AI Impact Analysis: Navigating Transformative Trends in Industry and Employment

ADSP 32018 IP01 Natural Language Processing and Cognitive Computing Final Project

Yifan Han

03/07/2024

Agenda

- Executive Summary
- Data Processing
- Topic Detection
- Sentiment Analysis
- Entity Identification
- Targeted Entity Sentiment Identification
- Timeline Analysis
- Conclusion

Executive Summary

Background

- Predictions indicate that by 2032, the AI market in the United States will soar to a staggering **$594 billion**

- **25%** of tasks could be automated by AI in the US and Europe, with particularly high exposures in administrative (46%) and legal (44%)

Exhibit 5: One-Fourth of Current Work Tasks Could Be Automated by AI in the US and Europe
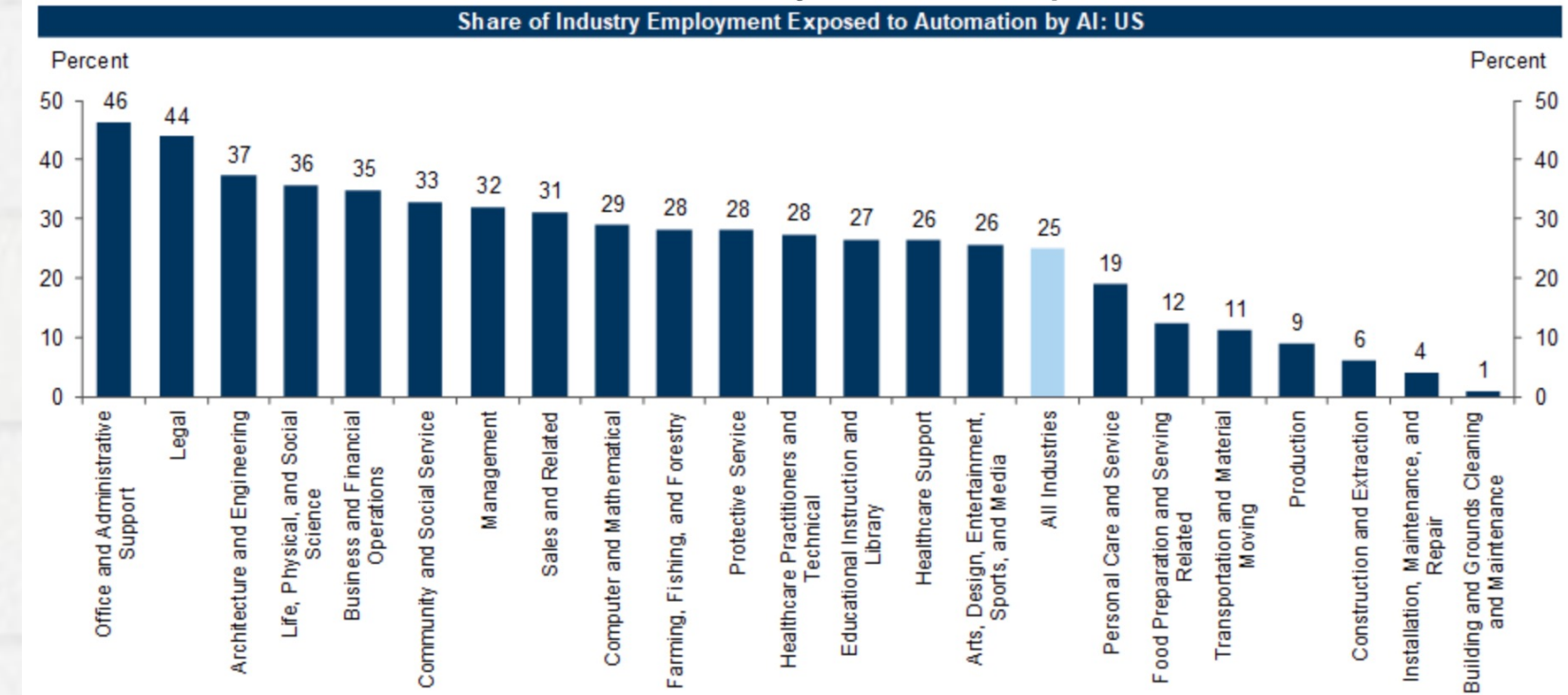
Source: Goldman Sachs

