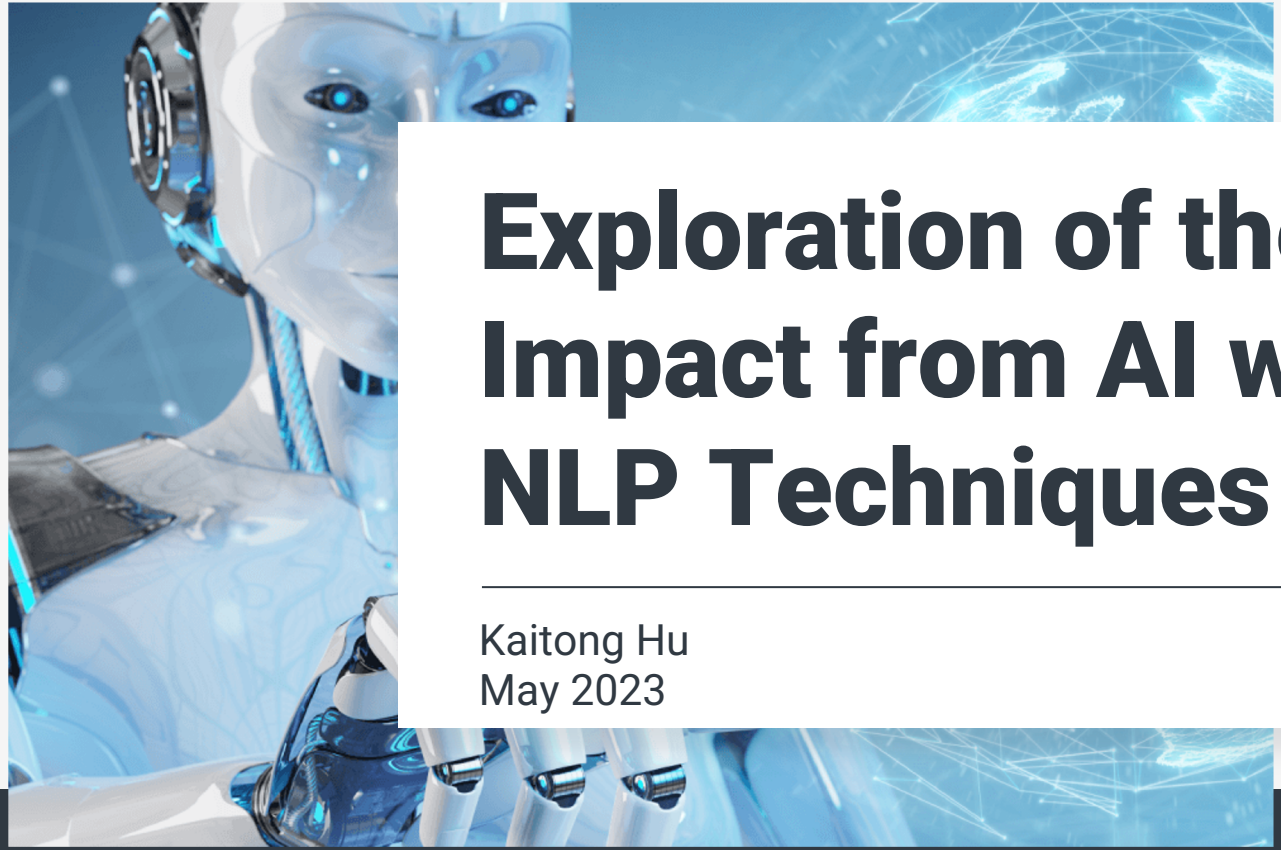


Exploration of the Impact from AI with NLP Techniques

Kaitong Hu
May 2023

Table of Contents

01 Executive Summary

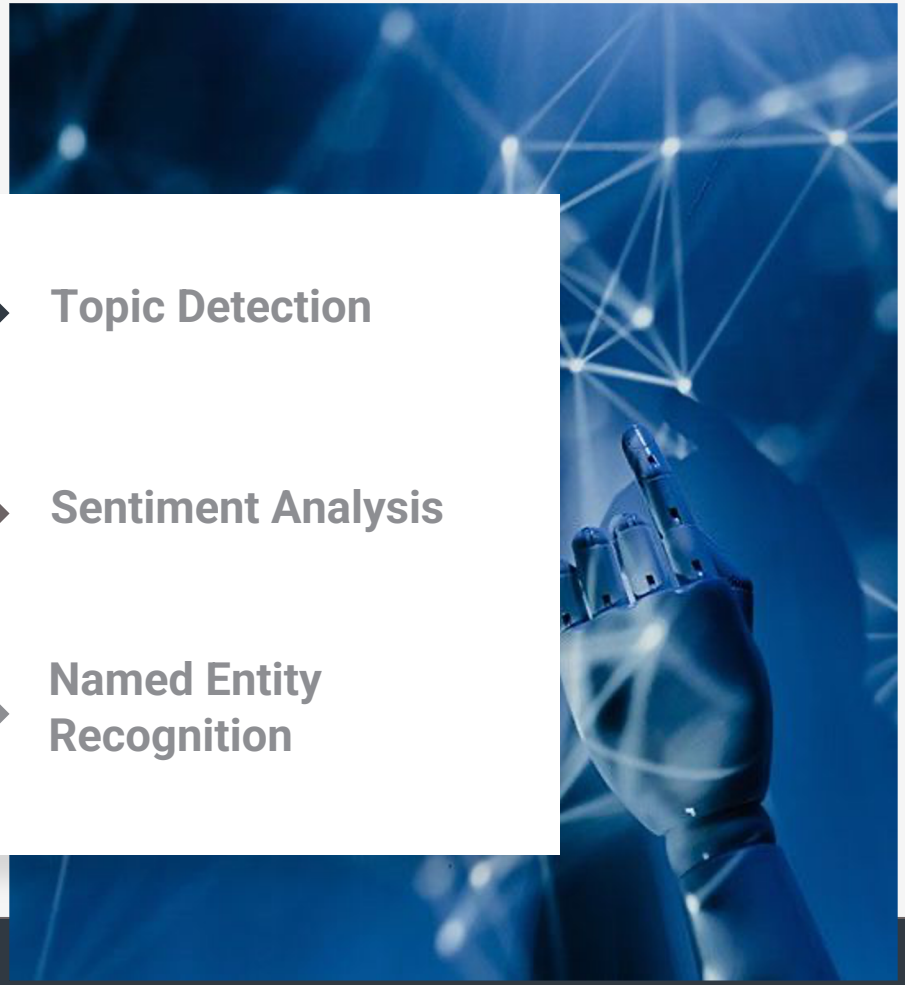
02 Recommendations

03 Data & Methodology

04 Topic Detection

05 Sentiment Analysis

06 Named Entity Recognition



Executive Summary

Background

The rapid development of AI technologies has transformed our life in numerous ways. These technological advancements have led to increased efficiency, automation of tasks, and improved well-informed decision-making process. However, this phenomenon has also sparked concerns and apprehensions across industries regarding the potential impact on job security and employment opportunities.

Objective

This project aims to employ a wide range of NLP techniques, including Topic Modeling, Sentiment Analysis and Named Entity Recognition, to identify types of tasks and jobs that are most likely impacted by AI. Further to this, to formulate actionable recommendations by leveraging insights derived from the analysis.

Conclusion

