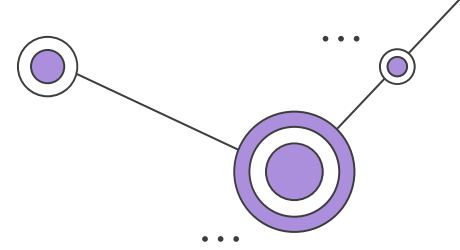


AI Impact Analysis: Transforming Industries and Job Markets

Boya Zeng

Executive Summary



Overview

This project will involve a comprehensive analysis of a large dataset of news articles to identify how AI is expected to impact various industries and job lines in the near future.

objective

Identify what types of tasks and jobs are most likely to see the biggest impact from AI by extracting meaningful insights from unstructured text.

Provide actionable recommendations on what can be done with AI to automate the jobs and / or improve employee productivity.

Insights

New technology like ChatGPT and Google Bard was discussed more often after Jan 2023.

Some jobs/industries impacted negatively are finance, art, etc, which need more advanced skills, interpersonal skills and creativity.

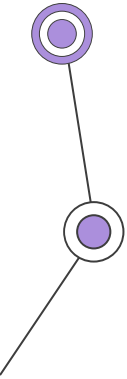
AI Integration has negatively impact Healthcare and Medicine, Finance, Technology, and Entertainment Industry over time

Recommendations

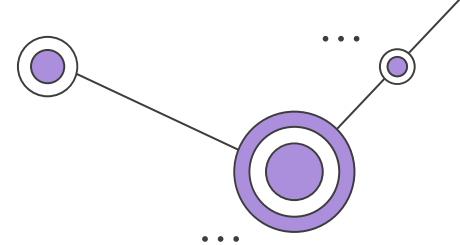
Academic institutions can educate the broader community and stimulate public interest.

Companies should invest in AI and machine learning for product enhancement.

Governments should make comprehensive laws and regulations about AI.



Methodology and Source Data Overview



Source Data

	url	date	language	title	text
0	http://galusaustralis.com/2020/02/486473/legaltech-artificial-intelligence-market-2019-technolog...	2020-02-26	en	LegalTech Artificial Intelligence Market 2019 Technology Advancement and Future Scope – Casetext...	LegalTech Artificial Intelligence Market 2019 Technology Advancement and Future Scope – Casetext...
1	http://newsparliament.com/2020/02/27/children-with-autism-saw-their-learning-and-social-skills-b...	2020-02-27	en	Children With Autism Saw Their Learning and Social Skills Boosted After Playing With This AI Rob...	Children With Autism Saw Their Learning and Social Skills Boosted After Playing With This AI R...
2	http://www.agoravox.it/Covid-19-un-messaggio-dai.html	2020-03-13	en	Covid-19: un messaggio dai ricercatori italiani ai colleghi stranieri - AgoraVox Italia	Covid-19: un messaggio dai ricercatori italiani ai colleghi stranieri - AgoraVox Italia

- ~200K news articles (about 900 MB)
- Topics: Data Science, Machine Learning and Artificial Intelligence
- Five column: url, date, language, title, text
- Time Span: 2020-2024

Methodology

- Article clean-up and filtering
 - Regular expression
 - Lemmatization
- Topic detection
 - LDA using ktrain
 - BERTopic
- Sentiment analysis (explicitly customized)
 - SpaCy
 - Customized Yelp Logistic Model
- Entity Identification
 - SpaCy
 - Regular expression
 - Named Entity Recognition (NER)
- Targeted sentiment identification
 - SpaCy
 - NER
 - Word Cloud Chart
 - Bubble Chart

