



# Unlocking the Power of AI

## New Article Analysis

NATURAL LANGUAGE PROCESSING FINAL PROJECT

Submitted by  
**Swathi Ganesan**  
**12372237**



# Executive Summary

## MISSION

To analyse trends in technology news articles by best leveraging NLP techniques in order to identify the potential of AI and drive a paradigm shift to promote adoption of novel technologies like AI and automation

## DRIVERS

Driven by the need for efficiency, innovation, and the transformative power of AI, we are dedicated to uncovering opportunities to automate tasks and enhance employee productivity

## CHALLENGES

Ensuring ethical and responsible AI adoption, and staying at the forefront of evolving AI technologies and algorithms

## Outcomes

Identified areas where task automation can be employed to enhance employee productivity, through the adoption of AI technologies while considering the negative sentiments associated with AI, such as privacy concerns, ethical considerations, cybersecurity, and environmental impact

# Analysis Process Flow

## Data Cleaning

De-duplication using **Yelp** trained **SVM** sentiment model



## Custom Sentiment Analysis

Using **twitter-roberta-base-sentiment** and **distilroberta-finetuned-financial-news-sentiment-analysis** with custom thresholds for Positive, Negative and Neutral sentiments



## Topic Modelling & NER

Identified major topics in Positive and Negative classes using **BERTopic** and analyzed top entities using **ktrain** text summarization with **pretrained BART**



## Targeted Sentiment Analysis

Extracted and analyzed sentiment trends for top entities using **TextBlob** assigned Polarity scores

# Ensuring Data Quality for Effective Analysis

01

## Eliminating Noise

- Newlines, tabs, whitespaces, special characters
- Email addresses, URLs
- Exhaustive list of stopwords (english and facebook)
- Web crawl remnants

02

## Text Pre-processing

- News articles and titles tokenized and lemmatized
- Single character tokens, nulls and token length > 20 removed
- **83.72%** of news articles have length <=1500 so truncating our news to limit at 1500 tokens

03

## Data Deduplication

Duplicates removed in the following combinations in order to reduce redundancy :

- URL - news text
- news title - news text
- news title – top level sentiment generated using YELP trained custom SVM

~200K

initial news dataset



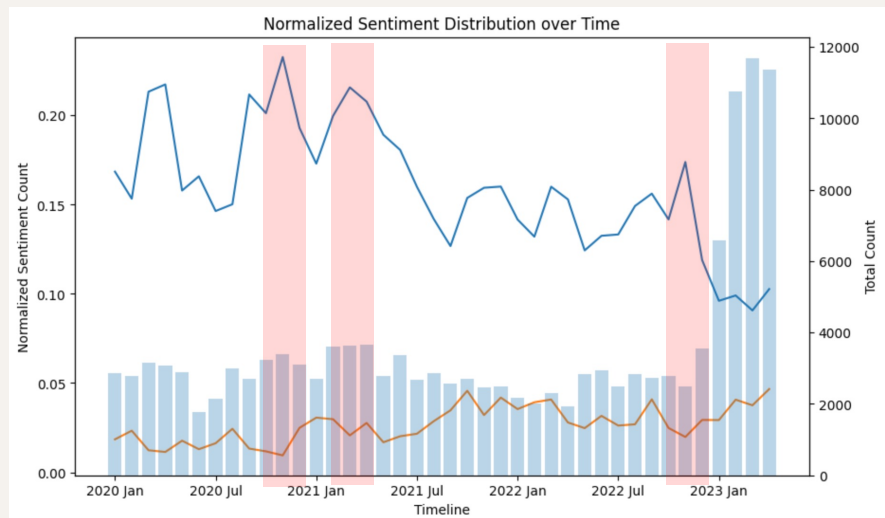
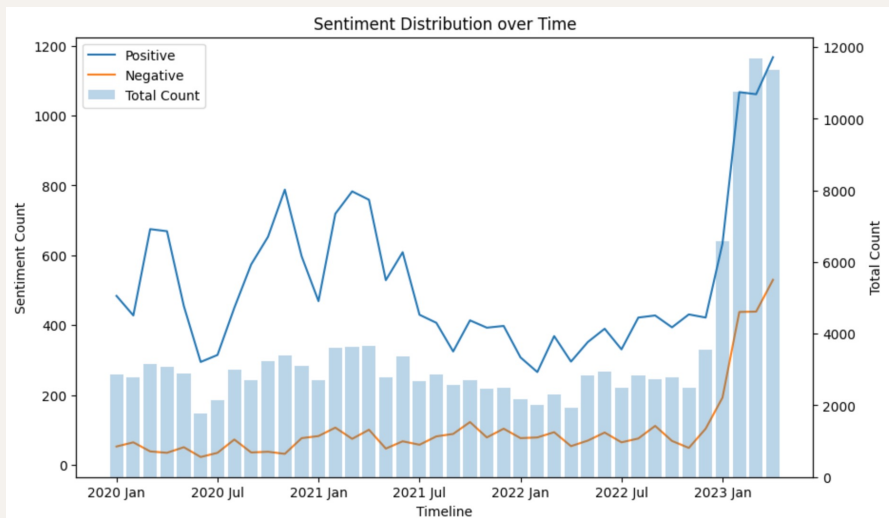
~140K

