Have you chatted with ChatGPT yet? Towards a dynamic framework of interaction between humans, AI, and society

Keywords: ChatGPT, artificial intelligence, AI affordances, BERTopic, Social responsibility

Extended Abstract

Background and research questions

On November 30, 2022, Open AI released a prototype chatbot ChatGPT (Chat Generative Pre-trained Transformer), which not only greatly fulfilled expectations and fantasies about artificial intelligence (AI), but also caused fear and apprehension. This "scary good" and "dangerously strong" (Musk, 2022) AI, is based on OpenAI's GPT-family of large language models (LLMs), and has been improved by utilizing both supervised and reinforcement learning algorithms. Despite its core function of mimicking a human conversationalist through prompting, the versatile ChatGPT not only can write poetry and song lyrics, compose music, write programming scripts, and debug computer programs, but it has also been used to generate academic essays (Bishop, 2023). It even nears the passing threshold of professional medical exams (Kung, et al., 2023), without any specialized training or learning. Due to its high degree of automation and uncontrollability, some educational institutions started to ban ChatGPT for raising fear of fraud and cheating on homework (Roose, 2023). The tensions between benefits and potential risks, automation and control, expectation and worry surrounding ChatGPT have sparked heated debate on social media, such as Twitter, providing researchers with an opportunity to better grasp, and then sketch, the public perception of interacting with ChatGPT and of its impact on society.

The massive and fast development of deep learning models in language-based AI of the last decade has brought not only major advancements in applications such as search engines, healthcare, recommender systems, and social media platforms, but also an increasing attention to their potential impact. It's been widely observed, for example, that the language models at the core of AI systems carry and reproduce cultural and societal biases, thus calling the need for fair, accountable, and transparent AI systems (Luchs, Apprich & Broersma, 2023). The advent of systems such as ChatGPT, which enable a fully explicit and direct human-AI interaction, opens up a whole new exploration opportunity into the dynamics of such interaction within society. The present research proposes to start this exploration by directly exploiting the public debate on ChatGPT and addressing two core research questions: (RQ1) What topics revolve around the interaction of people with ChatGPT as reported on social media? (RQ2) How is the dynamic interaction between humans, AI, and society portrayed and perceived by the public? We aim at answering such questions through an analysis of public discourse revolving around ChatGPT on Twitter, thereby proposing an extended framework to capture the dynamic interaction between humans, AI, and society.

Mixed methodological framework: computational grounded theory

To investigate the above research questions, we build on a mixed methodological approach, called computational grounded theory (Nelson, 2020), applying some modifications. Our novel approach combines expert knowledge of human beings with language processing techniques, yielding a more interpretable, reproducible, and stable approach to content analysis. Our five-step methodological framework includes corpus

construction, topic generation, topic refinement, topic confirmation, and theory generation in a flow illustrated in Figure 1.

Specifically, for this study, in the first step of corpus construction, the official Twitter API, Tweepy, was used to query and retrieve English tweets containing the keyword "chatgpt" over a two-month period (November 30th, 2022 to January 31st, 2023), and 646,054 tweets were retrieved. Then, the NLTK library in Python was utilized to preprocess the data:duplicates, short texts containing fewer than 5 words, hashtags, stopwords, and inflections were removed, resulting 584160 cleaned tweets. The second step of topic generation involves manual and computational exploration of 5000 sample tweets. Two annotators manually labeled each tweet to inductively generate topics. For the computational exploration, ChatGPT itself was employed to extract topics from the sample tweets in an unsupervised learning fashion. This involves around 100 prompts (prompt refers to the starting point of a conversation or question that a user provides to initiate a response from the ChatGPT), and each prompt involves the labeling of 50 tweets (restricted by the text length limitation of ChatGPT). If new topics are still generated after 100 iterations, the process continues by going back to the first step and expanding the sample. The process stops when no new topics are detected, which suggests theoretical saturation. The third step (topic refinement) triangulates and cross-checks the human and computational generated topics, forming a preliminary pattern. Step four, topic confirmation, confirms the preliminary pattern by utilizing BERTopic (a topic modeling technique that leverages BERT embeddings and TF-IDF to extract coherent topic clusters) to detect the topics within the whole dataset. The pattern is confirmed if all the topics identified by BERTopic can be classified and explained (Figure 2). The final step of theory generation refers to bringing the analysis together to build a more inductive, abstract, and theoretically solid understanding of human interaction with ChatGPT as reported and discussed on social media.