Analysis reveals that the Healthcare sector exhibits predominantly positive sentiment, followed by the Agriculture field. This can be attributed to the recognition of AI's potential in revolutionizing healthcare delivery and improving patient outcomes. In contrast, industries, such as publishing, art, and design, that rely heavily on human creativity as a core input are negatively impacted by this development in some ways.

Source Data Overview & Clean-up

DATA SOURCE

Original dataset contains 200332 rows * 5 columns since 2020-01

	url	date	language	title	text
0	http://en.people.cn/n3/2021/0318/c90000-983012...	2021-03-18	en	Artificial intelligence improves parking effic...	\n\nArtificial intelligence improves parking e...
1	http://newsparliament.com/2020/02/27/children-...	2020-02-27	en	Children With Autism Saw Their Learning and So...	\nChildren With Autism Saw Their Learning and ...
2	http://www.dataweek.co.za/12835r	2021-03-26	en	Forget ML, AI and Industry 4.0 – obsolescence ...	\n\nForget ML, AI and Industry 4.0 – obsolesce...
3	http://www.homeoffice.consumerelectronicsnet.c...	2021-03-10	en	Strategy Analytics: 71% of Smartphones Sold Gl...	\n\nStrategy Analytics: 71% of Smartphones Sol...
4	http://www.itbusinessnet.com/2020/10/olympus-t...	2020-10-20	en	Olympus to Support Endoscopic AI Diagnosis Edu...	\n\nOlympus to Support Endoscopic AI Diagnosis...

