Date: August 12th, 2023
To: Ken DeJarnette, Professor, Dominican University of California
From: Rana Demirer, Jackie Ocaña, Kohsuke Uchimura, Rahmat Ullah
RE: Autoregressive Language Modeling, ChatGPT and Bard

Autoregressive Language Modeling: Its Origins and Connections to Machine Learning and Natural Language Processing

Autoregressive language modeling is best defined in three different ways: breaking down the origins of the word, understanding machine learning, how it works in natural language processing.

The word autoregressive originates from Greek and Latin. The prefix auto- is based on the word Greek autos. According to the Online Etymology Dictionary, *autos* signifies "self, one's own, of oneself" (Online Etymology Dictionary). In addition, -regression, originating from the Latin word *regressio*, means "a going back, a return" (Online Etymology Dictionary). Based upon the etymology and combining the two terms together, it can be concluded the word means returning to oneself indicating a model that is self reflective.

Machine learning is a subset of artificial intelligence. Machine learning is particularly important in data science as it is used to build statistical models and algorithms often used to predict and classify information. (What is machine learning?). To build these machine learning models, it will need to be trained with information to predict and classify. Given the previous explanation of the etymology of autoregressive, the machine learning models must self-reflect on the information provided to give an output.

Another subset of artificial intelligence is natural language processing (NLP). NLP combines both etymology and machine learning together to understand what is being said linguistically. This mimics what is done in everyday life where words are associated with meaning, however the difference is this is done by computers.

Defining ChatGPT and Bard

In 2023, artificial intelligence took a leap and made autoregressive language models available to the public on a larger scale. Companies like Apple, Microsoft, Google, OpenAI, and DuckDuckGo have all launched their own chat bots built upon machine learning, autoregressive language modeling and natural language processing (McMillan, 2023). This memo will focus on OpenAI's ChatGPT and Google's Bard.

ChatGPT is currently known as the "original AI chatbot" as it was made publicly available and went viral on the internet. It requires an email address to use and offers a premium tier for a monthly fee. When interacting with ChatGPT, responses are formed based on data that was made available up until 2021. ChatGPT can give concise answers to factual questions and is more likely to give complex ethical reasoning compared to its competitor Bard. (III, 2023).

Bard released to the public shortly after ChatGPT. Bard can deliver real-time information from the internet and can give the sources from which it pulls the information. The experience on Bard is also more like conversing with another human being. However, Bard is less likely to engage in ethical questions, so it gives simpler answers. For example, upon asking Bard whether it is a sentient being, Bard will say it is unsure and cite the philosophical debates about the meaning of sentient. ChatGPT would simply answer no (III, 2023).

While both ChatGPT and Bard pose differences at the surface level, they show even more differences in ethical, privacy, and security matters.

ChatGPT and Bard Ethical Issues

Ethical issues with ChatGPT and Bard are not mutually exclusive. The deployment of AI chatbots like Bard and ChatGPT raises important ethical considerations and issues, such as privacy, data usage, transparency, and accountability. The conversations users have with AI tools like ChatGPT and Bard might contain personal or sensitive information, therefore ensuring the privacy and security of this data is crucial to prevent unauthorized access or misuse.

If we talk about the ethical concerns for Google's AI tool, the first thing is the huge amount of data that they have from their users since their deployment. Google has been collecting user data for many years to improve its services and develop AI tools (Bard vs ChatGPT). Google's collection of extensive user data can raise concerns about individuals' privacy. Users might not have always been fully aware of how their data is being used or might not have given informed consent for its use in AI applications. (Google Bard major ethical concerns)

Storing vast amounts of user data creates a potential risk of data breaches or unauthorized access, which can lead to personal information being exposed and misuse and this is a significant data security concern. If Google's AI tools are trained on biased or unrepresentative datasets, they can perpetuate and even amplify existing biases, leading to discriminatory outcomes or recommendations (Google Bard major ethical concerns).

The combination of user data and AI can be used to manipulate user behavior, such as influencing purchasing decisions or shaping opinions like politics through personalized content. Additionally, AI-powered recommendations can reinforce users' existing beliefs and preferences, limiting their exposure to diverse viewpoints and information. As known, Google's algorithms and AI models are often proprietary and not fully transparent, therefore this can hinder users' ability to understand how their data is being used to make decisions or generate content.

In addition to these considerations, AI models like ChatGPT and Bard can generate false or misleading information that might be mistaken for factual statements (Bard vs ChatGPT). If users rely heavily on AI models for decision-making or generating content, it could diminish critical thinking skills and creativity. Moreover, if the AI fails or generates inappropriate content, it can have real-world consequences. In addition, it sounds a little dystopian, however, As AI models become more sophisticated in mimicking human language, users might develop emotional connections. This raises questions about ethical implications if the AI's behavior is not transparent.

