AI Industry Impact Analysis: Trends and Strategic Insights

ADSP 32018: Natural Language Processing and Cognitive Computing

Author: Kaiyan Li

May 25, 2025

Agenda

01	Executive Summary	Summarizes the main insights from the analysis, identifies industries most affected by AI, and provides actionable recommendations for successful adoption.
02	Data Exploration & PreProcessing	Details the sourcing, filtering, and cleaning of ~200K AI-related news articles to create a reliable dataset for downstream analysis.
03	Modeling	Covers industry classification, topic modeling, and entity recognition to extract key topics, technologies, and sector-level insights from the AI news corpus.
04	Sentiment Analysis	Presents sentiment patterns across topics and entities, and visualizes trends over time to assess industry attitudes toward AI.
05	Conclusion & Strategic Outlook	Wraps up findings and highlights future opportunities, risks, and considerations for leaders navigating AI-driven transformation.

Executive Summary

Summarizes the main insights from the analysis, identifies industries most affected by AI, and provides actionable recommendations for successful adoption.

01

CENTRAL QUESTION

Which industries will be most impacted by AI, and how can adoption be made successful?

02

KEY INSIGHTS

AI's impact is uneven: some sectors are surging ahead while others lag due to ethical, regulatory, or infrastructure constraints.

03

PROPOSED SOLUTION

Use NLP on 200K+ articles to identify exposure, sentiment, and actionable strategies by sector.

Findings

Industry Exposure

Technology, Media, Healthcare, Education, and Communications are most impacted based on article frequency and topic alignment.

Findings

Sentiment Landscape

Neutral-to-cautious tone dominates public discourse, with spikes driven by major releases (e.g., ChatGPT).

Actions

Automate High-Frequency Tasks

In high-exposure sectors like Technology and Communication Services, deploy AI for customer support, scheduling, and routine workflows.

Actions

Boost Productivity with Copilots

In Education and Media, integrate AI copilots to support high-skill tasks like tutoring, writing, and content analysis.

Actions

Tailor AI Policies by Sector

In regulated sectors like Healthcare and Government, prioritize explainable AI and domain-specific risk frameworks.

Actions

Upskill Cross-Functional Teams

Topic trends show broad AI interest—successful adoption requires AI fluency across both technical and non-technical roles.

Executive Summary

Central Question

- What industries are most impacted by AI, and how can organizations leverage it effectively?

Most Affected Industries

- Technology – Rapid integration of foundation models, copilots, and infrastructure AI
- Media & Entertainment – Disruption from generative content tools; content authenticity concerns
- Healthcare – Growth in diagnostics and decision support, tempered by ethics/regulation
- Education – Interest in tutoring and personalization, but uneven adoption
- Communication Services – Enhanced service delivery via AI-driven automation

Content-Based Filtering

- After normalization, I applied rule-based filtering to remove irrelevant articles — specifically those not mentioning core AI-related keywords (e.g., “AI,” “artificial intelligence,” “machine learning,” etc.).
- This ensured that the dataset remained focused on genuinely AI-relevant content and reduced it slightly further to **119,354** articles.

Data Exploration & Preprocessing

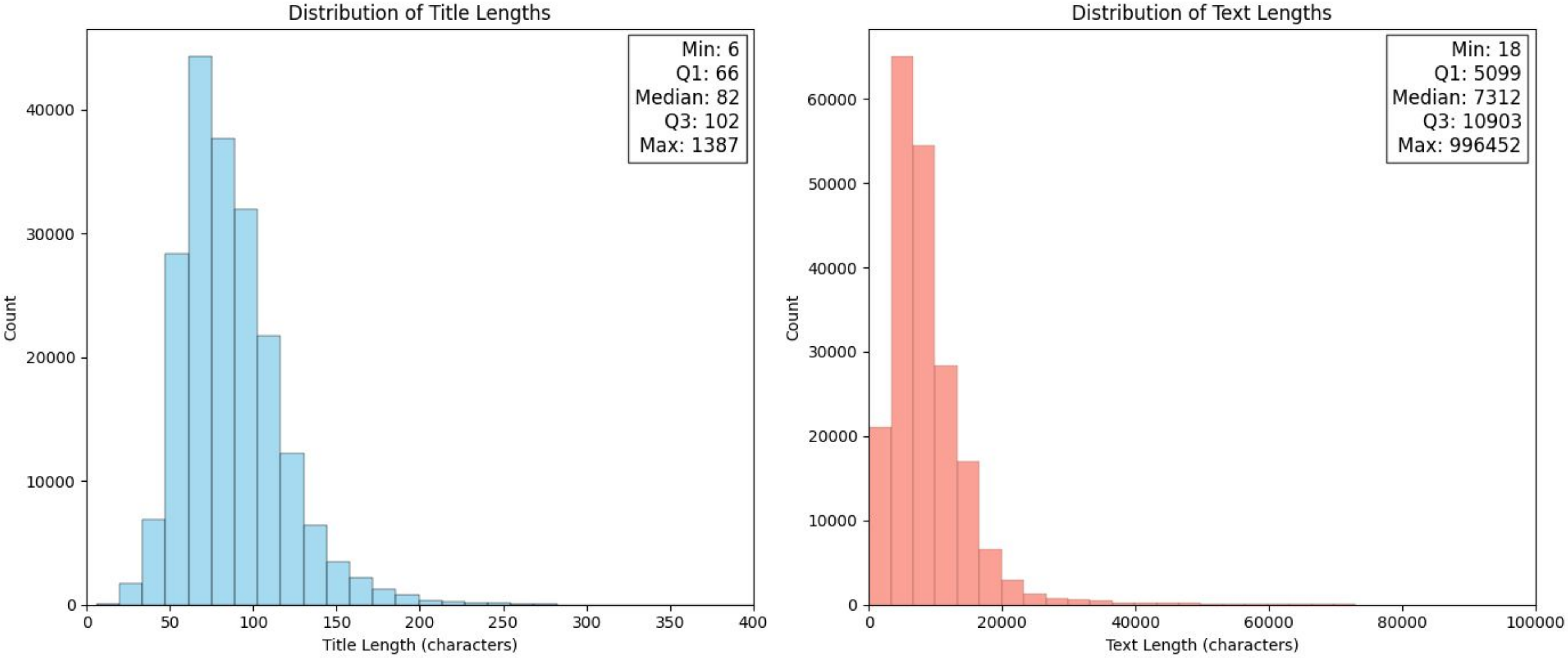
Details the sourcing, filtering, and cleaning of ~200K AI-related news articles to create a reliable dataset for downstream analysis.