Number of rows are reduced to 127k

Initial Clean-up

Drop duplicate rows, remove URLs with Regex, and drop irrelevant columns, such as language

Extract Main Text Body

Define keywords (e.g. “Related Stories”, “You May Also Like”, etc.) to locate the main text body, and remove redundant words

Check Relevancy

Drop articles that are irrelevant to ML, DS and AI

Tokenization

Use NLTK to lemmatize the words and remove stopwords from the text

Planning & Methodology

TOPIC MODELING » **SENTIMENT ANALYSIS** » **TARGETED SENTIMENT WITH NER** » **ANALYSIS**

LDA with KTrain Bertopic

Two topic modeling techniques are used to identify major topics

Customized Vader

SentimentIntensityAnalyzer() is updated with customized keywords

RoBerta Transformer

Fine-tuning a transformer pipeline pre-trained on financial news articles

SpaCy

SpaCy are used to extract 'ORG', 'PERSON', 'GPE' and 'PRODUCT' from the text

WordCloud

It is used to visualize the frequency of identified entities

Text Summarization

Distilbart-CNN transformer pipeline is used to synthesize insights

Topic Detection

Tools: LDA with ktrain, Bertopic

3 MAJOR TOPICS

TECHNOLOGY

Machine Learning
Artificial Intelligence
ChatGPT
GPT
LLM
Generative AI
etc.

INDUSTRY

Healthcare & Biotech
Agriculture
Military & Defense
Business & Finance
Media Creation
Automotive
Research & Education
etc.

LOCAL NEWS

Regulations & Legal
Climate Change
Local Events
etc.

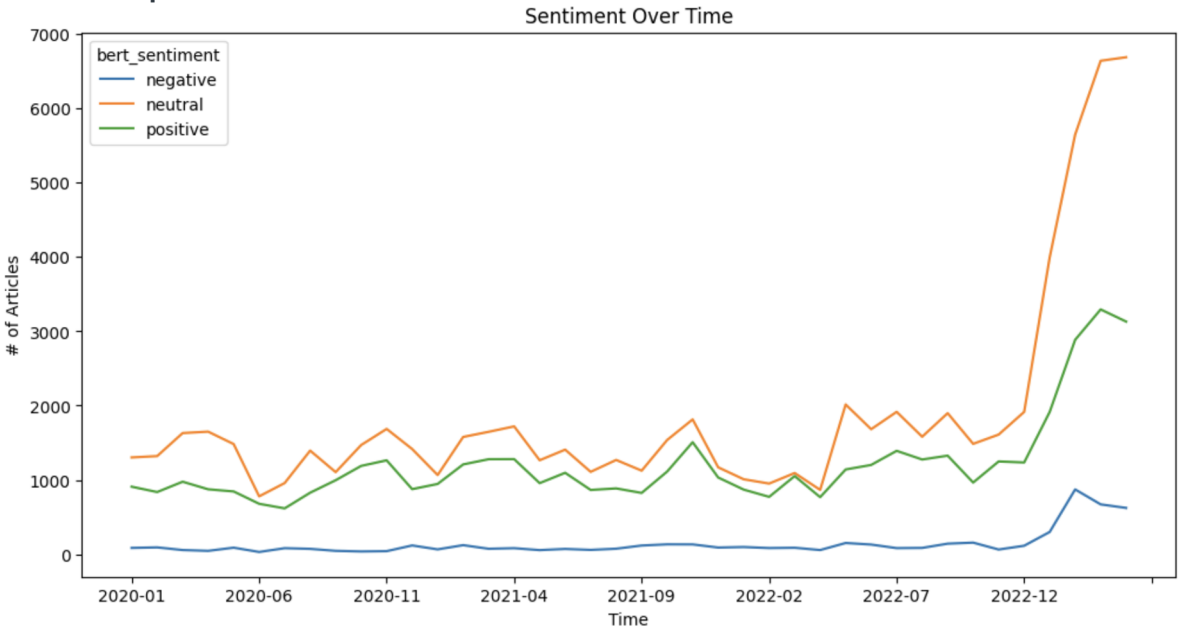
```
Topic -1: ai, new, model, company, technology, the, da
Topic 0: market, report, growth, analysis, global, fo
Topic 1: patient, cancer, clinical, health, medical,
Topic 2: market, analysis, report, growth, forecast,
Topic 3: china, chinese, baidu, beijing, ernie, alibal
Topic 4: watch, india, live, says, to, from, telecast
Topic 5: stock, investor, trading, billion, year, com
Topic 6: vehicle, car, road, traffic, driver, driving
Topic 7: customer, solution, nvidia, platform, edge,
Topic 8: und, zu, auf, die, sie, im, bewerten, stoxx,
Topic 9: trends, nyse, forecast, reports, dagoretti,
Topic 10: student, education, science, school, course
Topic 11: drone, military, defense, aircraft, air, do
Topic 12: eu, patent, regulation, law, inventor, euro
Topic 13: market, report, analysis, murphy, artificia
Topic 14: food, restaurant, recipe, grocery, chef, pi
Topic 15: healthcare, medical, medicine, market, imag
Topic 16: starfilled, course, filled, certificate, he
Topic 17: ibm, corporation, intel, nyse, microsoft, da
Topic 18: paid, brandvoice, program, richest, forbes,
Topic 19: funding, round, capital, ventures, company,
Topic 20: weather, inappropriate, post, comment, lehig
Topic 21: data, organization, business, ml, process,
Topic 22: student, school, teacher, chatgpt, essay, e
Topic 23: upthis, recaptcha, newslettersign, notice,
Topic 24: india, discourse, partnership, submit, adve
Topic 25: music, song, artist, lyric, sound, musician
```