U.S. Artificial Intelligence Market Size 2022 to 2032

Source: www.precedenceresearch.com

Problems

- Identify the **jobs and industries** that are susceptible to AI–driven disruption and those that hold the most promising prospects

- Offer **recommendations** to entities on strategies to expedite the advancement of these transformative capabilities

Next steps

- To answer the questions, **topic analysis**, **sentiment analysis**, **entity analysis** and **timeline analysis** will be conducted

Data Processing

1. Data Loading

Read Parquet file and load the data into Pandas **Dataframe**

2. Data Filtering

Took **10%** of the total data randomly for further analysis or processing

Dropped **non-English** result and articles with **less than 20 words**

Discarded irrelevant articles

3. Data Cleaning

Dropped **duplicate values**

Cleaned **newlines, tabs, special characters** and **links**

4. Tokenization

Tokenized and removed **stopwords** from the cleaned article texts

5. N-grams

Made **bigrams** and **trigrams** from tokens and combine original tokens and N-grams tokens together

6. Lemmatization

Lemmatized the combined tokens by only keeping nouns, adjectives, verbs and adverbs

ADSP 32018 Yifan Han

Topic detection (LDA)

By tuning N, the best N with highest Coherence score is **N=6**

Intertopic Distance Map (via multidimensional scaling)

Topic 1: Market and business intelligence

('0.061*"market" + 0.020*"report" + 0.017*"intelligence" + 0.015*"artificial" + 0.014*"analysis" + 0.013*"global" + 0.013*"growth" + 0.012*"ai" + 0.011*"industry" + 0.010*"forecast"')

Topic 2: AI technology and business solutions

('0.024*"ai" + 0.008*"datum" + 0.008*"technology" + 0.006*"customer" + 0.006*"business" + 0.006*"solution" + 0.006*"company" + 0.006*"use" + 0.006*"provide" + 0.005*"medium"')

Topic 3: Recent news and events

('0.013*"say" + 0.011*"ago" + 0.009*"ai" + 0.008*"hour" + 0.006*"use" + 0.006*"day" + 0.006*"new" + 0.005*"news" + 0.005*"public" + 0.004*"technology"')

Topic 4: AI in the stock market

('0.019*"ai" + 0.009*"company" + 0.009*"stock" + 0.009*"market" + 0.006*"technology" + 0.005*"datum" + 0.005*"business" + 0.005*"use" + 0.005*"intelligence" + 0.004*"say"')

Topic 5: Healthcare and communication

('0.009*"share" + 0.008*"patient" + 0.007*"communication" + 0.006*"health" + 0.006*"use" + 0.005*"digi" + 0.005*"company" + 0.005*"new" + 0.005*"ai" + 0.005*"report"')

Topic 6: General discussion about AI

('0.012*"ai" + 0.009*"use" + 0.007*"new" + 0.007*"say" + 0.005*"make" + 0.005*"get" + 0.004*"work" + 0.004*"chatgpt" + 0.004*"good" + 0.004*"also"')

Sentiment Analysis

## Trained Sentiment Analysis models using open-source data <u>Yelp</u>

Model	Accuracy
Naive Bayes Model	94.8%
Logistic Regression Model	97.3%
Support Vector Machine	97.4%