Data Exploration

Variable	Description	Example
url	Source link to the article (some URLs are no longer accessible, e.g., 404 errors)	https://allafrica.com/stories/202504250184.html
date	Article publish date (range: 2022-01-01 to 2025-04-28)	2025-04-25
language	Language of the article (only English retained; others were machine-translated or dropped)	en
title	Title of the article (length ranges from 6 to 1,387 characters due to scraping artifacts)	Africa: AI Policies in Africa - Lessons From Ghana and Rwanda - allAfrica.com
text	Full raw article content (length varies widely: 18 to 996,452 characters)	\nAfrica: AI Policies in Africa - Lessons From Ghana and Rwanda - allAfrica.com\nAllAfrica\nEnglish (current)\nEn Français\n\nToggle navigation\nMy Account\nToggle navigation\n\nAllAfrica\nMy Acco...

This dataset contains 200,083 records. Some titles failed to extract cleanly due to scraping issues (e.g., captured HTML metadata or page scripts), resulting in unusually long entries. Additionally, a subset of articles had extremely short or non-informative content, often due to incompatible webpage formats or missing body text.

Data Exploration



This plot shows the length distributions of article titles and texts. Titles are mostly concise, but text lengths vary widely, with some documents exceeding 900,000 characters due to scraping artifacts. I observed that many texts embed titles, and extremely short or long entries often lack usable content. Moving forward, I'll focus on the text field and apply normalization and length-based filtering to improve data quality.

Data PreProcessing

Length-Based Filtering

- I removed articles with text shorter than the **40th percentile** (6401 characters ≈ 1200 words), as they tend to lack meaningful content.
- I also excluded articles longer than **50,000 characters** (710 total), which are rare, resource-intensive, and unlikely to impact final results.
- This step reduced the dataset to **119,357** records, and was applied before text normalization to improve processing efficiency.

Text Normalization

- I applied a rule-based NLP pipeline to clean and standardize the text.
- This included:
    - Converting text to lowercase
    - Removing special characters
    - Eliminating common English stopwords
    - Applying lemmatization using WordNet to reduce words to their base forms (e.g., "running" → "run")
- This was done using an **NLTK-based** pipeline, which proved far more efficient than spaCy's small model for this dataset.
- <u>Cleaned text example</u>: *"africa ai policy africa lesson ghana rwanda allafrica com allafrica english current en fran ai toggle navigation account toggle navigation allafrica account english current en fran ai africa ai po…"*

Content-Based Filtering

- After normalization, I applied rule-based filtering to remove irrelevant articles — specifically those <u>not mentioning core AI-related keywords</u> (e.g., "AI," "artificial intelligence," "machine learning," etc.).
- This ensured that the dataset remained focused on genuinely AI-relevant content and reduced it slightly further to **119,354** articles.

Modeling

Covers industry classification, topic modeling, and entity recognition to extract key topics, technologies, and sector-level insights from the AI news corpus.

Modeling: Topic Detection (Zero-Shot)

Zero-Shot Classification

- To identify the industry focus of each article, I used zero-shot classification, which enables labeling without a pre-labeled training set, making it well-suited for my large, unlabeled corpus.
- Unlike traditional classifiers (e.g., logistic regression, SVM, or fine-tuned transformers), zero-shot models can generalize to **custom label sets** based on semantic understanding, allowing me to use a tailored list of industry categories without training a separate model.

DeBerta

- DeBERTa (*Decoding-enhanced BERT with disentangled attention*) is a transformer model developed by Microsoft that improves upon BERT by separating content and positional encoding, enhancing the model's ability to capture meaning in text.
- I selected the model MoritzLaurer/DeBERTa-v3-base-mnli-fever-anli for its strong performance on natural language inference tasks and its balance of speed, accuracy, and cross-domain generalization.
- Since classifying full texts would be computationally expensive, I used the **first 100 words** of each cleaned article for classification.
- I developed a set of **21** detailed, domain-specific industry labels, including:
  - Technology, Healthcare, Education, Media & Entertainment, Retail & Commerce, etc.
  - Each label was paired with a short descriptive prompt (e.g., "Technology: software, AI, machine learning, digital tools, IT") to improve classification accuracy.
  - 👉 The full list of classification labels is provided on the next slide. The top 5 industry labels, based on **article counts**, are enclosed in the red dashed box to highlight the sectors most prominently represented in the dataset.
- This step allowed me to assign each article to a specific industry sector, laying the foundation for downstream analysis of AI trends and sentiment within each domain.

Modeling: Topic Detection

21 Detailed Domain-Specific Industry Labels (from most to least prevalent)

1. Media & Entertainment: film, TV, publishing, content creation, social media
2. Technology: software, AI, machine learning, digital tools, IT
3. Healthcare: hospital, medical, patient care, treatment, clinical services
4. Communication Services: telecom, internet, digital communication, mobile networks
5. Education: schools, teaching, learning, academia, training
6. Transportation & Logistics: shipping, delivery, transit, supply chain, freight
7. Business & Finance: banking, stock, market, investment, corporate finance
8. Retail & Commerce: shopping, sales, e-commerce, consumer goods, merchandising
9. Hospitality & Food Services: restaurant, catering, hotel, food service, tourism
10. Government & Politics: public policy, legislation, governance, political systems
11. Agriculture & Environment: farming, sustainability, climate change, natural resources
12. Legal Services: law, courts, attorneys, legal advice, contracts
13. Administrative Services: clerical, office management, scheduling, administrative tasks
14. Materials: mining, raw materials, metals, manufacturing inputs
15. Utilities: electricity, water, energy, infrastructure, power grid
16. Real Estate: housing, property, real estate market, construction, leasing
17. Security & Protection: cybersecurity, surveillance, defense, law enforcement
18. Engineering & Sciences: engineering, research, innovation, R&D, technology development
19. Public Services: emergency services, public health, community programs, social services
20. Operations & Maintenance: logistics, facility management, repair, technical operations
21. Construction & Trades: building, carpentry, plumbing, electrical work, renovation

