

KABIR CHATURVEDI

Semester: Spring 2024

Decoding Public Sentiment: An Analysis of ChatGPT Discussions on Reddit (2022-2023)

(495 words)

Research Question:

How have public sentiments towards ChatGPT evolved on Reddit between 2022 and 2023, and what themes predominate in these discussions?

Background:

In the ever-expanding field of artificial intelligence (AI), public sentiment is critical, with platforms like Reddit becoming pivotal forums for debate and discourse on technological advancements. The foundational work by Taherdoost and Madanchian (2023) discusses the interplay between AI and sentiment analysis within competitive research, while Qi et al. (2023) analyzes the public perception of AI in the post-ChatGPT era through social media. These studies highlight the importance of dissecting Reddit conversations about ChatGPT from 2022 to 2023 to enrich our understanding of public engagement with AI technologies and the broader implications for the human-AI relationship.

Data:

This study will extract a dataset of 500 Reddit posts and comments from r/chatgpt, r/technology, and r/openai, spanning a spectrum of sentiments about ChatGPT. Data points will be selected using Reddit's "hot" algorithm, which surfaces popular and currently trending posts, to gather the most pertinent and discussed content. For the subreddits r/technology and r/openai, keywords such as "chatgpt" and "gpt" will be specifically used to refine the search and ensure relevance to the research question. The timeframe for data collection will start from the date of ChatGPT's introduction on November 30, 2022, until December 2023. This strategic selection method aims to clarify the general attitude toward AI, gauging the public's optimism for the potential of AI and apprehensions about its impact. By analyzing these discussions, the research will aim to ascertain if the sentiment toward ChatGPT is largely positive, celebrating its innovation, or negative, casting it as a source of concern.

Method:

The methodology unfolds in stages, beginning with the extraction of Reddit data via the PRAW library, followed by stringent preprocessing to ensure uniformity and readability. The main tool for sentiment analysis will be VADER, which will use a systematic approach to evaluate the text sentiment from Reddit

posts. It will assign a compound score to each text segment, covering the overall sentiment, along with separate scores to indicate the proportion of text that expresses neutral, positive, and negative sentiment. Meanwhile, EMPATH will identify more general thematic and emotional patterns. This combination aims to provide a comprehensive sentiment landscape of ChatGPT-related discussions, demonstrating the diverse perspectives held by Reddit users. An essential component of this research will be to differentiate sincere discourse from the sarcasm and humor frequently found in online interactions, nuances that could significantly influence the sentiment analysis of ChatGPT discussions on Reddit. Questions such as "Is there a correlation between the sentiment of a post about ChatGPT and the engagement it receives in terms of comments and upvotes?", "What specific aspects of ChatGPT are most frequently associated with positive or negative sentiments?", "How do sentiments towards ChatGPT vary across different subreddits dedicated to technology and AI?" will be considered while performing the analysis.

References:

- Taherdoost, H., & Madanchian, M. (2023, February 7). *Artificial Intelligence and sentiment analysis: A review in Competitive Research*. MDPI. <https://www.mdpi.com/2073-431X/12/2/37>
- Qi, W., Pan, J., Lyu, H., & Luo, J. (2023, July 11). *Excitements and concerns in the post-chatgpt era: Deciphering public perception of AI through social media analysis*. arXiv.org. <https://arxiv.org/abs/2307.05809>