

AI and the Workforce: Insights and Future Directions

ADSP 32018 Natural Language Processing and Cognitive Computing

Final Project

Mia Song



AGENDA

1. Executive Summary
2. Actionable Recommendations
3. Methodology Overview
4. Source Data Overview
5. Data Clean-Up and Filtering
6. Topic Detection
7. Sentiment Analysis
8. Topic Modeling with Sentiment
9. Entity Identification
10. Targeted Sentiment Identification

1. Executive Summary

AI's Impact on Industry and Employment

Objective

Leveraging AI analysis from 200K news articles to pinpoint tasks and jobs at the brink of AI transformation and to strategize on job automation and productivity enhancements.

Key Insights

AI Integration Hotspots:

Healthcare, Finance, and Business Intelligence as early-stage leaders in AI adoption.

While Finance and Business Intelligence show a cooling trend, Healthcare's engagement with AI continues to surge.

AI's Innovative Applications:

The introduction of **generative AI** like ChatGPT has revolutionized content personalization, expanding into sectors like **entertainment**, while also giving rise to crucial sectors such as **AI governance and digital security** due to emerging ethical and security challenges.

AI Development Accelerators:

Tech giants like **Google** and **Microsoft** are at the forefront of AI evolution, with their generative AI tools leading the charge. Influential institutions, such as **MIT**, and **NASA**, are vital to this advancement, showcasing a diverse ecosystem driving AI forward.

AI Investment Dynamics:

Tech giants (**Google, Microsoft, Amazon, Meta, IBM, Intel**) are set to advance AI integration, with **NVIDIA**'s semiconductors boosting computational needs. Notable players also include consumer goods (**Apple, Samsung**), social media (**YouTube, TikTok, Instagram**), and **Tesla** in the auto industry.

AI's Current Boundaries:

The technology faces challenges such as maintaining **content integrity**, preventing **bias**, regulating **image/video manipulation**, and defending against **cyber threats**. These areas underscore the vital necessity for human intervention alongside AI.

Recommendations

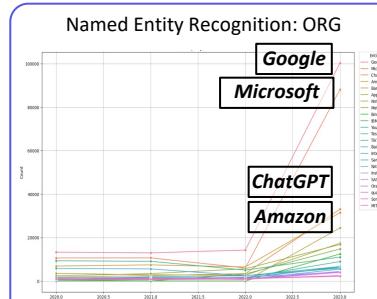
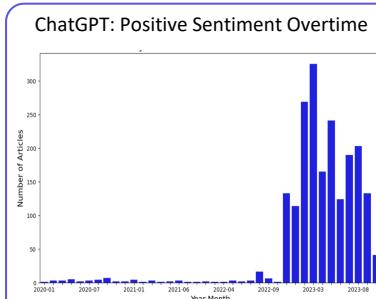
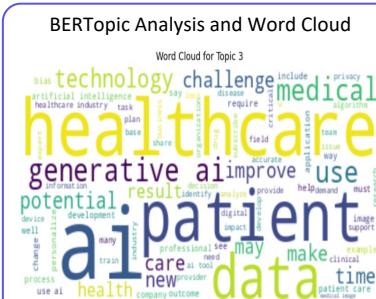
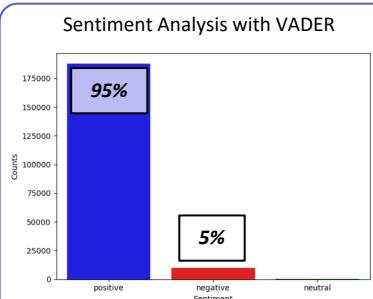
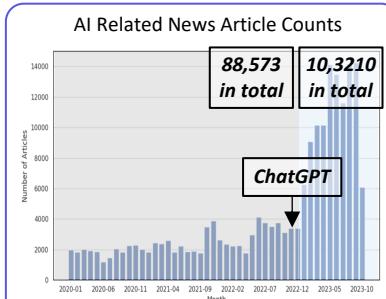
1. Prioritize AI in Key Sectors (Healthcare, Finance, Business Intelligence).

2. Expand into Generative AI (Especially in Entertainment).

3. Develop AI Risk Management and Compliance.

2. Actionable Recommendations

General Recommendations



Embrace AI Momentum

- Recognize the swift adoption of AI as a critical business driver and invest in AI technologies to stay ahead.

Leverage Positive AI Sentiments

Prioritize AI in Key Sectors

- Capitalize on the prevailing positive sentiment by promoting AI's benefits in your services and products.
- Focus AI development in healthcare, finance, and business intelligence for maximum impact and ROI.

Expand into Generative AI

- Incorporate generative AI, especially in fields like entertainment for innovative customer experiences.

Engage with AI Innovations

- Stay informed by leading AI firms and consider adopting their tools and platforms to your business operations.

Develop AI Risk Management

- Establish AI governance and risk management strategies to mitigate ethical and security challenges.

Industry-Specific Recommendations

Generative AI

- Automate content creation to reduce production times.
- Generate synthetic datasets while maintaining privacy.
- Offer customized user experiences in digital services.

Healthcare

- Deploy AI for swift and accurate medical diagnoses.
- Optimize patient scheduling and medication management.
- Enhance telemedicine with AI for initial assessments.

Finance

- Utilize AI for market analysis and risk management.
- Strengthen fraud detection with pattern recognition algorithms.
- Integrate AI chatbots for efficient customer service.

Business Intelligence

- Automate data analytics for insightful business reporting.
- Predict market trends to inform proactive strategies.
- Streamline business processes for increased productivity.

Entertainment

- Deploy AI for customized user and music recommendations on streaming services.
- Create AI-powered interactive media for personalized storytelling and music experiences.

AI Governance and Compliance

- Create ethical AI guidelines to guide responsible usage.
- Monitor compliance with automated AI tools.
- Regularly audit AI systems for bias and performance.

Generative AI holds the transformative potential to automate complex tasks, innovate solutions, and personalize experiences across all industries.

3. Methodology Overview

Our analytical framework employs a multi-stage NLP process designed to extract meaningful insights from news article concerning AI's *impact on employment*.

	Data Preprocessing	Topic Analysis	Text Summarization	Sentiment Analysis	Targeted Sentiment Analysis	Visualization
Objective	<ul style="list-style-type: none"> Prepares and cleanses data, ensuring the right format for in-depth analysis. 	<ul style="list-style-type: none"> Discovers and analyzes key themes related to AI and employment from large text collections. 	<ul style="list-style-type: none"> Distills extensive information into concise summaries to track the evolution of AI impact over time. 	<ul style="list-style-type: none"> Assesses emotions (positive, negative, neutral) in texts to determine AI's influence. 	<ul style="list-style-type: none"> Isolates and examines sentiments and topics tied to specific entities like companies. 	<ul style="list-style-type: none"> Converts complex data insights into easy-to-understand visual formats.
Methodology	<ul style="list-style-type: none"> Data Cleaning, Tokenization, Lemmatization Data Filtering with Keywords 	<ul style="list-style-type: none"> LDA with Gensim LDA with Ktrain BERTopic with sentiment analysis 	<ul style="list-style-type: none"> Summarizing Document with Ktrain with BERTopic 	<ul style="list-style-type: none"> Custom Model on Yelp TextBlob VADER 	<ul style="list-style-type: none"> Named Entity Recognition (NER) NER in sentiment analysis 	<ul style="list-style-type: none"> Line Chart Bar Chart Box Plot Bubble Chart Word Cloud