Modeling: Topic Detection

Top 5 Industry Labels Identified

Using zero-shot classification, I assigned each article to the most relevant industry label based on its content. Below are the top 5 industries identified and my interpretation of why they appear most frequently:

1. **Media & Entertainment**
   Many articles include words like “news,” “coverage,” and “headline,” which led to classification under media. This makes sense, as a large portion of the dataset comes from journalistic or commentary sources discussing AI: often general or outlier cases not tied to a specific domain.
2. **Technology**
   This is expected, given the high volume of technical articles discussing AI models, tools, and innovations. Keywords such as “data,” “model,” “algorithm,” “ChatGPT,” and “NVIDIA” frequently appear and drive classification into this category.
3. **Healthcare**
   Healthcare is a major focus area for AI, with discussions around diagnostics, treatment, and personalized medicine. Frequent terms like “patient,” “diagnosis,” and “medical” support the strong representation of this industry.
4. **Communication Services**
   This category includes telecom and internet infrastructure. Some articles may be misclassified here due to overlapping terms with technology: for example, “network,” “connectivity,” or “signal.”
5. **Education**
   Articles containing terms like “training,” “learning,” “student,” were often classified under education. However, many of these terms are also common in technical discussions about machine learning, which suggests this category may include more misclassified cases.

Modeling: Topic Detection

Cleaned Text	Label	Evaluation
newly published study demonstrates clinical effectiveness machine learning used achieve remission patient peanut allergy ein presswire different better work testimonial contact ein presswire news ...	Healthcare: hospital, medical, patient care, treatment, clinical services	✅ Correctly labeled: clinical, patient, treatment
sonata software integrates ai solution microsoft azure ai hindu businessline company market portfolio opinion budget free trial account subscribelogin menucloseepapersearchsearch company market co...	Business & Finance: banking, stock, market, investment, corporate finance	✅ Appropriate label: market, portfolio
gpt openai shifted direction companiesie supported optimal experience visit site another browser skip contentnbc news logopoliticsu newsbusinessworldculture trendstechhealthnbc news tiplineshare s...	Media & Entertainment: film, TV, publishing, content creation, social media	✅ Reasonably labeled: mixture of tech and media
inspur information ai server fully support newly announced nvidia h tensor core gpu business net skip content business net news professional primary menu business net business net search home marc...	Technology: software, AI, machine learning, digital tools, IT	✅ Clearly accurate: Nvidia, tensor core, AI server
global telehealth telemedicine market opportunity use blockchain ai analytics virtual assistant news wfmz com skip main content permission edit article edit close sign log dashboard logout account...	Education: schools, teaching, learning, academia, training	⚠️ Likely misclassified: little evidence of education, more relevant to healthcare

Overall, the model's performance is **acceptable** given the absence of labeled training data. Most classifications align well with domain-specific keywords, and even potentially misclassified examples are reasonable based on the presence of relevant terms (e.g., "assistant").
This supports the model's utility for high-level industry tagging in large, unstructured corpora.

Modeling: Topic Detection (LDA)

LDA vs BERTopic

- LDA (Latent Dirichlet Allocation): A probabilistic topic model that treats each document as a mixture of topics, and each topic as a distribution over words. Ideal for uncovering interpretable word clusters across large corpora.
- BERTopic: A more advanced model leveraging transformer embeddings and clustering. While powerful for semantic separation, its output was less visually intuitive for keyword exploration in my use case.

Keyword Exploration with LDA

- I used LDA on the cleaned text to explore prominent keywords and themes within the corpus.
- Although LDA topics don't directly align with industry labels, I approximated the **industry composition within each topic** by calculating proportions: e.g., Topic 3 = 70% Healthcare, 25% Technology.
- LDA's keyword-driven structure made it a better fit than BERTopic for supporting visual interpretation and downstream trend analysis.

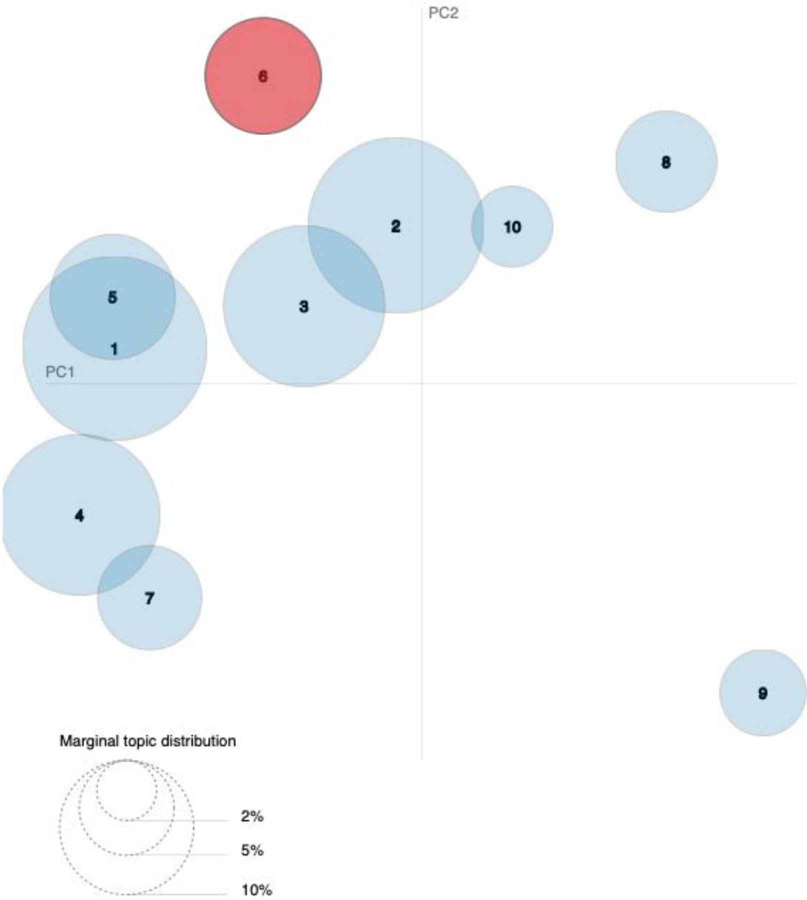
LDA Topic-To-Industry Mapping

- Topic 1: Media & Entertainment (46%) | Technology (31%)
- Topic 2: Media & Entertainment (41%) | Technology (24%)
- Topic 3: Media & Entertainment (33%) | Technology (29%)
- Topic 4: Media & Entertainment (73%) | Healthcare (9%)
- Topic 5: Media & Entertainment (49%) | Technology (28%)
- Topic 6: Technology (24%) | Business & Finance (16%)
- Topic 7: Media & Entertainment (44%) | Healthcare (18%)
- Topic 8: Transportation & Logistics (81%) | Communication Services (14%)
- Topic 9: Education (34%) | Healthcare (20%)
- Topic 10: Media & Entertainment (62%) | Technology (12%)