unique clean news articles

Jan 2021 – Apr 2023

# Article and Sentiment Trends Through Time

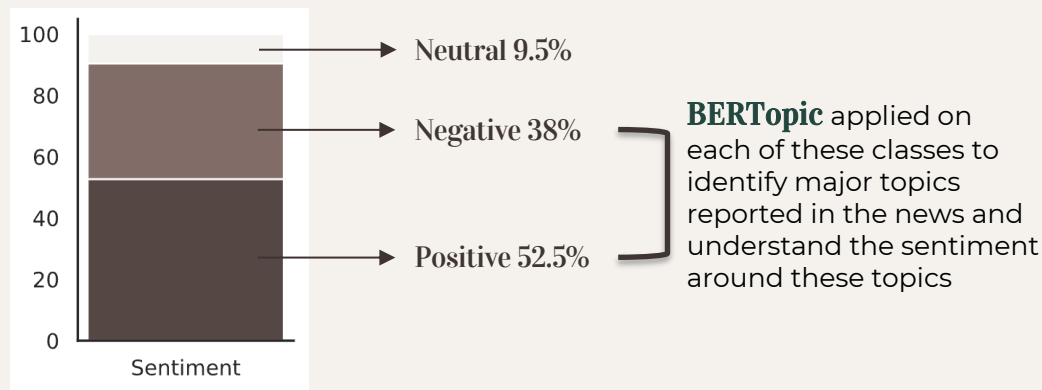
Using **distilroberta-finetuned-financial-news-sentiment-analysis**



Though we can observe an influx in the number of articles related to technology and development in 2023, normalized sentiment peaks can be observed in **late 2020, early 2021 and late 2022**

# BERTopic Modelling Overview

Sentiment Analysis is done using **twitter-roberta-base-sentiment** in order to identify and eliminate the Neutral class.



	Positive	Negative
Total BERTTopics	139	95
Reduced number of topics	30	30

## Next Steps :

- Identify major topics of discussion in each sentiment class
- Delve into the WHO, WHAT, WHERE details for each of these topics
- Select topics of interest and observe their sentiment trend over time

