SOCIETAL ATTITUDES TOWARD AI: A TEMPORAL STUDY

PROJECT REPORT

Amittai Siavava
Dartmouth College
amittai.j.wekesa.24@dartmouth.edu

Aimen Abdulaziz
Dartmouth College
aimen.a.abdulaziz.25@dartmouth.edu

Angelic McPherson
Dartmouth College
angelic.mcpherson.25@dartmouth.edu

November 22, 2023

ABSTRACT

Artificial Intelligence (AI) is a hot topic, especially recently. This uptick in public interest in AI is driven by the improved capabilities of deep learning architectures such transformers. The emergent capabilities of such models, including GPT [7], has immense utility in other fields. So why has this dawn of AI has been met with conflicted sentiments?

The very fact that recent neural network–based models have immense capabilities leaves professionals anxious about their jobs and creative products. The ethics of deploying AI models in critical, user-facing applications where errorneous predictions have real-world consequences is also a concern. In fact, a faction of industry leaders has even called for AI experiments to be paused until proper regulations are in place because of the risk of runaway AI.

There is also a lingering question around the ownership of work generated by AI models. When a model's output is clearly derived from a copyrighted artifact, who owns the output? The original creator of the imitated artifact may not want to be associated with the new imitation, since it may have inferior quality or even be offensive. The same concern applies to the creators of the AI models themselves. At the same time, imitations of copyrighted works *without permission* have their own legal implications.

This research project studies the societal attitudes toward AI, both currently and how they have evolved over the years, as a way to understand how different events have shaped the public's perception of AI. We use topic modeling, sentiment analysis, and procrustes analysis to analyze relationships across time periods and extract insight into the changing story of artificial intelligence.

Keywords AI · Machine Learning · Ethics

1 Introduction

We sought to find a *representative* yet *accurate* sample of the public's opinion on AI. We considered multiple potential sources of data, and their tradeoffs:

(i) *Research papers* are the most cutting-edge and factually correct, but they tend to dive into detailed exposition of novel model architectures and ideas, which is disconnected from the public's opinion.

(ii) *Social media posts* are most accessible to the public, yet they are often too short to offer a nuanced opinionand are not fact-checked so they are prone to inaccuracies.

(iii) *News articles* can be sensationalized and biased, but they are often longer (than social-media posts), fact-checked, and backed by current events and innovations. This keeps them (or at least the good ones) in touch with both the public's sentiments about AI/technology and in touch with new innovations.

We decided to use news articles as our target data source, since they are a good compromise. However, we eventually limited our domain to a few specific news sources that are both reliable and well-known, as explained in section 2. We then considered potential forms of analysis to use as a lens to study the data:

(i) *Topic modeling* can be used to identify the most common topics in the data. We can then focus on these topics and compare them across years or in individual years using methods such as procrusties analysis. It can also be insightful to see the most prominent topics in conversations in given periods, or how specific topics such as *ethics* became more or less emphasized after certain events, such as national elections.

(ii) *Sentiment analysis* can be used to identify sentiments or tones toward AI and how they change. This can be insightful in identifying when the public's attitude toward AI became more positive or negative.

(iii) *Procrustes analysis* can be used to identify shifts in conversation in specific time periods, and help highlight periods of particular interest.

2 Data Collection

As mentioned in the introduction, we sought to collect data from tech articles and blogs for our study, particularly centered around topics that are relevant to *machine learning* and *artificial intelligence*. To do this, we built a web scraper[1] that would perform a breadth-first crawl of the internet from a set of given seed URLs,collecting all the text from pages that contained a substantial count of the keywords we were interested in. We then compiled the collected data into a dataset of 17092 articles, with a total of 28 million words[2]

Year	**No. of articles**	**No. of Words**
2000	31	170567
2001	8	43041
2002	15	42792
2003	1	744
2004	9	54832
2005	5	19673
2006	14	33312
2007	26	39007
2008	17	69895
2009	20	75074
2010	31	52068
2011	52	122433
2012	141	294872
2013	124	253252
2014	588	621163
2015	743	1005004
2016	777	1078895
2017	1177	1467028
2018	1339	2055156
2019	1914	2611771
2020	1856	3298318
2021	2144	3552252
2022	2313	3841825
2023	3747	7383298
TOTAL	17092	28186272

Table 1: Count of articles and words per year

[1]The source-code is available on GitHub; see 2.
[2]The dataset is available on Huggingface; see 1.