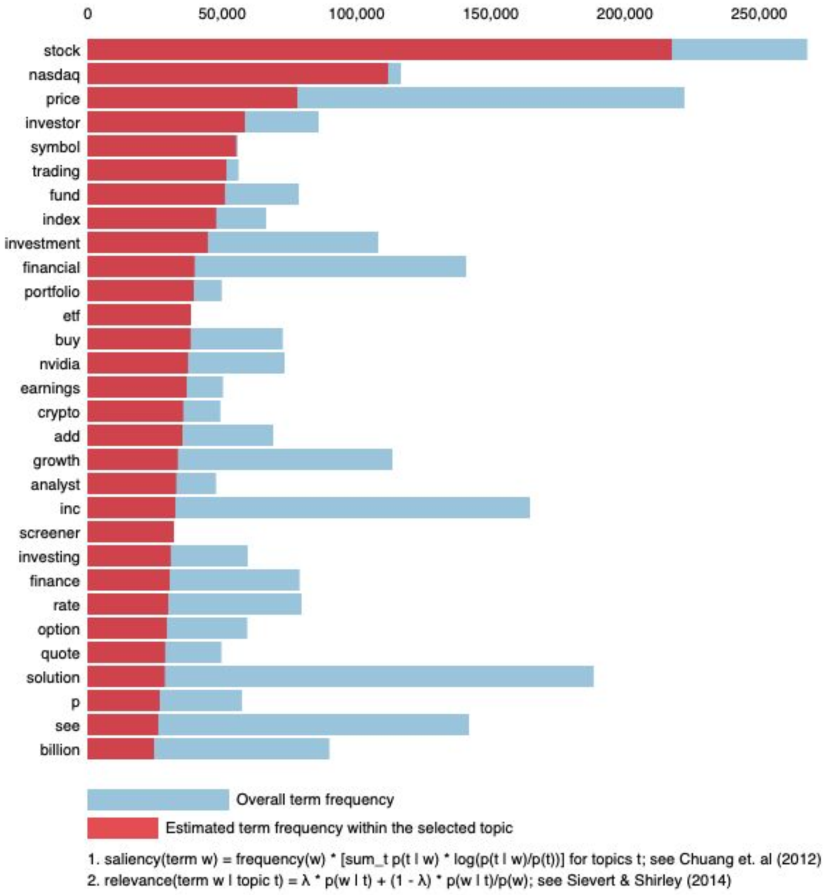
Note: Topics are ordered by prevalence, with Topic 1 being the most dominant and Topic 10 the least, based on the proportion of documents assigned to each.

Modeling: Topic Detection

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 6 (7.5% of tokens)



I'm particularly curious about finance and technology, so I further explored Topic 6 here. If you're interested, you can interact with the **full visualization** here.

This plot shows that Topic 6 is strongly associated with Technology (24%) and Business & Finance (16%).

On the left, the **Intertopic Distance Map** reveals that Topic 6 is loosely connected to nearby topics, which is reasonable given its overlap with industries like Media and Technology, as shown in the industry mapping on the previous slide. The **size** of each circle indicates the relative number of documents in each topic: the larger the circle, the more prevalent the topic.

On the right, the **Top 30 Relevant Terms** include stock, nasdaq, nvidia, and investment, clearly highlighting the topic's alignment with finance and tech themes. The red bars represent term frequency within Topic 6, while blue bars reflect their overall corpus frequency.

Modeling: Entity Extraction (NER)

Approach Overview

- I explored multiple NER strategies to extract entities from AI-related articles, with a focus on identifying technology names, organizations, and key individuals.

