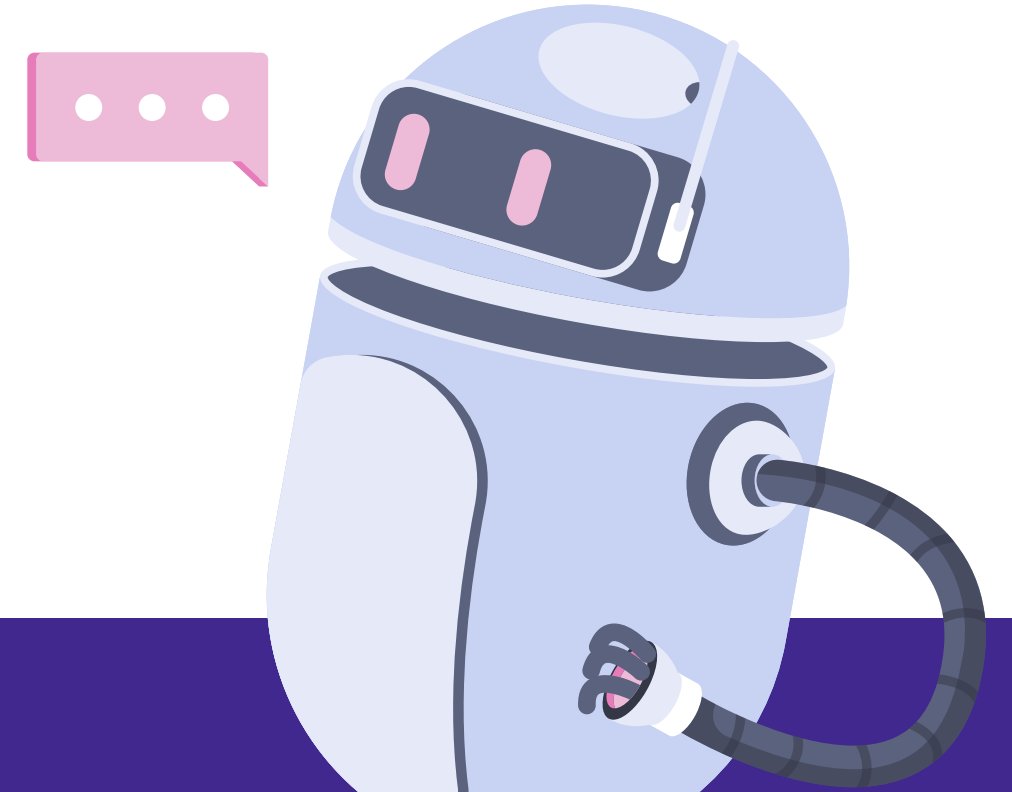
The Wave Has Come: Analyzing AI-related News

MSCA Machine Learning & Predictive Analysis



Han Jiang

May 22, 2023

THE UNIVERSITY OF CHICAGO

Executive Summary

Major Insights

- **Major AI Topics**: Key AI topics include generative AI, applications in business and healthcare, and large language models like ChatGPT. These have substantial impacts on tech firms and clinical research.
- **AI Strengths**: Success in AI hinges on data quality, sufficient R&D investment, and mature, automated tech infrastructure. These factors drive efficient operations and accurate predictions.
- **AI Limitations and Areas of Concern**: AI faces challenges including legal and regulatory constraints, ethical concerns, and tasks requiring human interaction. It's important to address these issues, especially in sensitive sectors like healthcare.
- **Leading Players**: Major tech firms lead AI development, with significant progress in fields like Natural Language Processing. A standout example is OpenAI's ChatGPT, indicating a promising direction for future AI exploration. For countries, the U.S. and China are the major leaders in AI development.

Future recommendations

- **Governments**: Foster AI literacy, support research, establish ethical regulations.
- **Investors**: Target high-growth AI sectors, be wary of highly-regulated fields.
- **Academics**: Integrate AI across curricula, stress on ethical education.
- **Traditional Industries**: Leverage AI for efficiency, foster tech partnerships.
- **Tech Firms**: Prioritize user-centric and ethical AI, collaborate for industry transformation

Actionable Recommendations

Governments and Regulators

Embrace AI revolutions! Governments should create supportive policies and regulatory frameworks that promote ethical AI use and protect citizen data. Investing in education and research, and encouraging innovation and multinational collaboration in AI.