A dynamic interaction between humans, AI, and society

This article proposes an extended framework for understanding human-AI interaction, encompassing a broader dynamic interaction between humans, AI, and society. The intended framework focuses four pillars that emerged from the analysis. 1) Human likeness, this dimension concerns the degree to which AI technology reflects characteristics of human interactions. 2) AI affordances revolve around the application and benefits of AI in specific fields of society. 3) Social impacts concern the impacts of AI on society and its corresponding legal and ethical issues, such as discrimination, bias, and disinformation. 4) Social responsibilities ensure that AI is transparent, interpretable, and comprehensible.

References

Bishop, L. (2023). A Computer Wrote this Paper: What ChatGPT Means for Education, Research, and Writing. Research, and Writing (January 26, 2023).

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health, 2(2), e0000198.

Luchs, I., Apprich, C., & Broersma, M. (2023). Learning machine learning: On the political economy of big tech's online AI courses. Big Data & Society, 10(1)20539517231153806.

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. Sociological Methods & Research, 49(1), 3–42.

Roose, K. (Jan, 2023). Don't ban ChatGPT in schools. Teach with it. The New York Times. Retrieved from https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html on February 20, 2023.

Swart, J. (2018). Haven't you heard? Connecting through news and journalism in everyday life. Rijksuniversiteit Groningen.
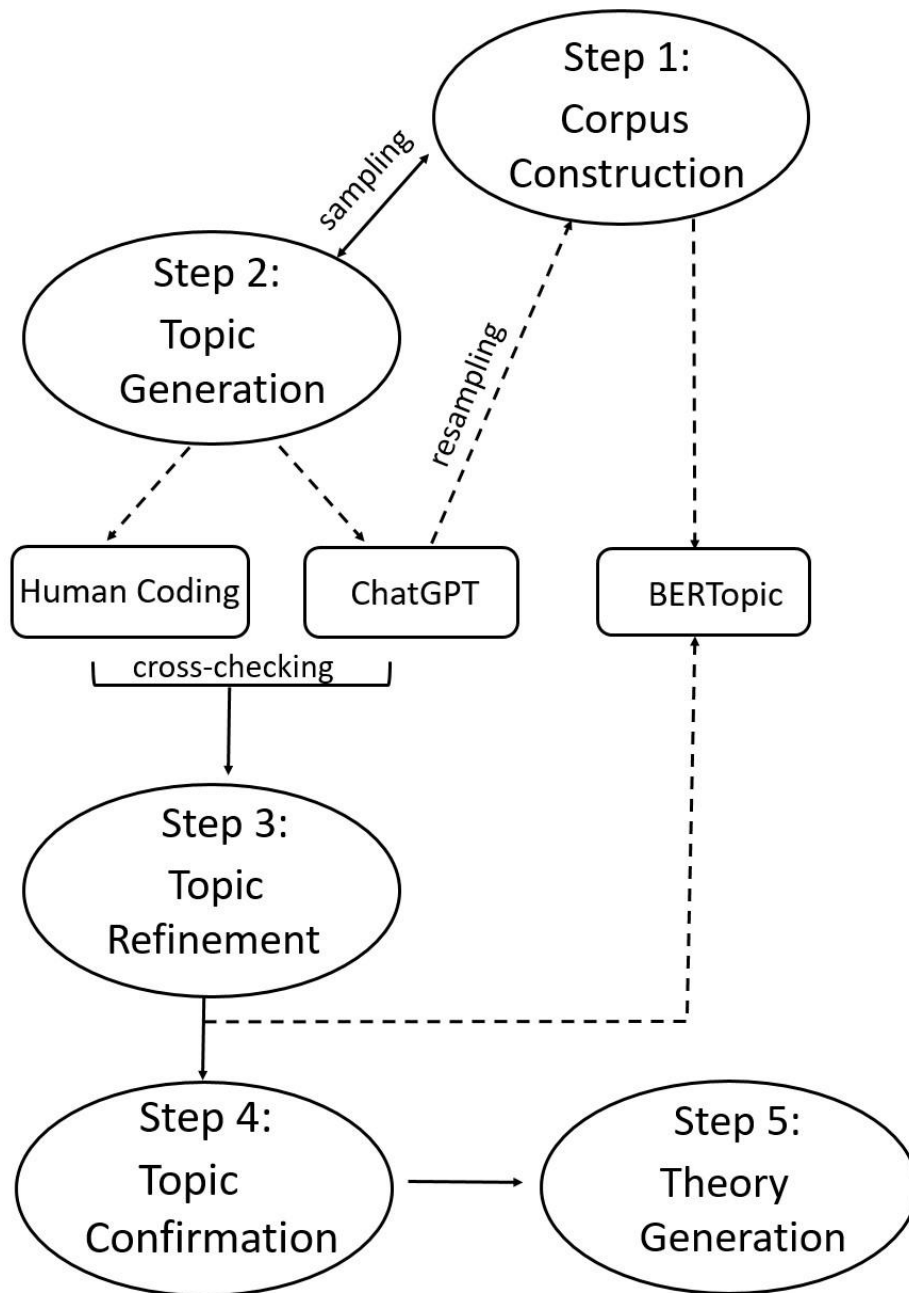
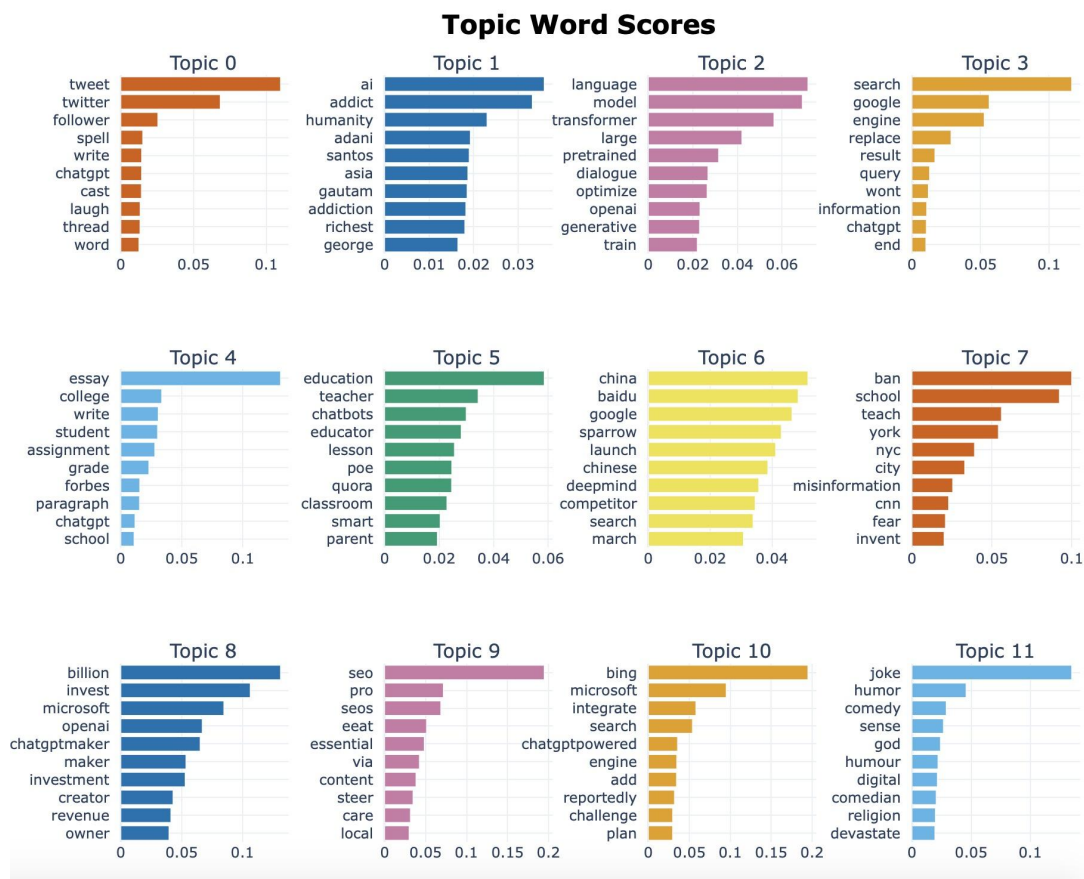Figure 1: Methodological flowchart: A novel approach to computational grounded theory

Figure 2: Preliminary results from BERTopic