ChatGPT: savior or terminator?: An online data study into ChatGPT's influence on global disparities in sentiment of AI news coverage

Ilse ten Boske
6787177
i.m.tenboske@students.uu.nl

Thomas van den Brink
6215173
t.k.vandenbrink@students.uu.nl

Disen Hu
2117096
d.hu1@students.uu.nl

Bram Swaanen
6683614
b.j.g.swaanen@students.uu.nl

Abstract

The general opinion of the population is both positively and negatively shaped and represented by news coverage. Opinions about topics surrounding artificial intelligence systems such as ChatGPT follow the same course. As ChatGPT is used worldwide, it is important to have a global picture of sentimentality regarding ChatGPT. To understand this we will do a sentiment analysis to find out whether different regions in the world show different responses in sentiment in news coverage of AI when ChatGPT was released. We will employ GDELT as our primary data source, which is a real-time database that has articles from all over the world. We will utilize the GDELT DOC 2.0 API to collect this data based on the keywords surrounding Artificial intelligence. We will subsequently compare the sentiment of the articles in different regions from before and after ChatGPT was released to answer our research question. Our research is low-risk both from a safety and ethical point of view, but there exists the chance that voices of people with a lack of access to newspapers will be underrepresented.

1. Introduction

When OpenAI released their LLM (large language model) called chatGPT in November of 2022, it had a large impact on society. It allowed people to write well formulated texts in a matter of seconds, where they would have previously taken hours to write something similar. It had both negative effects (making plagiarism much easier) and positive effects such as allowing for the automation of certain repetitive writing tasks. When new technological developments take place, it is important to not disregard the general public opinion on the new technology. An opinion that is too negative could lead to people becoming hesitant to use the technology.

The general opinion of the population regarding a certain subject is often both shaped by and represented in content produced by the media such as news agencies [1, 8, 10, 11]. Therefore, news coverage can serve as an adequate proxy for public opinion [10]. When researching and developing artificial intelligence (AI), it is important to keep in mind this public opinion on AI present in the media. This public opinion could effectively be gauged looking at the tone of news articles that contain certain keywords such as "Artificial intelligence". The tone of this subset of articles could be determined through a machine learning technique called sentiment analysis.

Previous research into the sentiment of news articles talking about artificial intelligence have successfully used sentiment analysis to determine that there was no negative bias against artificial intelligence [3]. This article took place however before the release for chatGPT, and 4 years is quite a long time in a rapidly developing field such as artificial intelligence, so these results might not be reproducible in the current environment. Previous research into the perception of ChatGPT in the media such as [7] has already shown that the zeitgeist regarding ChatGPT can succesfully been captured using sentiment, by exposing the most common positive/negative keywords used in tweets about artificial intelligence. Jasper Roe and Mike Perkins [9] have already explored the AI discourse in the media of the UK after ChatGPT was released. However, for academic research and multi-national business strategy it would be informative to have a global picture of the reception of ChatGPT and its influence on the public's perception of the field of AI, serving to approach AI-related subjects in a more context-sensitive way. This brings us to our research question: Did different regions show different responses in sentiment in their news coverage of AI when CHATGPT was released when analyzed with sentiment analysis?

2. Data Plan and Research Plan

Understanding how different regions perceive and react to these technological advancements is crucial for both policy makers and AI developers. This study aims to analyze the sentiment in news coverage of AI, with a specific focus on the timeline surrounding the release of ChatGPT, across different regions over the past three years (May 2024 at the time of writing). To achieve this, we employ GDELT (Global Database of Events, Language, and Tone) as our primary data source and analytical tool. GDELT provides a comprehensive, real-time database of global news events, allowing us to conduct a detailed sentiment analysis across various regions and time periods.

2.1. Data Source

GDELT [1], an expansive and dynamic global database, plays a crucial role in capturing and analyzing news narratives from across the world. It operates by continuously tracking a wide spectrum of news outlets—ranging from traditional media such as newspapers and TV broadcasts to digital platforms like online news portals and social media. By processing articles in more than 100 languages, GDELT achieves a truly global perspective, offering up-to-date insights into international media trends [6].

This system's ability to update in real-time is pivotal for tracking and interpreting ongoing global events swiftly. The rich metadata associated with each news item, including details on locations, involved parties, and thematic elements of the reports, helps in providing a nuanced understanding of the news. Furthermore, GDELT utilizes sophisticated natural language processing (NLP) techniques to assign sentiment scores to news content. These scores, which range from -1 (indicating a negative sentiment) to +1 (indicating a positive sentiment), are instrumental in quantifying and analyzing the emotional undercurrents of media coverage.

Through GDELT's comprehensive data collection and advanced analytics, researchers and analysts can conduct in-depth studies on the tone and context of news reporting worldwide, supporting a broad range of applications from academic research to market analysis.