4. Source Data Overview

Dataset Overview

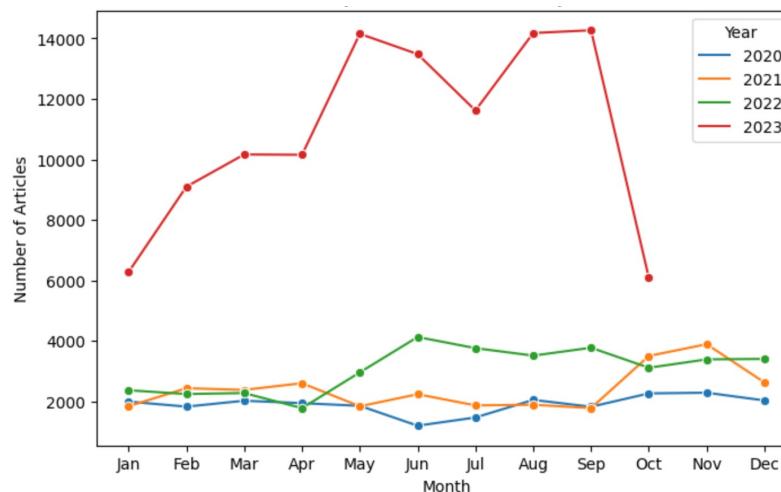
Total Entries: 199,677

Columns: URL, Date, Language, Title, Text

Data Types: Textual: Object (4 columns), Temporal: DateTime (1 column)

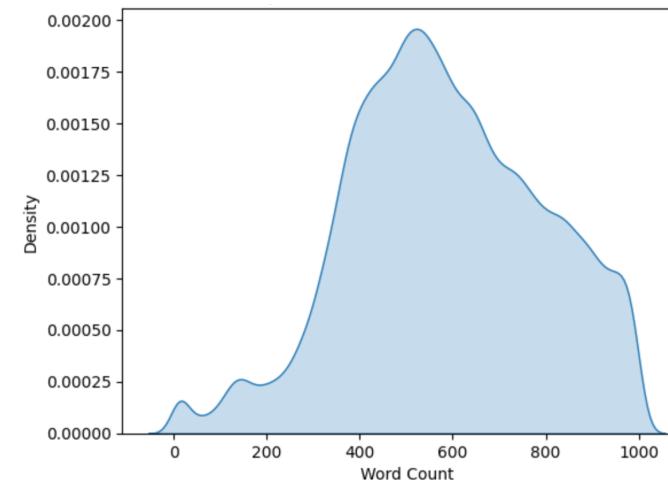
Coverage Period: January 1, 2020, to October 16, 2023 (1,095 days)

Monthly News Article Counts by Year



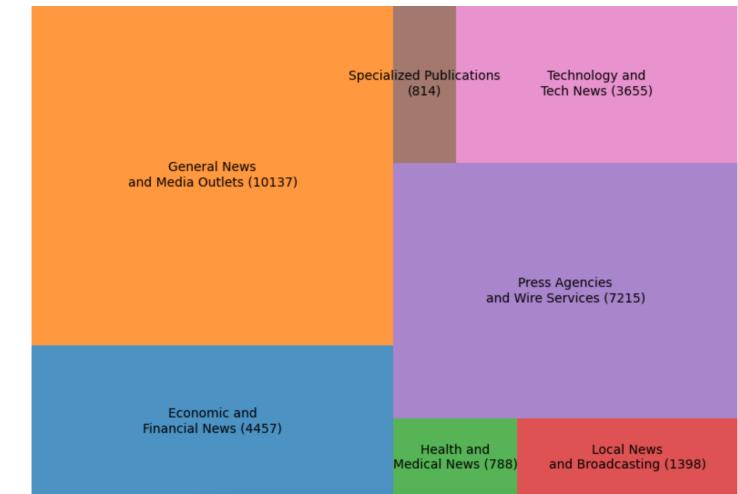
The news article count **increased** each year.

News Article Word Counts Distribution
(articles below 10,000 word counts)



News article word counts ranged from 3 to 29,325. To better understand the distribution, we capped it at 1,000 words, as the 75th percentile was 982. **Most articles were between 400 and 600 words.**

News Article Counts by Press Category



Most of the collected news articles, 10,137 in total, came from **general news and media outlets**, followed by 7,215 from press agencies and wire services.

4. Source Data Overview

Text Summarization*

Topic	Summarization
Healthcare	<p>Healthcare Artificial Intelligence Market is Booming Worldwide Investigated in the Latest Research by KSU The Sentinel Newspaper. The research report study the market size, share, key drivers for growth, major segments, and CAGR. The report studies Major Industry Key Players such as AiCure Cyrcadia Health Google IBM Microsoft Atomwise Lifegraph Modernizing Medicine.</p>
Finance	<p>Artificial Intelligence Is Helping Improve Breast Cancer Detection For Women With Dense Breasts. Using artificial intelligence, the researchers were able to look at a total of 9,200 MRI scans of dense breasts. The artificial intelligence flagged off all the scans with lesions, and dismissing of those without lesions, all without missing any cancers.</p>
Business Intelligence	<p>The Business Research Company The Financial Services Market Sees A Rise In Demand For Artificial Intelligence. The global financial services market size is expected to grow from 23.31 trillion in to 25.83 trillion in at a compound annual growth rate CAGR of 10.8. The increasing demand for artificial intelligence AI in financial services is an emerging trend. Interest rates globally are forecasted to rise in most of the developing and developed economies.</p>
Entertainment (Music)	<p>Report of execs say increasing business process efficiency is top benefit of AI. The need to boost customer service, employee efficiency, and acceleration of innovation are the three main factors driving an increase in AI adoption. The majority of survey respondents are planning to adopt AI to solve these business challenges within the next months.</p>
Image Generation	<p>How pattern used text to audio AI to make an entire album We're at the precipice of a fundamental shift in how we think about making music. Generative AI is about to redefine creativity. We speak to the producer pushing the boundaries of this extraordinary new technology.</p>
AI Governance and Compliance	<p>The Weird Rise of AI Music Rolling Stone. From voice cloning wars to looming copyright disputes to a potential flood of nonhuman music on streaming, AI is already a musical battleground. Editor's picks The Worst Decisions in Music History The Greatest Singers of All Time The Greatest Songs of Alltime The Greatest TV Shows of All time.</p>
	<p>Deepfake porn could be a growing problem amid AI race. Deepfakes are videos and images that have been digitally created or altered with artificial intelligence or machine learning. Porn created using the technology first began spreading across the internet several years ago. Some AI models say they're already curbing access to explicit images.</p>
	<p>Misinformation machines AI chatbots can spew falsehoods, even accuse people of crimes they never committed.</p>
	<p>New York City school officials this week started blocking the impressive but controversial writing tool that can generate paragraphs of human-like text. The decision by the largest U.S. school district to restrict the ChatGPT website on school devices and networks could have ripple effects on other schools. Many school districts are still scrambling to figure out how to set policies on if and how it can be used.</p>
	<p>CryptoRom Scammers Add AI Chat Tool, Like ChatGPT, and Fake Hacks on Crypto Accounts to Their Toolset, Sophos Finds. In, investment fraud caused the highest losses of any scam reported by the public to the US FBI's Internet Crimes Complaint Center IC3.</p>

*Used ktrain for text summarization, leveraging Representative_Docs from BERTopic.

5. Data Clean-Up and Filtering

Following initial data preparation, we implemented various levels of data cleansing, resulting in *three distinct datasets*, each tailored for specific analytical objectives.