In earlier versions of the dataset, we realized that a substantial fraction of the data collected was, in fact, blurbs for research articles. This was understandable, since we had not restricted our scraper to only look at web domains that are news outlets, and many news outlets will have citation links to blurbs for relevant research articles. To fix this, we decided to explicitly limit our scraper to web domains that are news outlets, such as Wired, The Verge, and MIT Technology Review. We also included a few domains for research companies that are prominent drivers of the conversatiosn around AI, such as Google DeepMind, OpenAI, and Cohere.

3 Sentiment Analysis and Topic Modeling

We sought to find out how the sentiments of news articles about AI have changed over time. We first used `nltk` to extract sentiment intensity aggregates for each year. We focused on four main sentiments: *positive*, *negative*, *neutral*, and *compound*.

3.1 Overall Sentiment Shifts Over Time

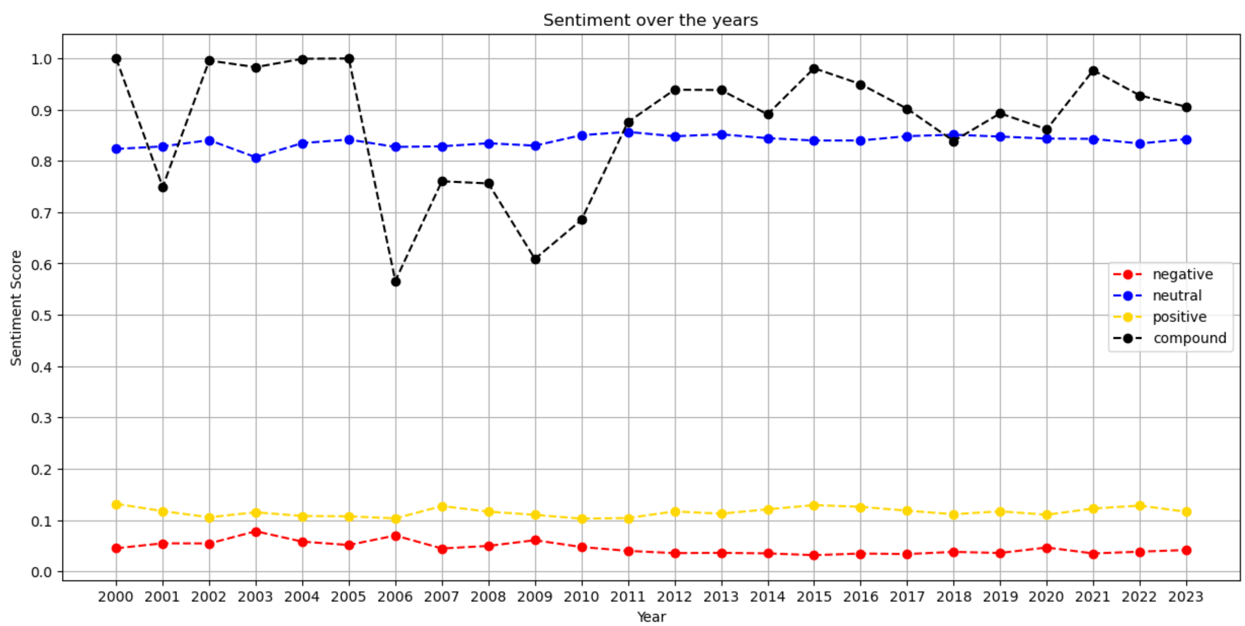


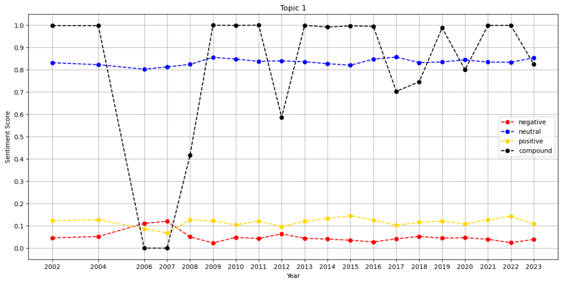
Figure 1: Sentiment analysis of articles about AI from 2000 to 2023.