Venture Funds and Investors

Venture capitalists should focus on AI-driven solutions in healthcare, fintech, edtech, and autonomous vehicles due to their high growth potential. However, they should tread cautiously in sectors like nursing or school systems where personal care and human interaction are paramount and regulatory scrutiny is high.

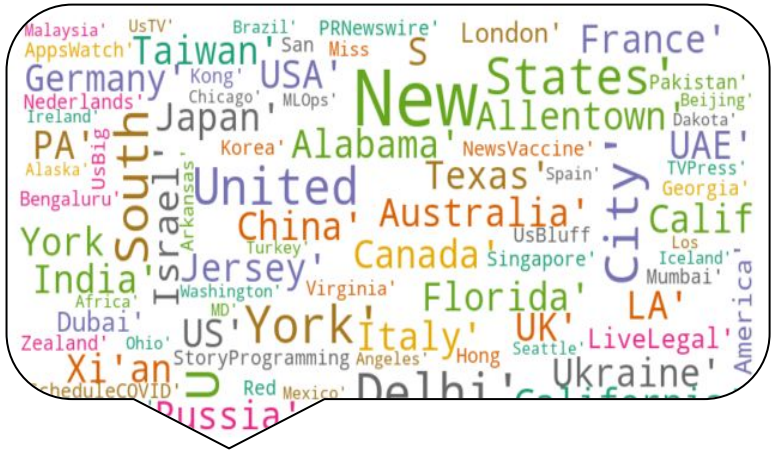


Figure: Word Cloud of Geopolitical Entities with positive sentiment associated with AI

Academic Institutions

Academic institutions should foster interdisciplinary AI research, integrate AI literacy, and partner with industry for practical exposure. Ethical AI education should be prioritized, and ensure to cultivate an innovative and inclusive AI culture.

Traditional Industries

Traditional industries can leverage AI to drive efficiency and innovation. For example, manufacturers can use AI for predictive maintenance and supply chain optimization. The service and commerce industries can use AI chatbots and data analytics to improve customer services.