# Exploring Positive Sentiment

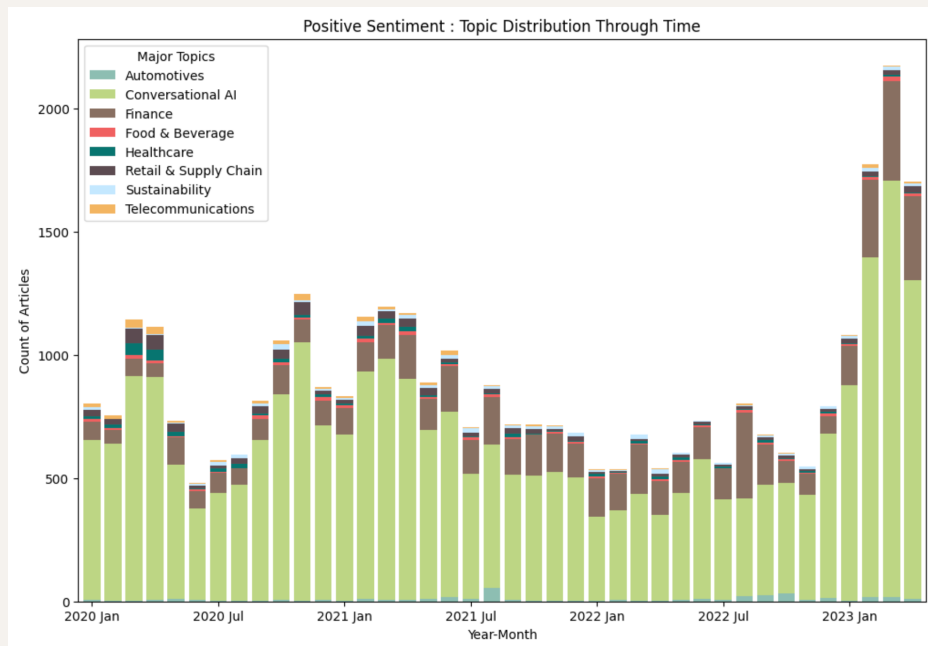
Findings (Summarized for top entities\* using **ktrain**) :

## Topic 1 : Automotives

- Autonomous driven **Lyft robotaxi** powered by **Argo AI's** technology in **Ford** vehicles in Austin
- **Elon Musk** debuts 'Friendly' humanoid **Tesla Bot, Optimus** during company's AI day
- **Spin** announces cutting edge on-vehicle AI platform to bring sidewalk riding and parking detection technology to cities across the **US** and **UK**
- **Audi** using AI to design new wheels and **Porsche** using AI to suggest good driving roads
- **Self-driving delivery vehicle** hits **UK** roads as government funds AI car projects

## Topic 2 : Conversational AI

- AI in **Behavioral and Mental Health Care** market projected to witness astonishing growth by 2026
- **Google** building a 1000-language AI model to beat **ChatGPT** and **Apple** acquires AI Voice Startup **Voysis** to make **Siri** better understand our queries
- **OpenAI** scraps waitlist For Its **AI Art Generator Dall-E**



*Influx in Finance news is observed in 2023 and a strong positive sentiment can be observed in news about AI and technology through the years*

*\* Top entities for each topic in Positive class illustrated in Appendix I and text summarisation available in 07\_BERTopic\_NER\_Positive.ipynb*

# Exploring Positive Sentiment

## Conversational AI contd.

- **Facebook** is creating an AI that can view the world and interact with it like a human and **PDFgear's** new **ChatGPT-** Powered feature allows users to chat with PDFs
- Using **AI** to digitally replicate human voices, **Anthony Bourdain's** Voice Recreated For New Documentary

## Topic 3 : Finance

- **SoundHound Voice AI** investors are lined up for success as stocks soar to new heights
- AI in **Medical Diagnostics Market** worth \$3,868 million by 2025
- UK based **TrueCircle**, a compute vision startup scoops \$5.5M to use AI to drive recycling efficiency
- **PreciTaste** raises \$24M to redefine foodservice operations with Vision AI technology

## Topic 4 : Food & Beverage

- **Gastrograph**, an AI platform models human sensory perception uses AI to help companies predict and personalize flavor for new food products
- **Brightseed** finds nutrients inside plants with AI to improve human health

## Food & Beverage contd.

- **Tastry AI** recommends perfect wine acting as a personal sommelier based on Artificial Intelligence
- **IntelligentX Brewing Co.** has created a premium beer that uses Machine Learning algorithms to improve itself from feedback
- **Kroger** launches **Chefbot**, a first-of-its-kind AI Twitter Tool that delivers personalized recipe recommendations based on ingredients already in users' kitchens.

