NAVIGATING AI REVOLUTION

ANALYSIS OF INDUSTRY IMPACT AND FUTURE WORKFORCE

Minh Vo | November 23, 2023

EXECUTIVE SUMMARY

In recent years, Artificial Intelligence (AI) has reshaped various industries, influencing not only the current landscape but also future innovation. This project focuses on the multifaceted impacts of AI, identifying tasks and jobs susceptible to transformation through AI advancements. By analyzing a substantial corpus of news articles, the project aims to understand AI's influence in recent years and project its trajectory in the coming decades.

Key Findings & Insights:

1. **Rise of Conversational AI**: ChatGPT and Bard have a significant influence on different industry, including education,. These tools offer personalized learning experiences while simultaneously challenging academic integrity. Their dual impact indicates the need for balanced integration within educational frameworks.

2. **Job Automation**: With AI predicted to automate over a quarter of tasks in the US and Europe, our findings supports Goldman Sachs' projections, particularly for professions in office administration, legal sectors, and architecture. Contrarily, job roles demanding hands-on skills, as suggested by Moravec's paradox, appear less susceptible to AI disruption.

3. **Impacts on Healthcare:** Healthcare is witnessing significant advancements with AI, from personalized medicine to efficient diagnostics. Innovations by entities like Google's DeepMind Health Technology is one of the typical example that demonstrates its ability to enhance patient outcomes and operational efficiencies within the healthcare ecosystem.

4. **News & Media Engagement**: AI's role in enhancing user experience is clearly reflected in the news and media sectors, shaping media consumption and user engagement as well as optimizing personalized content delivery and consumer insights. However, this also raises some discussions around data ethics.

5. **Data Security and Ethical Governance:** The integration of AI also requires for the need of the robust ethical frameworks security measures to address the growing concerns around privacy and ethical AI deployment.

ACTIONABLE RECOMMENDATIONS

Leverage AI's transformative potential to streamline operations, foster innovation, and enhance employee productivity while upholding ethical principles and safeguarding security.

Entity-based recommendation

Companies

- Implement conversational AI tools like ChatGPT and Bing AI to enhance customer service and personalized support while ensuring that data handling complies with ethical standards and privacy laws.
- Evaluate repetitive tasks that can be automated using AI, such as data entry, scheduling, etc., to increase efficiency and allow employees to focus on higher-value work.
- Invest in training programs to build a workforce capable of developing and managing AI solutions.

Academic Institutions

- Enhance the teaching and learning experience by incorporating AI tools that provide personalized learning paths for students and administrative automation for educators.
- Provide students with AI/ML courses as well as AI ethics and literacy lessons to help them improve their skillsets and understand the importance and impact of AI.

Governments

- Set forth regulations that balance AI-driven progress with citizens' rights and privacy protections, establishing benchmarks for responsible AI usage.
- Drive AI applications in public services with transparency and accountability, ensuring that AI's impact on society is positive and in line with public welfare.

Industry-based recommendation

Healthcare

- Refine diagnostic procedures, enabling healthcare providers to offer accurate, timely, and tailored patient care, while protecting patients' personal information and privacy.
- Deploy AI in pharmaceutical research to accelerate drug discovery and to personalize treatment plans.

News & Media

- Identify media trends and consumer behavior to gain insights into audience engagement and preferences, allowing for more strategic content planning and targeted advertising.
- Leverage AI to provide content that aligns with user interests and preferences, while still meeting ethical standards to prevent bias and misinformation.

Finance

- Analyze real-time market insights, risk assessment, and provide personalized financial services.
- Embrace predictive analytics for financial forecasting and maintain the accuracy and responsiveness to market changes.

Technology

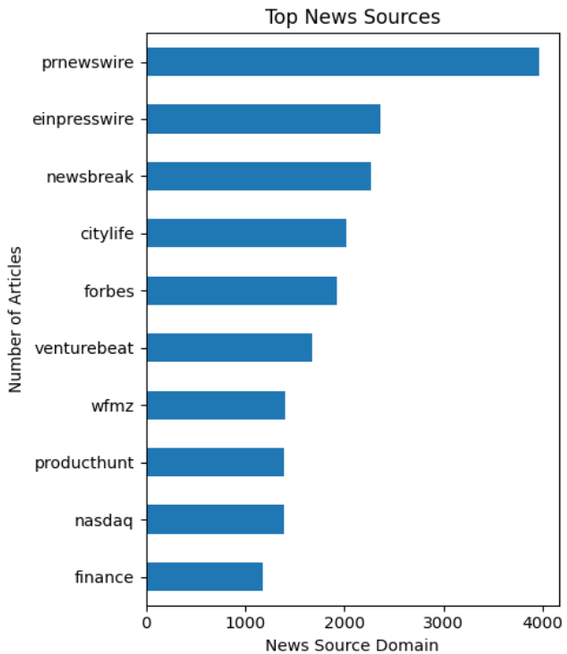
- Invest in advanced research and development, particularly in areas like conversational AI and image generation.
- Prioritize cybersecurity and ethical considerations in all AI applications, ensuring that AI tools are secure, fair, and transparent, and comply with privacy standards.

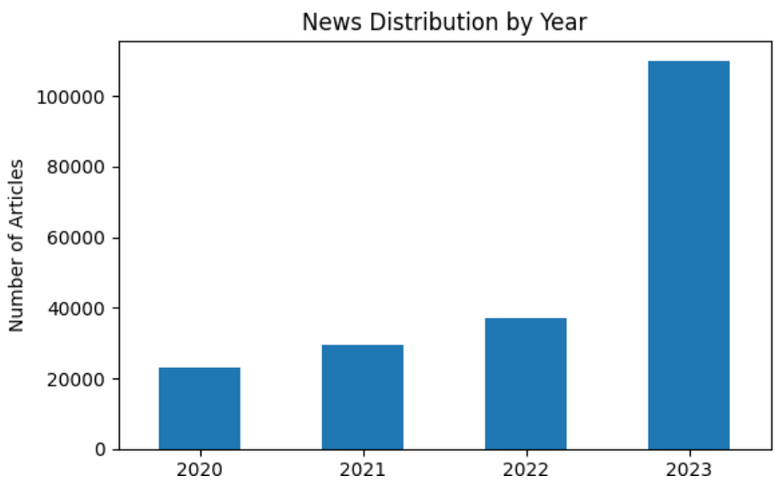
DATA OVERVIEW & METHODOLOGY

Approximately 200K AI-related news articles were preprocessed on GCP Vertex AI

Data Overview

- There are **199,677** news articles collected from multiple online sources, covering topics related to Data Science, Machine Learning, and Artificial Intelligence.
- Time frame: **January 2020 – October 2023**.
- There is a significant increase in the number of articles starting in 2023.
- Data includes URLs, published date, language, title, text of each article.
- All articles are in **English**.
- **PRNewswire** is the leading source for AI-related articles in the corpus, with the presence of technology and finance-focused domains contributing to the overall distribution of news coverage.





Methodology

Topic Detection

- Applied LDA on lemmatized news text to identify major themes and topics in the text corpus.

Sentiment Analysis

- Detected sentiment (Positive/Negative) of news text using pre-trained sentiment analysis models like VADER and TextBlob.
- Performed topic modeling post sentiment analysis using BERTopic to detect topics and industry by sentiment.

Timeline Analysis

- Analyzed distribution of positive & negative news articles over time.
- Utilized K-Train to find common topics in AI yearly.

Entity Identification

- Applied SpaCy large model on cleaned news text to build Named-Entity Recognition (NER) for entity identification.
- Cleaned NER entities manually to optimize more accurate results.