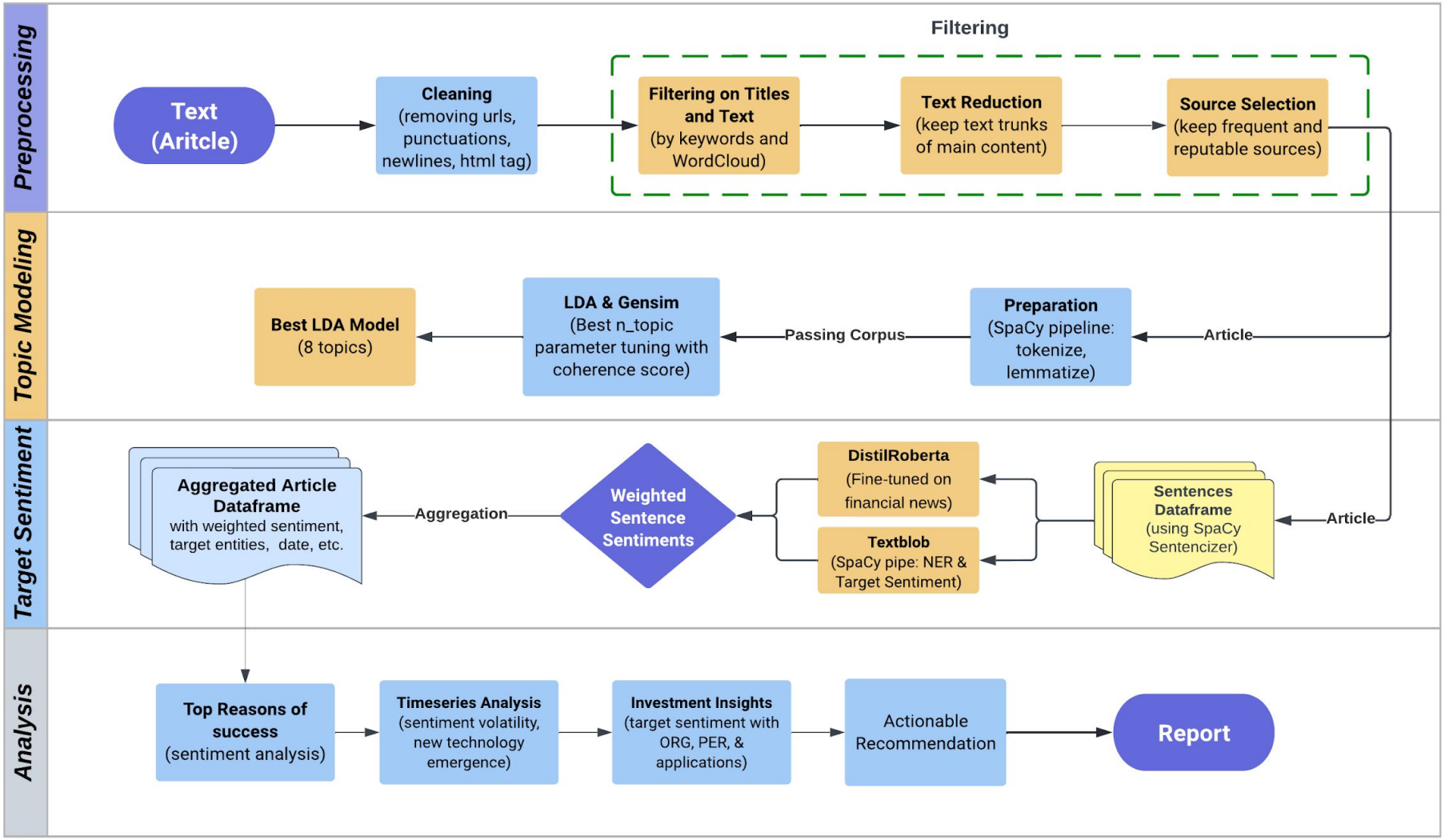
Technology Firms & Industries

Tech firms should continue to innovate while ensuring AI and data ethics, which are critical for user trust. They should focus on advancing AI capabilities, refining user-centric AI applications, and promoting transparent and explainable AI. Collaboration with traditional industries for digital transformation can also open new avenues for growth and societal impact.

Methodology and Data Source Overview

Data Source

- 200,332 news articles
- All in English
- Title, text, & url
- AI-related Topics
- Period:
  - Start: Jan, 2020
  - End: Apr, 2023



More on Text Preprocessing, EDA & Sentence-Article Aggregation

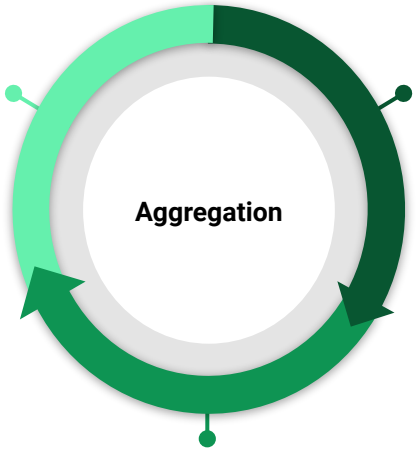
Data Preprocessing

- **Cleaning**: removing urls, html tags (e.g. /n /a), special punctuations (e.g. √·§¶)
- **Filtering**:
  - **By Keyword**: title and text should contain keywords like AI, Data Science in a customized list
  - **By trunk length**: split* article to trunks of sentences and remove trunks with extreme length (either too long or too short) to reduce irrelevant information (e.g. advertisement)
  - **Source Filtering**: use domain name from url as source; only keep frequent publishers**; also keep reputable sources (e.g. FT Times, Wall Street Journal)

Basic Exploratory Data Analysis

- Null value check: no null values
- Length: (per article): averagely 54 trunks, 50% percentile is 27 trunks
- Domain: 5k+ domains (sources) in total; 180k+ articles in total; 352 frequent publishers (about 6% of total sources), but they produce about 70% of the articles