## Topic 5 : Healthcare

- In 2022, the **FDA** authorized 91 **AI** and **ML** enabled medical devices
- Google's **DeepMind** created an AI for spotting breast cancer that can outperform human radiologists and is more trustworthy than experts at reading medical scans
- **Lunit's** AI Solution for Breast Cancer Detection wins commercial approval in **Taiwan**
- In virus-hit **South Korea**, AI monitors lonely elders
- **Paige** raises \$5 million more from Goldman Sachs to detect cancer with computer vision. launching the first FDA-Authorized AI Software for Digital Pathology





# Exploring Negative Sentiment

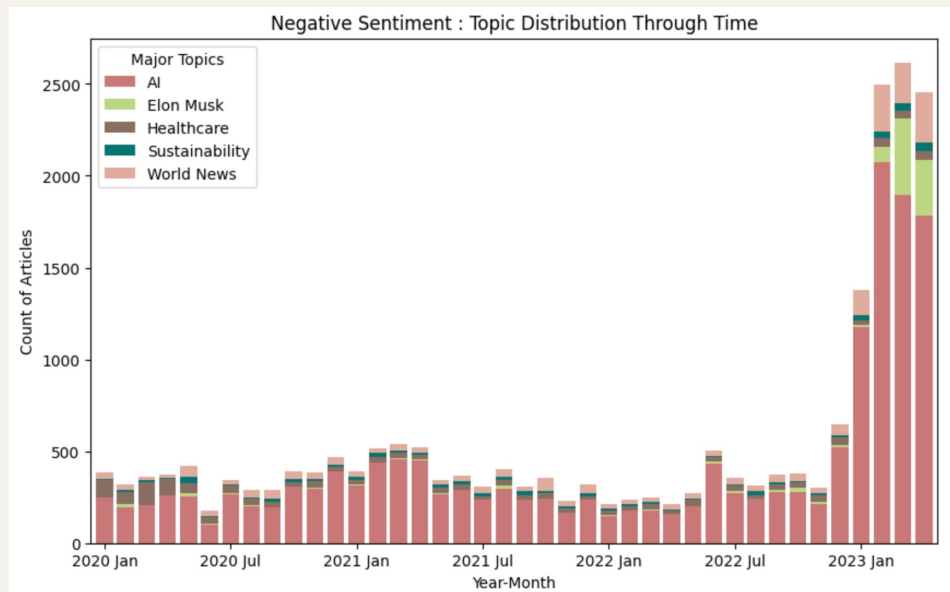
Findings (Summarized for top entities\* using **ktrain**) :

## Topic 1 : AI

- **OpenAI** addresses **Italy's** privacy watchdog action against **ChatGPT** due to a data breach
- Sci-Fi magazine to suspend submissions due to flood of **AI-generated stories**
- **Google** AI's new chatbot **BARD**, made a \$100B mistake in a demo ad
- **Timnit Gebru**, **Google's** Co-Head of Ethical AI, claimed that she was fired for expressing concerns in an email regarding the company's AI practices
- **Microsoft** bakes ChatGPT-like tech into search engine **Bing**
- **Microsoft** hits the brakes on **Bing** AI Chatbot after it compares reporter To Hitler
- Growing adoption of AI in the **Telecommunication** industry projected for 2025

## Topic 2 : Elon Musk

- **Elon Musk** takes a dig at netizens criticising futuristic tech
- **Elon Musk** demand immediate pause of Advanced AI Systems citing 'risks to society', claims **TruthGPT** will counter AI 'bias'
- **ChatGPT** lists **Donald Trump**, **Elon Musk**, **Kim Kardashian** and **Kanye West** as 'controversial'



