The Impact of Artificial Intelligence

Newspaper Article Analysis

Akash Sannidhanam

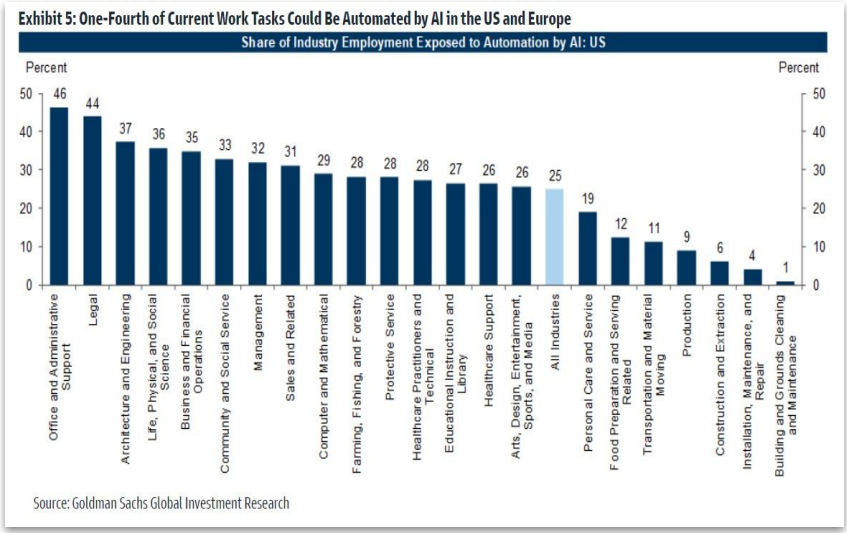
Agenda

1. **Executive Summary**
2. Actionable Recommendations
3. Data Preprocessing
4. Topic Detection
5. Sentiment Analysis
6. Entity Recognition

Current Artificial Intelligence Landscape

In March 2023, **Goldman Sachs** released a report estimating that about **25% of jobs in the U.S. and Europe could be automated by AI**, though the impact will vary widely across industries. Jobs that involve office work, legal tasks, architecture, and social sciences could see **automation rates above 30%**, while roles like construction, maintenance, and installation are expected to remain mostly unaffected.

Furthermore, a **Facebook Research paper** referencing **Moravec's Paradox**, which argues that tasks requiring physical and sensorimotor skills are much harder for AI to perform than those involving abstract thinking.



Both of these reports are shaped by the recent surge in **Large Language Models (LLMs)**. In this analysis, a **corpus of ~200,000 news articles** focused on Data Science, Machine Learning, and AI. The goal is to analyze this collection to **identify which industries and job roles are most likely to be impacted by AI in the coming years** based on real-world conversations and reporting.

Key Insights about AI and Its Perception

- **Techniques Used**: NLP filtering (removal of noise like links, emails), tokenization of text, sentiment and time-series analysis, entity recognition, and **topic extraction using LDA** — showing **a predominantly positive narrative around AI**, suggesting public concerns of AI replacing jobs may be overstated in the media.
- **Sentiment Analysis** using VADER and a custom Yelp model revealed that **successful AI applications** are linked to terms like *research, innovation, and communication*, while **failed initiatives** are tied to *risk, privacy, and ethics*.
- **Overall sentiment** toward AI and data science remains **highly positive over time**, despite ongoing **privacy, legal, and ethical concerns**.
- **NER (SpaCy)** identified **Microsoft and Google** as the most mentioned entities — driven by their advancements in generative AI (e.g., ChatGPT, Bard). However, both also face negative sentiment over **privacy, data collection, and monopolistic practices**.
- **Top negative themes** include AI's role in *deepfake images, cancer research concerns, poor market, ethical risks, and AI news trends*, though these are **less frequently mentioned**, indicating a stronger focus on AI's positive impact.

Agenda

1. Executive Summary
2. **Actionable Recommendations**
3. Data Preprocessing
4. Topic Detection
5. Sentiment Analysis
6. Entity Recognition

How to Approach an AI-Driven Future

1. **Human-AI Collaboration**: Reimagine roles to leverage AI as a tool that enhances human skills. Focus on tasks where AI can assist, not replace, and invest in cross-functional collaboration to ensure AI solutions are designed with human input, increasing acceptance and effectiveness.
2. **AI Upskilling and Training**: Governments and companies should invest in widespread AI and data literacy training, including AI ethics and bias, to prepare workers at all levels for AI integration.
3. **Career Pathways and Reskilling**: Create clear AI-related career paths and support transitions into AI-augmented roles through mentorship, flexible learning, and hands-on experience.
4. **AI Governance and Regulation**: Develop strong, adaptive AI regulations focused on privacy, ethics, and media transparency, while educating the public to build trust in AI systems.
5. **Continuous Learning and Innovation**: Foster a culture of ongoing AI learning and experimentation through workshops, AI labs, and public-private collaboration to ensure AI remains human-centered.