SVM model has the highest accuracy

Confusion Matrix
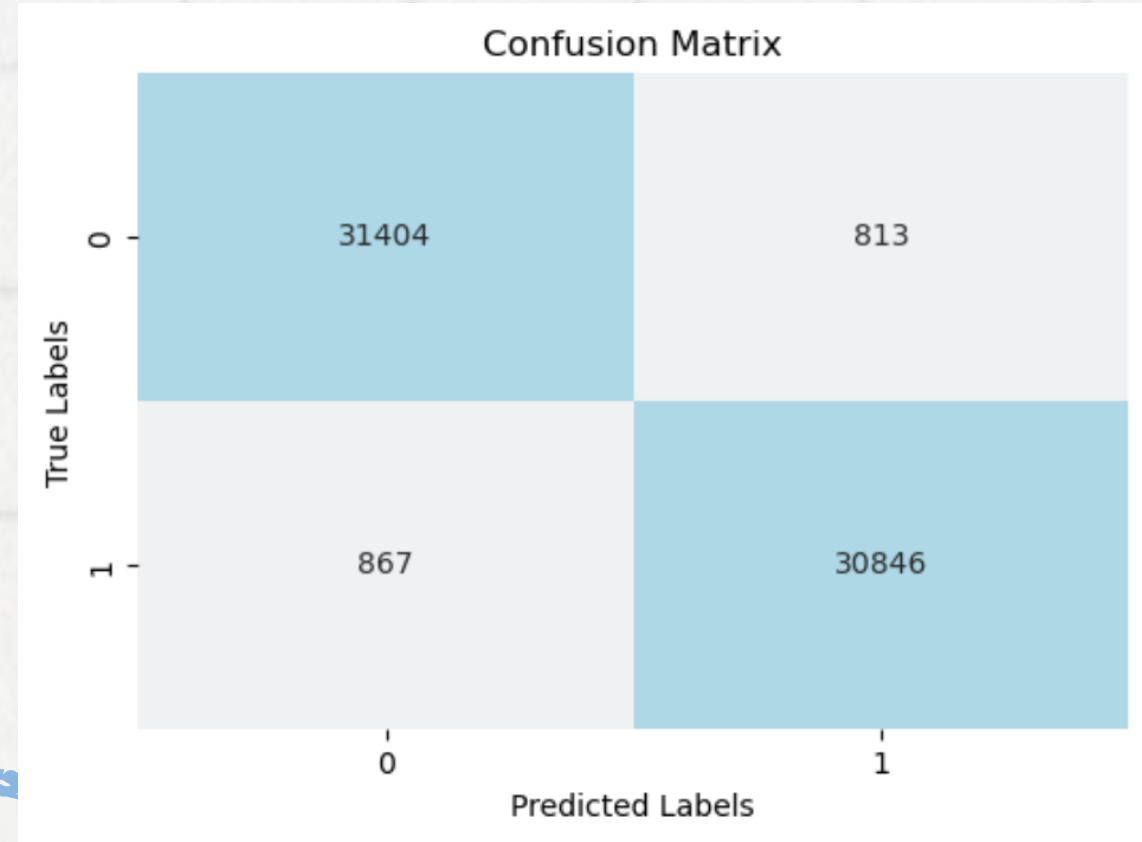
AI Discussions Surge in 2023, but So Does the Divide in Opinions

- Discussions over AI have **surged** since 2023, indicating an increased awareness and interest in the field

- The gap between the number of articles expressing negative and positive sentiments towards AI began to **widen rapidly** from 2023, possibly due to some ethical issues