We notice little gradual changes in the *positive*, *negative*, and *neutral* sentiments over time. However, the *compound* sentiment shows significant jumps and dips over time.

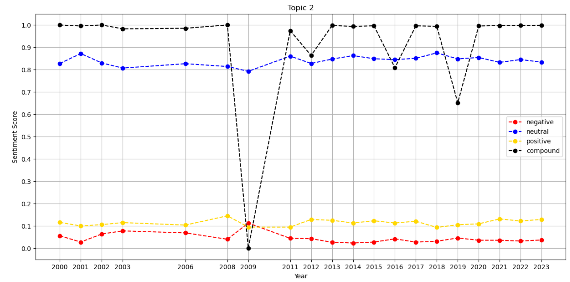
3.2 Sentiment Shifts for the Top Ten Topics Over Time

Since the composite sentiment score is a weighted average of other sentiment, we are probably missing an important nuance in our data that is not captured by the three sentiments.

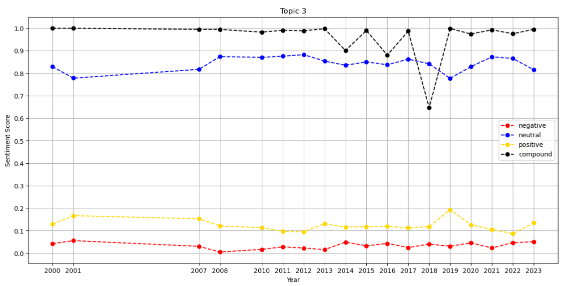
To try to understand this better, we used *Latent Dirichlet Allocation* (LDA) to extract the top-10 topics across the combined dataset then analyze how the sentiments specific to each of those topics shift over time.



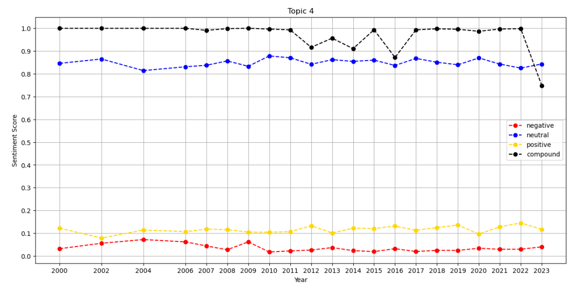
(a) *chatbots, agents, twitter, privacy, ...*



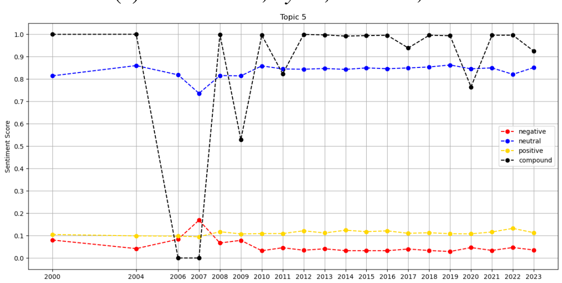
(b) *google, data, intelligence, privacy, China...*