Targeted Sentiment Identification

- Leverage the results from NER and BERTopic to find typical entities among positive and negative articles.

ARTICLE CLEANING & FILTERING

190,487 articles were finalized for analysis after the cleaning & filtering process

Data Cleaning

<u>Cleaning up news title and news text:</u>

- Removed URLs, emails, newlines, tabs, numbers, punctuation, multiple spaces, whitespace.
- Removed special characters other than words (except hyphens and underscores).
- Separated words by "_" and "-".
- Removed words that are longer than 14 characters.
- Removed web remnant words such as *Ad, Refresh, Search, Login, Logout, Register, Subscribe, Sponsored, You may also like, Join us, All rights reserved, Leave a comment', Privacy rights, Follow, Reply, Subscribe, etc.*

Article Filtering

- Dropped irrelevant articles (which are not related to AI) by filtering articles that contain keywords: **Data Science, Machine Learning, Artificial Intelligence, ML, AI, ChatGPT.**
- Detected and dropped **2649** duplicated news articles.
- Kept articles with maximum **2500** words length.

<u>Text Standardization:</u>

- Tokenized and lemmatized the news title and text, and only kept nouns, adjectives, verbs, and adverbs using SpaCy.

TOPIC DETECTION

Utilized LDA with tunning to generate insights of AI applications across different industry

Applied Latent Dirichlet Allocation (LDA) to the entire text corpus.

Performed hyperparameter tunning to identify the best LDA model for topic detection.

Best LDA Model Parameters:

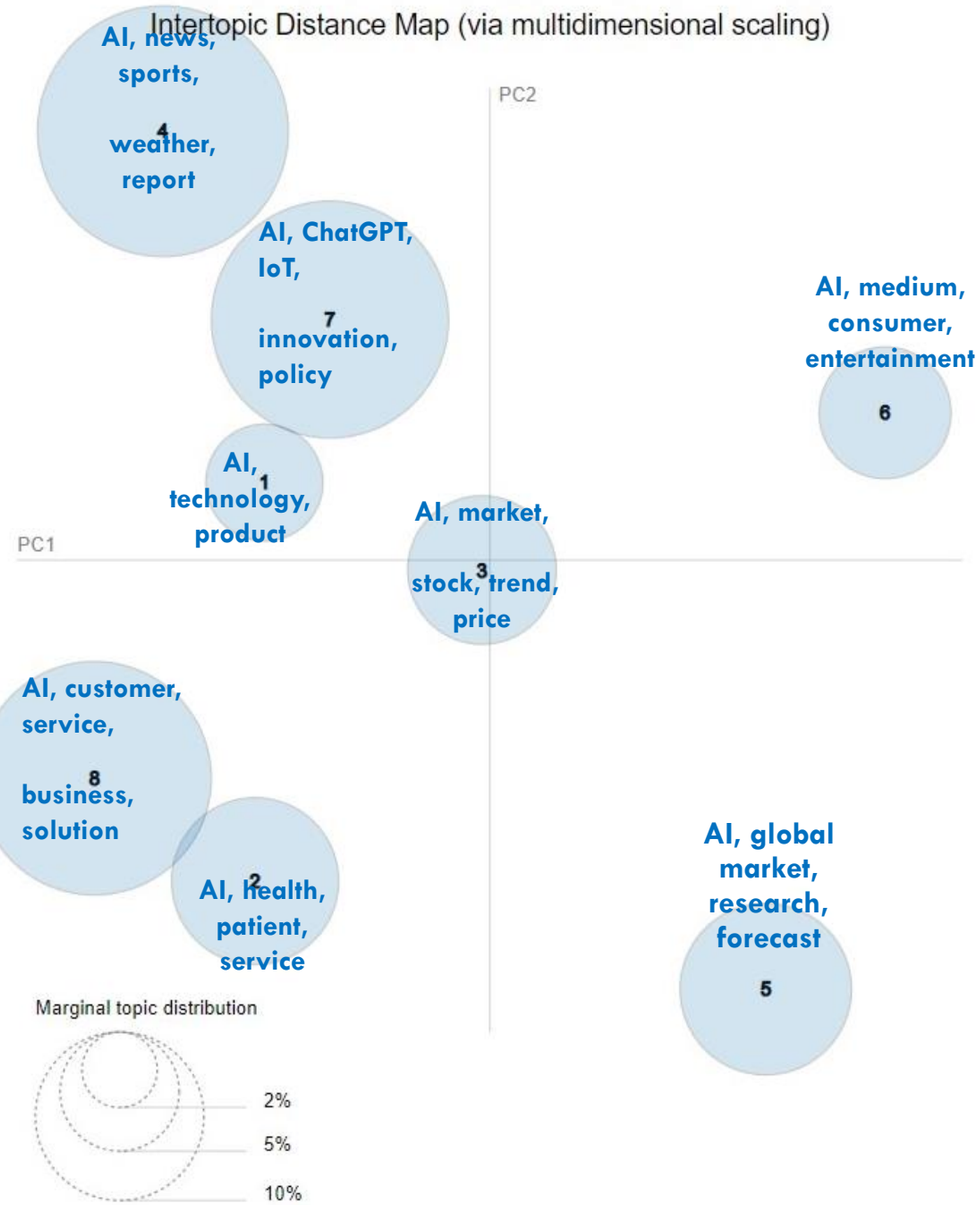
- Topics: 8
- Alpha: Asymmetric
- Beta: Auto
- Coherence Score: 0.38

Some Major Topics Detected:

- AI Integration and Innovation (e.g. ChatGPT, IoT)
- AI in Healthcare
- AI in News
- AI in Finance & Global Business
- AI in Media & Entertainment

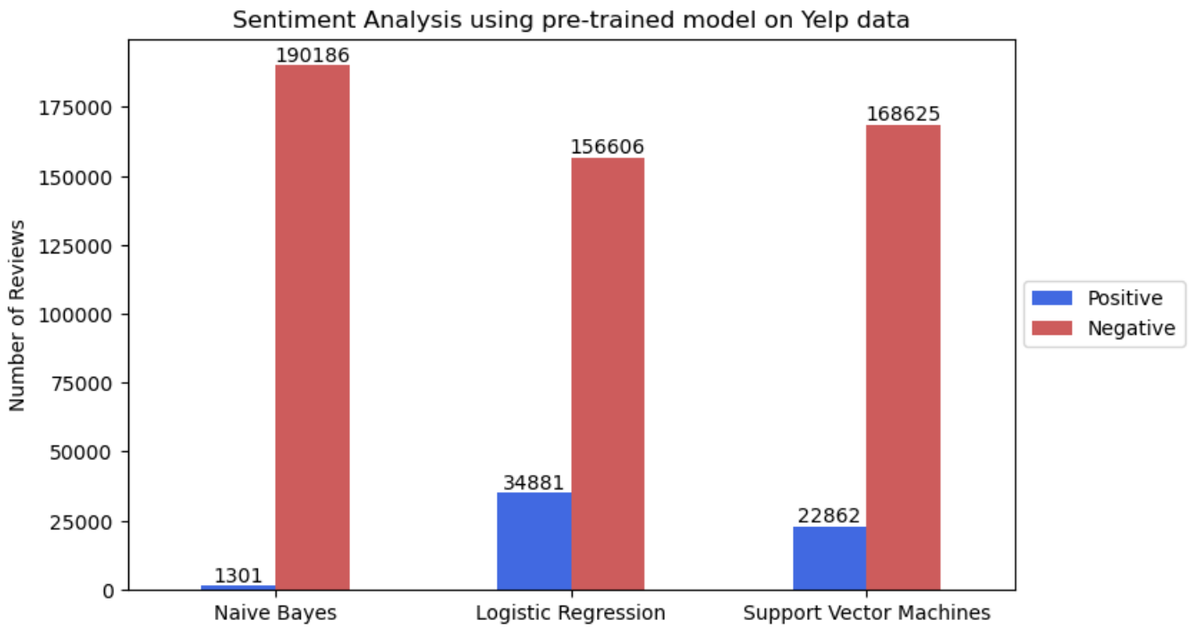
Insights from the topics:

- Surge in **AI innovation,** with tools like ChatGPT becoming integral to technological ecosystems, indicating a future of integrated AI solutions.
- **Healthcare** is witnessing a transformative shift with AI, leading to more efficient management of health services.
- AI contributes to **global market** analysis, enabling stakeholders to make informed decisions based on trend analysis and forecasts.
- The **news landscape** is evolving with AI, facilitating more content delivery and real-time reporting across sports, weather, and other current events.
- **Media & Entertainment** sectors are utilizing AI to tailor experiences, demonstrating the technology's impact on consumer engagement.