Percentage of Topic Frequency in Positive Sentiment Articles
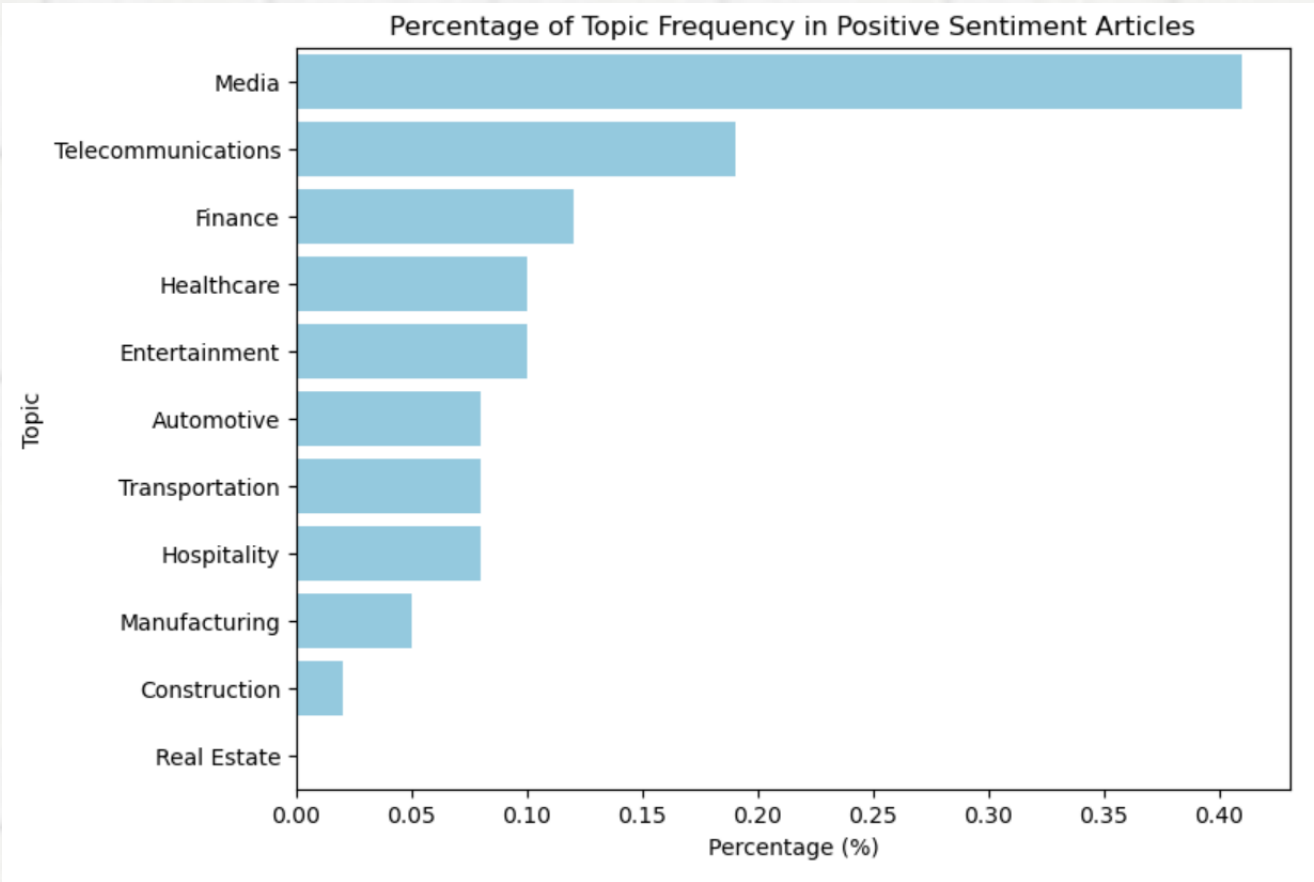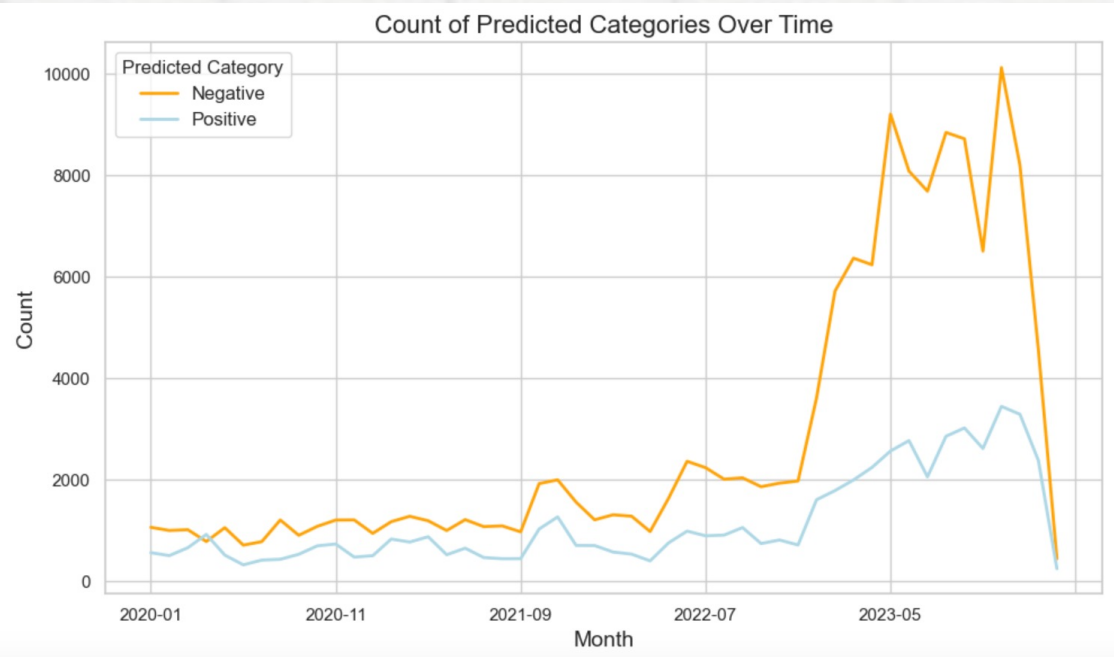
Promising AI field