Data Preparation

Language

- Select English news articles only

Data Types

- Convert 'Date' from object to DateTime format

Nulls

- Check for and confirm no null values

Filtering

- Remove duplicate news articles (1,464 articles, 0.7%)
- Discard Irrelevant Articles using Keywords such as AI, ML, and NLP (146 articles, 0.07%)

Data Cleaning

Initial Data Cleaning

- Clean Regex (Remove HTML remnants, Remove newlines, tabs, and emojis)
- Normalize space
- Remove numeric values
- Restrict text length from 1 to 20

Data Cleaning Refinement

- Convert text to lowercase
- Remove stopwords
- Strip special characters and punctuation

Data Normalization

- Apply lemmatization to reduce words to their base or dictionary form`

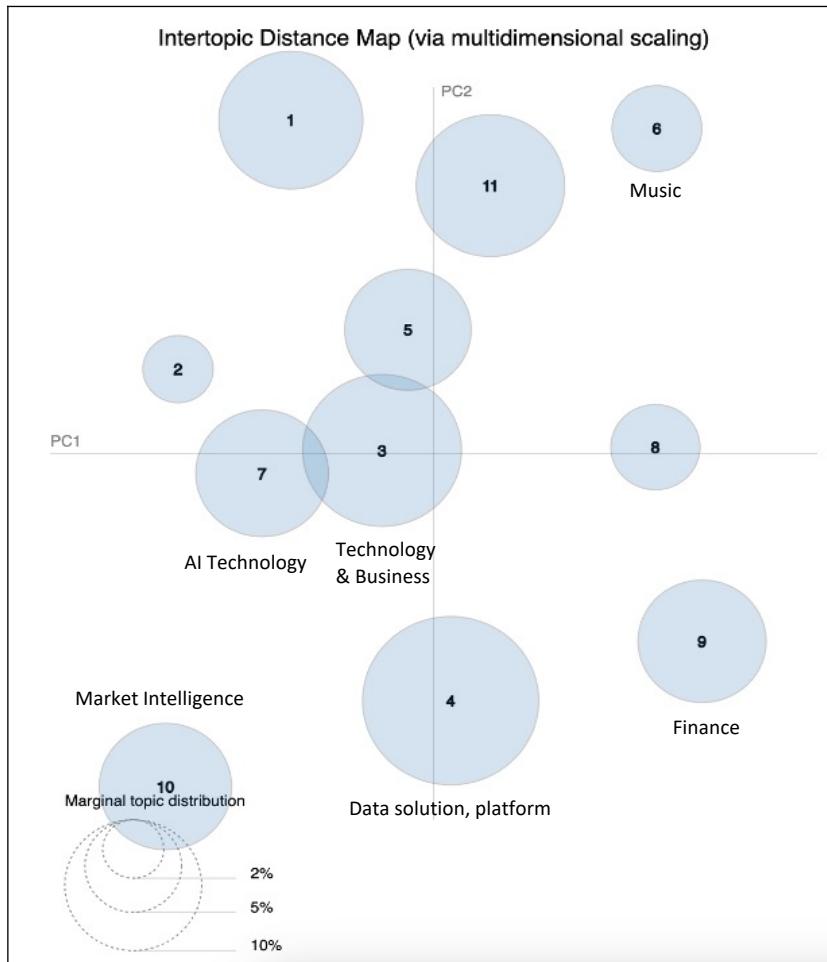
1. Texts for NER

2. Texts for General Purpose

3. Lemmatized Texts

6. Topic Detection

Topic Clustering Map from LDA Analysis



We observed an *extensive impact of AI* on both the job market and industrial landscape through Latent Dirichlet Allocation (LDA) analysis with Gensim.

LDA using Gensim with Model Fine Tuning

Objective: Fine-tuning the Gensim LDA model

Method: Iterative hyperparameter adjustment

Data Sample: Subset of 9,900 news articles (5% of total)

Optimal Parameters:

- Number of Topics: 11
- Alpha (Document-Topic Density): Asymmetric
- Beta (Word-Topic Density): Auto

Performance: Coherence Score: 0.42399 (sample data), **0.42378 (full data)**

Detected Keywords

- ai, data, technology, business, company, human, work, chatgpt, solution, platform, research, forecast, intelligence, stock, financial, investment, fund, health, school, policy

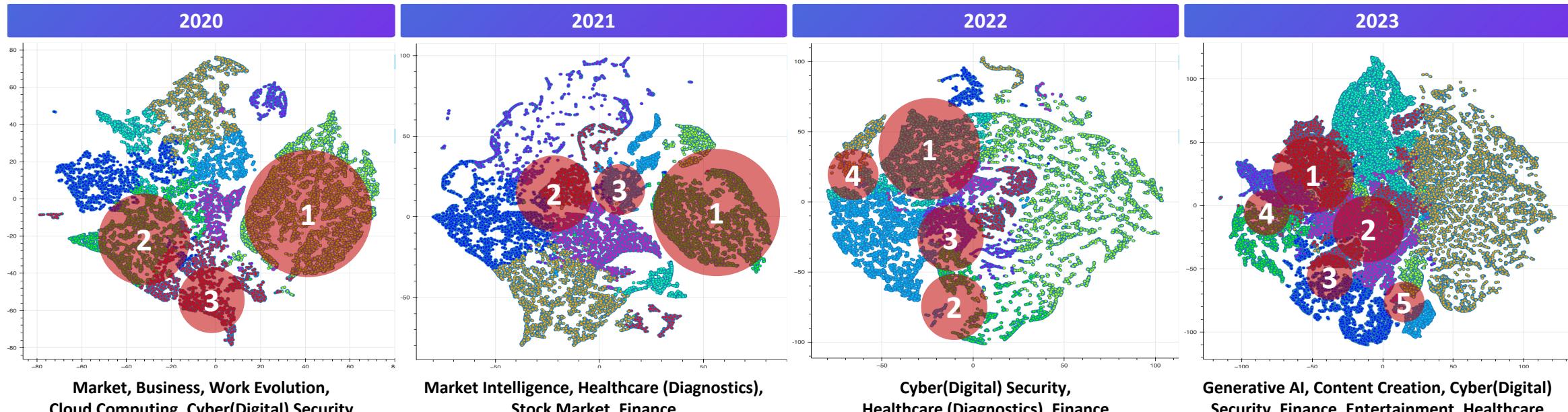
Detected Topics

- AI and Workforce
- AI and Technology
- AI and Business, AI and Market Intelligence
- AI and Finance
- AI and Healthcare
- AI and Education

6. Topic Detection

Key industrial trends in AI's impact on *healthcare, finance, and security*, and the rising role of *generative AI in content creation*, are evident through time-based analysis using LDA with Ktrain*.

*Parameters: n_topics=10, n_features=10000, min_df=5, max_df=0.5, stop_words='english', model_type='lda', lda_max_iter=5



**Market, Business, Work Evolution,
Cloud Computing, Cyber(Digital) Security**

1. AI's impact on market analysis and business strategy.
2. AI's effect on the job market and the evolution of work.
3. AI's integration in cloud computing and security.

**Market Intelligence, Healthcare (Diagnostics),
Stock Market, Finance**

1. AI's role in market intelligence and predictive analytics.
2. AI's advancement of healthcare diagnostics and patient care.
3. AI's automation of stock market trading and financial analysis.

**Cyber(Digital) Security,
Healthcare (Diagnostics), Finance**

1. AI's strengthening of digital security and user privacy.
2. AI's enhancement of medical diagnostics and treatment.
3. AI's automation of financial trading and market analysis.