Agenda

1. Executive Summary
2. Actionable Recommendations
3. **Data Preprocessing**
4. Topic Detection
5. Sentiment Analysis
6. Entity Recognition

Transformed and Filtered Newspaper Articles for an Accurate Analysis

Data Cleaning

- Refined **Title and Text** columns
  - Removed URLs, hashtags, mentions, new lines, symbols, and publications
- Filtered out **Stopwords and numbers**
- Only retained the **5th - 95th** percentile of records based on article length

Feature Engineering

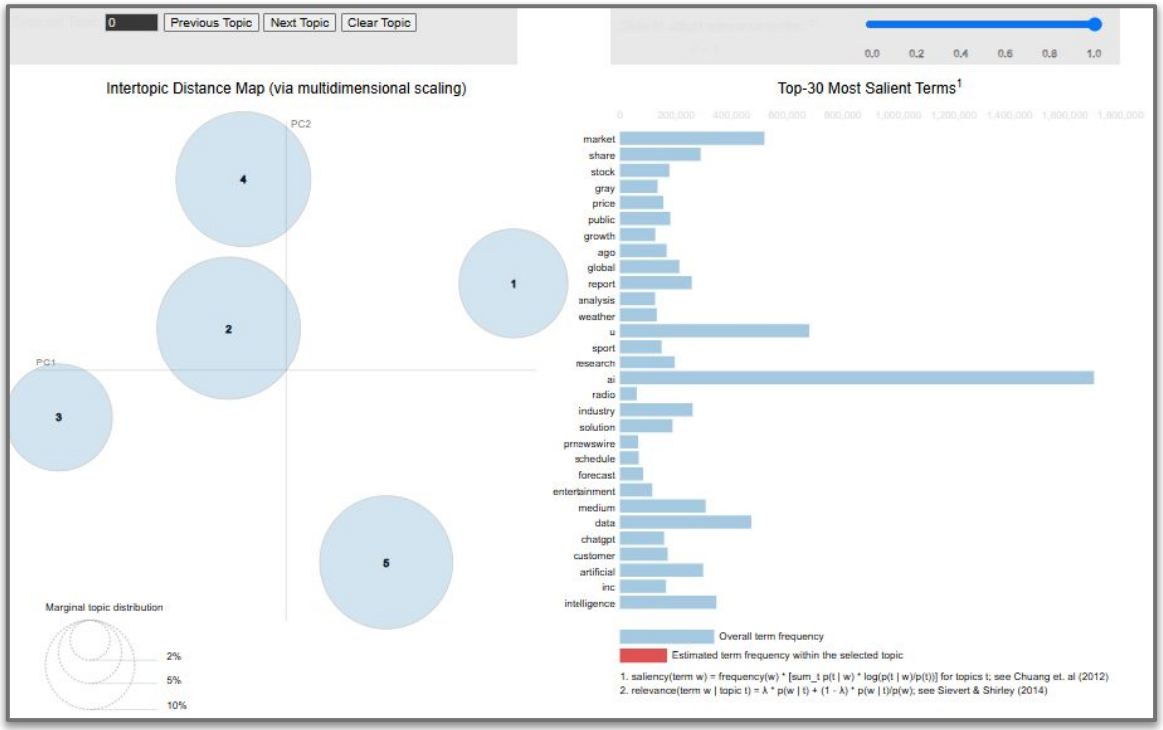
- Expanded on **Time Variables**
  - Year, month, day, and day of the week
- **Tokenized** cleaned Title and Text columns
  - Split individual words
  - Turned all words into **lowercase**
  - **Lemmatized** to get the base form of words

After all the preprocessing steps, there were **159,769** articles

Agenda

1. Executive Summary
2. Actionable Recommendations
3. Data Preprocessing
4. **Topic Detection**
5. Sentiment Analysis
6. Entity Recognition