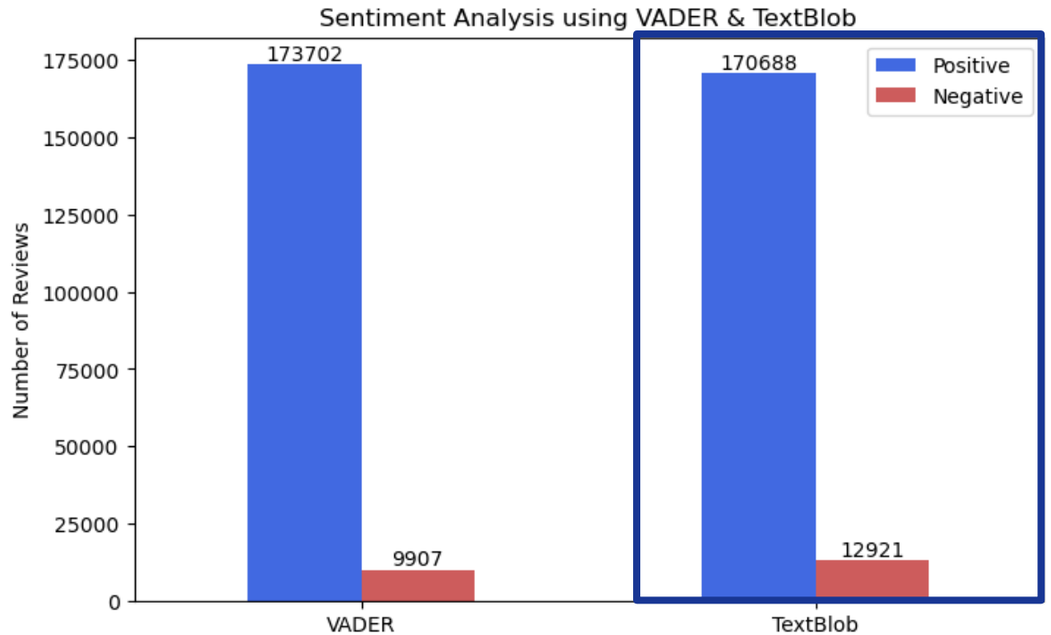


SENTIMENT ANALYSIS

The sentiment analysis was conducted using pre-trained models on open-source data like, TextBlob and VADER. TextBlob was selected for gauging accurate sentiment distribution of news articles.



- To determine the sentiment analysis for news text corpus, three binary classifier models were trained on Yelp reviews dataset (1 = Positive; 0 = Negative). Logistic Regression achieved the highest accuracy for sentiment classification (96%).
- However, when these pre-trained models were applied to news articles data, majority of news text were detected as "0" (Negative). This is probably due to the contextual difference, particularly the complex structures in news articles text, and the more opinionated and straightforward expressions in Yelp reviews text.

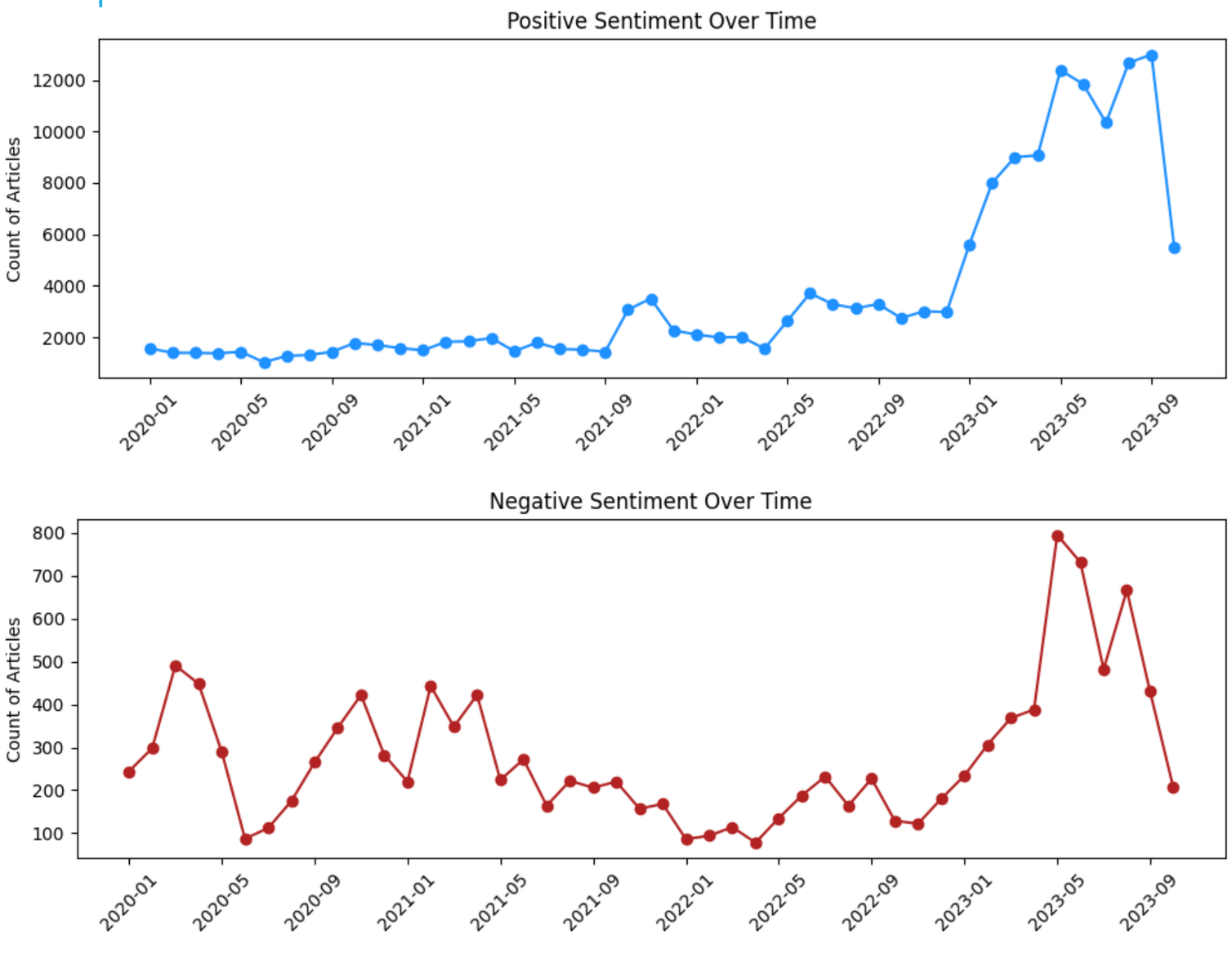


- On the other hand, VADER and TextBlob provided a more nuanced understanding of the sentiments expressed within the articles, with a greater number of articles being classified with positive sentiment.
- Considering these two models, TextBlob uses a more traditional, lexicon-based approach, so it will be better suited for news text that use more formal language. Meanwhile, VADER is often applied to detect sentiments in social media since it can handle slang, emoticons, and other informal communication forms well, yet it is also less effective for formal, longer texts.