Sentiment Analysis with RoBerta Transformer

Tools: Customized Vader, Roberta Transformer Pipeline

The sentiment analysis will be based on results from **the RoBerta transformer pipeline**, because manual inspection of the results shows that it outperforms the customized Vader

Generally, the public holds predominantly **favorable opinions** towards the advancement of AI. However, it is noteworthy that there has been **a surge in the number of positive and negative articles since November 2022**, which coinciding with the release of ChatGPT and other AI technologies.



Sentiment Overtime (Cont.)

Definition:
pos_rate = # of positive articles/ # of related articles
neg_rate = # of negative articles/ # of related articles

Upon closer examination of the attitudes towards AI across various industries, it is interesting to note that AI advancement is evidently welcomed in the **healthcare and biotechnology** field as indicated by its highest positive rate. Conversely, sectors that often rely on the creative input, critical thinking, and artistic expressions of individuals, for instance, **the publishing and media creation** industries, have received comparatively lower positive rates.

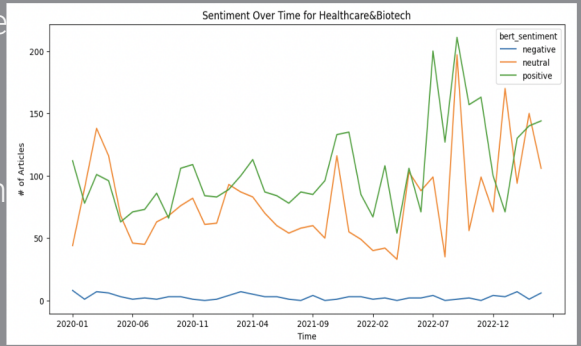
Industries such as automotive and telecommunication exhibit a moderate level of sentiment regarding the applications of AI. a comprehensive examination of the attitudes held by these industries will be discussed in subsequent slides.