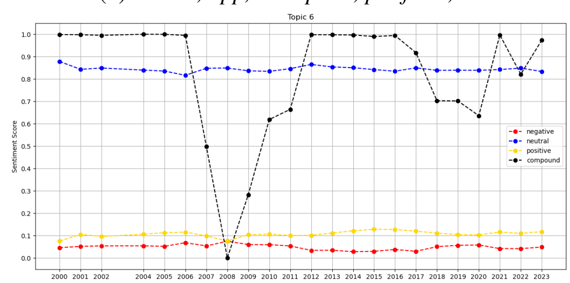
(c) *autonomous, cyber, chatbots, ...*



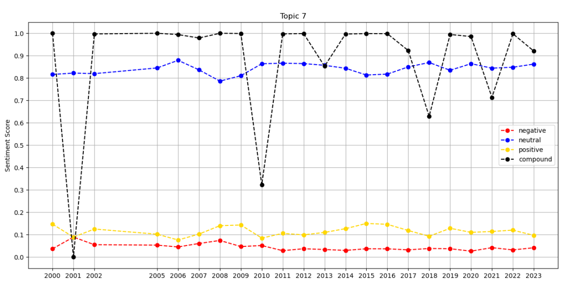
(d) *mobile, app, enterprise, platform, ...*



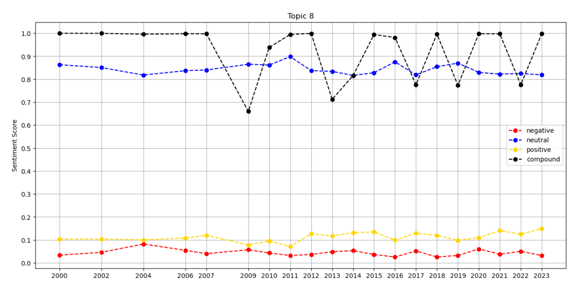
(e) *ai, microsoft, human, future, microsoft, apple, ...*



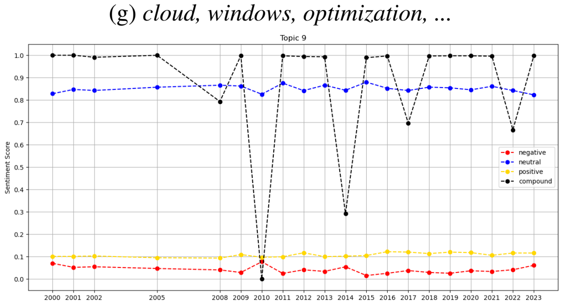
(f) *deep, learning, deepmind, mit, google, ...*



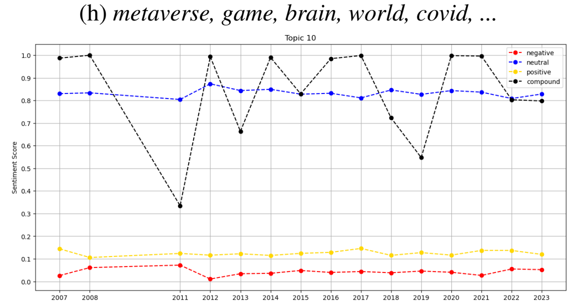
(g) *cloud, windows, optimization, ...*



(h) *metaverse, game, brain, world, covid, ...*



(i) *developers, web, google, codex, raises, ...*



(j) *security, software, real, health, ...*

Figure 2: Sentiment shifts across the top 10 topics from 2000 to 2023.

The composite sentiment score for multiple of the top topics show significant dips to (occassionally) zero in the periods between 2005 and 2011. We believe this might be a result of certain concepts prominent in the top topics not being as prevalent in the news articles during that period. For example, fig. 2f has *deep, learning, deepmind, mit*, and *google* among its top words, but the *deepmind* research lab was established in 2010 [4]. It is perhaps more interesting to look at why the topic has high sentiment scores some earlier years (2000–2004), but we leave that for future work.

4 Procrustes Analysis

Beyond just studying how individual sentiments have changed over time, we also sought to study how the *overall* conversation shifts. To do this, we used *procrustes analysis* to compare the word embeddings of conversatiosn across different years.

Figure 3: Procrustes Comparison By Years.

Although some standout shifts are visible, the above plot is too dense to be useful. Thus, we narrow down the study to consecutive years, as shown in fig. 4, and to years relative to 2022, as shown in fig. 5. We thought these two studies would be useful to identify both gradual shifts and shifts relative to the introduction of large language models (LLMs) in 2020.