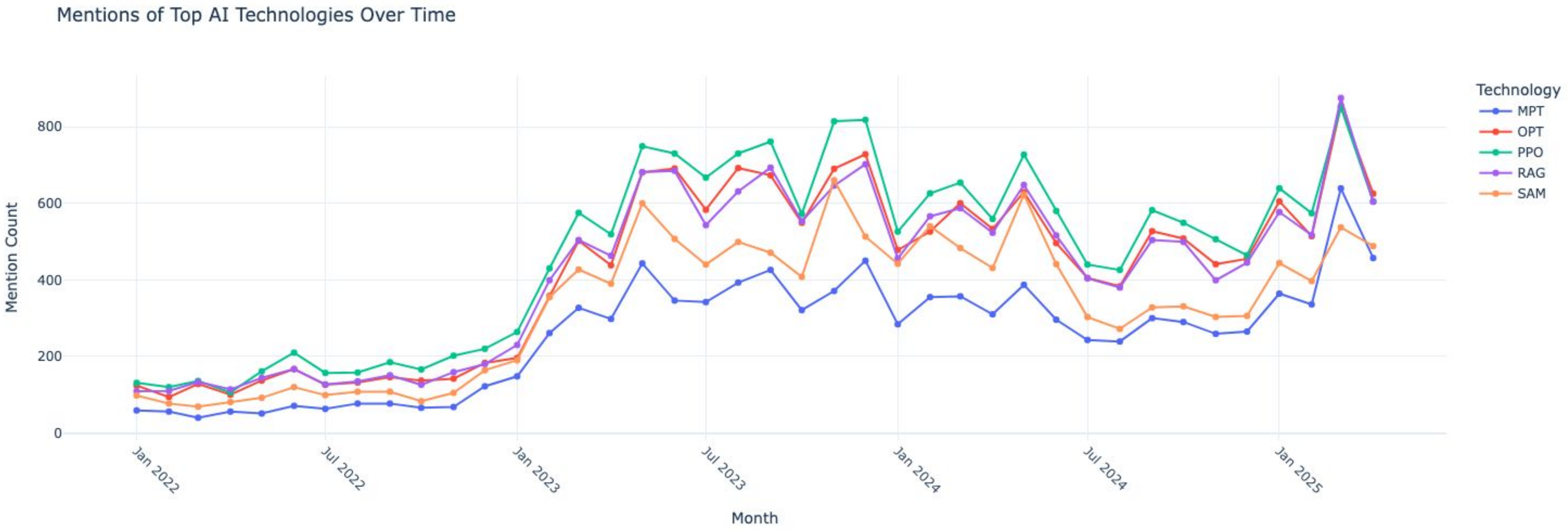
Other Methods ❌

- Standard NER models (e.g., spaCy, Flair) cannot directly recognize modern AI models or tech-specific entities (e.g., GPT-4, Midjourney).
- I attempted to train a model-based extractor using a custom BIO-tagged dataset (~1000 GPT-generated samples with both synthetic text and annotated labels), but performance was poor due to:
    - Small training size
    - Poor generalization
    - Lack of high-quality, domain-specific annotations
- I also considered using GPT-4 via API for entity extraction, but it was too costly to apply across the entire article corpus.

Keyword-Based Extractor ✅

- Used `en_core_web_sm` from spaCy for **original text** normalization (to preserve model names like GPT-4)
- Designed custom keyword lists to extract:
    - Technologies (e.g., GPT-4o, Midjourney, AutoGPT): ~1000 GPT-generated relevant keywords
    - Organizations (e.g., OpenAI, Google, Anthropic)
    - People (e.g., Sam Altman, Elon Musk)
- Due to high processing time and project focus, I limited extraction to the **Technology industry** only, resulting in **25,236** records.

Modeling: Entity Extraction



“

The top plot shows the monthly mention trends for the five most frequently referenced AI technologies (e.g., MPT, OPT, PPO, RAG, SAM). Mentions surge around **early 2023**, corresponding to increased public discourse and product announcements. Trends remain relatively stable with periodic spikes, reflecting ongoing innovation cycles.

The bottom plot displays mentions of top geopolitical entities (GPE), such as the US, India, and China. The **US dominates** in coverage volume throughout the timeline, while other countries like India and China follow at a steady rate. The increase around **early 2023** mirrors the broader AI attention spike seen in the technology mentions. This pattern is reasonable, as all articles are in English and appear to be scraped primarily from English-language news websites, which tend to emphasize US-centric content.

Modeling: Entity Extraction (NER)

Model Results

- Top 10 AI Technologies: PPO, OPT, RAG, SAM, MPT, ChatGPT, Flow, Ray, GPT-4, Capilot.
    - PPO: Reinforcement learning algorithm used in RLHF (e.g., for ChatGPT)
    - OPT: Meta’s open-source LLM series
    - RAG: Combines retrieval + generation for question answering
    - SAM: Vision foundation model for segmenting anything
    - MPT: Efficient LLM developed by MosaicML
    - ChatGPT / GPT-4: Conversational AI models by OpenAI
    - Copilot: GitHub's AI coding assistant
    - Ray / Flow: Infrastructure tools for AI scaling and orchestration
- Top 10 Locations (GPE): US, India, China, UK, **New York**, Australia, Canada, Japan, France, Germany.
- Top 5 Organizations: Apple, Google, Meta, Microsoft, OpenAI.
- Top 2 Persons: Elon Musk & Sam Altman.

The only city in the top 10, likely due to media company presence and frequent news coverage.

Pros and Cons

- ✅ High accuracy in recognizing geopolitical entities (GPE), which aligned well with real-world trends.
- ✅ Flexible customization using domain-specific keywords for technology entities.
- ❌ Limited person recognition due to insufficient keyword coverage.
- ❌ Some organizations (e.g., OpenAI, Generative AI, Twitter) were misclassified as persons.
    - Likely due to sentence structures where these orgs appear before human-behavior verbs (e.g., OpenAI said, Twitter responded), matching typical syntactic patterns of people in NER.
- ❌ Expanding keyword coverage significantly increases processing time for the extractor, making it more time-consuming at scale.

Sentiment Analysis

Presents sentiment patterns across topics and entities, and visualizes trends over time to assess industry attitudes toward AI.

Sentiment Analysis

Transformer-Based Models

I evaluated three transformer-based sentiment models before selecting the final one:

- ✅ cardiffnlp/twitter-roberta-base-sentiment **(Chosen)**
  - Fine-tuned on millions of English tweets and designed to detect sentiment in short, social-style text.
  - Captures nuanced sentiment, supports positive / neutral / negative labels, and generalizes well across diverse article types.
  - Given its contextual accuracy, sentiment granularity, and strong generalization to AI-related news content, I selected this model as the most suitable for my analysis.
- 🔎 ProsusAI/finbert
  - Fine-tuned on financial documents and earnings reports, making it ideal for finance-related sentiment classification.
  - Designed for financial text; results were similar to CardiffNLP but slightly biased toward finance-specific language.
- 🔎 siebert/sentiment-roberta-large-english
  - Trained on a large English corpus from various domains, but optimized for binary sentiment tasks.
  - Only provides positive / negative labels; lacks granularity and underperforms on mixed or neutral content.