Solutions to ChatGPT's and Bard's Ethical Issues

There is research for ethical concerns with ChatGPT and Bard like refining training data, improving response guidelines, allowing user customization of AI behavior within ethical limits, and soliciting user feedback to enhance models.

In addition to those technical solutions, educating the public about the potential ethical implications of emerging technologies can be an effective solution. As AI models become more prevalent in our lives, individuals should be provided with accessible information that highlights the risks and benefits associated with these tools. Initiatives that foster digital literacy and ethical AI awareness can empower users to critically assess AI-generated content, discern misinformation, and maintain a healthy level of skepticism.

User customization within ethical boundaries can be another effective solution. Along with public education and user customization combined, it forms a robust strategy to address and manage ethical concerns.

Privacy Issues with ChatGPT and Bard

The privacy issues of autoregressive language models pertain to how the company collects and uses data for the model. In terms of this issue, both Google and OpenAI seem to have the same issue. Nowadays, we can collect any public information on the internet using data scraping. In terms of this issue, both Google and OpenAI seem to have the same issue. Of course, both Google and OpenAI use this technique for training Bard and ChatGPT. But can they use any information on the internet? Even if they mention that in their privacy policy and ask users to consent to their policy, do we have to allow them to collect all our information on the internet such as social networking services, websites and so on? Also, there is much information about people not using their service. But they still have collected as much information as they can. How do they know which data they should not use? California law firm concerned and claimed Google and OpenAI about their collection of data.

Bard

On July 13, a California law firm claimed "secretly stealing" vast amounts of data from the web to train its AI. (Cecily, 1) According to a California law firm, Google silently updated their privacy policy claiming any public information can be used to train its AI products like bard. Law firms are questionable about this standpoint, this may invade privacy without consent. For example, if someone is arrested wrongly as we have seen in the class, Google may use this information without consent by scraping data from the website. Even if they recognize that arrest would be wrong, it may take a time to update this information. The issue of this, Google forces us to choose whether to use the internet and consent that Google's AI use all our personal and copyrighted information or avoid the internet forever (Cecily, 18).

ChatGPT

OpenAI also claimed a similar privacy issue by a California law firm. They alleged that ChatGPT used stolen private information from hundreds of millions of internet users, including children of all ages, without their informed consent or knowledge. (Cecily, 3) ChatGPT is also using data scraping techniques to collect data from social media such as Twitter and Reddit. They can use this information if they@ comply with protocol, but if this information was used outside of the owner's intention, there is a possibility that it is an invasion of privacy. In this case, OpenAI used their scraped data without owner's consent in the context of ChatGPT.

Solution For the Privacy Issues

This issue is exactly associated with the transparency issue of privacy policy. Both OpenAI and Google don't show how they collect data, and when we use the data. One solution I think is that they need to describe from which service they collect data and describe all possible cases that they use this information. However, it should be very difficult to all case because what data auto regressive model uses depends on the user's question. If privacy policy for this model would be too strict, the model should be less accurate and no more useful. Whether we accept to be collected our information by Google and ChatGPT to seek more convenience, or strict heavily to protect our information, which is more important for you?

Security Issues with ChatGPT and Bard

The deployment of autoregressive language models like ChatGPT and Bard introduces significant security concerns that demand careful consideration. These concerns revolve around safeguarding user data, preventing unauthorized access, and maintaining the integrity of the AI systems. Bard's and ChatGPT's ability to generate human-like text raises concerns about privacy and security, as sensitive user data could be inadvertently disclosed or misused. Additionally, ChatGPT could be used to create deep fakes or other forms of misinformation, further exacerbating concerns about trustworthiness and digital content integrity. Addressing these concerns requires robust data protection measures and mechanisms to prevent the misuse of the technology (Ray 145).

Data Vulnerability: One of the primary security challenges lies in protecting the vast amounts of user data that power these models. As these models are trained on diverse and extensive datasets, the potential for data breaches and leaks becomes a critical issue. Adversaries could exploit vulnerabilities in the AI systems to gain unauthorized access to sensitive user information, raising concerns about privacy and data protection.

Adversarial Attacks: Autoregressive language models are susceptible to adversarial attacks, where input data is carefully crafted to deceive the model and generate unintended outputs. These attacks can lead to AI-generated content that contains malicious or harmful information. Ensuring the robustness of the models against such attacks is crucial to maintain their reliability and prevent potential harm.