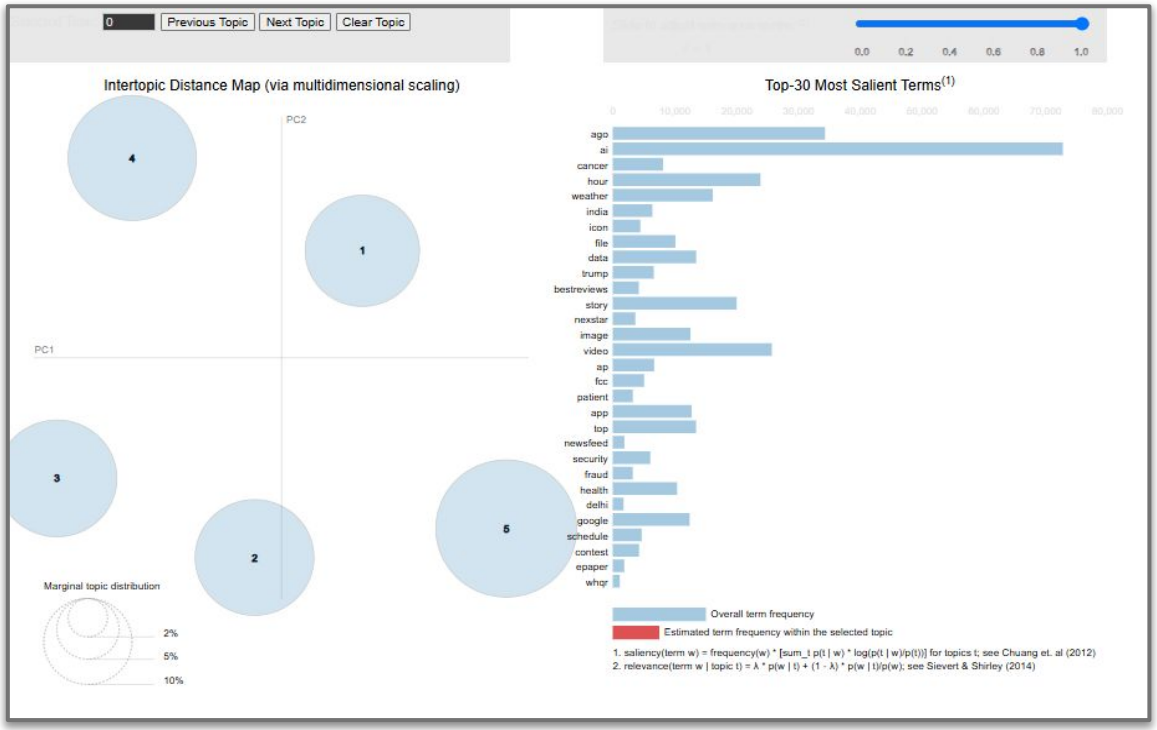
Topics for Positive Sentiment



Relevant Topics

1. AI-driven innovations in global markets and business growth.
2. AI and technology in news, media, and society with a focus on innovation and communication
3. AI-driven innovations in public services, media, and business communication.
4. AI and tech innovations from leading companies (Google, Microsoft, OpenAI).
5. AI solutions and enterprise technology for business and customer innovation.

Topics for Negative Sentiment



Relevant Topics

1. Concerns about AI risks, data privacy, security
2. AI's role in social media, misinformation, political manipulation, and the broader impact on public discourse and elections
3. Concerns about AI's rapid advancements, its impact on humans, ethics, misinformation in media, and AI-generated content
4. Discourse about AI's role in healthcare
5. AI's role in generating and spreading misinformation through media and apps, concerns about AI's influence on public life (schools, sports, news)

Agenda

1. Executive Summary
2. Actionable Recommendations
3. Data Preprocessing
4. Topic Detection
5. **Sentiment Analysis**
6. Entity Recognition

The Overall Sentiment towards AI is Astoundingly Positive

Sentiment	Count
Extremely Positive	145,470
Positive	3,318
Neutral	2,188
Negative	988
Extremely Negative	7,805