Traditional Models

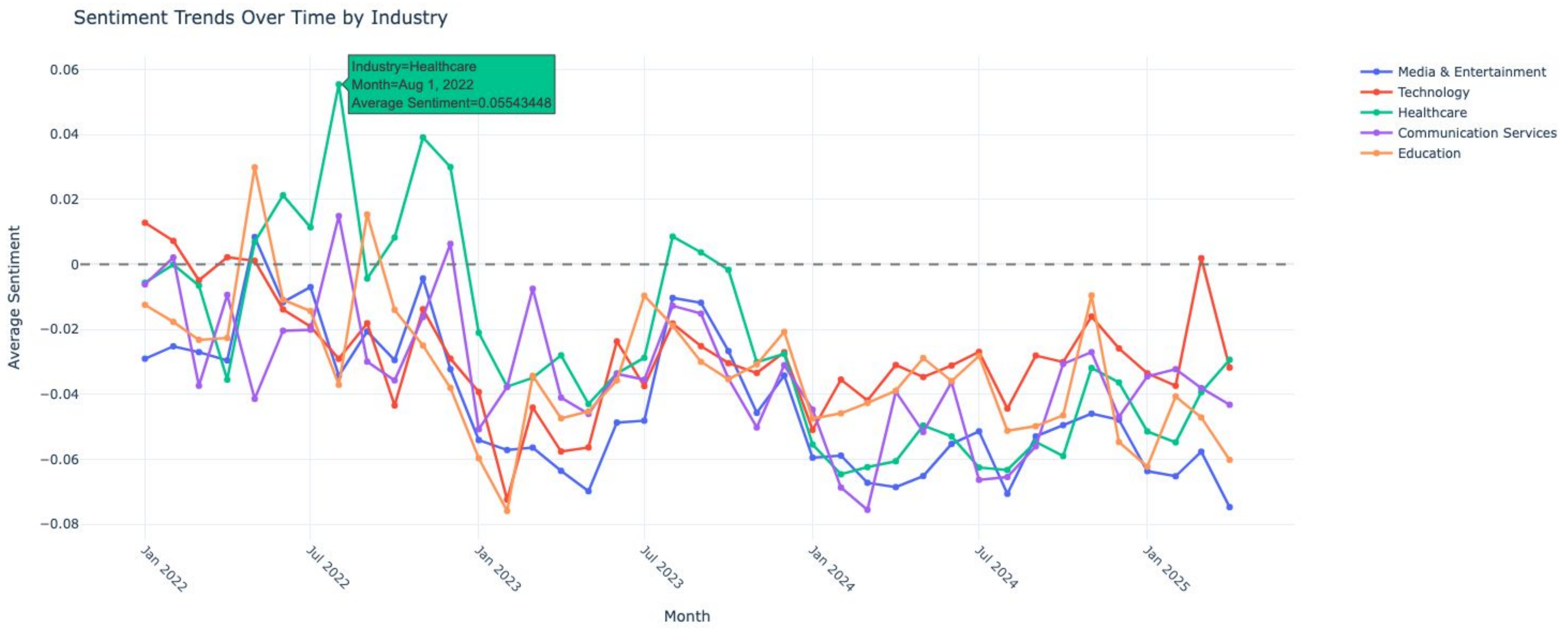
- Lexicon-based models like VADER or traditional Naive Bayes classifiers tend to:
  - Over-predict positive sentiment
  - Struggle with technical or formal language
  - Miss subtle cues (e.g., sarcasm, negation, conditional tone)

Sentiment Analysis

Cleaned Text	Context	Cleaned Text	Context
microsoft keep betting ai window arm pc get even better xda menu sign close news deal submenu news deal review submenu laptop cpu graphic card ssds product award best guide submenu best laptop best chromebooks best cpu best graphic card tutorial submenu window macos linux chromeos thread topic submenu acer apple dell hp lenovo microsoft razer sign newsletter trending copilot snapdragon x series window update day forum close microsoft keep betting ai window arm pc get even better window forecast …	Success 0.6072	… always partner pytorch tensorflow jax others certainly continued growing seeing service group seeing nvidia also partner optimize latest platform offer generative ai seeing opportunity less moving workload even still cpu using little bit ai simplified ai using much intelligent larger model improve quality service better interaction device even talking hockey puck kitchen counter cloud also partner optimize model …	Success 0.7214
artificial intelligence call center market expected grow exponentially ibm corporation oracle corporation amazon web service artificial solution international ab thenelsonpost skip content thenelsonpost news home thenelsonpost main menu artificial intelligence call center market expected grow exponentially ibm corporation oracle corporation amazon web service artificial solution international ab leave comment stats n data january new york november comprehensive market analysis report …	Neutral -0.0462	… advance medical imaging technology radiologist healthcare provider expected diagnose patient quickly accurately annalise ai fuse highest quality imaging data best computer science produce comprehensive ai clinical decision support solution empowering clinician make accurate faster decision patient first approach proudly clinician led come deep understanding challenge faced medical imaging ai solution provide clinician second set eye allowing detect confidence drive better health outcome patient annalise …	Neutral 0.1663
shibarium partner bad idea ai bad score new listing price jump ad ad yuri molchan bad idea ai token added yet another crypto exchange pushing bad price significantly news thu cover image via www freepik com read u today google news bad idea ai project native token bad scored new listing time bad support added rubic crypto exchange according tweet offer cross chain swap user buy bad binance coin bnb well use token blockchains ad ad bad new listing spur price …	Failure -0.6342	… complaint citing instance chatbot hallucinating outright giving wrong response wale concern chatbot accuracy new unique july study researcher revealed demonstrated chatgpt getting dumber amid claim openai verge bankruptcy might explain increasing number complaint ai generated article inaccurate information come relatively insensitive recent one ai generated poll featured microsoft start msn alongside story guardian highlighting unfortunate …	Failure -0.5190

The sentiment results generally reflect the tone of the articles **well**. To be more specific, highly positive texts emphasize innovation and collaboration, while negative ones highlight issues like misinformation or failure. Despite occasional noise or repetition in the raw text, the model is able to capture the overall emotional direction effectively.