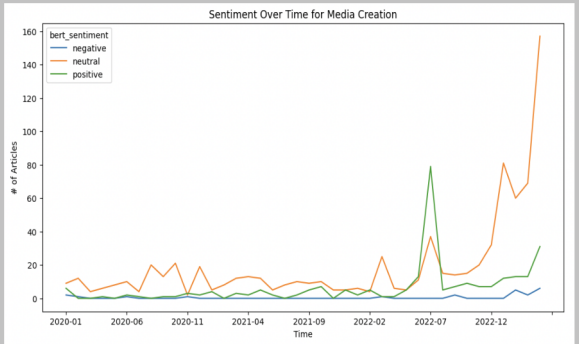
Industry	average_sentiment	pos_rate	neg_rate
Healthcare&Biotech	0.694068	0.558262	0.014263
Agriculture	0.900165	0.428130	0.015456
Others	0.771219	0.388248	0.039194
Military & Defense	0.638020	0.356274	0.108835
Automotive	0.774084	0.344052	0.034468
Telecommunication	0.745128	0.278901	0.056184
Technology	0.500261	0.277656	0.134127
Education	0.789510	0.252174	0.029130
Media Creation	0.748846	0.244860	0.019626
Finance	0.855573	0.223977	0.082795
Publishing	0.597318	0.185026	0.059811

Sentiment Overtime (Cont.)

Different industries are impacted in various ways

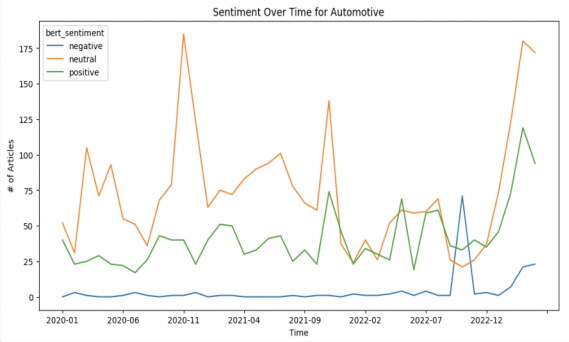
The quantity of positive healthcare-related articles consistently surpasses that of negative articles, which implies a prevalent favorable attitudes within this sector.

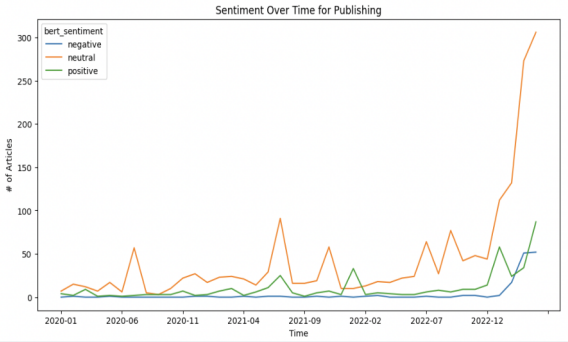




The media creation sector showed an optimistic sentiment in July 2022, which coincided with the release of DALL-E2.

The automotive sector exhibited a relatively negative sentiments in October 2022





The publishing industry demonstrated an increasing pessimistic sentiment since december 2022, potentially impacted by the release of ChatGPT

Sentiment Overtime (Cont.)

Text summarization technique, such as Distilbart-CNN, are applied on articles with the highest positive/negative rate to identify reasons for success/failure

```
[{'summary_text': ' The global artificial intelligence in healthcare market '
                  'reached a value of nearly $7,833.6 million in 2021 . The '
                  'market is expected to grow from $7.6million in 2021 to $ '
                  '33,107.6m in 2026 at a rate of 33.4% . The software market '
                  'was the largest segment of the artificial'}]
[{'summary_text': ' Naver's automatic AI response system 'Clova Care Call' '
                  'automatically calls those who are being actively monitored '
                  'twice a day at 10 am and 3 pm . The service will be '
                  'preemptively started in six gus and guns . The call will '
                  'check for abnormalities of the subject such as fever and '
                  'respiratory symptoms and'}]
```

- **Clear Use Cases and Measurable Outcomes**
- **Established Regulatory Frameworks**
- **Availability of Diverse Data**

Healthcare & Biotech

SUCCESSFUL

```
[{'summary_text': ' ChatGPT has provided a global hit by demonstrating an '
                  'ability to generate essays, poems and conversations from '
                  'the briefest of prompts . But Italian regulators said last '
                  'Friday that the firm had no legal basis to engage in '
                  'massive data collection . European authorities including '
                  'those of France, Ireland'}]
[{'summary_text': ' The Defense Department has been on a journey in earnest to '
                  'take advantage of artificial intelligence capabilities '
                  'since it launched the Joint AI Center (JAIC) in 2019 . In a '
                  'multi-domain environment, DoD leaning on data, AI to make '
                  'faster decisions.'}]
```