Model Security: The security of the autoregressive language models themselves is another area of concern. Unauthorized access or manipulation of the models' parameters could lead to the generation of biased, misleading, or inappropriate content. Safeguarding the models' integrity through encryption, access controls, and continuous monitoring is vital to mitigate this risk.

Solution to Security Issues

Addressing the security issues associated with autoregressive language models requires a multi-faceted approach that combines technical solutions, best practices, and user education.

Robust Model Training: Conversational AI models can be vulnerable to adversarial attacks or malicious inputs; enhancing Bard's and ChatGPT's robustness and security can ensure its reliable performance in various environments. Implementing robust model training techniques, such as adversarial training and data augmentation, can enhance the resilience of autoregressive language models against adversarial attacks. By exposing the models to potential vulnerabilities during training, they can learn to better handle deceptive input data.

Regular Security Audits: Conducting regular security audits and vulnerability assessments on the AI systems can help identify and patch potential weaknesses. Periodic evaluations of the models' code, architecture, and data sources can enhance their overall security posture.

User Awareness and Control: Empowering users with greater control over the AI-generated content and their data is essential. Providing clear and transparent options for users to customize the behavior of the models while setting ethical boundaries can enhance user trust and data security.

Collaboration with Security Experts: Collaborating with cybersecurity experts and researchers can provide valuable insights into potential security vulnerabilities and solutions. Engaging external professionals can lead to more robust security measures and proactive threat mitigation.

Ensuring the security of autoregressive language models like ChatGPT and Bard is crucial to their responsible deployment. By addressing data vulnerabilities, guarding against adversarial attacks, and implementing rigorous security practices, we can create a safer AI environment that benefits users while minimizing risks.

Conclusion

In the world of AI chatbots like Bard and ChatGPT, there are significant ethical issues such as privacy, unfairness, and transparency. We need to find a good middle ground between making a strong technology and using it responsibly. This means being open about how things are done, having rules to follow, helping users understand, and all agreeing to use AI in a fair and good way. If we put together a combination of transparent ways of handling data, providing informational programs to educate the public, letting users control how things will work while they use AI tools, and always staying open to feedback, we can work together to solve the ethical problems that AI brings. If we use these ideas, companies can make sure AI grows in a good way that respects what users want, builds trust, and follows the right rules in this time of AI technology.

Works Cited

Auto+regressive: Search online etymology dictionary. Etymology. (n.d.-a). https://www.etymonline.com/search?q=auto%2Bregressive

Cecily, Mauran. “Google slapped with a lawsuit for ‘secretly stealing’ data to train Bard” Article of mashable, July 13, 2023. SpringerLink, https://mashable.com/article/google-lawsuit-ai-bard

Cecily, Mauran. “OpenAI is being sued for training ChatGPT with ‘stolen’ personal data”

Hetler, A. (2023, July 3). *Bard vs. chat GPT : How are they different? (2023)*. WhatIs.com. https://www.techtarget.com/whatis/feature/Bard-vs-ChatGPT-Whats-the-difference?Offer=abt_pubpro_AI-Insider

III, R. E. W. (2023, June 13). Chatgpt vs. Bard: What’s the difference? Lifewire. https://www.lifewire.com/chatgpt-vs-bard-7504876

Mauran, C. (2023, April 19). *Google launched Bard despite major ethical concerns from its employees*. Mashable. https://mashable.com/article/google-bard-ethics-employees

Mauran, C. (2023b, June 30). *OpenAI is being sued for training CHATGPT with “stolen” personal data*. Mashable. https://mashable.com/article/openai-chatgpt-class-action-lawsuit

McMillan, M. (2023, April 20). 7 best CHATGPT alternatives I’ve tested. Tom’s Guide. https://www.tomsguide.com/features/chatgpt-alternatives

Regression: Search online etymology dictionary. Etymology. (n.d.-b). https://www.etymonline.com/search?q=regression

What is machine learning?. IBM. (n.d.). https://www.ibm.com/topics/machine-learning

What is Autoregressive Model | Deepchecks. (2022, December 25). Deepchecks. https://deepchecks.com/glossary/autoregressive-model/#:~:text=An%20autoregressive%20language%20model%20islanguage%20processing%20and%20machine%20translation

M. Gupta, C. Akiri, K. Aryal, E. Parker and L. Praharaj (2023), "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy," in IEEE Access, vol. 11, pp. 80218-80245, 2023, doi: 10.1109/ACCESS.2023.3300381. https://ieeexplore.ieee.org/abstract/document/10198233

P. P. Ray (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S266734522300024X