**Generative AI, Content Creation, Cyber(Digital)
Security, Finance, Entertainment, Healthcare**

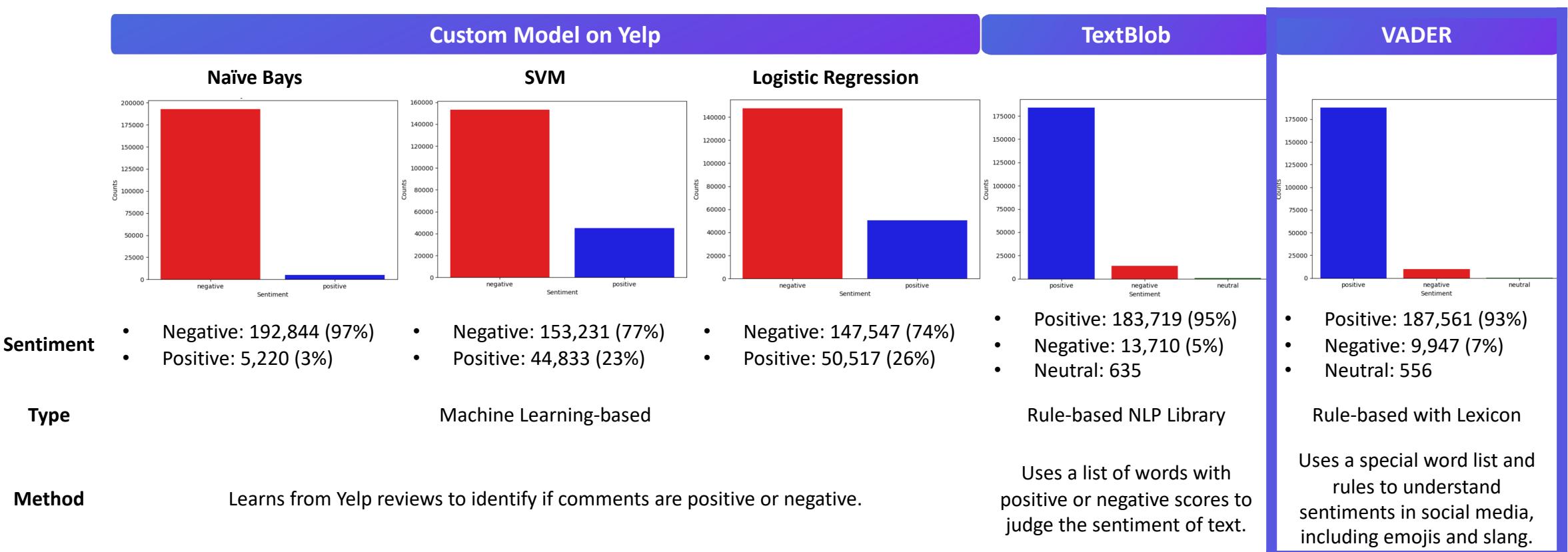
1. Generative AI redefines content creation.
2. AI shapes legal and cybersecurity norms.
3. AI revolutionizes financial analytics and investing.
4. AI personalizes entertainment experiences.
5. AI improves healthcare communication.

7. Sentiment Analysis: Model Comparison

VADER was chosen for its better handling of *diverse expressions* in news articles, compared to **TextBlob** and **Yelp** models.

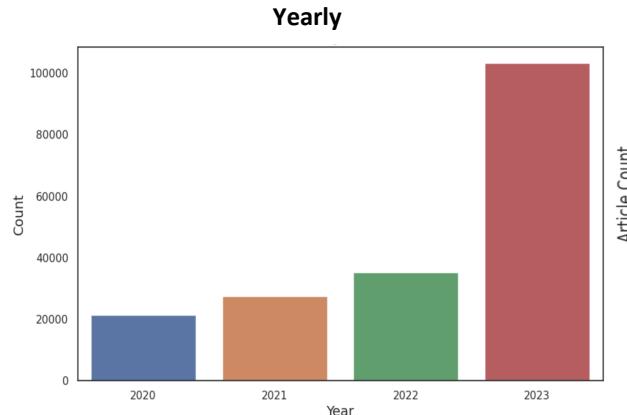
A model trained on Yelp reviews may not perform well with news articles due to different language styles and contexts.

TextBlob is limited by its predefined lexicon and may struggle with complex language and context in news.