*Note: split with patterns (e.g. “all right reserved” and “\n\n”; normally we have 5+ sentences per trunk
**Note: If one domain/source has more than 100 articles, we call it frequent publisher



	text	title	date
	Decentralized Machine Learning Reaches Market Cap of 15,919.00 DML Enterprise Leader Daily Ratings News for Decentralized Machine Learning Complete the form below to receive the latest headlines and analysts' mendations for Decentralized Machine Learning with our free daily email newsletter: Follow EnterpriseLeade Recent Posts GE Announces Its Plan To Sell The Distributed Power Business To Advent International Ebang Communication Resorts To The Filing Of An IPO In Hong Kong How to Open DAA, ...	Decentralized Machine Learning Reaches Market Cap of $15,919.00 (DML) - Enterprise Leader	2020-01-01
	Artificial Intelligence AI in Marketing Market Scope, Size, Share, Trends, Forecast by 2026 Market Reports Observer Skip to content 488 E 11th Ave, Eugene, OR 97401, USA 1 541 687 2347 email protected Market Reports Observer Artificial intelligence AI is the ability to process information in a way similar to the thought process of humans in learning, decision making, and problem solving. Enterprises are now identifying the value associated with integrating AI into their business processes. ...	Artificial Intelligence (AI) in Marketing Market Scope, Size, Share, Trends, Forecast by 2026 – Market Reports Observer	2020-01-01

Major Topics: LDA with Gensim

Major Topics:

- 8 major topics are identified with second highest LDA coherence score*
- Most obvious patterns: generative AI, ChatGPT, Tech Firms, Application of AI in business & services, AI & Analytics in healthcare
- Two example topics are chosen for validation
  - topic# 6 & 1: major tech firms and applications of AI (especially in ChatBot)
  - topic#3 & 5: application of AI in research (especially in clinical science)

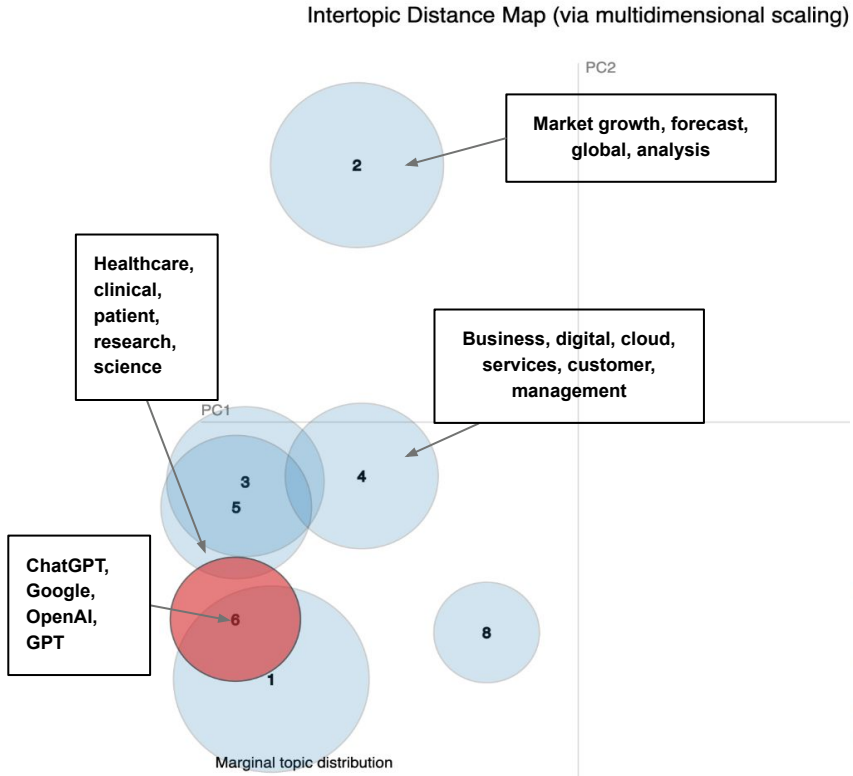


Figure: Visualized LDA (removing outlier topic; only show selected boxes of topics due to limited space