4.1 Procrustes Analysis of Consecutive Years

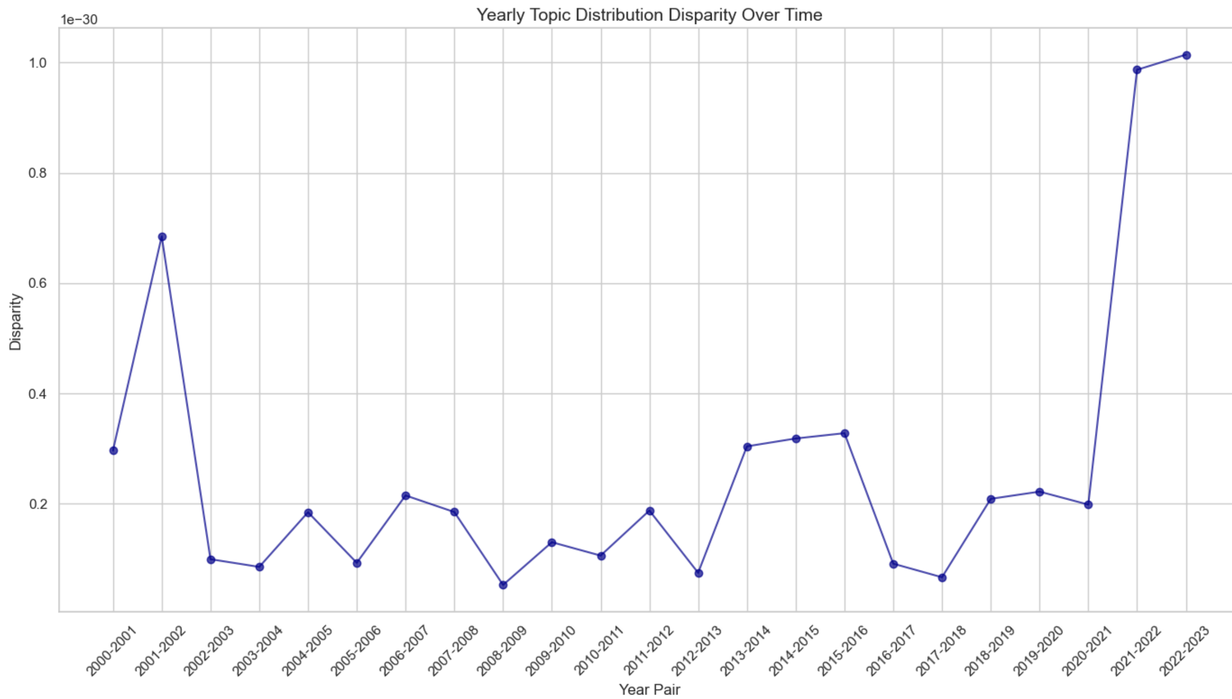


Figure 4: Procrustes analysis of consecutive years.

A lower disparity value suggests a higher similarity in topic distributions (the 10 topics for LDA) from one year to the next, while a higher value denotes a more pronounced change.

We notice a few interesting trends:

(i) In the early 2000s, there is a relatively minor flucations in disparity. However, there appears to be significant shifts in the periods 2000–2001 and 2001–2002. This period corresponds to the period of the dot-com bubble burst, which may have influenced a shift in the discourse around AI, and technology in general.

(ii) From 2002 to around 2013, there is relatively minor fluctuations in disparity.

(iii) There is noticeably higher disparities in the period from 2013 to 2016. These years could correspond to multiple significant events, such as technological advancements, presidential elections, and social-media becoming more mainstream.

(iv) From 2018 onward, we see a *significant* increase in disparity, with the highest disparities in 2021–2022 and 2022–2023. This three-year period corresponds to the significant advancements in language models, such as GPT-3 [3], which have been shown to be capable of generating text that is nearly indistinguishable from human-written text. This could be due to the introduction of LLMs, which have been shown to be capable of generating text that is indistinguishable from human-written text. This could have caused a shift in the discourse around AI, as people became more aware of the capabilities of AI.