*Influx in news about Elon Musk observed in 2023 and news about AI and technology have been increasing through the years*

*\*Top entities for each topic in Negative class illustrated in Appendix II and text summarisation available in 08\_BERTopic\_NER\_Negative.ipynb*

# Exploring Negative Sentiment

## Topic 3 : Healthcare

Text summarization shows us that **healthcare** is **highly positive** and does not really have negative sentiment. News is tagged as negative due to words like cancer, pandemic, etc.

## Topic 4 : Sustainability

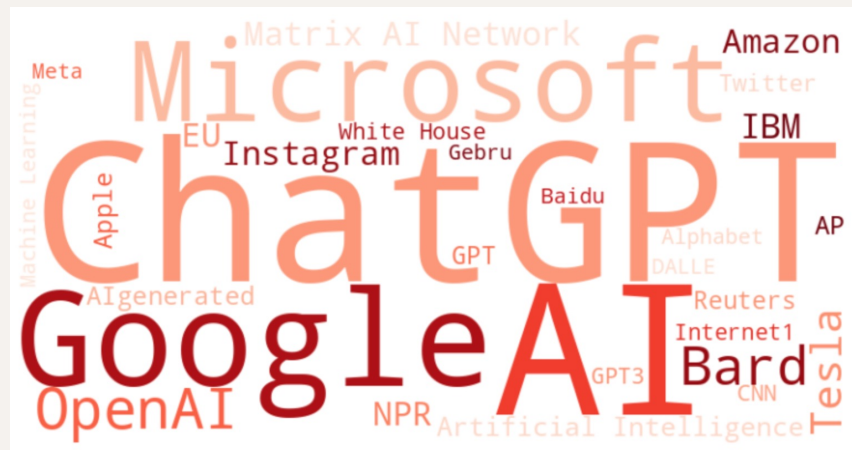
- **Greenpeace** studies show **Microsoft**, **Google** reliant on fossil fuels despite 100% renewable energy pledges
- **Elizabeth Jardim**, campaigner for **Greenpeace USA** reports that Google backs off on AI for Oil and Gas Extraction pledging not to build more custom AI to aid Fossil Fuel Extraction
- **ChatGPT** data centres are consuming a staggering amount of water. Though research shows the carbon footprint of **AI** models, water consumption to run them on a large scale has remains under the radar

## Topic 5 : World News

- **Artificial Intelligence** Assistants To Outnumber People in **China** where **Digital humans** like **Xiaoice**, an AI-powered love bot are revolutionizing human-computer interactions with their immersive experiences blending the digital and real worlds.
- **Baidu's** launch of new **AI** product **Ernie Bot** raises questions about cybersecurity with shares falling after disappointing **AI** chatbot debut

## World news contd.

- **Indian PM Narendra Modi** emphasizes need to safeguard the world from the weaponization of **AI**, highlighting the importance of responsible AI development and usage.
- **Saudi Arabia** signs MoUs with **IBM**, **Alibaba** and **Huawei** on **AI** whereas **IBM** Stock goes down after weak growth in cloud and AI revenue



## Top 3 Entities of Interest



From the 2 **WordClouds** of our **Positive** and **Negative** classes we can select the most talked about topics and perform targeted sentiment analysis over time