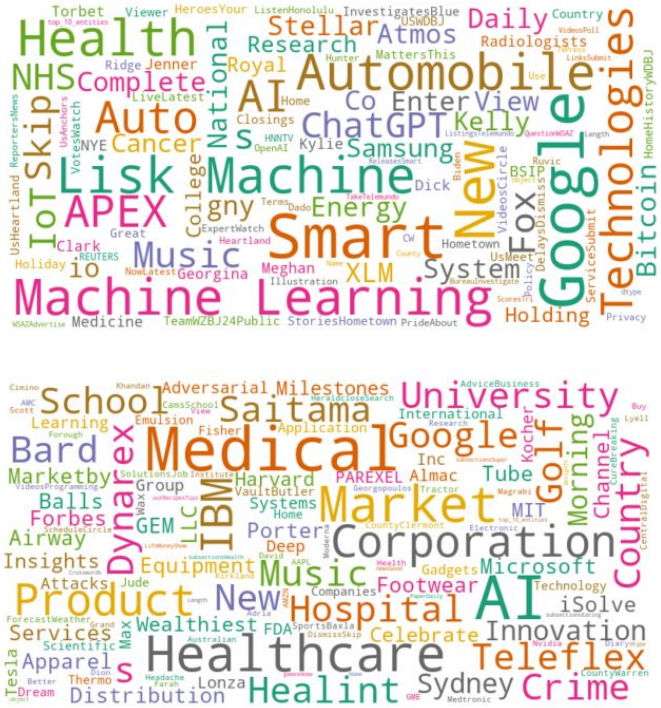
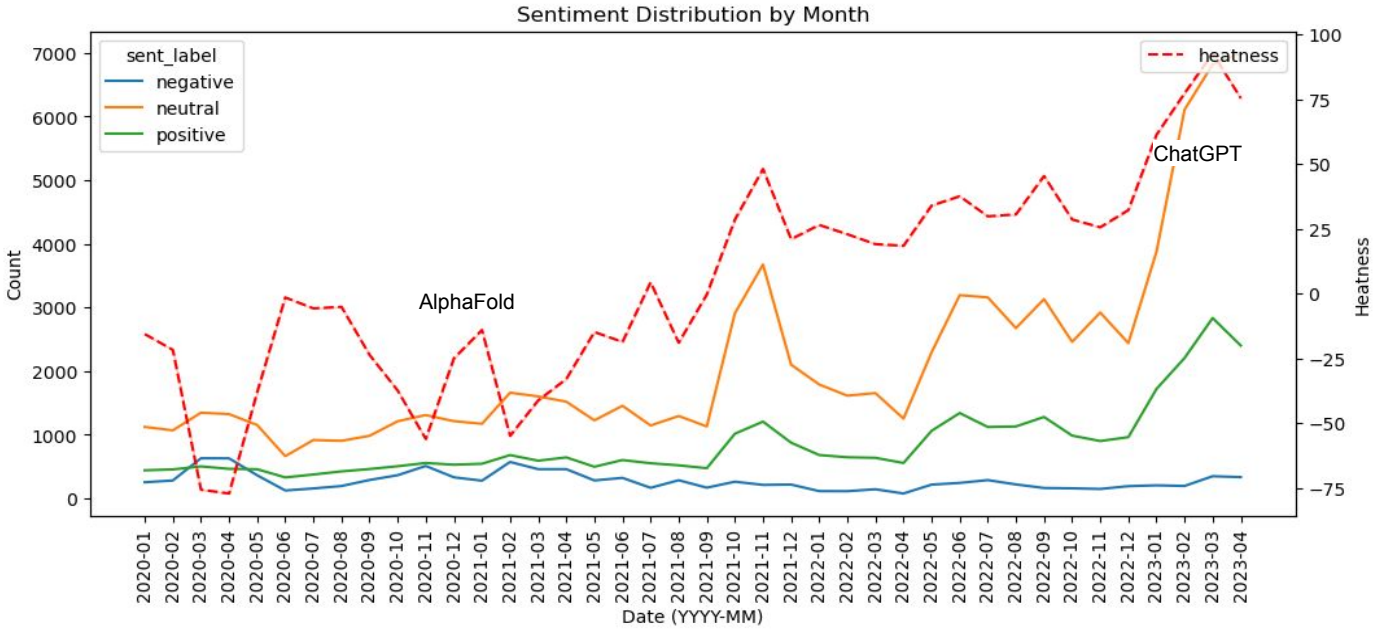
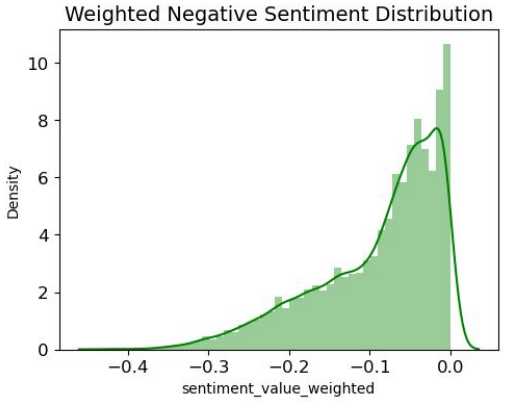