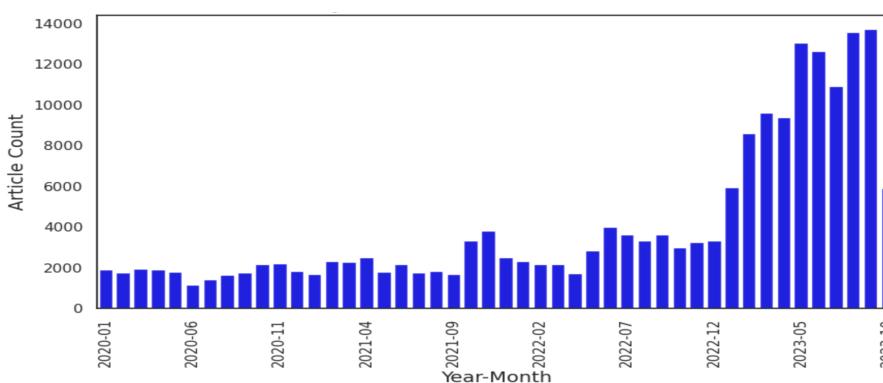


7. Sentiment Analysis: Overtime View

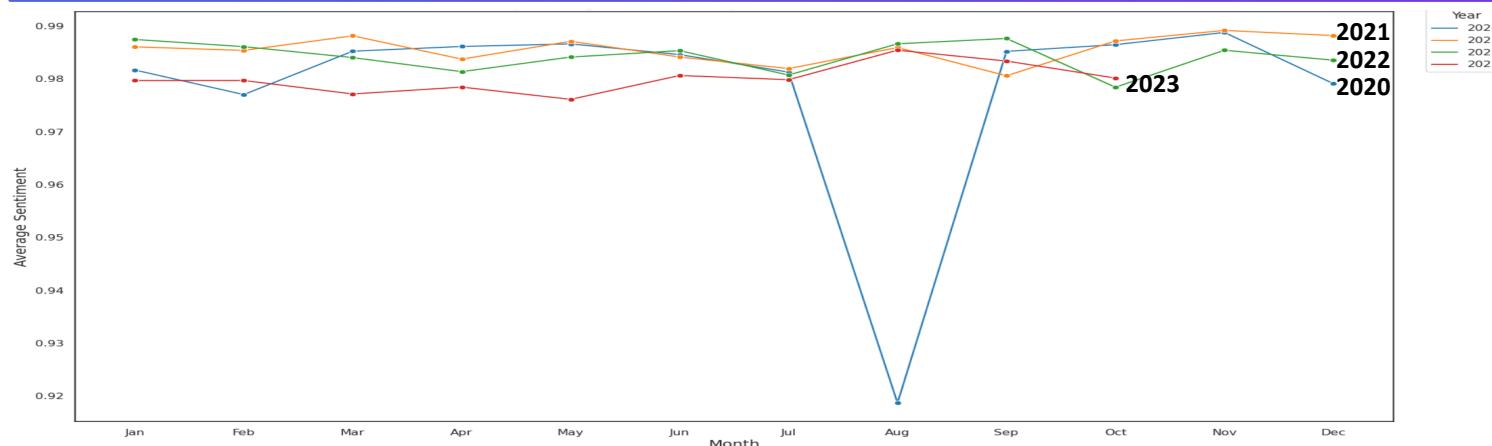
Positive News Article Counts with Vader



Monthly



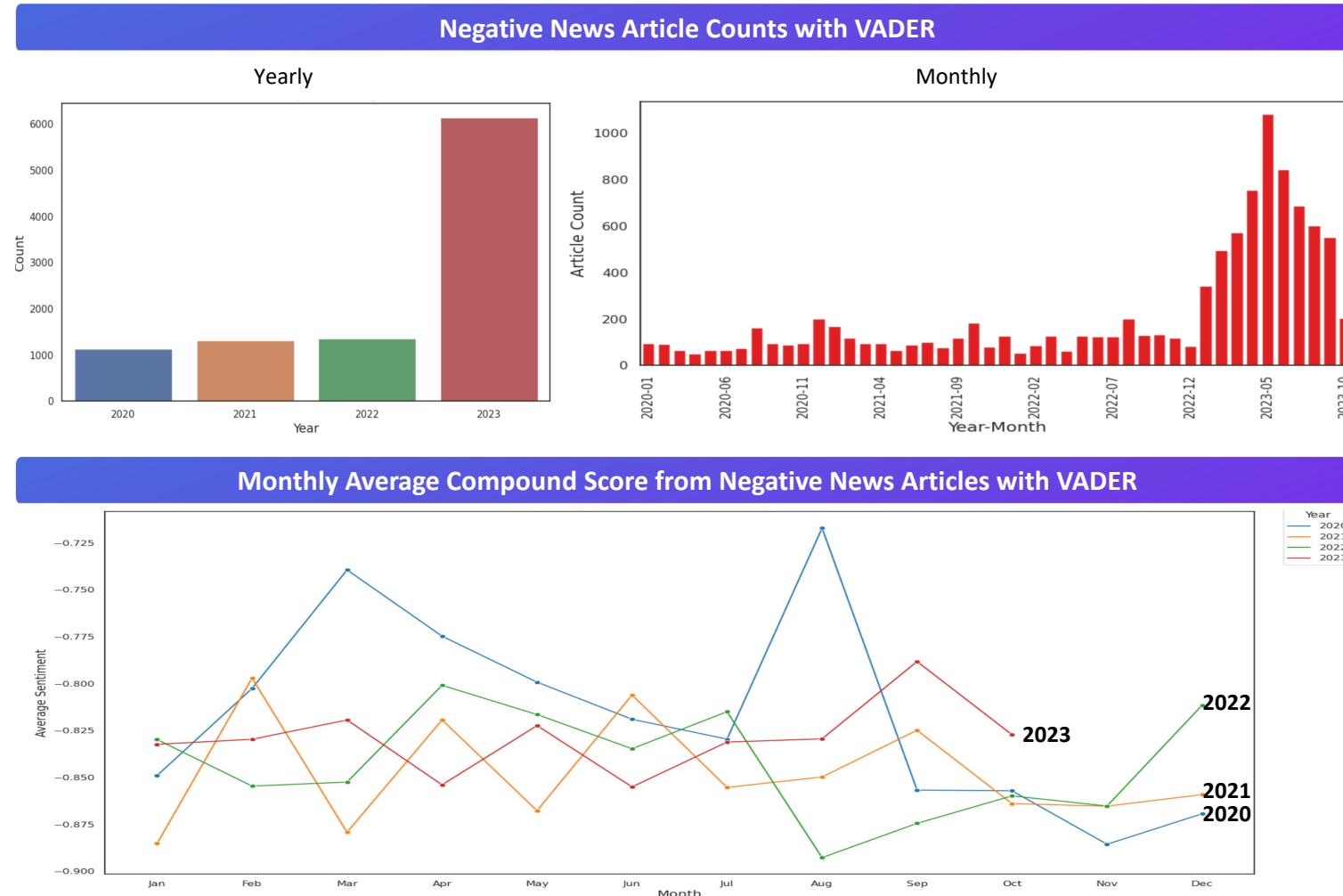
Monthly Average Compound Score from Positive News Articles with Vader



Key Takeaways

- News articles with positive sentiment have shown a increase over time, surging by 192% from 2022 to 2023.
- This uptrend is likely linked to the emergence of Generative AI, such as ChatGPT, and its growing applications in sectors like finance and healthcare, as revealed by our LDA analysis.
- Averaging positive and negative scores separately can provide insights into the intensity and prevalence of each sentiment type within the data.
- The monthly average compound scores from news articles, generally between 0.97 and 0.99, indicate a strong positive sentiment over the years. However, a yearly downward trend has emerged since 2021.

7. Sentiment Analysis: Overtime View



Key Takeaways

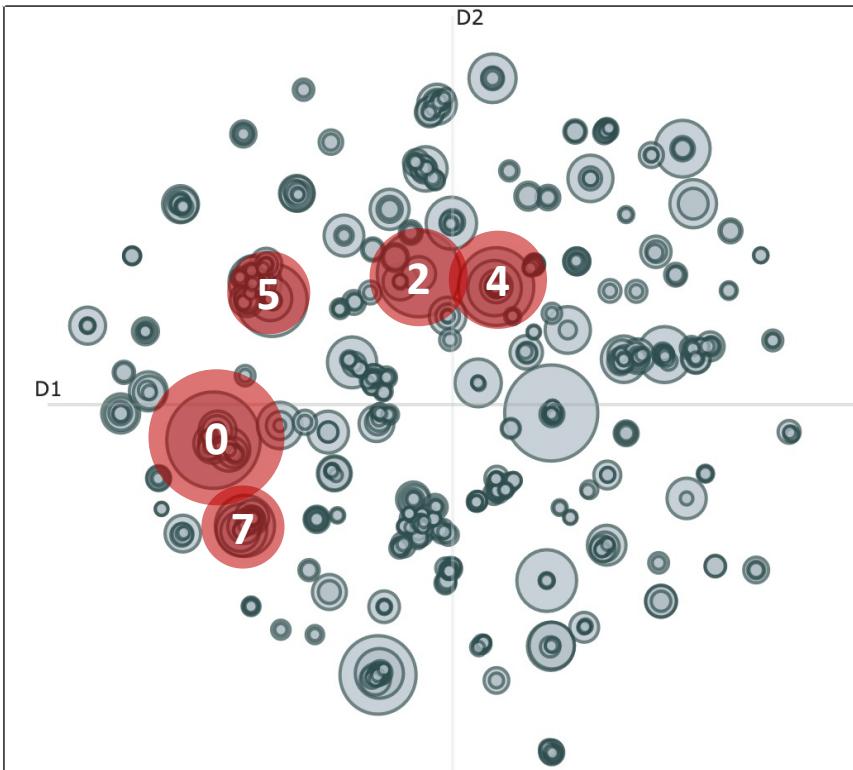
- The proportion of negative sentiment in the data is small, accounting for 5% compared to 95% for positive sentiment.
- Despite this, the incidence of negative sentiment in news articles has risen sharply, increasing by 350% from 2022 to 2023.
- Compound scores closer to -1 indicate stronger negativity. Interpreting the monthly average compound score for negative sentiment in our data is not clear, yet the recent 2023 trend suggests a weakening of negative sentiment.

Sentiment analysis over time reveals key patterns, suggesting the need to combine it with topic modeling to uncover the root causes of these sentiment trends.

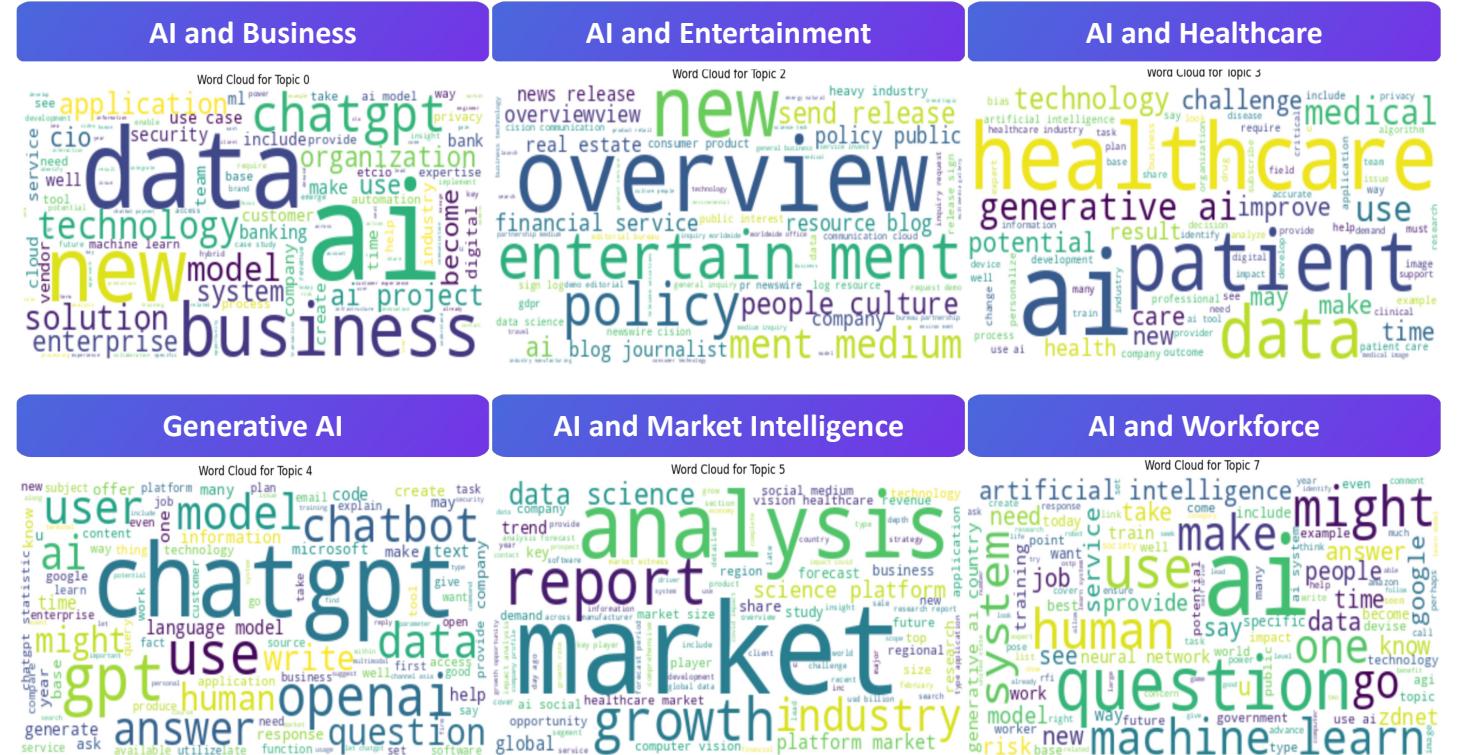
8. Topic Modeling with Positive Sentiment