Google



# Microsoft

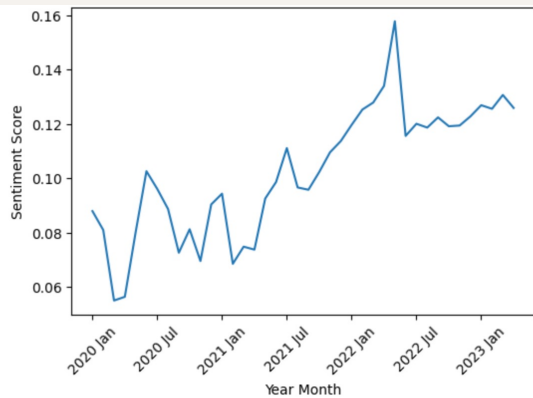


# ChatGPT

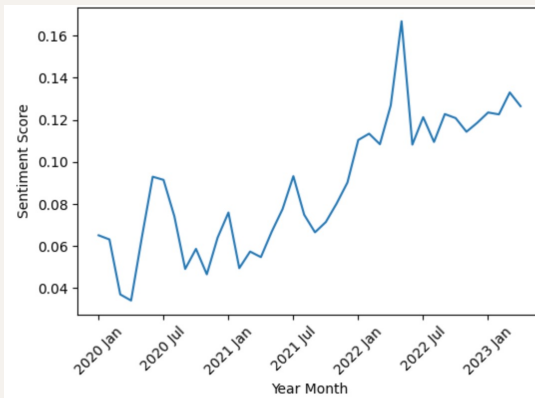
# Targeted Sentiment Trend

## Google

- Both Google and Microsoft show similar trends in their sentiment timeline
- Peaks in **Q2 2022** can be seen due to Google firing Blake Lemoine, the engineer who claimed AI chatbot LaMDA was sentient
- Sustainability concerns raised around the fact that Microsoft, Google are reliant on fossil fuels despite 100% renewable energy pledges
- Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI

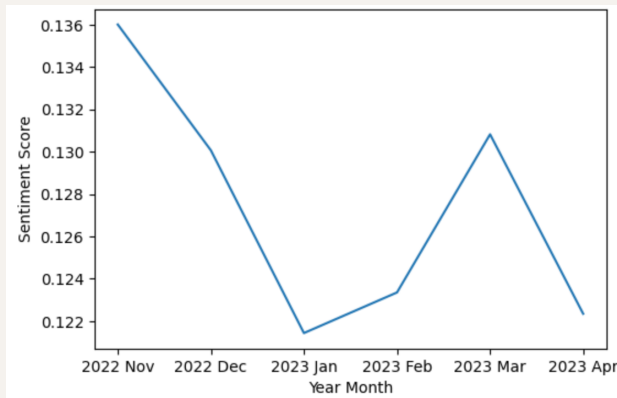


## Microsoft



## ChatGPT

- Major buzz caused by the successful launch of ChatGPT in Late 2022
- Negative sentiment as Apple, Goldman Sachs, Samsung and other market leaders ban ChatGPT at work due to concern around privacy risks



# Actionable Recommendations

- **Customer Service:** Continuously innovate conversational AI models to enhance natural language understanding and improve user interaction. Foster the growth of AI in **behavioural and mental health care** to provide personalized and accessible support to individuals. **Conversational AI** and **Chatbots** can handle customer inquiries and support, reducing the need for human agents in routine interactions
- **Taxi, Ride-hailing Services and Public Transportation:** Advance **Autonomous Driving** technology and invest in AI-powered vehicle design and optimization
- **Finance Audits:** Utilize AI in **financial diagnostics** and analysis, and leverage vision AI technology for **fraud detection** and **risk assessment**
- **Food & Beverage Industry:** **Optimize flavour profiles, curate menus, meal planning** and personalize products using AI platforms, and implement **AI-driven recommendation systems**. AI studies weather patterns, and market trends, to **optimize inventory levels, reduce stockouts, and ensure timely delivery** of food products
- **Healthcare & Therapy:** Embrace AI-enabled medical advances in **medical diagnostics, personalized medicine, drug discovery and development, remote patient monitoring, telehealth administrative efficiency** along with **high precision robotic surgery and assistance**. Conversational AIs can be leveraged for **AI-Powered Virtual Therapy**
- **Retail & Supply Chain:** Implement AI solutions for **inventory management** and **on-shelf availability**, and utilize AI in fashion for personalized recommendations and AI assist in **demand forecasting, inventory management, and logistics planning**
- **Sustainability:** Utilize ML and AI for **disaster prediction** and resource management, and implement AI tools for **energy optimization and waste reduction**. AI can be used for **automated construction, safety monitoring and planning automation**