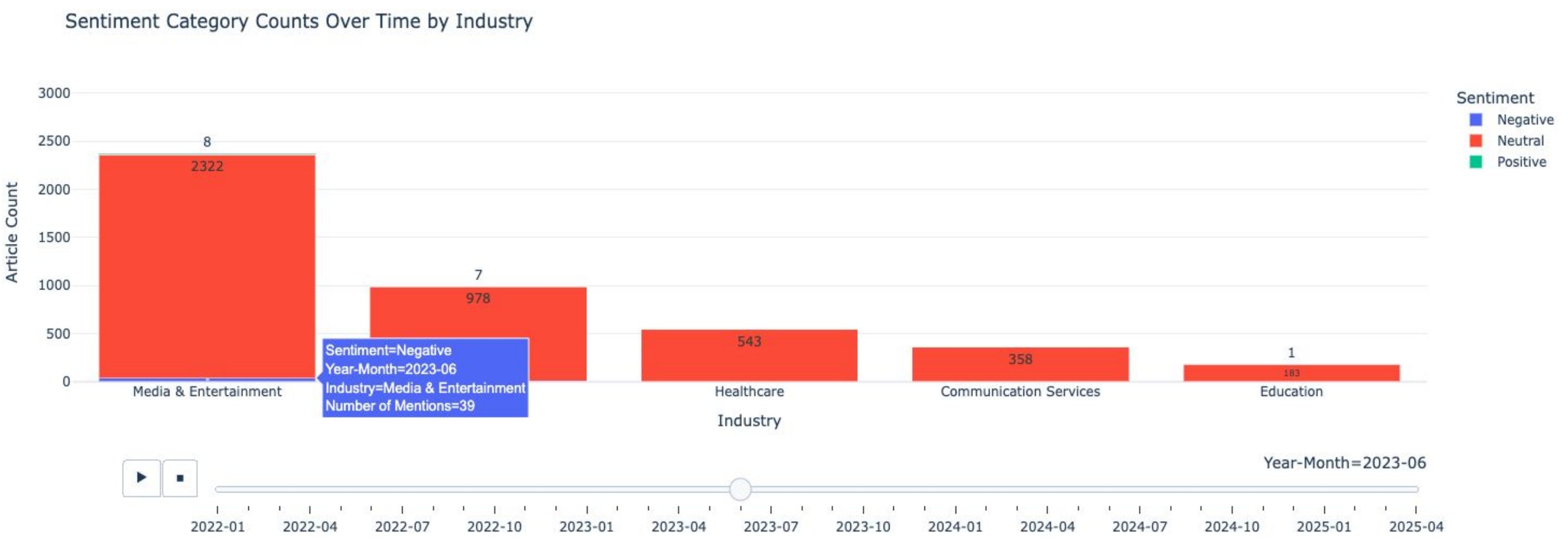
Sentiment Analysis: Topic-Level

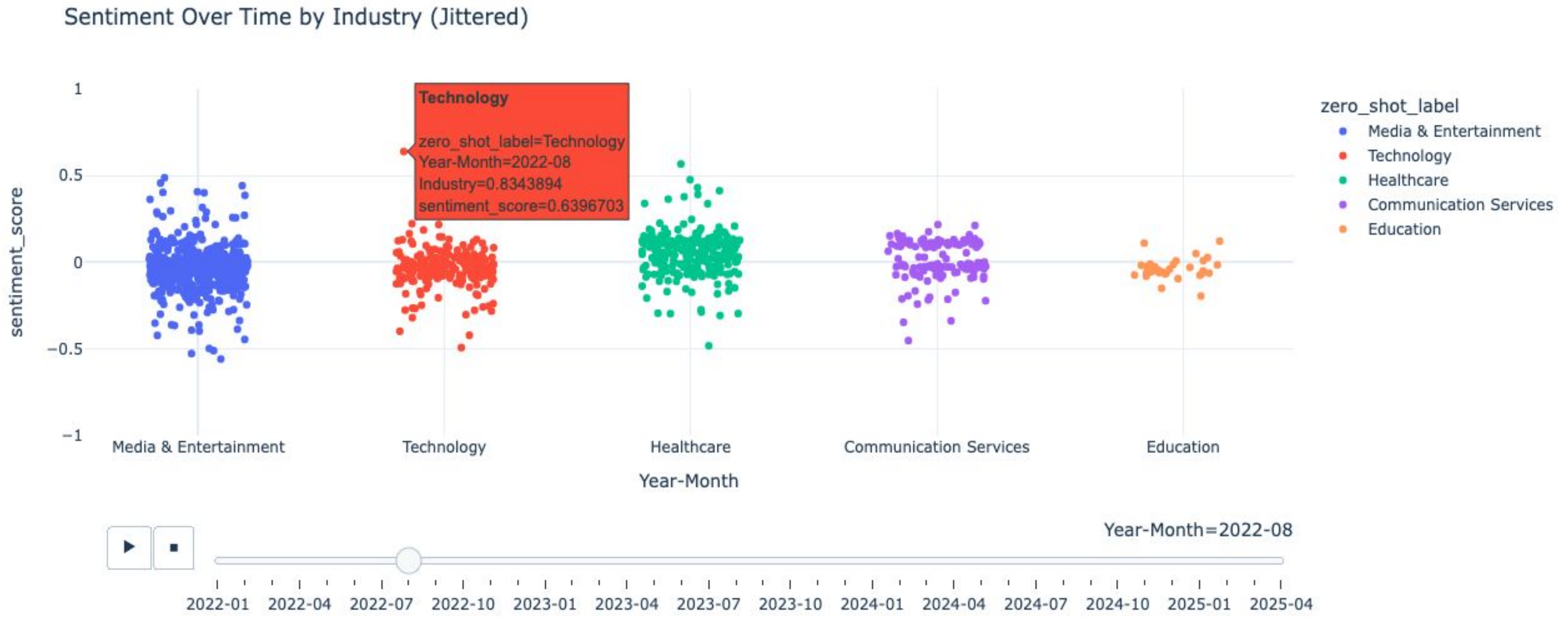


This interactive line chart shows sentiment trends over time across the top 5 industries: Media & Entertainment, Technology, Healthcare, Communication Services, and Education. Overall, sentiment fluctuates **around neutral**, with most scores between -0.08 and +0.06—indicating a measured or cautious tone rather than strong positivity or negativity.

A notable increase in article volume appears around **early 2023**, coinciding with AI's surge in public attention and the release of tools like ChatGPT and Stable Diffusion. This influx brought broader coverage and a wider range of opinions—initially marked by optimism, then gradually shifting toward more critical or skeptical tones as discussions around AI's limitations, ethical risks, and societal impact (e.g., employment, misinformation) intensified.

Spikes in sentiment may be driven by particularly strong articles, and differences across industries are also evident: Healthcare tends to show higher variability—likely reflecting both excitement about AI applications and concern over clinical risks—while Media & Entertainment often dips lower, potentially due to criticism of AI's impact on creators and content integrity.