2.2. Data Collection Strategy

The data collection for this study focused on retrieving news coverage related to "artificial intelligence" over the past three years, particularly emphasizing the period surrounding the release of ChatGPT. We utilized the GDELT DOC 2.0 API [4], which allows for comprehensive querying and extraction of news articles along with their sentiment scores. The specific parameters included the keyword "artificial intelligence," a time span of the last 36 months, and English as the language of the articles. To capture a global perspective, we extracted data from multiple countries, including the United States, United Kingdom, Canada, Australia, India, China, Japan, Netherland, German, France, etc. Despite setting English as the search language, GDELT's advanced translation capabilities enabled the inclusion of relevant articles in other languages, translating them into English using their proprietary model [5].However, it is important to note that the data, being translated from various languages, may not fully capture the original sentiment, introducing potential biases into the sentiment analysis. The API was set to return data directly in a visual format using the "Timeline Tone" mode, which provided an immediate graphical representation of the sentiment trajectory over time. This structured approach ensured that we gathered relevant data to analyze the sentiment shifts associated with ChatGPT's release in various regions.

[1] If at any point in our analysis GDELT's functionality proves to be insufficient to answer our research question, we will supplement the analysis with data from Google API.

2.3. Data Processing

The data from the API, specifically provided by GDELT, included visual representations of monthly sentiment scores ranging from -1 to 1, where positive values indicated positive sentiment, negative values indicated negative sentiment, and values around zero suggested neutral sentiment. These visuals facilitated direct analysis, eliminating the need for extensive preprocessing. By presenting the data graphically, we were able to efficiently compare and analyze trends over the specified period, significantly streamlining the examination of sentiment fluctuations following ChatGPT's release across various regions.

2.4. Comprehensive sentiment analysis

The investigation commenced by employing the graphical outputs from the GDELT API, which facilitated a meticulous examination of sentiment trends particularly surrounding the launch of ChatGPT. The analysis focused on monthly average sentiment scores, which were instrumental in unveiling discernible patterns and substantial shifts in public sentiment. This approach not only present the general sentiment trends but also identified specific periods of notable sentiment fluctuation, thereby providing a understanding of the shifts in public opinion over time.

Further enhancing the analysis, geographic metadata provided by GDELT was utilized to segment the data based on specific locations, enabling a targeted comparative analysis across various regions, including the United States, United Kingdom, Canada, Australia, India, China and etc. This segmentation allowed for an exploration of regional variations in sentiment towards artificial intelligence, specifically in the context of significant technological introductions such as ChatGPT. This comprehensive methodology highlights the nuanced interplay between cultural, eco-

nomic and technological dynamics in shaping public perceptions of emerging technologies.

3. Ethical Considerations

The Ethic and Privacy Quick Scan of the Utrecht University Research Institute of Information and Computing Sciences classified this research as low-risk with no fuller ethics review or privacy assessment required. Furthermore, we used the DEDA framework [2] to focus our attention on possible risks to academic integrity our research might pose. We found that since our research uses news articles as a proxy for public opinion, voices of people that do not buy news papers, for example for financial reasons or because of illiteracy, will be under-represented. This is a bias that will be further addressed in the discussion section of this paper.

References

[1] Cindy T. Christen and Kelli E. Huberty. Media reach, media influence?: The effects of local, national, and internet news on public opinion inferences. *J&MC Quarterly*, 84(2):315–334, 2007. 1

[2] Aline Shakti Frankze, Iris Muis, and Mirko Tobias Schäfer. Data ethics decision aid (deda): a dialogical framework for ethical inquiry of ai and data projects in the netherlands. *Ethics and Information Technology*, 23(1):551–567, 2021. 3

[3] Colin Garvey and Chandler Maskal. Sentiment analysis of the news media on artificial intelligence does not support claims of negative bias against artificial intelligence. *Journal of Integrative Biology*, 24(5):286–299, 2020. 1

[4] Kalev Leetaru. Gdelt 2.0: Our global world in realtime, 2024. Accessed: 2024-05-17. 2

[5] Kalev Leetaru. Gdelt translingual: Translating the planet, 2024. Accessed: 2024-05-17. 2

[6] Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer, 2013. 2

[7] Reuben Ng and Ying Yu Joanne Chow. Powerful tool or too powerful?: Early public discourse about ChatGPT across 4 million tweets. *PLoS ONE*, 19(3), 2004. 1

[8] Bianca Oehl, Lena Maria Schaffer, and Thomas Bernauer. How to measure public demand for policies when there is no appropriate survey data? *Journal of Public Policy*, 37(2):173–204, 2016. 1

[9] Jasper Roe and Mike Perkins. 'What they're not telling you about ChatGPT': exploring the discourse of AI in UK news media headlines. *Humanities and Social Sciences Communications*, 10(753), 2023. 1

[10] Paul Statham and Howard Tumber. Relating news analysis and public opinion: Applying a communications method as a 'tool' to aid interpretation of survey results. *Journalism*, 14(6):737–753, 2013. 1

[11] Weisha Zhang, Jiazhong Lu, and Yuanyuan Huang. Reserach on the dissemination of public opinion on the internet based on the news channels. *IEEE*, 18(1):485–488, 2021. 1