After analyzing sentiments with VADER, we used BERTopic for topic modeling. BERTopic's transformer-based approach provides deeper, context-aware insights compared to traditional bag-of-words models.

Topic with Positive Sentiment Clustering

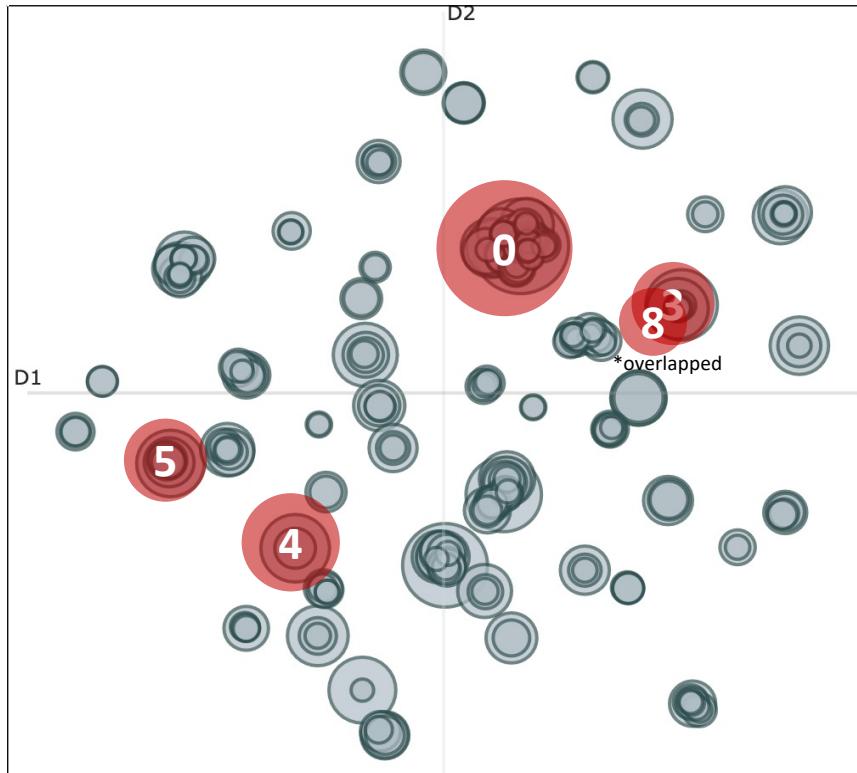


Word clouds were generated using the 'Representative_Docs' from each topic identified by BERTopic.



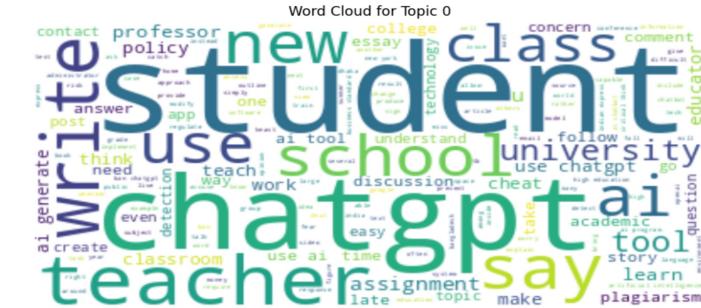
8. Topic Modeling with Negative Sentiment

Topic with Negative Sentiment Clustering



Word clouds were generated using the 'Representative_Docs' from each topic identified by BERTopic.

AI and Ethics (Content Creation)



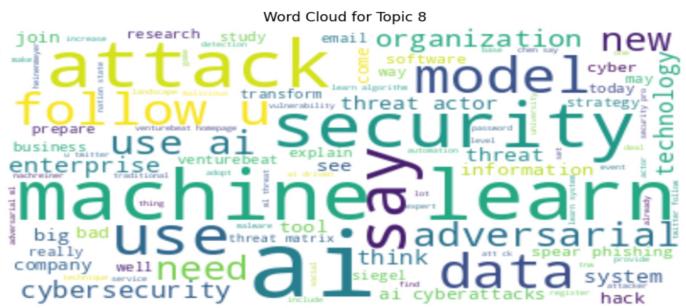
AI and Ethics (Model Bias)



Image, Video Manipulation



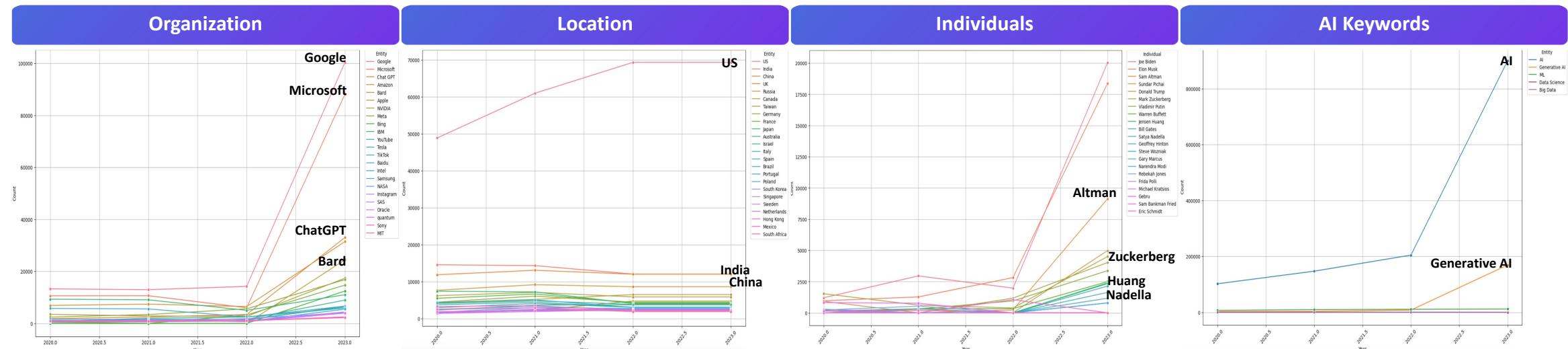
AI and Security



Combining BERTopic with sentiment analysis, we gained clearer insights into major topics: Positive sentiments focused on AI's transformative impact on industries, while negative sentiments centered on ethical and moral concerns.

9. Entity Identification

Through Named Entity Recognition (NER with spaCy), we effectively pinpointed crucial entities pertaining to *organizations (ORG)*, *geographical locations (GPE)*, and *individuals (PERSON)*. Each entity type reflects a growing interest in generative AI, especially ChatGPT.



- Google, Bard, Microsoft, ChatGPT, Meta, NVIDIA, IBM, Amazon, YouTube, Apple, Tesla
- The frequency of organizations named about generative AI has significantly increased since 2022.

- US, India, China, UK, Russia, Canada, Taiwan, Germany, France, Japan
- The United States leads by a significant margin, followed by India and China. Notably, both ChatGPT and Bard, leading Generative AI models, were developed in the U.S.