Therefore, TextBlob would be more suited for this analysis because of its ability to handle formal language typically found in news articles.

TIMELINE ANALYSIS – SENTIMENT OVER TIME

Leveraged the TextBlob sentiment result to visualize the distribution of AI-related articles from 2020 to 2023



- Most news articles in the corpus tend to be positive, accounting for approximately 90%, reflecting a generally optimistic perception of AI advancements.
- Both sentiments exhibit a strong increase post-2023, suggesting a pivotal period where AI might be at the forefront of public discourse, possibly due to breakthroughs and controversies of AI technologies.
- Economic factors such as AI's contribution to stock market dynamics and financial services, could also influence sentiment swings in news reporting.

Increase in Positive Articles:

- The surge in positive sentiment articles since 2023 could be attributed to the rise of AI technologies, particularly in Conversational AI, which has improved user experience.
- The continuous development AI in healthcare industry over time can contribute to optimistic news.

Negative Articles Considerations:

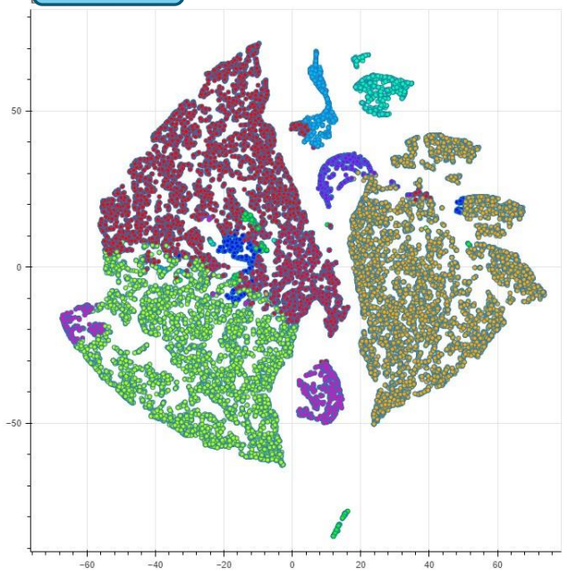
- A rise in negative sentiment articles might be due to public concern over job automation, with AI's impact on human employment.
- Ethical issues around AI could lead to critical news coverage

TIMELINE ANALYSIS – TOPIC DETECTION YEARLY

Performed topic modeling using K-train to discover the prevailing AI themes influencing different sectors each year.

- Common topics over time: **Healthcare, Market Analysis, Finance, Customer Experience, Advancement Technology, and Security.**
- Notably, 2023 has witnessed a surge in the development of Generative AI & Conversational AI (ChatGPT), indicating its impact on various industry.

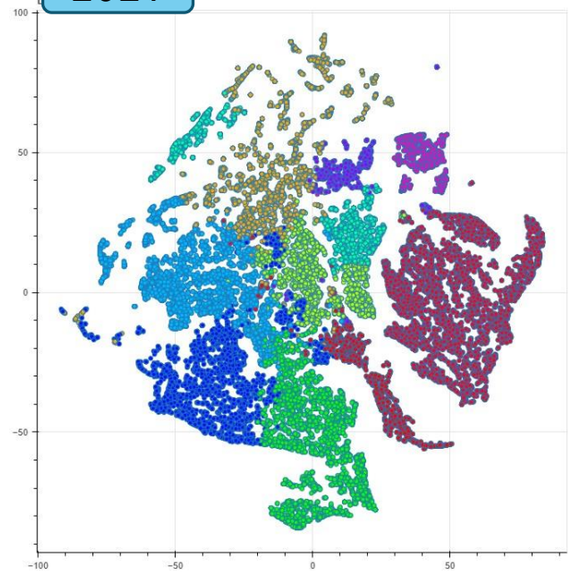
2020



Major Topics:

- AI in Market Analysis & Growth
- AI's Role Amidst COVID-19
- AI-driven Solution for Customer Service
- AI in Financial Analysis
- Advancements in AI Research

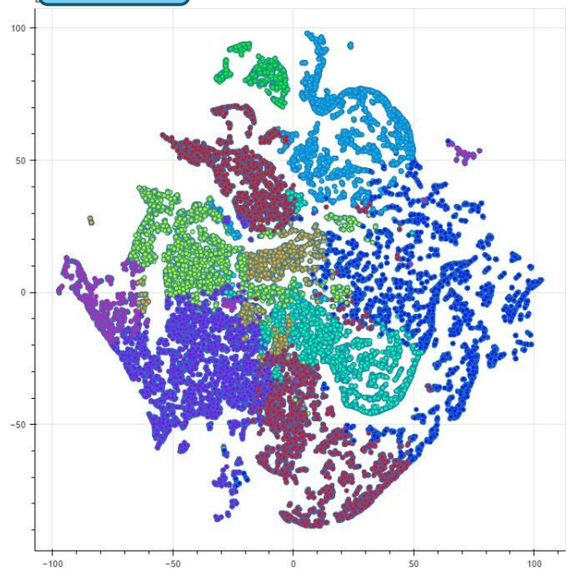
2021



Major Topics:

- AI in Strategic Market Analytics
- AI in Governance and Security
- AI-powered Customer & Enterprise Solutions
- AI's Impact on Media & Entertainment
- Healthcare and AI Synergy

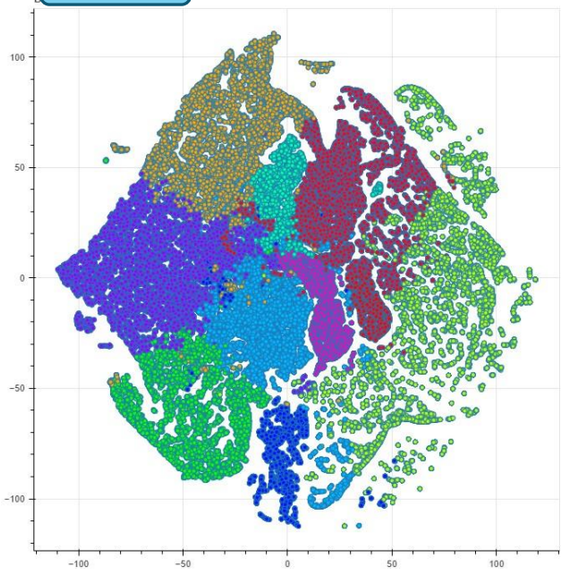
2022



Major Topics:

- Digital Customer Interaction Platforms
- AI in Financial Services
- AI-driven Image Research
- AI's Contribution to Healthcare
- Advancements in Computing & Security

2023



Major Topics:

- AI in News and Communication
- Generative AI in Business Applications
- AI-powered Interaction Tool

(e.g. ChatGPT)