4.2 Procrustes Analysis Relative to 2022

To identify gradual shifts, we studied how all years compare to a single year. We picked 2022 for this study since it is the year when significant advancements large language models such as GPT-3 [3] started having impacts in how people do their work and interact with technology, especially with the commercialization of chatbots such as ChatGPT [6].

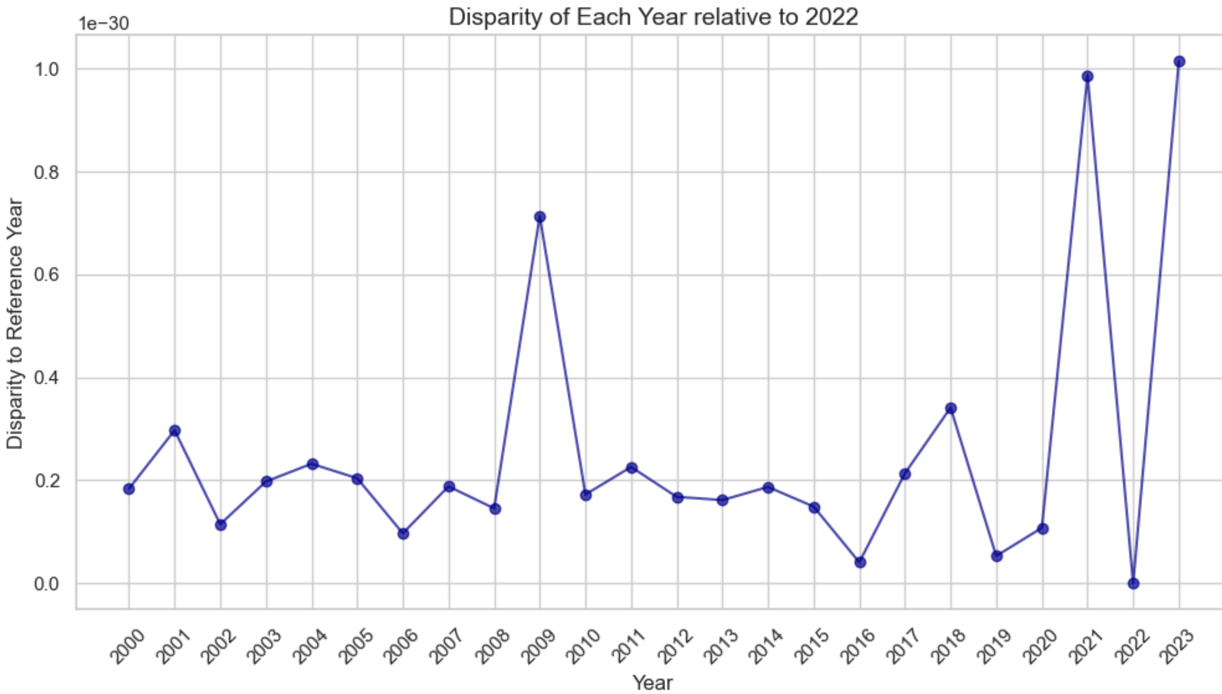
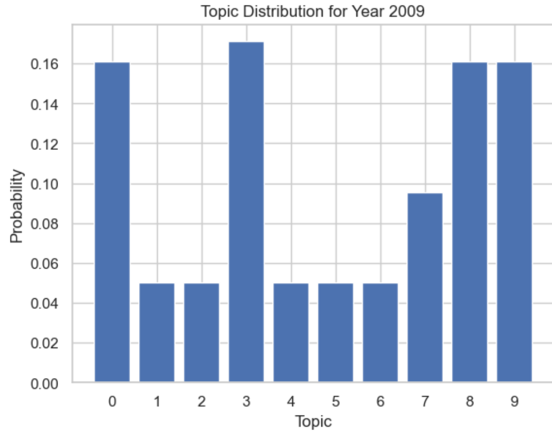