- Elon Musk (Tesla CEO), Sam Altman (OpenAI CEO), Mark Zuckerberg (Meta CEO), Jesen Huang (NVIDIA CEO), Satya Nadella (Microsoft CEO)
- The increased prominence of Sam Altman of OpenAI and Satya Nadella of Microsoft correlates with the rise of ChatGPT and generative AI.

- AI, Generative AI, ML, Data Science, Big Data*
- The emergence of AI in discussions has spiked following the introduction of generative AI.

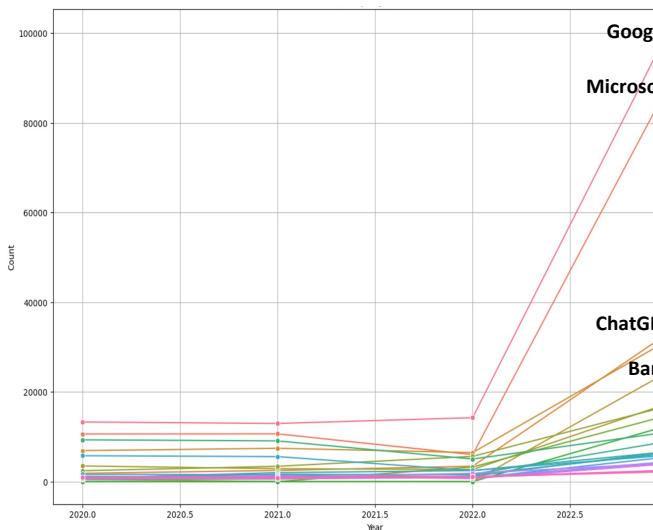
*AI-related terms, even those incorrectly categorized as other entities in the named entity recognition process, were collected to analyze their trend.

9. Entity Identification

Utilizing NER, we identified key companies, academic institutions, and government entities driving AI advancements that reshape employment.

Organizations Accelerating AI Development that Impacts Employment

Organization



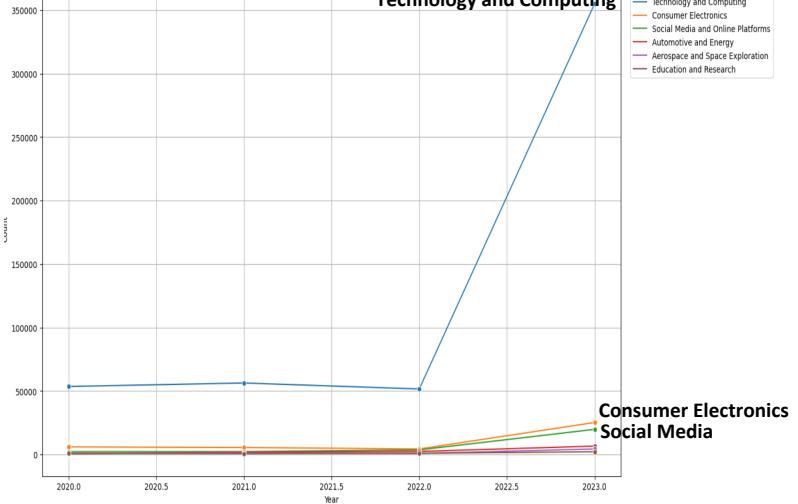
Organizations*

*Mapped ChatGPT as Microsoft, Bard as Google

Industry*

*Industries Identified Based on Recognized Organization Names

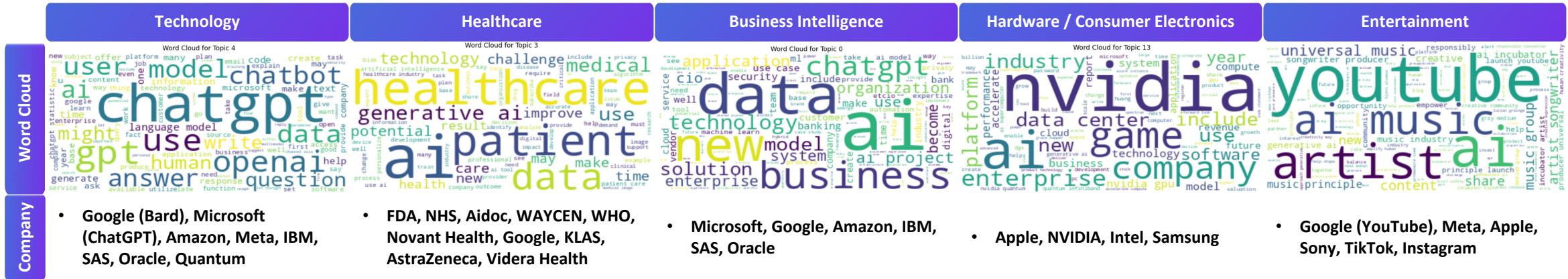
Technology and Computing



- 1 Microsoft and Google, with their innovations ChatGPT and Bard respectively, are poised to be pivotal in driving change within **the generative AI landscape**.
- 2 Leading the **technological charge** are **companies** like NVIDIA, IBM, SAS, Oracle, and Quantum, whose contributions are shaping the transformation.
- 3 NASA and MIT are enhancing generative AI progress through their respective **governmental and academic roles**, underscoring a multifaceted industry impact.

9. Entity Identification with Topic Modeling

Success stories: What types of companies are planning to invest in AI?

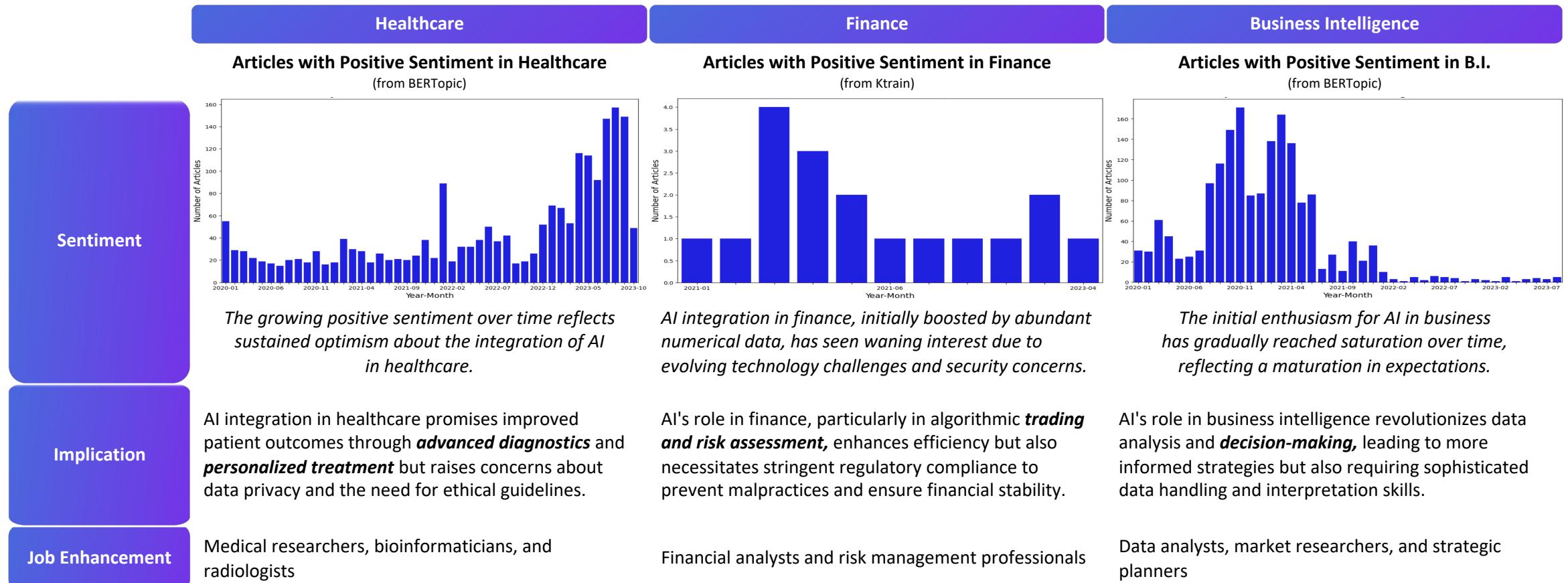


Failures: What types of applications cannot currently be transformed by AI, based on today's state of technology?