- Utilizing Natural Language Interface (**NLI**), it has been demonstrated that the most promising fields for AI are **Media**, **Telecommunications**, and **Finance**
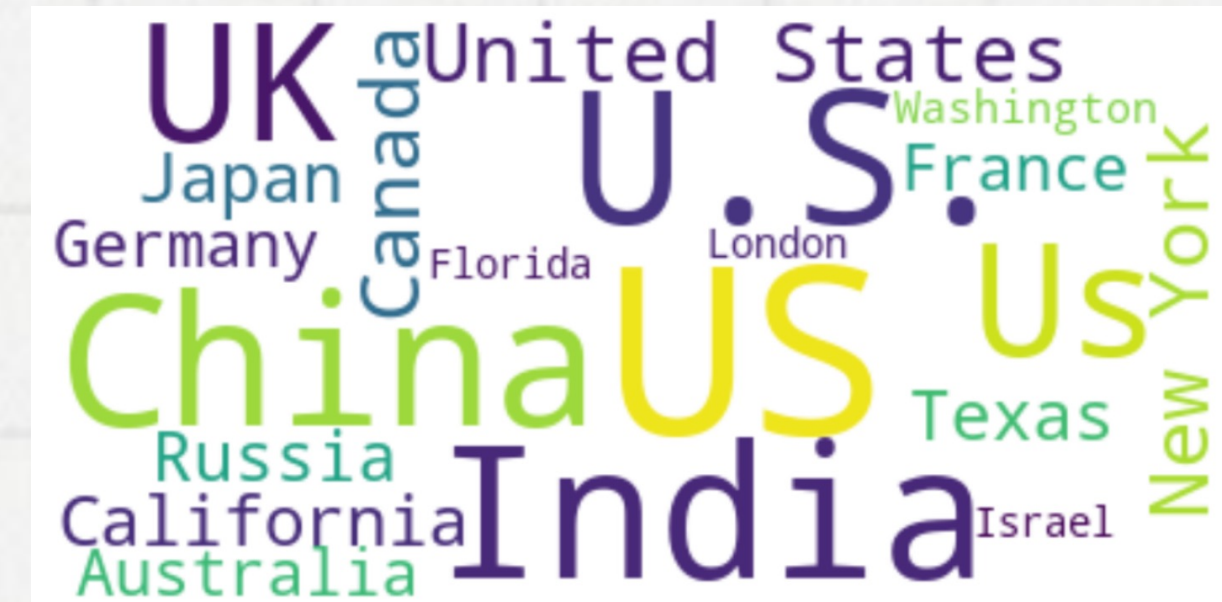
Entity Identification
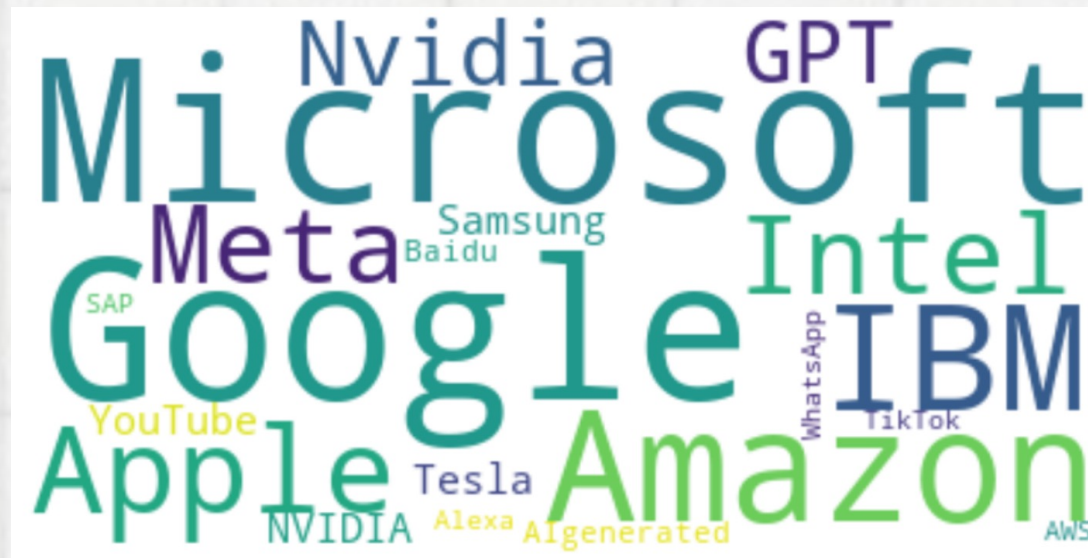
Organizations
- **Tech companies** like Microsoft, Google, Apple Amazon, and Nvidia are leading the development of AI technology