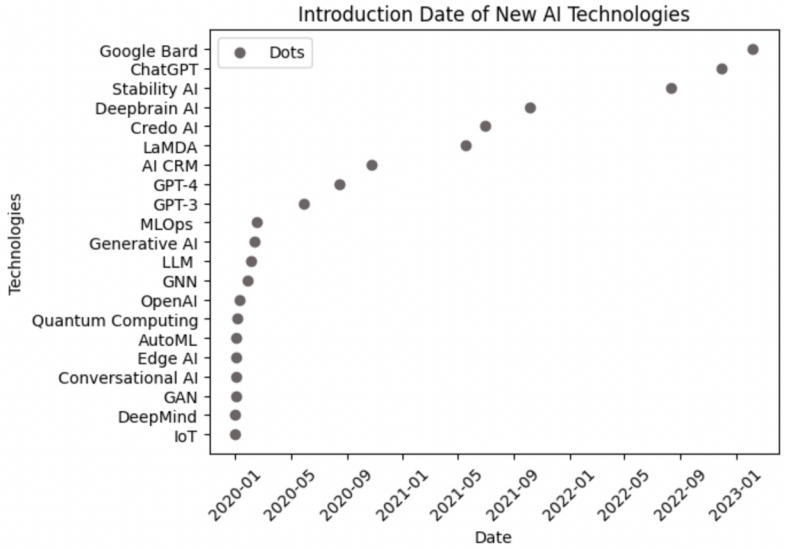
- **Ethical Considerations**
- **Insufficient Data Quality**
- **Intricate Nature of Content Creation**

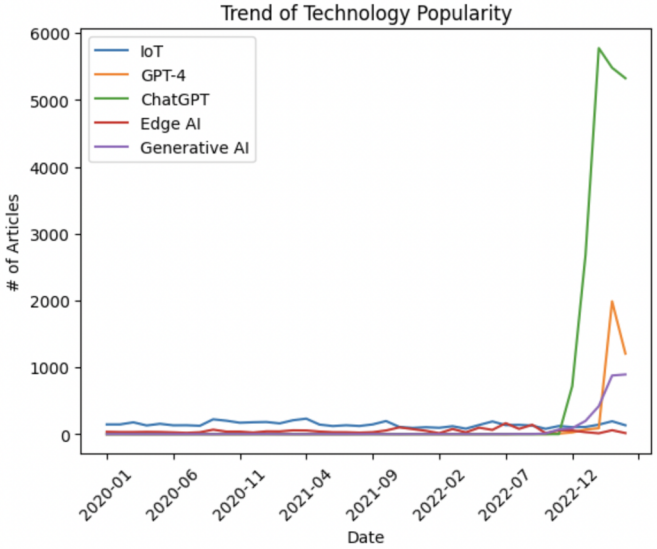
Publishing

FAILURE

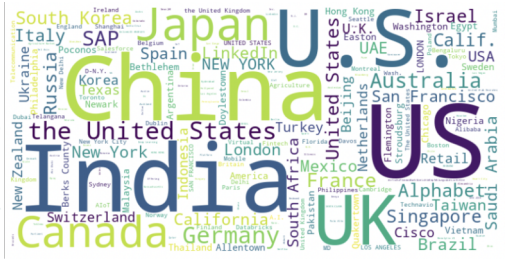
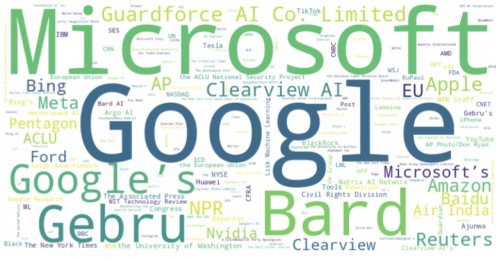
Entity Identification

More mature technologies, including IoT, have been a topical discussion for several years, and the volume of related articles has remained relatively stable. In contrast, **generative AI technologies**, exemplified by ChatGPT, have started to gain public attention since late 2022. Algorithms that laid the ground for such technologies (e.g. LLM) have been discussed for a considerable period, but not until 2021 have these been put into practical implementation.