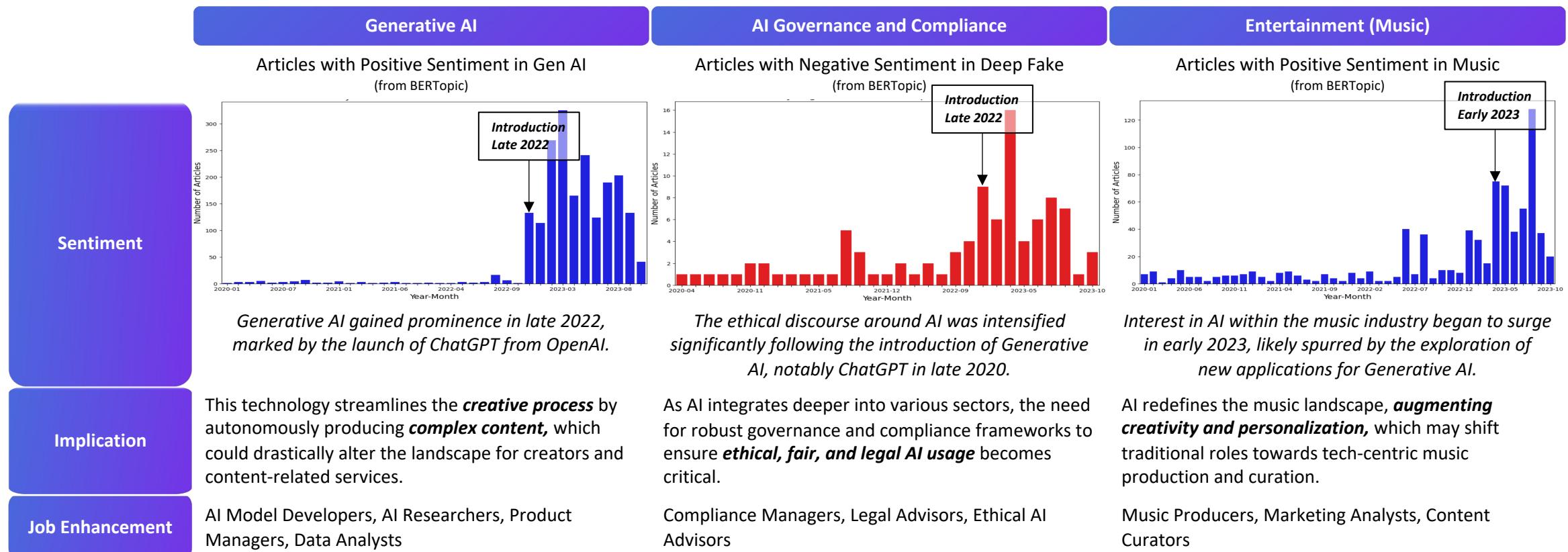
10. Targeted Sentiment Identification

Leveraging topics identified by BERTopic / Ktrain and associated articles, we pinpointed three prime sectors for AI integration: *Healthcare, Finance, and Business Intelligence*.



10. Targeted Sentiment Identification

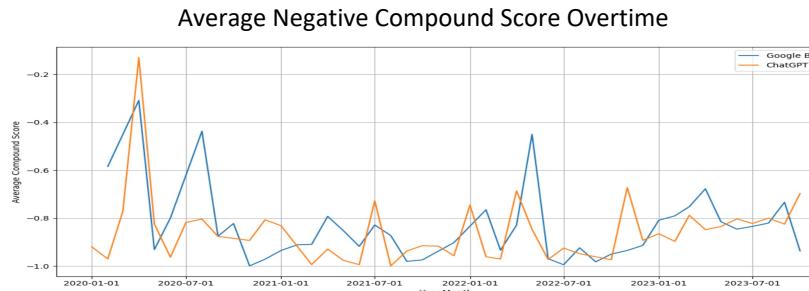
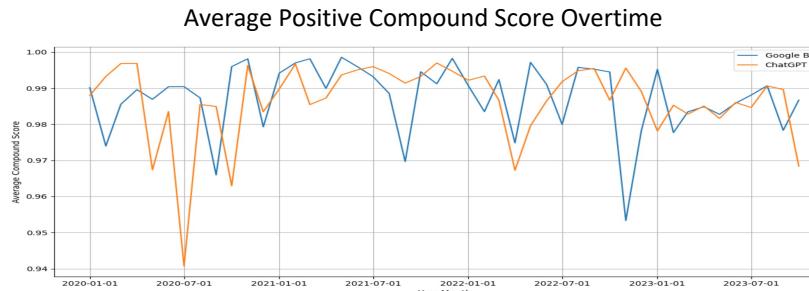
We identified key AI integration technologies: *Generative AI, AI Governance and Compliance, and Entertainment*. AI Governance and Compliance, uniquely analyzed with negative sentiment like cheating, bias, and manipulation, is anticipated to expand as it tackles these specific challenges.



10. Targeted Sentiment Identification

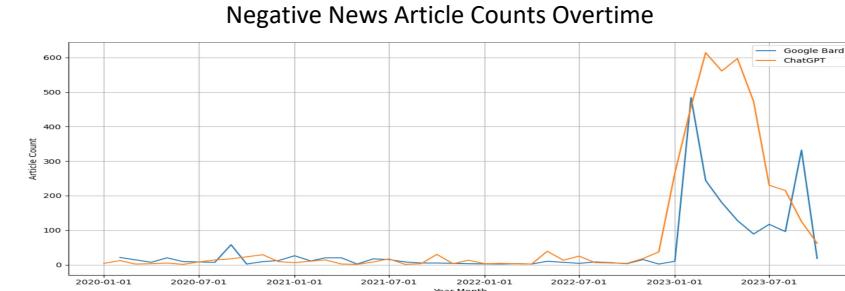
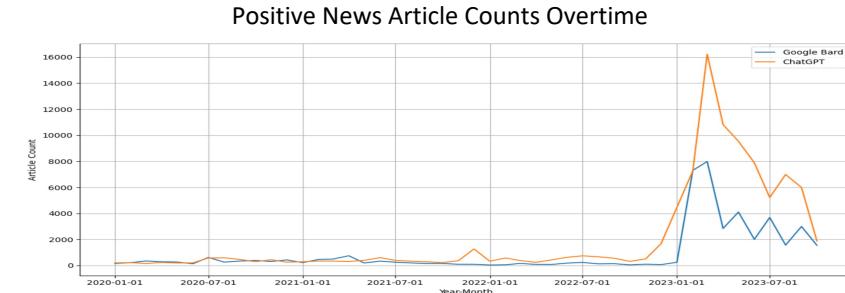
In our BERTopic analysis, we recognized generative AI as pivotal. Integrating NER, we pinpointed *ChatGPT* and *Bard* as key players in this domain. Leveraging these key entities, we advanced our analysis to track sentiment trends over time, focusing specifically on these industry leaders.

Compound Score Comparison by Sentiment



- Despite some outliers, the compound score trends for ChatGPT and Bard generally move in tandem.
- Both show similar variations in the strong intensity of sentiments close to 1 or -1 over time.

News Article Counts Comparison by Sentiment



- ChatGPT has garnered more media attention, both positive and negative, since its launch in November 2022 than Bard.
- Of late, negative coverage of Bard has risen as ChatGPT's has gradually fallen.

The average compound score shows public sentiment towards ChatGPT and Bard is similar.

However, ChatGPT has a more significant media presence, both positive and negative, suggesting higher potential to lead in generative AI. Meanwhile, Bard may face reputation challenges, evidenced by increasing negative media coverage.

Thank You