Article Clean-up and Filtering

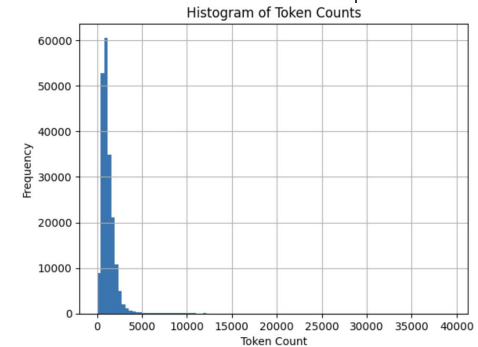
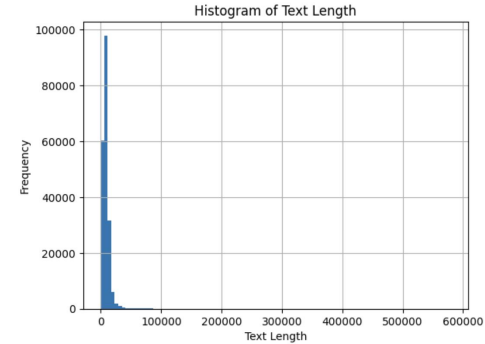
- Remove non-English News Article
- Remove duplicates articles
- Filter out too long/short article (Outliers fall off over 1.5 IQR)
- Use Regular expressions
- Remove newlines, tabs, white space, and remnants of web crawls,
- Remove Remove stopwords
- Remove Irrelevant Articles: based on keywords (relevant key word found by N-grams that appeared in the first 20% of the article) and context

For Topic Modelling

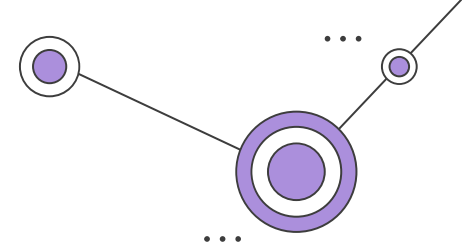
- Remove single-character (mostly punctuation)
- Lowercaseing
- Use Lemmatization to normalize text

For NER

- No Further Clean-up



Topic Detection & Summarization



LDA using Ktrain

topic:10	count:29276	gray press release prnewswire alert platform solution weather music country
topic:29	count:12586	like say people make work way thing image think human
topic:11	count:11293	market global analysis growth research forecast key trend player size
topic:17	count:11191	ago hour video weather story day app file sport local
topic:21	count:8966	government student risk school university national public president regulation country
topic:13	count:7697	email subscribe facebook comment sign newsletter twitter article search share
topic:23	count:7368	human model research researcher algorithm potential computer future development task
topic:24	count:7310	chatgpt model user google tool language chatbot generative image search
topic:5	count:7278	india video sport entertainment latest live lifestyle health indian story
topic:12	count:6938	generative organization process employee enterprise job customer work automation management
topic:0	count:6814	best review google apple mobile tech feature gaming game pro
topic:26	count:6755	sport home game woman food day team player family star
topic:2	count:6293	public schedule radio program community music npr facebook search twitter
topic:25	count:6251	digital event innovation startup partner global tech award future leader
topic:14	count:6205	health patient medical healthcare care clinical cancer disease hospital imaging
topic:22	count:5741	customer experience product user platform marketing brand conversational cooky sale
topic:27	count:5424	cloud nvidia edge application chip solution model computing customer software
topic:28	count:4801	share price bank market finance day stock china fund rate
topic:19	count:4600	stock share market investment trading fund investor price million analyst
topic:18	count:3543	openai microsoft ceo altman november musk march december google january
topic:3	count:3430	product entertainment release consumer resource general cision overview health public
topic:9	count:3428	security statement risk forwardlooking cybersecurity threat cyber result release future
topic:7	count:3087	republic release newswires south north press island ein presswire distribution
topic:16	count:2797	science energy project engineering design management supply climate development software
topic:1	count:2537	insurance legal law program paid credit best online quote court
topic:6	count:2000	vehicle camera smart car autonomous robot vision sensor safety robotics
topic:8	count:1932	nasdaq market icon symbol stock quote add solution available investing
topic:4	count:1591	drug discovery science medicine research protein cell workday die und
topic:15	count:540	cooky illinois maine pulse average iowa pennsylvania utah law360 marketing
topic:20	count:430	learning machine education student model course deep using learner training

These topics indicate a diverse interest across different domains with a strong emphasis on the integration of AI and technology into various aspects of different industry, business, society, and personal life.