- AI in Education
- AI and Ethical Considerations in Governance

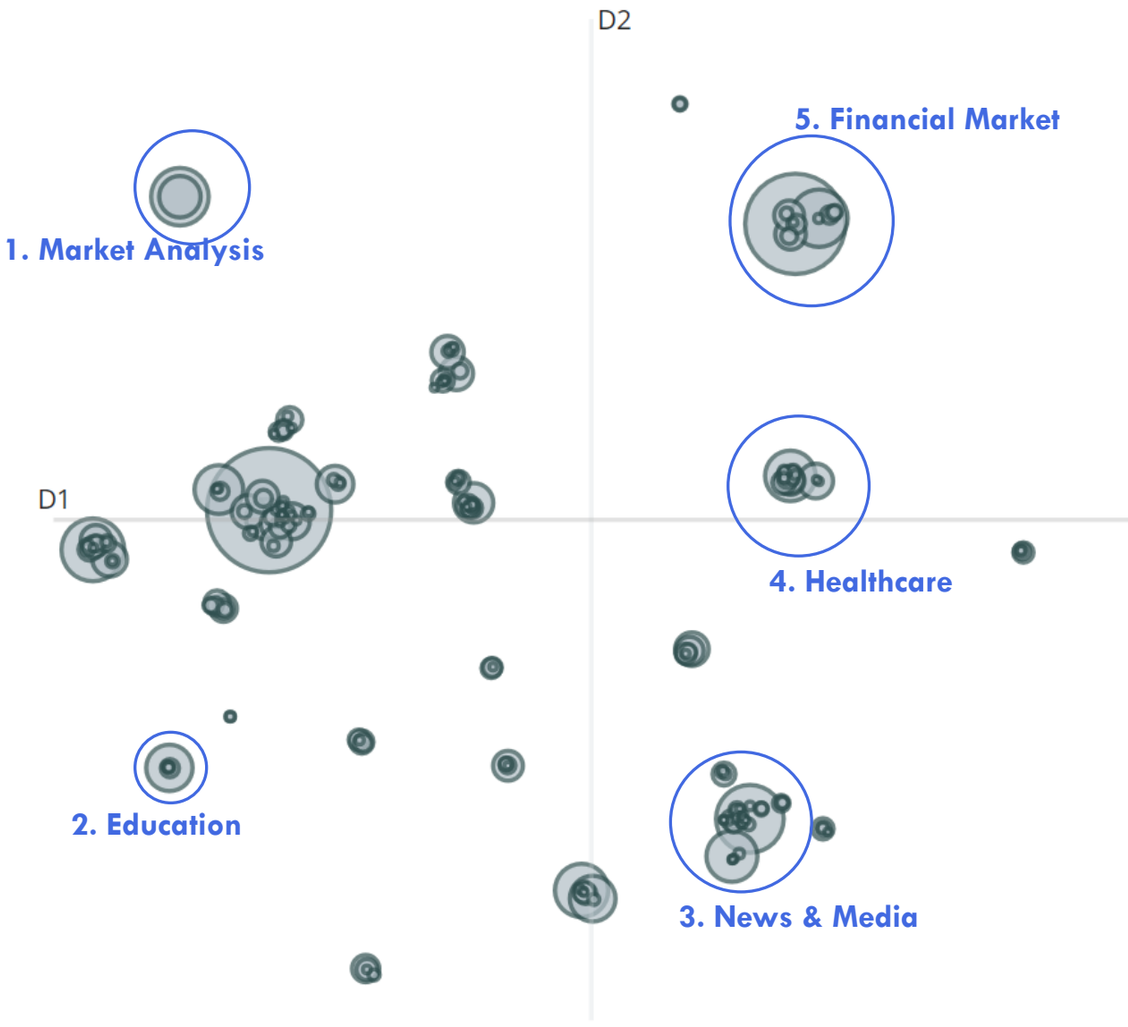
SENTIMENT ANALYSIS – TOPIC DETECTION

Applied BERTopic to investigate AI-related topics by groups of Positive & Negative articles

643 positive topics and 250 topics were detected in the corpus. After tunning and using HDBSCAN to cluster the topics using each c-TF-IDF representation, the number of positive and negative topics reduced to 150 and 50, respectively.

Examples of topics representing Positive sentiment:

1. AI contributes to **global market** analysis to guide strategic decisions based on trend analysis and forecasts.
2. The **education** sector is reshaped with adaptive learning platforms, facilitates personalized learning experiences and supports educators through intelligent systems.
3. AI is prominent in the **news and media,** optimizing content delivery and enhancing audience engagement.
4. AI-driven **healthcare** were a key focus, particularly in more accuracy diagnostics and personalized medicine.
5. AI's integration in **financial market** drives smarter investments through complicated algorithmic trading and stock analysis.



SENTIMENT ANALYSIS – TOPIC DETECTION

Applied BERTopic to investigate AI-related topics by groups of Positive & Negative articles

643 positive topics and 250 topics were detected in the corpus. After tunning and using HDBSCAN to cluster the topics using each c-TF-IDF representation, the number of positive and negative topics reduced to 150 and 50, respectively.

Examples of topics representing Negative sentiment:

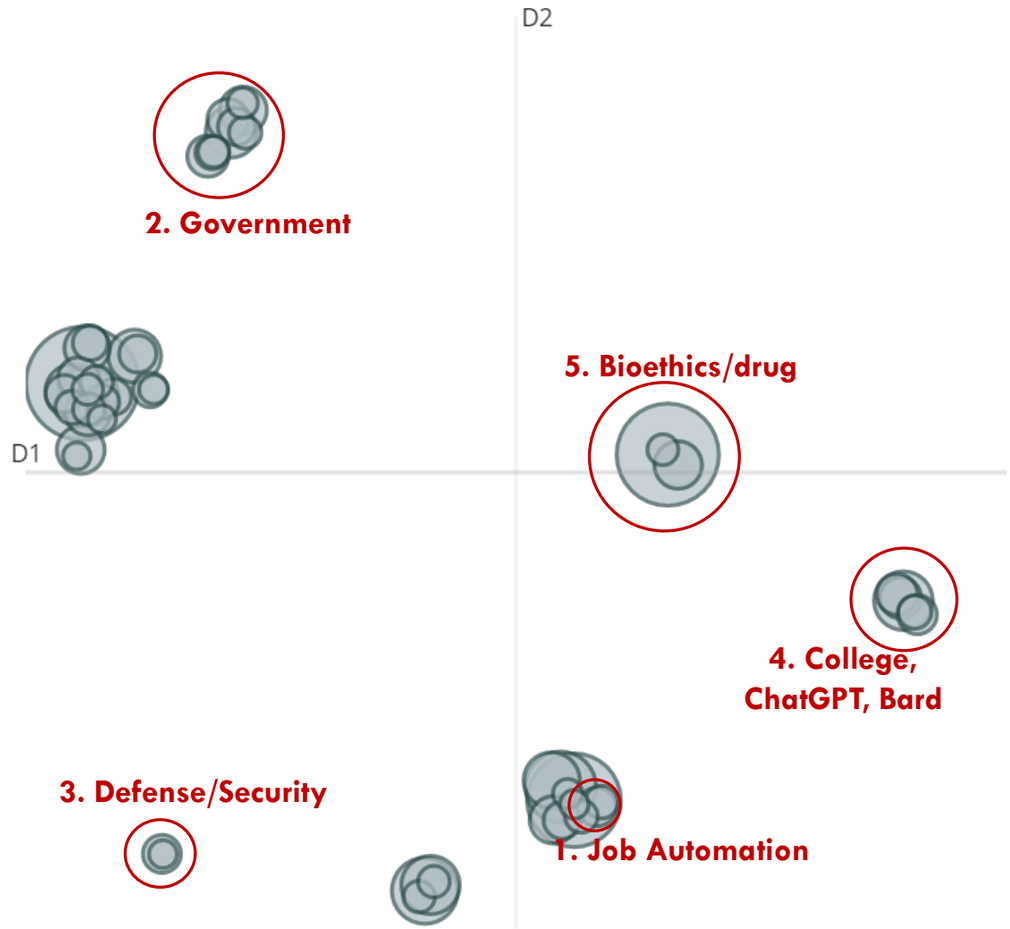
1. The rise of AI also raises the concerns about **job automation,** leading to apprehension about job displacement.