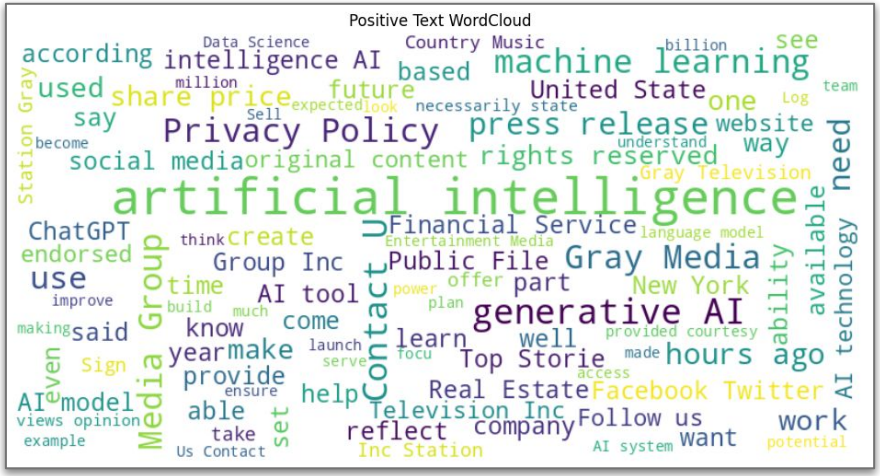
One method involved building a custom sentiment prediction model, trained on labeled Yelp review data using a classifier, specifically logistic regression.

Another method used VADER (Valence Aware Dictionary and Sentiment Reasoner), a dictionary-based tool for sentiment analysis. VADER assigns sentiment scores to words and phrases, which makes it possible to classify text into different sentiment categories.

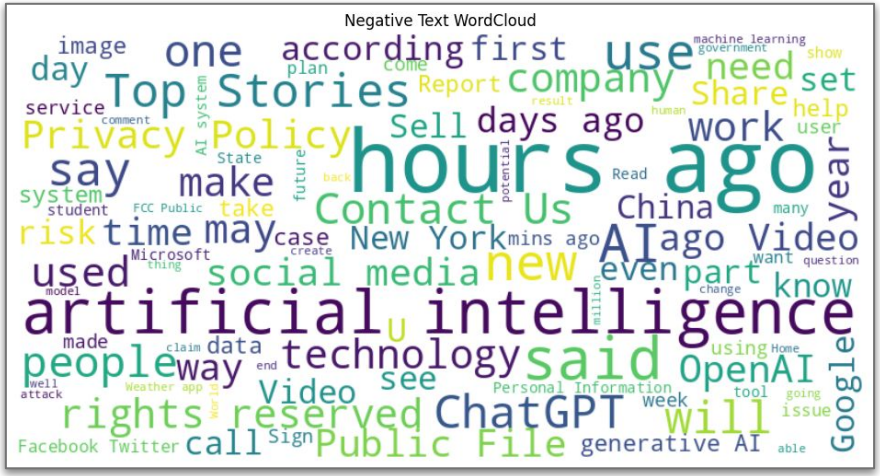
Both models were tested independently by running them on the dataset. After comparing the results, decided to move forward with **VADER for sentiment analysis**, as it proved to be more accurate and reliable than the custom Yelp-based model, which is binary in its nature.