Topic Summarization using LexRank

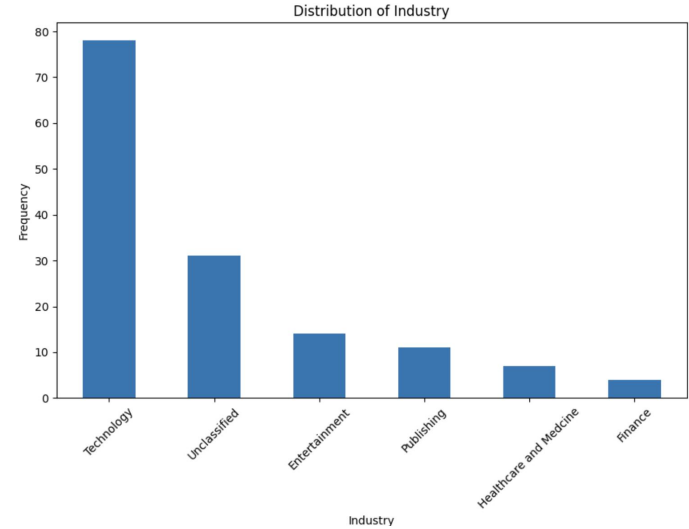
Example: topic 3

The document encompasses a myriad array of subjects, prominently featuring the announcement of Huawei's victory in a sci-fi AI application competition, underscoring the fusion of technology and creative storytelling. It delves into the evolving landscape of business and technology, spotlighting advancements in AI, data analytics, and green technology. The narrative extends to the realm of environmental consciousness, with discussions on sustainable transportation and AI's role in environmental stewardship. Noteworthy mentions include partnerships and innovations aimed at harnessing AI for enhancing fleet management, healthcare, and cybersecurity, pointing towards a future where AI not only revolutionizes industry practices but also contributes to societal and environmental well-being."

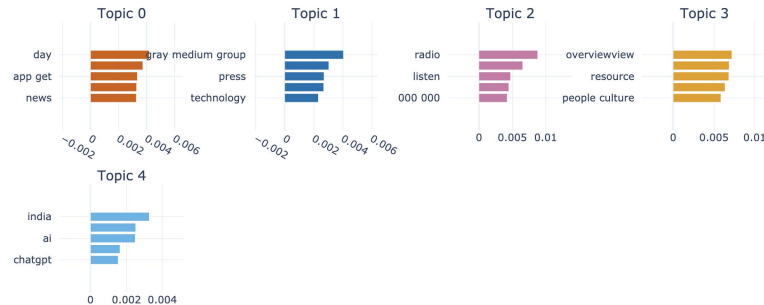
Topic Detection & Selection

BERTopic

	Topic	Name
0	-1	-1_ai_technology_2023_data
1	0	0_day_bestreviews_app get_hour ago video
2	1	1_gray medium group_pnewsire_press_inc
3	2	2_radio_donate_listen_volunteer
4	3	3_overviewview_consumer_resource_entertainment medium
...
294	293	293_ceo surveyed_generative ai_meet ceo artificial_ceo artificial intelligence
295	294	294_first alert weather_hubbard_news team_thing jefferson
296	295	295_3351212fcc_alert weather_voteselection_teamcontact usstation historygray
297	296	296_cnn_log cnn_cnn account_log cnn account
298	297	297_result view result_breaking news today_technology news automotive_gaming technology news