2. **Government** and **defense**-related AI applications might raise debates on warfare and surveillance.

3. The intersection of AI and **cybersecurity** underscores the escalating arms race in **digital defense,** necessitating advanced protective measures against AI-augmented threats.

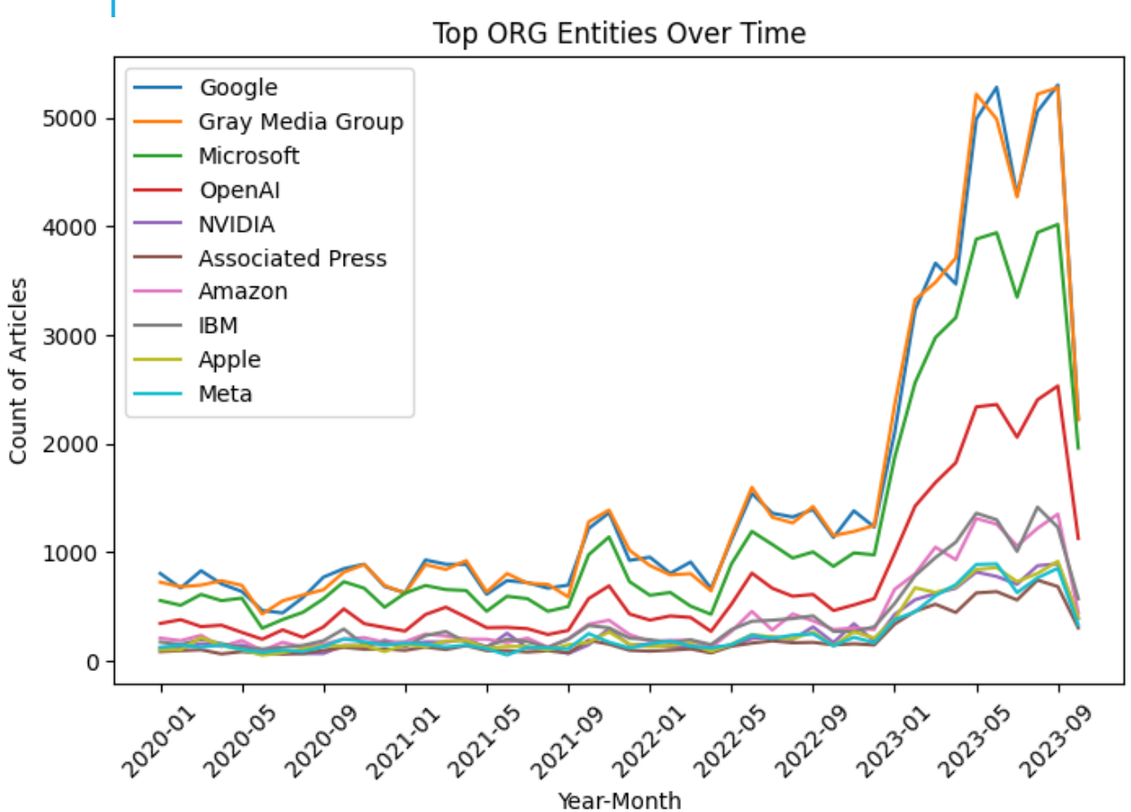
4. The impact of AI in **education** tends to have dual sentiment implications: positive due to enhanced learning tools and negative because of challenges in academic integrity and the nature of learning.

5. AI's rapid development necessitates a parallel advancement in **bioethics** to address emerging moral dilemmas.

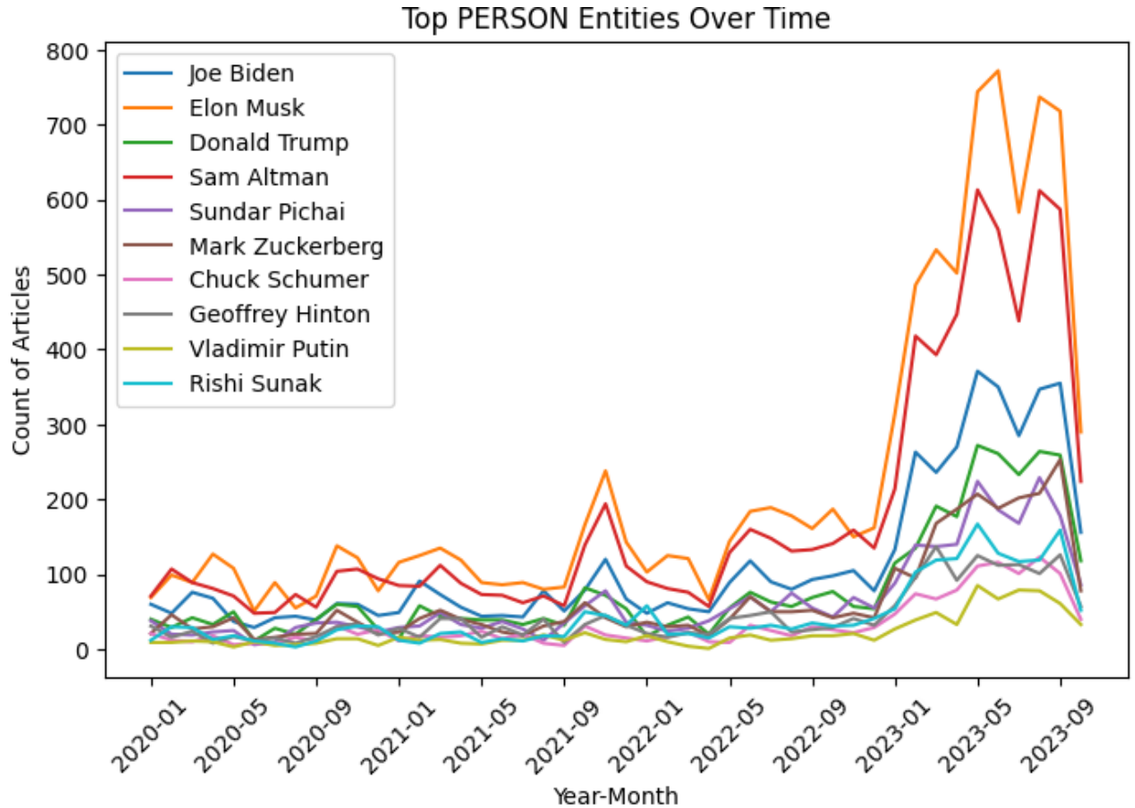


ENTITY IDENTIFICATION – ORGANIZATION & PERSON

Used spaCy large model with NER for entity identification across four labels: Organization, Person, GPE (Location), and Product. Each label reflects the top AI-related entities mentioned the most since 2020. These line plots visualizes the top 10 entities of the four labels over time.