How Words and Phrases Compare Based on Sentiment

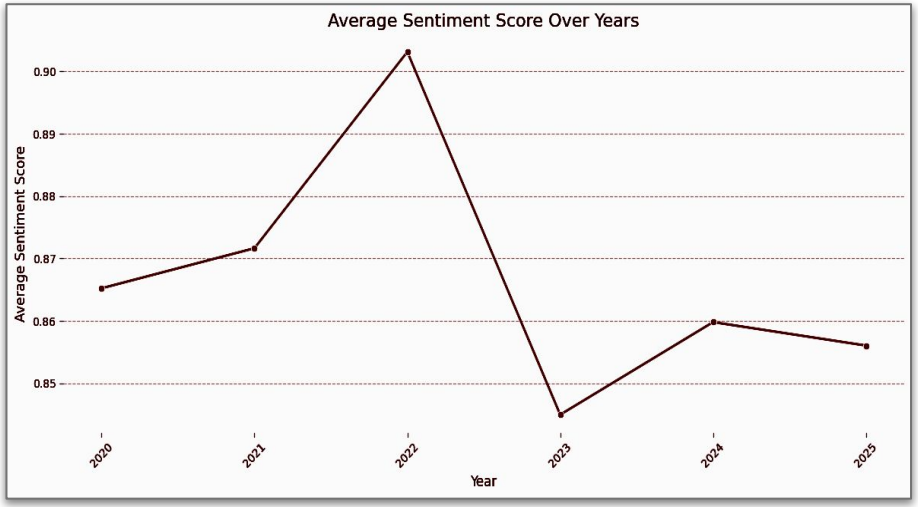
Top Positive Words for AI

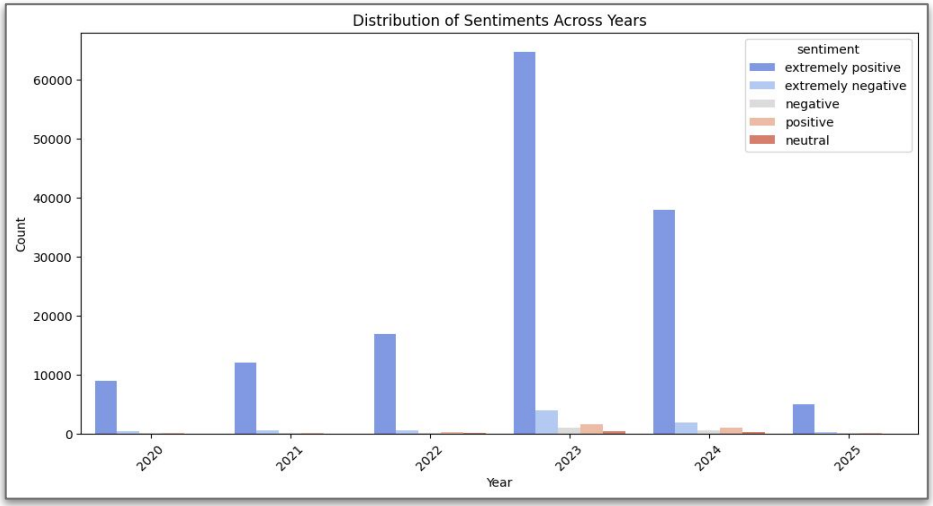


Top Negative Words for AI



Sentiment Analysis Over time





The average sentiment toward AI has shifted over the years from 2020 to 2025. Sentiment steadily increased from 2020 to 2022, **peaking in 2022 with the highest positivity**, perhaps due to the introduction of ChatGPT. However, there is a sharp decline in 2023, suggesting a **significant shift toward more negative or cautious attitudes**. After 2023, sentiment slightly rebounds in 2024 but starts to decline again in 2025, indicating ongoing mixed feelings or emerging concerns about AI. Although positive sentiment remains strong, the increased negative and neutral responses in later years suggest **growing public debate and mixed feelings about AI advancements**.

Agenda

1. Executive Summary
2. Actionable Recommendations
3. Data Preprocessing
4. Topic Detection
5. Sentiment Analysis
6. **Entity Recognition**

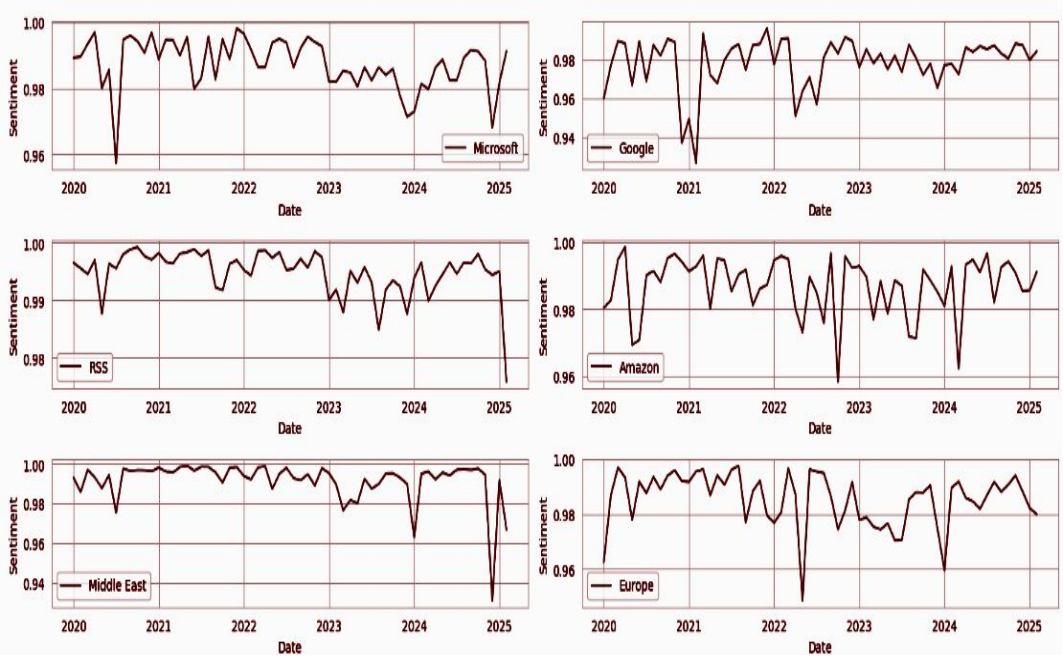
Top Entities by Sentiment

Top Positive Entities for AI	
Entity	**Count**
Microsoft	71,705
Google	57,092
Gray Media Group	49,215
Apple	26,851
Amazon	22,811