Topic Word Scores

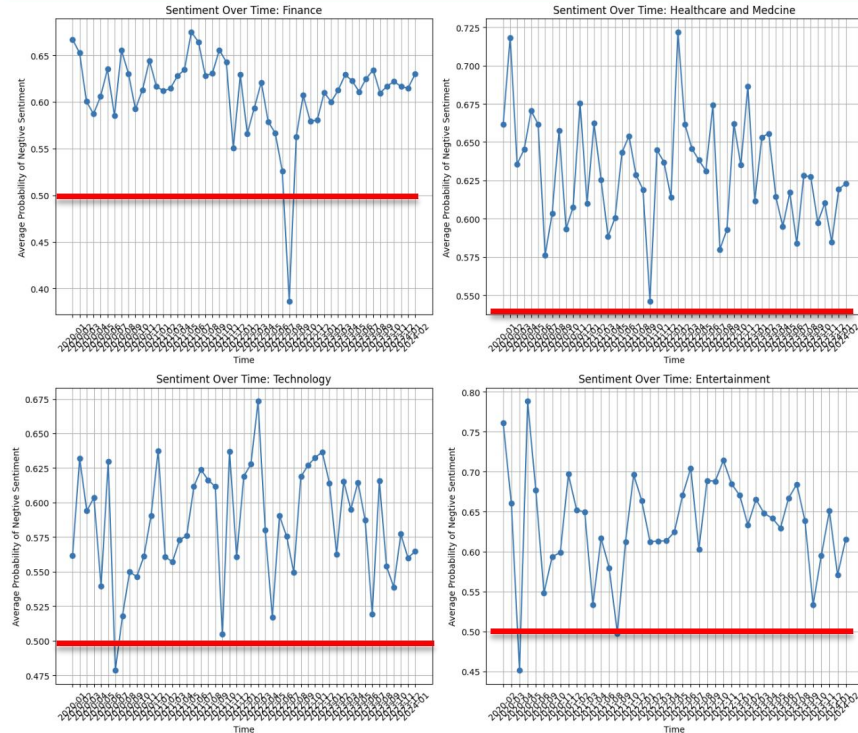


- Find total 299 topics
- Technology is the most prominent topic in the dataset, likely reflecting a high volume of discussion and coverage in the news articles.

Next Step :

- Identify topics related to AI integration, new technologies, and AI solutions & Industry

Sentiment Over Time for Industry AI Integration



Probability of Negative Sentiment over Time for Each Industry

Customized Sentiment Analysis

- Summarizing news content with ktrain for industry news
- Yelp+ Logistic model to analysis sentence-level sentiment
- Aggregate sentence-level Sentiment for each document
 - Mode sentiment class
 - Average of probabilities positive\ negative sentiment

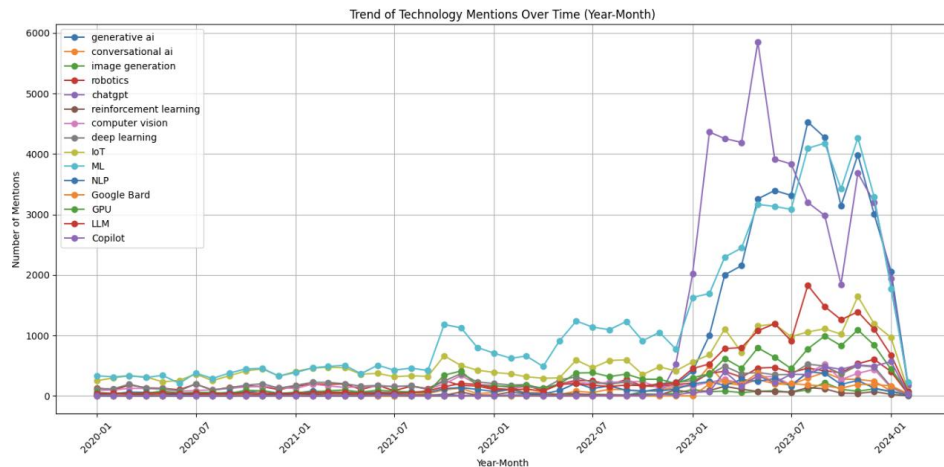
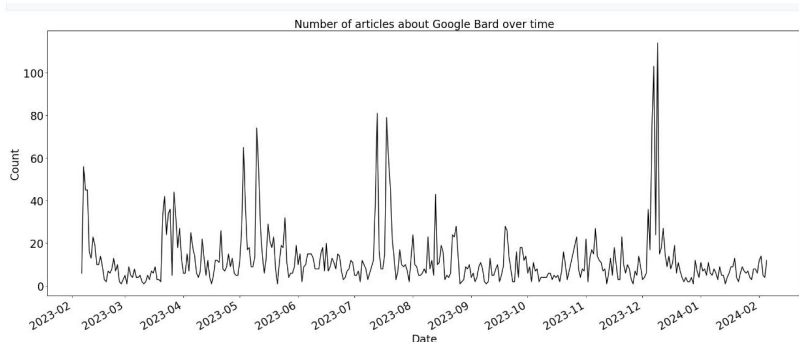
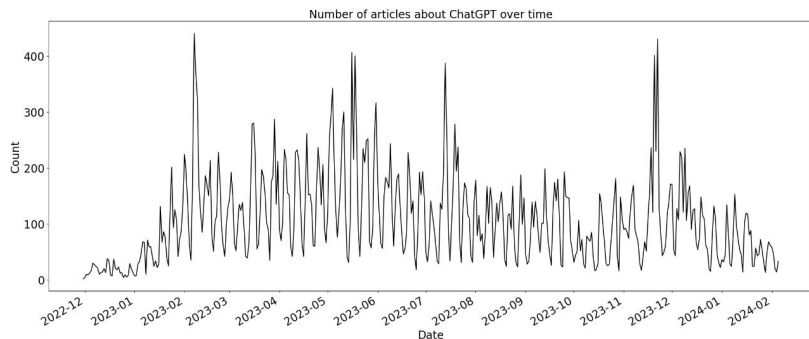
Sentiment Probabilities