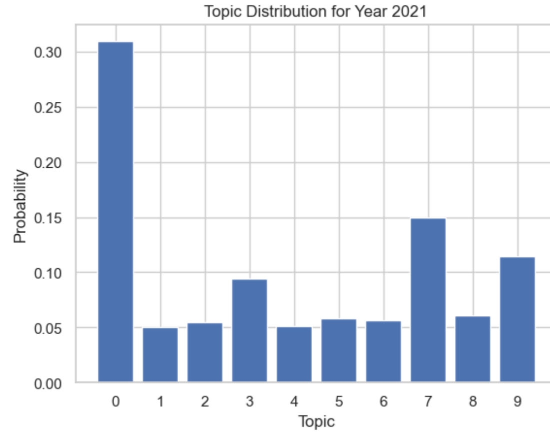
Figure 5: Procrustes analysis of all years relative to 2022.

While the disparities are consistently low (compared to the study in section 4.1), it is noticeable that 2009, 2021, and 2023 each have substantial disparities with 2022.

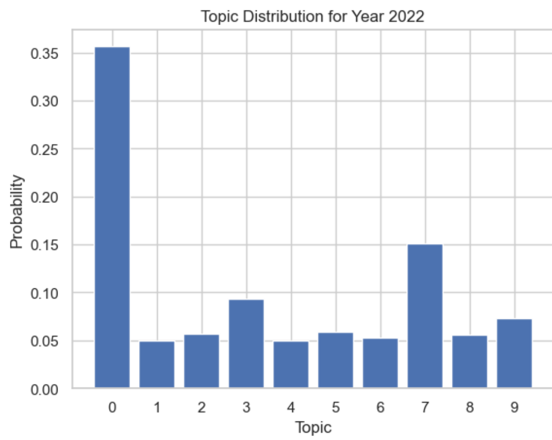
To understand the reasons for these disparities, we looked at the most-relevant topics for each of these years, as shown in fig. 6.



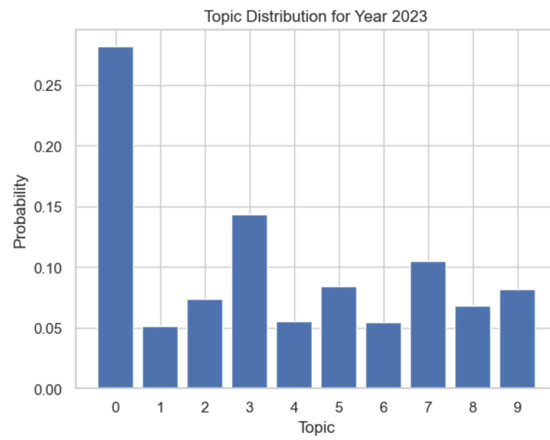
(a) Procrustes analysis of 2009 relative to 2022.



(b) Procrustes analysis of 2021 relative to 2022.



(c) Procrustes analysis of 2022 relative to 2022.



(d) Procrustes analysis of 2023 relative to 2022.

Figure 6: The Most-Relevant Topics for Each of 2009, 2021, 2022, and 2023.

We notice a few trends:

(i) *Topic 0* remains nearly most-relevant across both periods. When we look at the top words for this topic, they include: *model, machine, algorithm, train, research, . . .*. It is not surprising that these would occur consistently across both periods, since they are fundamental to AI.

(ii) *Topic 3* and *Topic 7* also feature consistently among the top topics across the years. Their top words include *openai, human, research, think, explain*, and *sam*.

5 Possible Extensions and Improvements

Our study is not exhaustive. Here are some potential extensions and improvements.

1. Our dataset is not perfect. It is evident in table 1 that our data is skewed towards the later years. For instance, 2023 has 3747 articles, while 2003 has only 1 article. A potential improvement would be augmenting the dataset with more articles from the earlier years. However, this will require some strategy and perhaps access to a corpus of archived articles from the prior years, since a naive web crawler like ours will ultimately end up visiting a lot of articles from the recent years. It is estimated that 90% of the data on the internet was created in the last two years [8].