Entity Identification (Cont.)

	ORGANIZATION	PERSON	LOCATION
POSITIVE			
NEGATIVE			

A notable observation is that certain entities appear in both the positive and negative word clouds, indicating that these terms are discussed and interpreted differently across various contexts. It becomes challenging to determine the precise attitudes of organization, individuals, or locations towards AI without taking the specific topics addressed in the corresponding articles into consideration.

Targeted Sentiment

Healthcare & Biotech

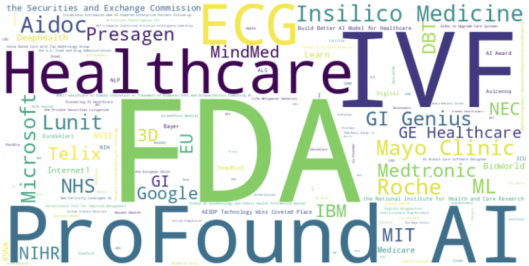
SUCCESSFUL

The word clouds generated provide some insights into what types of companies in the Healthcare & Biotech field should invest in Data Science/ AI initiatives. Specifically, organizations that:

1. Have vast amount of patient data
2. Actively participate in drug discovery
3. Are specialized in medical device/digital health
4. Are at the forefront of medical research and innovation

On the right are some technologies these organizations can potentially invest in to sustain their competitive advantages and create social values.

Positive Organizations in Positive Healthcare-related Articles



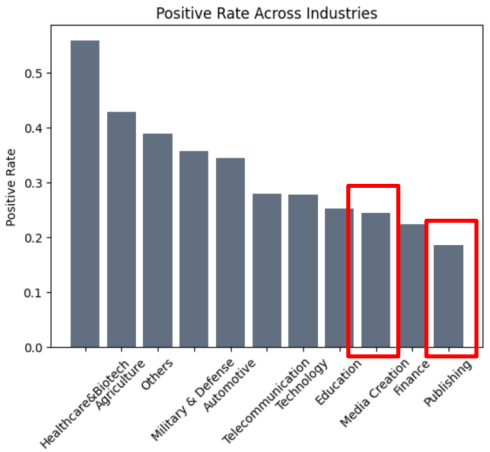
Positive Products in Positive Healthcare-related Articles

Targeted Sentiment (Cont.)

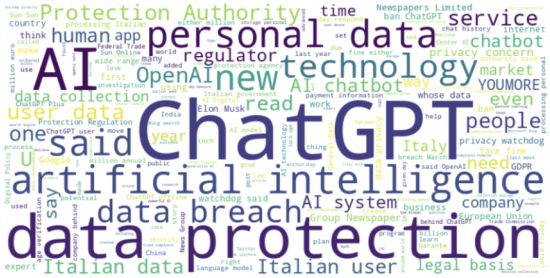
Publishing

FAILURE

Publishing, along with industries such as Media Creation, thrives on human creativity and originality to generate content and deliver it to the audience. While Generative AI has made noticeable advancements in content generation, it cannot completely transform these industries in the following ways:



- AI generated contents lack unique personal experiences, cultural nuances, and individual storytelling techniques.
- Contents generated by these type of technologies sometimes fail to establish emotional connection with readers
- AI technologies might not be able to meet the changing reader preferences and societal trends



Actionable Recommendations

As indicated by the below graph, ethics use of AI technologies has been a controversial topic for years, and it can possibly hinder the advancement of AI. Appropriate measures should be undertook to address these problems.

Generated from Topics Related to the Concerns

ORG.

Establish Ethical and Regulatory Frameworks
Draft and implement clear ethical guidelines for the responsible use of AI technologies to foster public trust and ensure AI technologies are deployed in a manner that aligns with social values, consequently boost worker productivity.

GOV.

Promote International Collaboration
International collaboration and knowledge sharing should be encourage to accelerate AI development. This can be achieved through joint research programs and technology exchange initiatives.

PER.

Tailored Education & Training
Design AI/DS-integrated curriculums to train students in AI fundamentals and practical skills to produce skilled professionals who can contribute to the development of AI.