Locations
- **U.S., China**, and **India** are significant contributors to AI development


Person
- **CEOs** of tech companies are mentioned in AI articles frequently because of their significant roles in driving and leading AI development
- **Presidents** are mentioned in AI articles due to their influence on AI policy and regulation, as well as their role in shaping the national strategy and direction for AI development

Products
- **GPT4, DeepMind, and DeepBrain AI** are the most popular AI products

Targeted Entity Sentiment Identification

Sentiment Distribution of Entities (ORG)

Sentiment Distribution of Entities (PRODUCT)
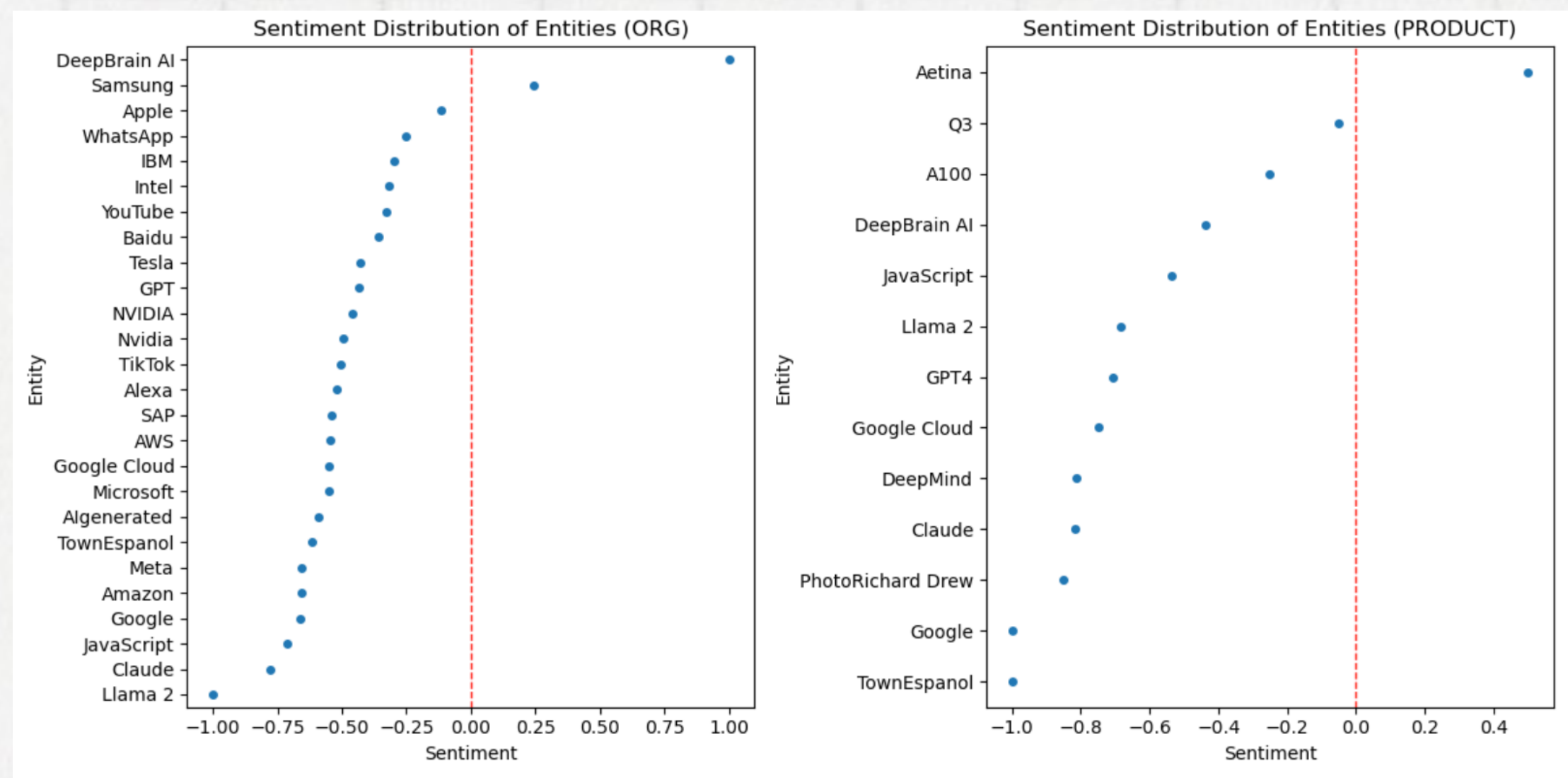
Negative Sentiments Dominance

- There is a prevalence of **negative** sentiments across both organizations and products, as indicated by the negative sentiment scores associated with most entities

Entity Perception

- Entity Perception: Some entities like Claude and Llama 2 have very negative sentiment, suggesting potentially significant issues with perception or reception