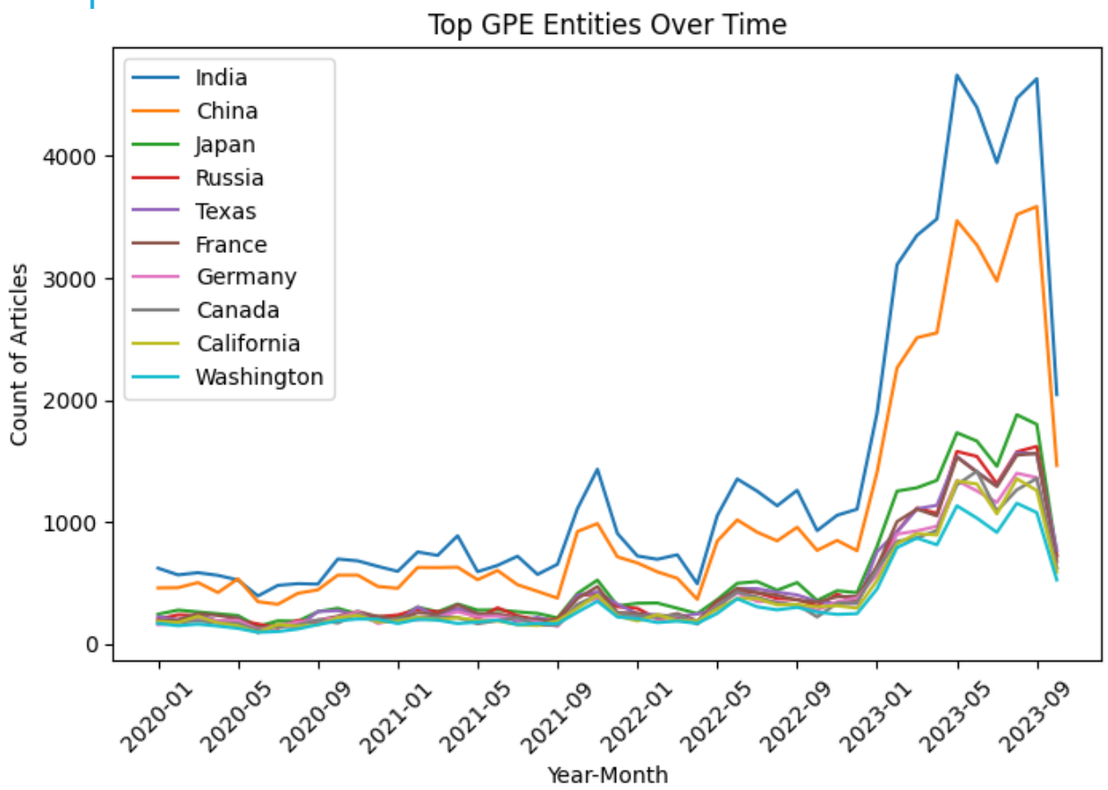
- **Google, Microsoft,** and **OpenAI** are the leading companies, which showcases their pivotal roles in advancing AI technology. Besides tech giants, **NVIDIA,** a semiconductor company, is also showing their active involvement in AI domain.
- **Gray Media Group** stands out as the media organization which extensively writes about AI advancement, followed by Associated Press.



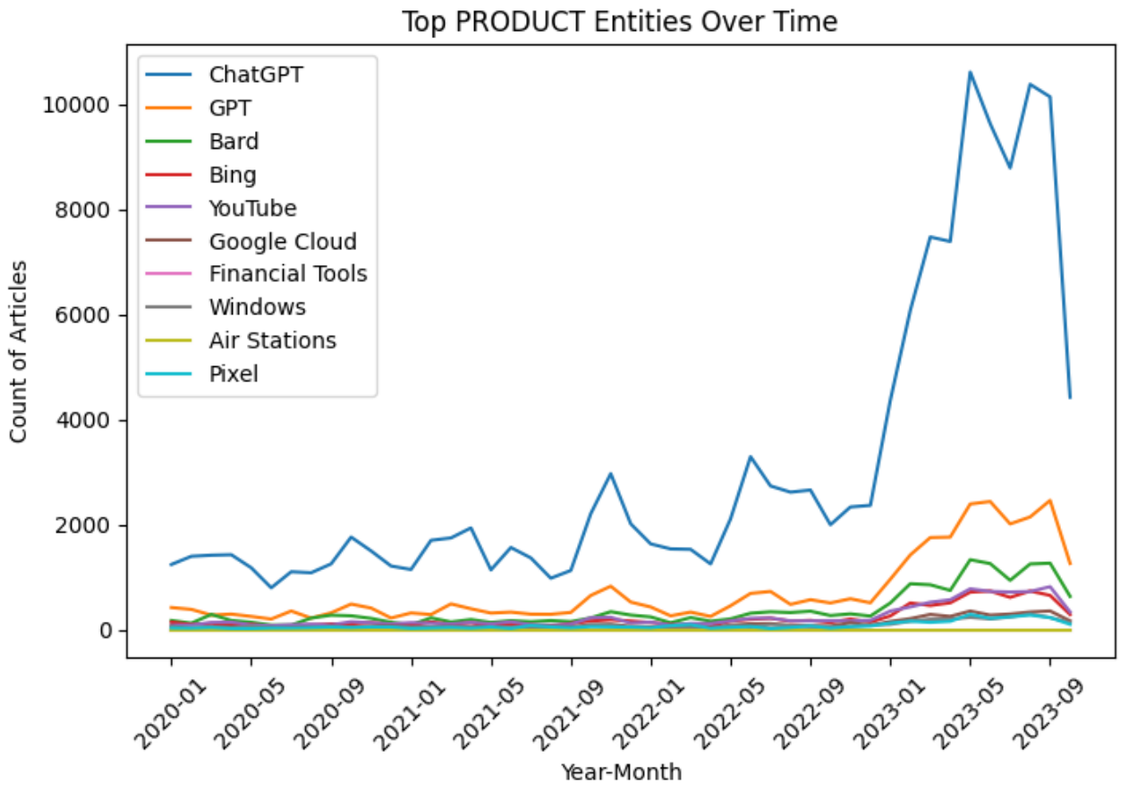
- **Elon Musk** and **Sam Altman** are notable for their speeches and commentary on AI, reflecting in the surges of articles over time corresponding to their announcements.
- Apart from tech company CEOs, national leaders and policymakers like **Biden, Trump, and Putin,** are also participating in AI discourse, suggesting impact of AI on societal and political engagement.

ENTITY IDENTIFICATION – GPE & PRODUCT

Used spaCy large model with NER for entity identification across four typical labels: Organization, Person, GPE (Location), and Product. Each label reflects the top AI-related entities mentioned the most since 2020. These line plots visualizes the top 10 entities of the four labels over time.



- The chart highlights the prominent locations in AI news coverage.
- **India, China,** and **United States** (shown as states' names like Texas, California, Washington, etc.) lead the discourse, reflecting their investments and advancements in AI. Such increases may also correspond with AI-related events or policy announcements in these countries.



- **ChatGPT, Bard,** and **Bing** AI are prominent Conversational AI solutions, advancing search engine and content generation.
- Google and Microsoft products **(Pixel, YouTube, Google Cloud, Windows, Excel)** showcases the foundations of these platforms in supporting AI development across different industries.

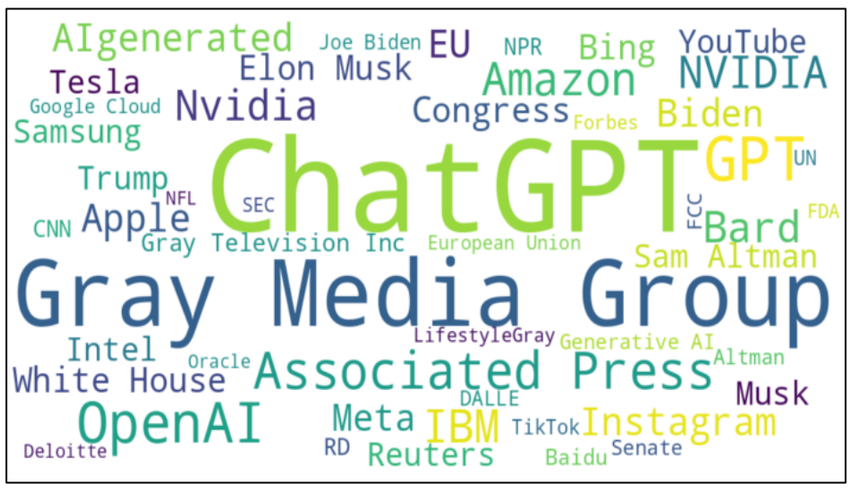
ENTITY IDENTIFICATION BY TARGET SENTIMENT

Leveraged NER and **Positive** BERTopic to reveal industries and company ready to lead AI revolution, uncovering success stories.