- Probabilities of negative sentiment
 - when $> 0.5 \Rightarrow$ Negative Sentiment
 - when $< 0.5 \Rightarrow$ Positive Sentiment

Industry Insight

- AI Integration has negatively impact Healthcare and Medicine, Finance, Technology, and Entertainment Industry over time
- Those industries impacted negatively are finance, art, etc, which need more advanced skills, interpersonal skills and creativity, or like tech industry human positions could rapidly replcae by AI.

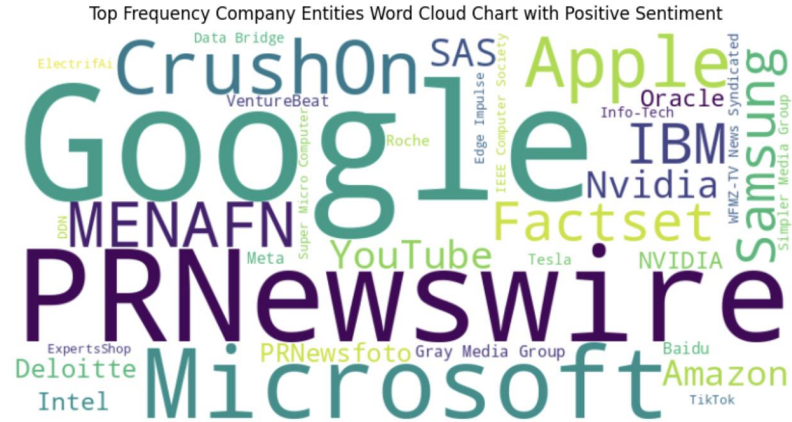
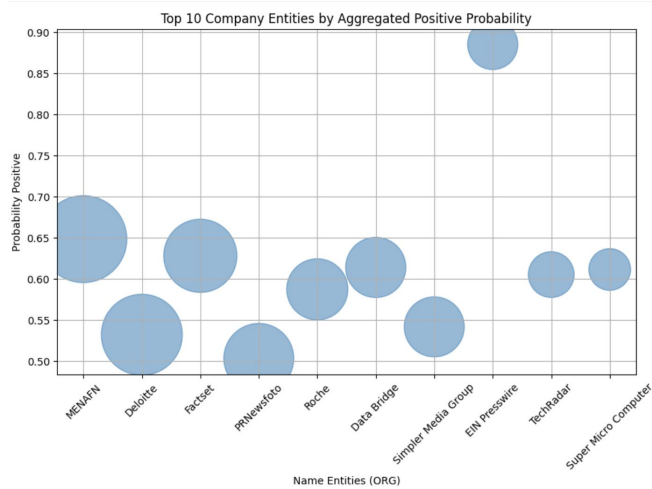
Introduction of New Technologies & AI Solutions



AI related Technologies & Solutions have gained significant media attention over the observed period :