Timeline Analysis



Timeline of Popular AI Entities by Maximum Count
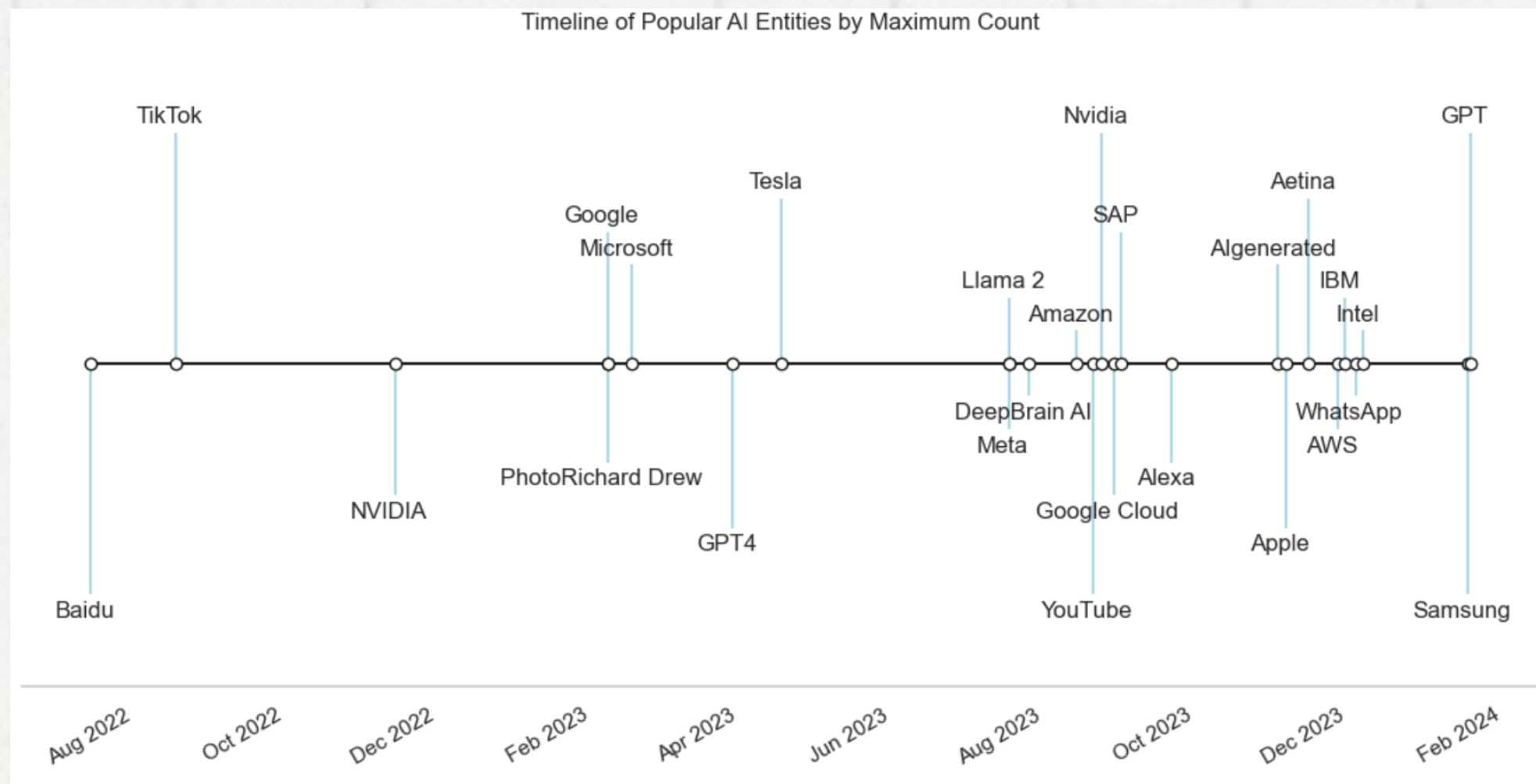
The Evolution of AI: Impact of New Technologies and Corporate Engagement

- From 2023, a huge amount of new AI technologies and solutions are surging, such as **GPT4** and **Llama 2**

- An increasing number of prominent companies, including **Tesla**, **Nvidia**, **Apple**, and **Samsung**, are actively venturing into the field of AI

- In 2024, **Samsung** has emerged as a key player in the AI arena, demonstrating significant potential and poised to continue exerting a substantial impact. Meanwhile, technologies like **GPT** remain pivotal in driving innovation within the industry

Conclusion

Despite the surge in AI discussions and advancements, there remains a widening gap between positive and negative sentiments, reflecting ongoing **ethical concerns** and **societal apprehensions** about AI's impact
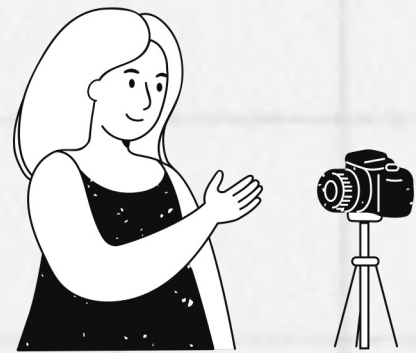
Employment within the **Media**, **Telecommunications**, and **Finance** sectors is anticipated to experience lesser disruption from AI advancements. Besides, the proliferation of AI is generating increased job opportunities in key contributing regions such as the **U.S., China, and India**

As AI technology evolves with the introduction of groundbreaking solutions like **GPT4** and **Llama 2**, it becomes increasingly crucial for stakeholders, including governments, corporations, and society at large, to prioritize ethical frameworks and collaborative efforts to ensure responsible AI development and deployment

Tech giants such as Microsoft, Google, Apple, Amazon, and Nvidia continue to spearhead AI development, while emerging players like **Samsung** are quickly making their mark, highlighting the dynamic landscape of AI innovation