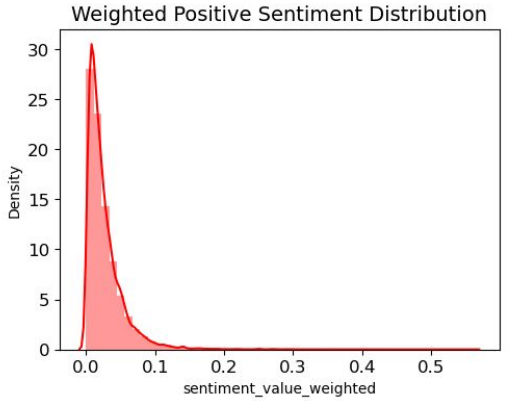
Figure: Word Cloud of Topics (above: Topic 6 & 1, below: Topic 5)

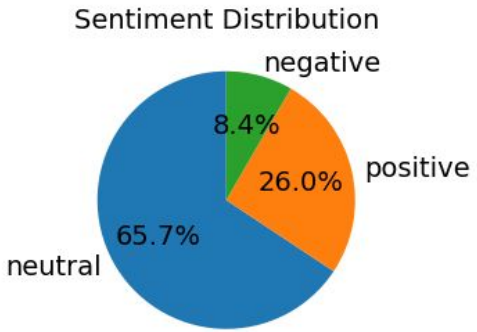
*Note: the model with highest coherence has fewer topics, thereby being less interpretable

Sentiment Volatility and Distribution









Sentiment volatility over time

- The ratio of market sentiment is mostly neutral and a small proportion of positive sentiment; only a few has negative sentiment
- The trend of market sentiment is upward, based on aggregate sentiment and volume of discussions (aka. count of articles)
- Heatness
  - Defined as Count-of-Articles * Average-Monthly-Sentiment
  - Works well to capture market heatness (e.g. COVID-19 impact & Microsoft's 36% layoff at early 2020; ChatGPT's viral launch at Nov, 2022)

Distribution of sentiment and Aggregation Methodology

- The breakdown of overall market sentiment distribution is: neutral (65.7%), positive (26.0%), and negative (8.4%), which is in accordance with the market heatness figure above

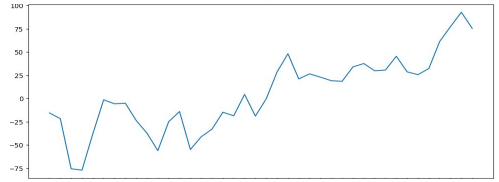
Aggregation of Sentiment:

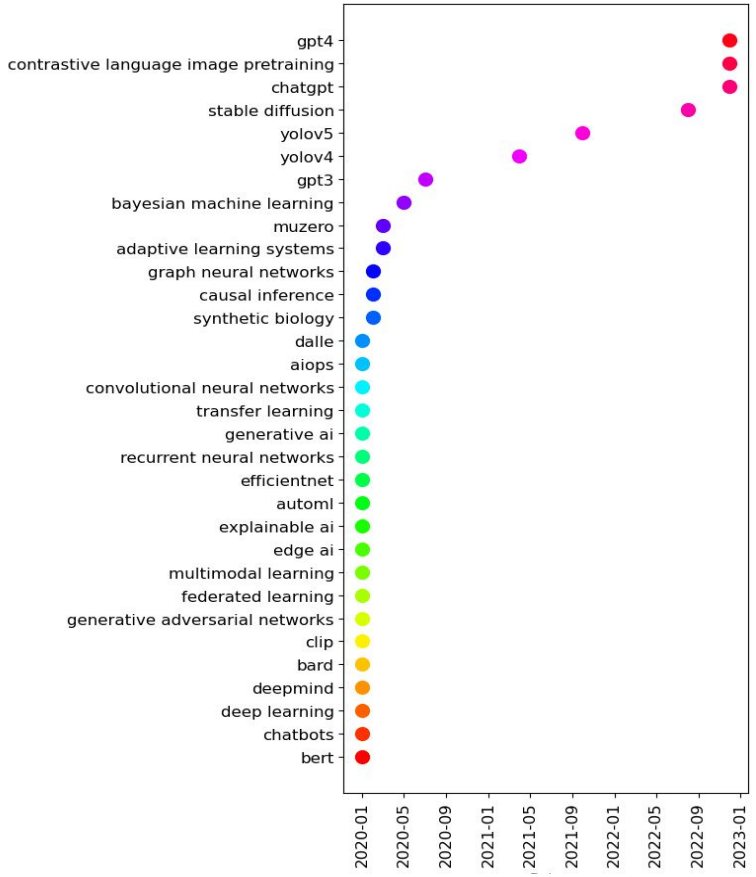
- Articles are split into sentences with Spacy Sentencizer)
- Two sentiment models are used: Textblob for target sentiment and Distil-Roberta fine-tuned on financial news
- Article sentiment is weighted average of the aggregated sentence sentiments with the two models

Descriptive tones in news article affects the overall sentiment

Emergence of New Technologies: A Timeline

Figure: Timeline of new technology related to AI (above: market heat graph with same-scale time axis





Timeline of Newly introduced AI technologies and solutions since January, 2020

- New technologies are continuously entering as trending topics in both academia and industry

Takeaways:

- Market heat: as defined in last page, measured by average monthly sentiment and counts of discussion (aka. news publication)
- **Asynchronous**: the time that new technologies are introduced often has a lag before it becomes a hot topic in public discussion (i.e., GPT3 was introduced in 2021)
- **Delays from theory to product**: There tends to be a period for the industry to follow-up with the theoretical development in AI; it takes extra time for converting theories to business-grade products
  - For example, for deep learning (i.e., generative neural networks, recurrent neural networks, convolutional neural networks), the technologies are introduced before Jan, 2020. But its popularity came up with a peak in 2022 due to the launch of Stable Diffusion.
- **Technical explosion of Natural Language Processing (NLP) and Large Language Models (LLMs)**:
  - although the field of NLP and LLMs are continuously evolving (i.e., GPT3, Bert), the market went viral about this topic in late 2022 as the product ChatGPT was introduced by OpenAI; **GPT-based business explosion** occured due to its outstanding capabilities to understand natural language and generalize in different occasions like writing, conversation, and even coding.
- **Interoperability**: the success of *transformer* architecture not only improved NLP, but also extended to computer vision
- **Potential "rising stars"**: based on LDA topic modeling and this timeline, some potential hot topics of AI solutions & technologies might associated with Multi-modal Model with language and image preprocessing, interpretable machine learning, beyansian causal inference

Investing in AI: Stories of Risks and Opportunities

Reasons for Successful AI Solution and Technologies

- **Data Availability & Data Quality:** Access to vast, high-quality datasets is crucial. The more relevant, accurate data AI has, the better it can learn, make predictions, and deliver effective solutions.
- **Sufficient Fundings in R&D:** Adequate investment in research and development allows for exploring innovative approaches, employing top talent, and accelerating the progress in AI technology and applications.
- **Level of Technical Maturity and Automation:** Successful AI relies on advanced, reliable tech infrastructure and the capacity for automation, reducing human error and increasing efficiency.