Financial Market

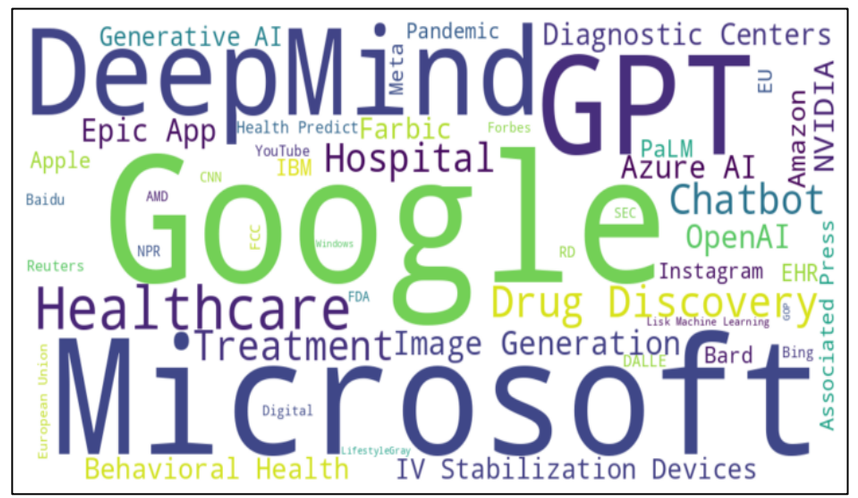
- **Google** Assistant has become valuable tool for seeking personalized financial advice and planning.
- **Microsoft** Azure AI helping financial institutions automate compliance tasks and monitor regulatory changes to remain compliant with privacy regulations.
- **OpenAI**'s GPT-3 generates accurate and informative financial reports, saving time and resources for financial analysts and accountants.

News & Media



- The presence of **Gray Media Group** as the leading press for the AI news, followed by Associated Press, reflects the media's role in disseminating AI-related information, highlighting the media's interest in the field's development.
- **NVIDIA**'s AI platform, Maxine, allows for real-time translation of video content, helping breaking down the language barrier and allowing people from different cultures to connect and share information.
- **OpenAI, ChatGPT** remain hot topics in this industry.

Healthcare

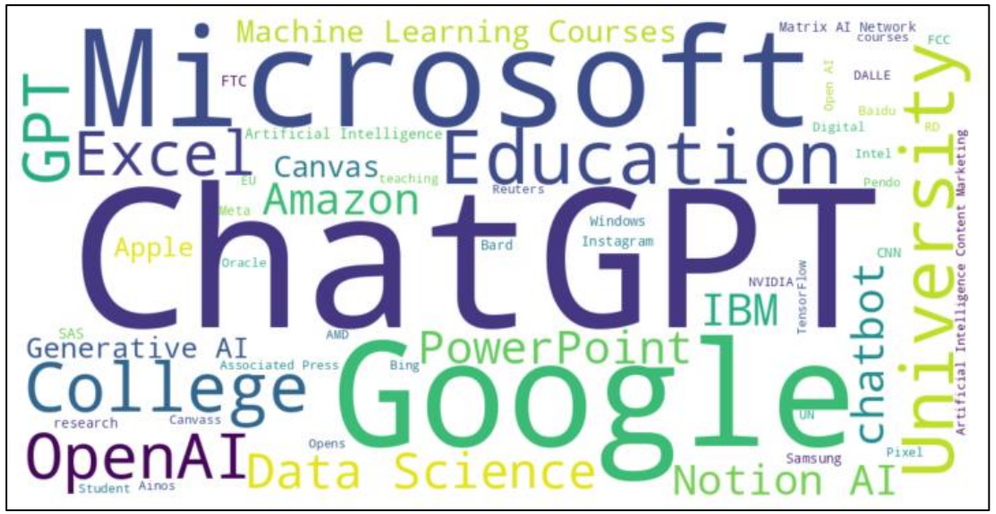


- **Google's DeepMind Health Technology** have been developing AI and LLM products to support human health. They are also exploring the use of AI to help diagnose cancer and prevent blindness.
- **Microsoft** has introduced healthcare-specific AI tools (e.g. **Fabric and Azure AI**) to standardize electronic health records and improve patient care.
- **Drug discovery** and **clinical diagnosis** are notable applications of AI in healthcare.

ENTITY IDENTIFICATION BY TARGET SENTIMENT

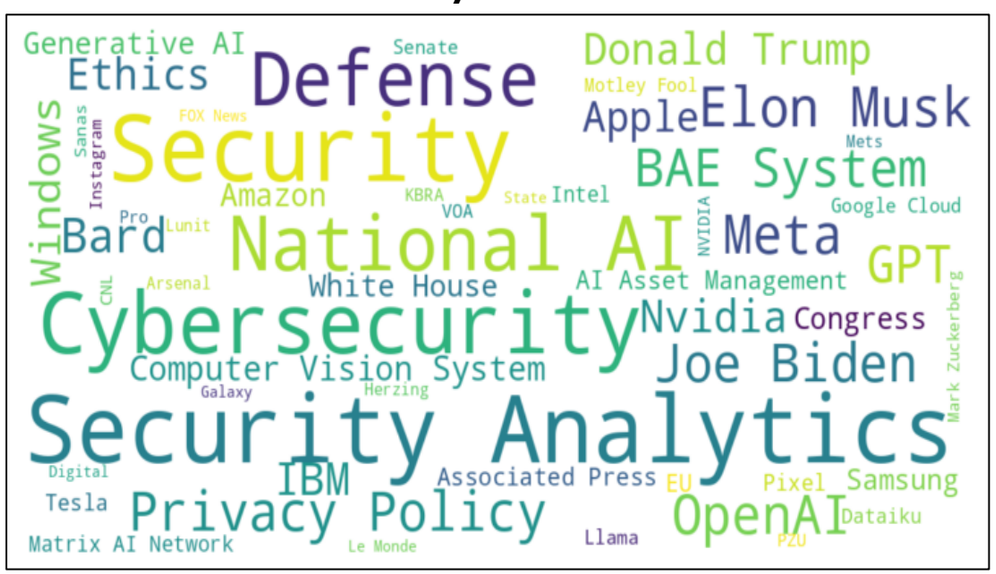
Leveraged NER and **Negative** BERTopic to discover industries impervious to AI transformation, constrained by current technological limitations.

Education



- Generative AI tools (like **ChatGPT, Notion AI**) may make it easier for students to cheat on assignments and exams. These tools can also help them write essays that pass plagiarism checks, which challenges the authenticity of students' work.
- AI cannot replicate the relationship-building aspect of teaching, which is essential to create engaging learning environment. Besides, AI can provide answers, but it does not teach the methodologies or the reasoning behind those answers.

Security & Defense



- AI systems require large datasets for training, which can include sensitive personal and national security information, raising concerns about privacy breaches and data misuse.
- While AI can enhance threat detection and response times, it also provides new tools for cyber attackers, creating an arms race between security professionals and malicious entities.
- AI systems may not be completely efficient in making complex ethical decisions that require human judgment, particularly in defense scenarios where life-and-death decisions must be weighed with moral and ethical implications.

THANK YOU!