (2019). Says who? The effects of presentation format and source rating on fake news in social media. Mis quarterly, 43(3), 1025-1039.

Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, et al. (2023) Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health 2(2): e0000198. https://doi.org/10.1371/journal.pdig.0000198

Kurban, C., Peña-López, I., & Haberer, M. (2017). What is technopolitics? A conceptual schema for understanding politics in the digital age. IDP. Revista de Internet, Derecho y Política, (24), 3-20.

Lindrea, B. (2023, April 6). OpenAI faces Canadian privacy probe alleging personal info harvesting. Cointelegraph. https://cointelegraph.com/news/openai-faces-canadian-privacy-probe-alleging-personal-info-harvesting

Macready, H. (2023, November 16). How to use CHATGPT for social media + 46 prompts to get started. Social Media Marketing & Management Dashboard. https://blog.hootsuite.com/chatgpt-social-media/

McClain, C. (2024, March 26). Americans' use of Chatgpt is ticking up, but few trust its election information. Pew Research Center. https://www.pewresearch.org/short-reads/2024/03/26/americans-use-of-chatgpt-is-ticking-up-but-few-trust-its-election-information/#chatgpt-and-the-2024-presidential-election

Orbix AI. (n.d.). Revolutionary GPT technology for streamlined AI Solutions. Orbix AI. https://www.orbixai.com/

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35, 27730-27744.

Partadiredja, R. A., Serrano, C. E., & Ljubenkov, D. (2020, November). AI or human: the socio-ethical implications of AI-generated media content. In 2020 13th CMI Conference on Cybersecurity and Privacy (CMI)-Digital Transformation-Potentials and Challenges (51275) (pp. 1-6). IEEE.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber Physical Systems.

Satariano, A. (2023, March 31). Chatgpt is banned in Italy over privacy concerns. The New York Times. https://www.nytimes.com/2023/03/31/technology/chatgpt-italy-ban.html

Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023, July). Whose opinions do language models reflect?. In International Conference on Machine Learning (pp. 29971-30004). PMLR.

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT–Exploring the ethical issues of an emerging technology. International Journal of Information Management, 74, 102700.

Wang, B., Chen, W., Pei, H., Xie, C., Kang, M., Zhang, C., ... & Li, B. (2023). Decodingtrust: A comprehensive assessment of trustworthiness in gpt models. arXiv preprint arXiv:2306.11698.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica, 10(5), 1122-1136.

Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., ... & Irving, G. (2019). Fine-tuning language models from human preferences. arXiv preprint arXiv:1909.08593.