---

# Conclusion

Analysing the *negative sentiments* around AI, key challenges in adoption of AI are identified as follows

- Robust data privacy and security measures are crucial to prevent data breaches and address regulatory concerns in AI applications
  - Responsible AI development is essential to address potential biases and risks associated with AI-generated content
  - Rigorous testing and quality control measures should be implemented to avoid costly mistakes in AI-powered systems and applications
  - Embracing AI technologies in the telecommunication industry requires considering potential challenges and prioritizing data privacy and security
  - AI adoption in healthcare should carefully consider ethical implications and prioritize human-centered approaches in patient care
  - Responsible use of resources is vital to minimize the environmental impact of AI applications
-

---

# Areas of Improvement

- New articles can be cleaned better by using robust cleaning methods in order to remove web crawl remnants like news agency info, news reporter names, recommended articles, social media tags (Instagram, YouTube) and other noise as they led to noisy NER entities
  - To implement Sentiment Analysis, pre-trained transformer models were used but inconsistencies spotted in topics generated during topic modelling (esp. negative healthcare topic). Some level of customisation and manual labelling required to correctly identify the sentiment of the text
  - With access to more computational power, there would be flexibility for more sophisticated and advanced models and NLP techniques to be implemented for the above tasks
-



# Appendix I

Top named entities in the **Positive** class by Topic

AUTOMOTIVES

ORGANIZATION	PRODUCT	PERSON
(Tesla, 1731)	(AI, 817)	(Musk, 441)
(AI, 295)	(ADAS, 54)	(Elon Musk, 278)
(Optimus, 175)	(EV, 50)	(Dojo, 43)
(BMW, 159)	(Model, 31)	(Optimus, 40)

CONVERSATIONAL AI

ORGANIZATION	PRODUCT	PERSON
(AI, 22435)	(AI, 54238)	(Instagram, 2729)
(ChatGPT, 13979)	(Reddit VKontakte Share, 1197)	(Su Tierra Tiempo, 1304)
(Microsoft, 11112)	(Bing, 1193)	(Bing, 953)
(Google, 10086)	(Logout, 992)	(Costello, 773)
(IBM, 6419)	(Aldriven, 938)	(Jim Thorpe, 666)

FINANCE

ORGANIZATION	PRODUCT	PERSON
(AI, 7240)	(AI, 15487)	(Vectorspace AI, 884)
(Nvidia, 4213)	(CouponsPurple Squarespace, 394)	(Forbes, 470)
(Lisk Machine Learning, 3301)	(JavaScript, 370)	(Espaol Franais, 383)
(Machine Learning, 2866)	(A100, 335)	(Phil Mackintosh, 347)
(NVIDIA, 2865)	(HighVolume Penny, 306)	(Andy Serwer, 339)

FOOD & BEVERAGE

ORGANIZATION	PRODUCT	PERSON
(AI, 241)	(AI, 621)	(Costello, 24)
(Artificial Intelligence AI Food Beverages, 200)	(JavaScript, 11)	(Prabha Samakalika, 13)
(Artificial Intelligence AI Food Beverage, 136)	(MaxAi Buhler AG, 11)	(Malayalam Edex, 13)
(ChatGPT, 103)	(Seq, 11)	(Chefbot, 13)

HEALTHCARE

ORGANIZATION	PRODUCT	PERSON
(AI, 423)	(AI, 1039)	(Paige, 44)
(Google, 160)	(DeepMind, 17)	(Xray, 20)
(FDA, 91)	(Lunit, 17)	(Trump, 16)