Reasons for Failure in Adopting AI Solutions and Technologies

- **Legal or Regulatory Constraints**: Legal barriers or strict regulations can hinder AI adoption, particularly when data privacy laws limit access to necessary data.
- **Ethical Concerns**: Ethical issues, like AI bias or lack of transparency, can pose significant challenges to the successful integration of AI into a company's operations.
- **Reliance on Human Interaction**: Some tasks require a human touch for nuanced understanding or empathy, which AI may struggle to fully replicate, affecting its successful adoption.

Case 1: Success

VIDA

VIDA Diagnostics

VIDA's growing recognition as a leader in AI-based imaging is founded on its suite of software solutions, which empower healthcare organizations and physicians in the fight against lung and respiratory disease.

+

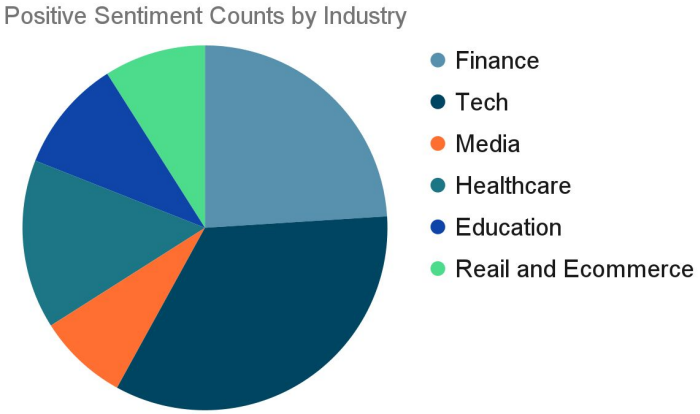
Remote Driving Faces Ban in UK

Delivery of rental cars using remote driving could be outlawed after Law Commission recommendations in England and Wales

Case 2: Regulatory Risk

Target Sentiment: Successful AI Firms and Fields to Avoid

Top 100 firms with positive aggregated sentiments



Successful AI Firms:

- **Example**: Google, Intel, OpenAI, Baidu, Microsoft, Amazon, Tesla, Facebook, Apple, Nvidia
- **Rationale**: these firms are pioneering developers in either software-level AI solutions or hardware (e.g. GPU supplier Nvidia), or even in both fields. In other words, they are the cornerstones

Fields to Avoid

- **Example***: Healthcare(hospital, med equipment), School, Innovation (music, media)
- **Rationale**: current state-of-art AI is not sufficient to handle tasks that heavily rely on:
  a. human senses and physical interaction (e.g. nursing at hospital)
  b. communication with intellectual capability (e.g. teaching)
  c. creative and innovative work (e.g. music, art)

*Note: healthcare is a huge industry so it is possible to see AI-related terms have both positive and negative sentiment in healthcare industry

Top 100 fields with negative aggregated sentiments

Showcase Of How Target Sentiment Is Calculated

name	label	cumulative_sentiment	frequency	avg_sent_per_article
Artificial Intelligence Systems	ORG	-244.8	26	-9.415385
Artificial Intelligence Advisory Service	ORG	-215.4	36	-5.983333
Artificial Intelligence in Construction	ORG	-133.8	27	-4.955556
Automotive Artificial Intelligence	ORG	-445.2	90	-4.946667

Three Most-mentioned Names with potential reasons:

- Elon Musk (Tech Firm CEO & celebrity), Joe Biden (Politician), Sam Altman (OpenAI)

References

1. Romero, M. (2022). Mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis. Hugging Face. Retrieved May 27, 2023, from https://huggingface.co/mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis
2. Ghanoum, T. (Dec, 2021). Topic modelling in Python with Spacy and Gensim. Towards Data Science. Retrieved May 27, 2023, from https://towardsdatascience.com/topic-modelling-in-python-with-spacy-and-gensim-dc8f7748bdbf