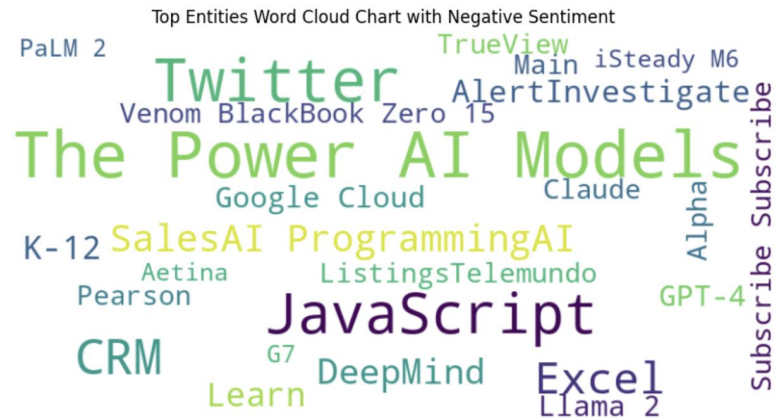
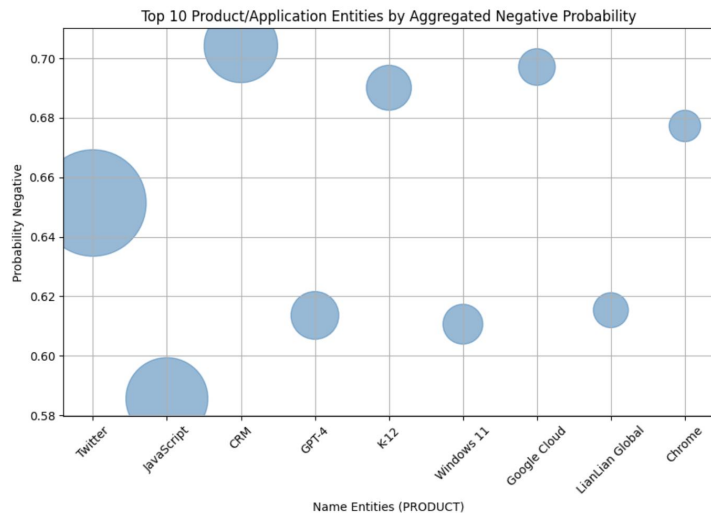
- ChatGPT became extremely heated topic after 2023-01 which is after OpenAI released its demo in November 30, 2022.
- Similar to "ChatGPT", "Google Bard" has several peaks, with a sharp increase and subsequent decrease, which may reflect market reactions or significant announcements.

Entity Identification & Sentiment



- Companies such as "MENAFN" and "Deloitte" have both a high frequency of mentions and high positive sentiment, suggesting they may be viewed favorably in news reports related to technology investment or success. It might indicate their role in facilitating technological advancements or digital transformations for other businesses.
- Tech giants like "Google," "Microsoft," "Apple," and "IBM" appear prominently, implying they are frequently cited in positive stories about technology investments or successes.
- There is a significant representation of tech companies, consulting firms, and new media organizations, suggesting these sectors are actively investing in new technologies.

Entity Identification & Sentiment



- Entities like "Twitter," "AI," "JavaScript," and "CRM" stand out, suggesting these are common topics of discussion when it comes to the challenges or limitations in the current state of AI or technology.

Actionable recommendations

Companies

Topic 0:
professional mobile partner web community build stories platform product free
Topic 1:
gray group prnewswire media data 2023 platform release press generative
Topic 2:
market analysis growth report global 2020 machine research learning key
Topic 3:
medical dr serviceprivacy endorsed ve music gray ceo advertise usmeet
Topic 4:
serviceprivacy endorsed music ve advertise usmeet hour captioning ceo gray
Topic 5:
news chatgpt best 2023 world business tech india like data
Topic 6:
computer vision 2022 data gray learning machine platform group prnewswire
Topic 7:
learning researchers science data human machine research models model scientists
Topic 8:
serviceprivacy endorsed data ve music advertise usmeet hour ceo science
Topic 9:
medical gray care health healthcare dr group prnewswire research media

Academic Institutions

Topic 0:
market analysis growth report forecast global players size 2020 machine
Topic 1:
city dr professor digital world people 2021 marketing virtual data
Topic 2:
customer marketing usmeet data teams join endorsed experience instagram founder
Topic 3:
usmeet women teams endorsed founder instagram dc art started necessarily
Topic 4:
professionals clients member applied skills generative collaboration institutions training network
Topic 5:
usmeet endorsed teams founder officer instagram subscribe dc edit ve
Topic 6:
computer vision conference 2022 society recognition june program 24 technical
Topic 7:
virtual software smart company devices hardware worldwide platform experiences global
Topic 8:
news chatgpt 2023 data best learning science human world intelligence
Topic 9:
gray group data prnewswire media 2023 platform press release solutions