Top Negative Entities for AI	
Entity	**Count**
Google	5,091
Microsoft	5,041
Android	2,648
Trump	2,592
Biden	2,017

Using **SpaCy NER,** the majority of the entities extracted are associated with positive sentiment, indicating that AI is generally welcomed by the public. Interestingly, **Google** and **Microsoft** also appear among the top negative entities, suggesting they receive both praise and criticism in AI discussions. Other negatively associated entities include Android, Trump, and Biden, indicating political figures and tech platforms also raise controversy regarding AI.

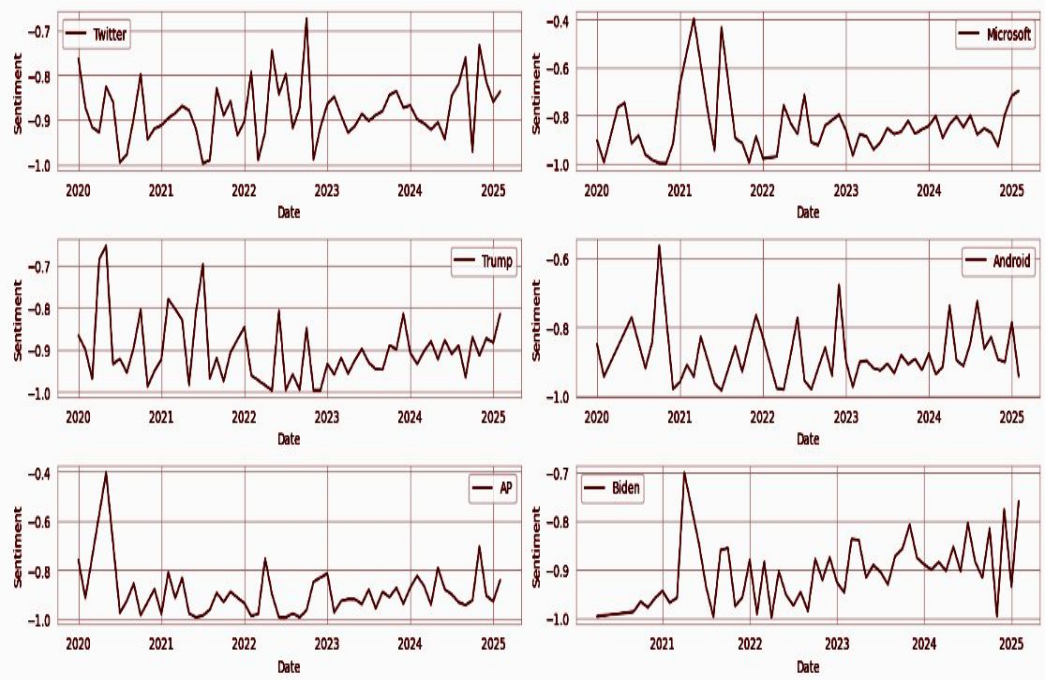
Targeted Sentiment Analysis for Positive Entities



Microsoft, Google, and Amazon each show consistent positive sentiment with a few noticeable dips, likely tied to specific events or controversies that temporarily affected their public image, such as COVID-19. The **RSS feed**, which represents a broad mix of AI-related news, stays relatively stable until it takes a dip in 2025, indicating a consistently positive narrative around AI in general media.

Regions like the **Middle East and Europe** also maintain a strong positive sentiment, although they exhibit some minor fluctuations, suggesting the effects of events that impact on a global level. Ultimately, these entities and regions are mostly viewed in a positive light when it comes to AI.

Targeted Sentiment Analysis for Negative Entities



Twitter, Trump, and Biden show consistently high levels of negative sentiment, with occasional spikes that may reflect specific political events or controversies. **Microsoft and Android**, though largely viewed positively in other contexts, appear here with persistent negative mentions, showing that public opinion is mixed most likely due to data and privacy concerns. **AP (Associated Press)** also carries steady negative sentiment, perhaps reflecting criticism in AI-related news coverage.

While the intensity of negative sentiment fluctuates, these entities have been consistently tied to AI discussions with a critical or unfavorable tone.