RETAIL & SUPPLY CHAIN

ORGANIZATION	PRODUCT	PERSON
(AI, 618)	(AI, 1137)	(Sameer Joshi, 58)
(Artificial Intelligence Retail, 531)	(AI Fashion Market, 69)	(Costello, 41)
(IBM, 467)	(SAS, 56)	(Sanjay Jain, 36)
(Microsoft, 351)	(AI AR, 52)	(Levi, 33)
(Amazon, 259)	(Microsoft Nvidia, 51)	(Tiliter, 32)

SUSTAINABILITY

ORGANIZATION	PRODUCT	PERSON
(AI, 371)	(AI, 1118)	(Twitter Instagram, 34)
(BrainBox, 147)	(C3 AI, 29)	(Bidgey, 25)
(ML, 128)	(EV, 22)	(NABusiness, 22)
(Google, 106)	(Argonne, 21)	(Watson, 16)

TELECOMMUNICATIONS

ORGANIZATION	PRODUCT	PERSON
(IBM, 266)	(AI, 308)	(Costello, 18)
(Artificial Intelligence Telecommunication, 237)	(Cloud Telecommunication AI, 238)	(Sanjay Jain, 10)
(Onpremises Telecommunication AI, 175)	(Cloud Telecommunication AI Market, 42)	(Sameer Joshi, 10)
(Google, 166)	(Powered WordPress, 22)	(Aleje Jerozolimskie, 10)
(Cisco Systems, 147)	(AI Telecommunication Market, 21)	(Porter, 10)

# Appendix II

Top named entities in the **Negative** class by Topic

AI

ORGANIZATION	PRODUCT	PERSON
(ChatGPT, 26450)	(AI, 39343)	(Bing, 2275)
(AI, 20278)	(Bing, 2317)	(Bard, 1221)
(Google, 18026)	(Air Stations Air, 552)	(Biden, 1218)
(Microsoft, 12345)	(Air Stations Air Playing, 517)	(Elon Musk, 1194)
(Bard, 5318)	(Aldriven, 469)	(Trump, 1139)

ELON MUSK

ORGANIZATION	PRODUCT	PERSON
(Tesla, 2313)	(AI, 3694)	(Musk, 3804)
(ChatGPT, 2294)	(SpaceX, 141)	(Elon Musk, 2354)
(AI, 2047)	(Teslas, 97)	(Trump, 1249)
(OpenAI, 1508)	(Open AI, 53)	(Higgins, 752)
(Microsoft, 853)	(Get Android, 48)	(Donald Trump, 614)

HEALTHCARE

ORGANIZATION	PRODUCT	PERSON
(AI, 2047)	(AI, 4060)	(Judy Faulkner, 104)
(Google, 636)	(ReviewHospital, 84)	(IssuePrint IssueCurrent, 88)
(ChatGPT, 281)	(Aldriven, 41)	(Trump, 86)
(Amazon, 252)	(YouTube, 36)	(healthcareGreat, 64)
(FDA, 219)	(JavaScript, 33)	(Barrett, 64)

SUSTAINABILITY

ORGANIZATION	PRODUCT	PERSON
(AI, 629)	(AI, 1544)	(Tom, 39)
(Google, 312)	(Android, 28)	(Elizabeth Jardim, 27)
(ChatGPT, 189)	(DeepMind, 25)	(Gene, 26)
(Greenpeace, 124)	(Aldriven, 24)	(Biden, 25)
(Microsoft, 117)	(JavaScript, 18)	(nan, 24)

WORLD NEWS

ORGANIZATION	PRODUCT	PERSON
(AI, 2901)	(AI, 5282)	(Ernie Bot, 275)
(ChatGPT, 2341)	(HelpSupport, 158)	(Biden, 178)
(Google, 967)	(Bing, 93)	(Narendra Modi, 143)
(Baidu, 636)	(ChatGPTstyle, 82)	(Aria, 140)
(Microsoft, 502)	(JavaScript, 82)	(Donald Trump, 132)