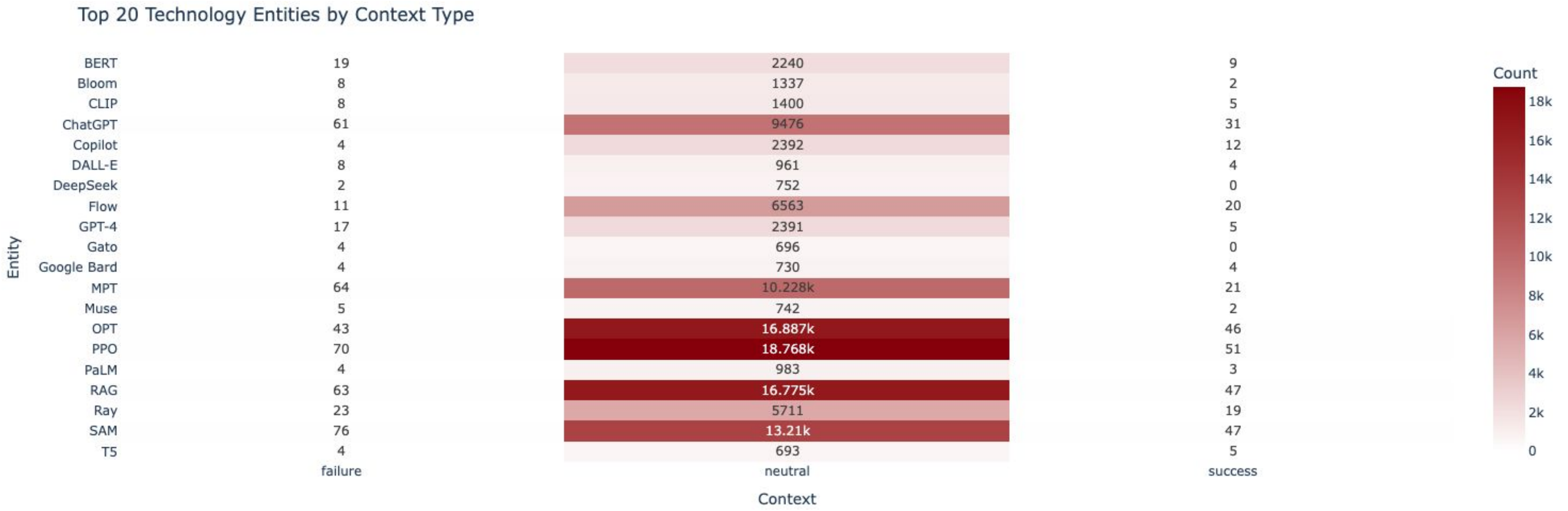


These two **interactive animated** plots visualize sentiment trends across the top 5 industries.
👉 Full animations (highly recommended): Top figure | Bottom figure
Tip: If the animation plays too quickly, pause it and then click play again to view transitions more gradually.

The top plot is a stacked bar chart displaying monthly article counts by sentiment label—stacked from bottom to top as Negative, Neutral, and Positive. As illustrated by a one-month snapshot, the sentiment distribution of the entire dataset is **highly skewed**: 99.24% of articles are labeled Neutral, with only 0.49% Negative and 0.27% Positive. When further looking at the continuous sentiment scores, 33.42% of articles have non-negative sentiment (≥ 0), while 66.58% are negative (< 0), suggesting a generally cautious or critical tone in public discussions of AI.

The bottom plot uses a jittered scatter layout to show individual article sentiment scores over time and across industries. This visualization helps reveal **outliers** — articles with exceptionally positive or negative tone — that are often obscured in aggregated charts. For instance, in August 2022, there's a high-scoring Technology article, and a sentiment spike in Healthcare that aligns with earlier observations. If you watch the animation over time, you'll notice a notable increase in article volume around early 2023, reflecting AI's growing public visibility and expanded cross-industry discourse.

Sentiment Analysis: Entity-Level



These two heatmaps show the **top 20 most mentioned entities** in the technology industry, broken down by context type: Success, Neutral, and Failure.

The top plot includes a mix of technologies, organizations, and locations, while the bottom plot focuses specifically on technology models. Most entities appear predominantly in neutral contexts, though models like PPO, OPT, and RAG also receive notable mentions in success contexts.

These plots highlight which technologies dominate AI-related discussions, and how they're generally perceived in public narratives.

“

These two line charts illustrate **sentiment trends** over time for the top mentioned AI technologies (top) and GPE entities (bottom).

The top plot shows that public sentiment toward **leading AI models** like MPT, OPT, PPO, RAG, and SAM generally fluctuates around neutral, with occasional dips into negativity. Despite minor differences, all technologies follow a broadly similar trend, indicating consistent public tone across models.

The bottom plot focuses on **geopolitical entities** such as China, India, the U.S., UK, and US. Sentiment here tends to be more negative overall, especially after mid-2022. The initial optimism seen for some locations fades over time, likely reflecting geopolitical tensions or societal concerns tied to AI deployment.

Together, these plots highlight how sentiment evolves differently for technologies and locations, and offer insights into the tone of AI-related discourse across different types of entities.

Conclusion & Strategic Outlook

Wraps up findings and highlights future opportunities, risks, and considerations for leaders navigating AI-driven transformation.

🔚 Conclusion

- AI is not impacting all industries equally: Technology, Media, Healthcare, Education, and Communication Services are leading in adoption, but motivations and challenges vary widely.
- Public sentiment remains mostly neutral to cautious, suggesting both optimism and concern exist across stakeholders.
- Entity and topic-level analyses show that successful AI integration is often tied to transparent communication, trust, and alignment with business needs.

🔭 Strategic Outlook

- Text Cleaning: Use spaCy pipeline for more reliable lemmatization and tokenization. Current setup led to some noisy artifacts (e.g., "contentnbc").
- Data Quality: Consider excluding invalid URLs. Articles with dead links may indicate low-quality or withdrawn content.
- Industry Classification: Enhance zero-shot prompts for edge cases (e.g., Education, Communication); explore multi-label classification and use confidence scores.
- NER Scope & Coverage: Expand keyword lists for organizations and individuals, and apply NER across the entire dataset, not just Technology.
- Sentiment Refinement: Reduce neutral clustering by testing alternative models or hybrid methods (e.g., rule-based + transformer). Calibrate thresholds for clearer sentiment separation.

🏛️ Broader Strategic Implications

- Companies: Prioritize explainable AI, deploy task-specific tools (e.g., copilots), and use sentiment monitoring to guide rollout messaging
- Academic Institutions: Expand AI literacy across disciplines; conduct independent impact studies on AI adoption in vulnerable sectors
- Governments: Fund responsible AI research, establish sector-specific regulatory frameworks, and promote access to trusted infrastructure

THANK YOU