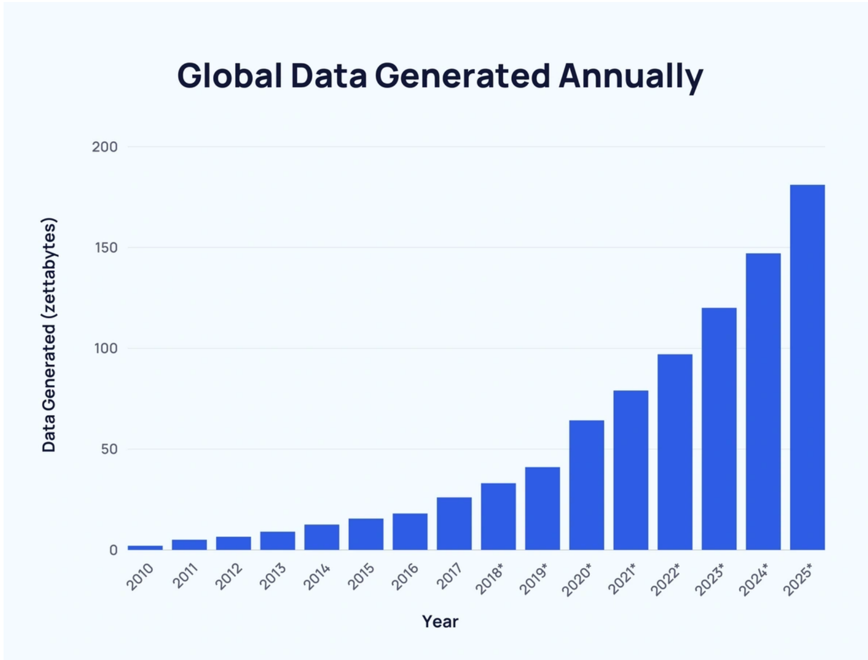


Figure 7: Exponential explosion of data on the internet. [8]

2. In cases where we employed topic modeling using LDA, it is sometimes noticeable that multiple topics have highly-related top words. For examples, see figs. 2a and 2c. Using a more sophisticated topic modeling technique such as BERTopic [5] could help group such topics together, and pick out other relevant topics from the dataset.

3. It is also important to note that we are in the middle of a shift in how humanity views and talks about AI. This means that our study is necessarily an incomplete snapshot of the discourse — in fact, a very significant event (the firing and later rehiring of OpenAI CEO Sam Altman [9]) happened while we were working on this project, and we had to update our dataset to reflect this. A future study could look at how the sentiments and trends reflected in our study pan out over the next few years.

6 Credits

We are particularly grateful to Professor Soroush Vosoughi for giving us a few ideas on how to approach this project, which kinds of analyses to perform, and how to interpret the results.

References

[1] Amittai Siavava, Aimen Abdulaziz, Angelic McPherson. *AI Tech Articles Dataset*. 2023. DOI: 10.57967/hf/1324. URL: https://huggingface.co/datasets/siavava/ai-tech-articles.

[2] Amittai Siavava, Aimen Abdulaziz, Angelic McPherson. *Web Scraper for AI Tech Articles*. 2021. URL: https://github.com/siavava/functional-scraper.

[3] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL].

[4] DeepMind. *About DeepMind*. 2023. URL: https://deepmind.google/about.

[5] Maarten Grootendorst. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. 2022. arXiv: 2203.05794 [cs.CL].

[6] OpenAI. *Introducing ChatGPT*. 2022. URL: https://openai.com/blog/chatgpt.

[7] Alec Radford et al. "Language models are unsupervised multitask learners". In: (2019).

[8] Exploding Topics. *Ninety Percent of the World's Data Created in Last Two Years*. 2023. URL: https://explodingtopics.com/blog/data-generated-per-day.

[9] Wired. *Sam Altman's Second Coming Sparks New Fears of AI Apocalypse*. 2023. URL: https://www.wired.com/story/sam-altman-second-coming-sparks-new-fears-ai-apocalypse.