Insight from the positive sentiment topics associates companies, academic institutions and government entities using LDA :

Companies

- **Market Research and Machine Learning:** Topics related to market analysis and machine learning (Topic 2, Topic 7) suggest focusing on growth through data-driven insights. Investing in machine learning research and market analysis can help companies identify new opportunities and trends.
- **Innovate:** Invest in AI and machine learning for product enhancement.
- **Collaborate:** Form strategic partnerships with academia and government.

Academic Institutions

- **Enhanced Collaboration and Training:** The emphasis on collaboration and skills training (Topic 4) suggests that institutions should foster networks that support professional development and applied learning, particularly in emerging technologies like generative AI.
- **Research in Computer Vision and AI:** The focus on computer vision and AI (Topic 6) underlines the importance of research in cutting-edge technologies. Academic institutions should prioritize research in these areas, potentially through partnerships with industry to drive innovation forward.
- **Public Lectures and Seminars:** Host public lectures, seminars, and workshops on AI advancements and implications to educate the broader community and stimulate public interest.

Actionable recommendations

Government

Topic 0:
data public 2022 science ways facebook air 2023 community music
Topic 1:
china market global energy travel world chatgpt india artificial 2021
Topic 2:
credit program small market personal best business china advertising said
Topic 3:
research market daily rss articles africa asia europe press financial
Topic 4:
times india 2024 world daily learning subscribe live generative credit
Topic 5:
world finance apple capabilities cloud education daily india performance 2023
Topic 6:
2023 chatgpt best world tech data openai said business video
Topic 7:
white house meet white house risks experts opportunities president science biden
Topic 8:
market analysis global growth report research key intelligence 2021 trends
Topic 9:
gray gray media prnewswire media group data defense group media press release company

Insight from the positive sentiment topics associates companies, academic institutions and government entities using LDA :

Government Entities

- **Focus on Data Science and Public Services:** The mention of data science and public community services (Topic 0) highlights the potential for using data to improve public services. Government entities should invest in data science capabilities to enhance everything from healthcare to urban planning.
- **Global Cooperation and Market Understanding:** Topics related to market analysis and global trends (Topic 8) suggest a need for understanding and engaging with global markets. Government policies should support international cooperation and foster environments conducive to business and technological innovation.
- **Funding and Grants:** Provide funding and grants for AI research, particularly in areas that have significant potential for social good but might not attract sufficient private investment.
- **Policy Development:** Develop and implement policies that encourage the ethical development and deployment of AI, including privacy protection, data governance, and accountability standards.

Conclusion

- The topics identified via LDA and BERTopic models indicate a broad range of interests in the dataset, with particular emphasis on the integration of AI into various industries, businesses, and social aspects. The detailed topic clusters suggest an extensive discourse on how AI is not only a technological revolution but is deeply enmeshed with daily human activities and industries.
- Custom sentiment analysis has revealed varying perceptions of AI across industries over time. It shows that there might be concerns regarding AI's impact on healthcare, finance, technology, and entertainment, potentially due to fears of job displacement or ethical implications.
- There has been significant media attention on new AI technologies, particularly on tools like ChatGPT and Google Bard. These spikes in media attention might correspond with product releases, updates, or significant market impacts.
- Entities such as MENAFN and Deloitte are viewed positively, which could correlate with successful AI adoption or innovation. Major tech companies like Google, Microsoft, and Apple are frequently mentioned, which could indicate their leadership in the AI space.
- The mention of entities like "Twitter," "AI," "JavaScript," and "CRM" alongside negative sentiment might highlight the public's concerns or the challenges faced by these platforms or technologies.

A decorative network diagram with purple nodes and lines. The nodes are represented by concentric circles, with some having a solid purple center and others being hollow. They are connected by thin grey lines. There are three main paths: one in the top right corner, one in the bottom left corner, and one in the bottom center. Each path starts with an ellipsis (...).

Thanks