

NLP Final Project

AI Development & Integration in
Future Employment Landscape

Instructor: Nick Kadochnikov

Presenter: Yezi Liu



Executive Summary 1

Article Cleaning & Filtering

- Used **regex** to narrow 200,435 articles down to 192,323 AI-relevant articles
- Applied **extensive cleaning** methods on article text to prepare for topic modeling

Topic Modeling

- Used **ktrain** package to identify initial 25 raw topics
- Applied **topic keyword analysis, article sampling, and text summarization** to narrow down to 11 highly AI-relevant topics to discover **top candidates for AI integration**

Sentiment Model & Sentiment Over Time Analysis

- Customized **logistic regression model** trained on **Yelp** data with **96.5%** test accuracy
- Industry candidates impacted by AI
 - Positive sentiment: Healthcare, Sports, Education, Art, App Development
 - Negative sentiment: Government, Police
 - Unsure: Tech
- Jobs impacted by AI(sentiment over time)
 - Educator: recent positive sentiment
 - Artist: recent positive sentiment
 - Technicians & Doctors: sentiment fluctuations

Executive Summary 2

New AI Technology Introduction Timeline

- Applied **moderate cleaning & NER** from **spacy** package
- Aggregated entities by **counts** to identify popular technologies
- **Aggregated article mentions** as the sign of technology introduction
- GPT: 2022 Winter | OpenAI: 2022 Winter | Google Cloud: 2023 Spring | Generative AI: 2023 Spring | Alexa: 2023 Spring | NLP: 2021 Winter

Recommendations to Accelerate AI Development

- **Companies:** use AI to build/scale applications and for content generation
- **Academic Institutions:** use AI chat box to replace human educators
- **Governments:** use AI-enabled decision intelligence solutions

Targeted Entity Sentiment Identification

- Company types planning to invest in AI(successes):
 - **Automotive** Companies
 - **Social Media & Online Service** Companies
 - **Technology Consumer Electronics** Companies
- Applications currently cannot be transformed by AI(failures):
 - **Environmental issue** domain
 - **Airline** domain

Data Source & Initial Article Cleaning/Filtering

Original 200,435
news articles

Clean text for topic
modeling

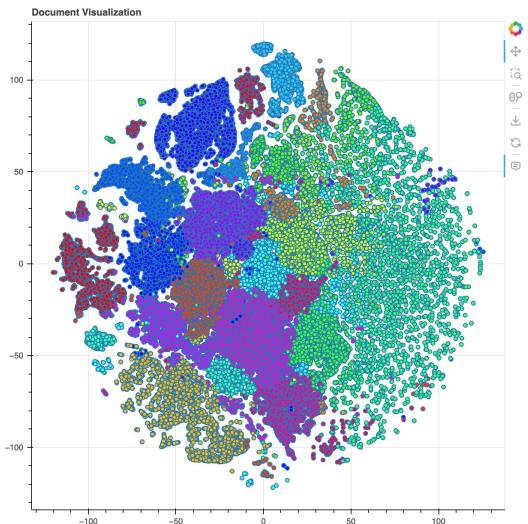
Articles contain 'AI', 'artificial
intelligence', 'machine
learning', 'deep learning', 'data
science', etc.

Remove urls, new lines, html,
punctuations, numbers,
stopwords, too long/short tokens;
lowercase, lemmatization.

Use regex to filter & narrow down
to 192,323 news articles

AI Integration Top Candidates From Topic Modeling

Initial 25 raw topics from *ktrain* package



Topic keywords analysis



Article sampling & text summarization

11 highly AI-relevant topics and top keywords

- Topic 21: chatgpt image people work tool human model
- Topic 5: customer cloud generative enterprise digital
- Topic 12: model machine learning process algorithm
- Topic 16: openai microsoft google tech ceo chatgpt altman icon billion musk
- Topic 7: health patient medical healthcare clinical
- Topic 23: science learning student research university education machine scientist
- Topic 2: nvidia chip edge computing vision application
- Topic 14: insurance crypto casino credit program bitcoin
- Topic 18: energy vehicle global car robot supply digital
- Topic 8: law security legal financial share cybersecurity transaction
- Topic 15: stock market nasdaq share price trading investor symbol fund etf

Sentiment Analysis Methodology

Step 1: Built **Customized** Sentiment Analysis Model

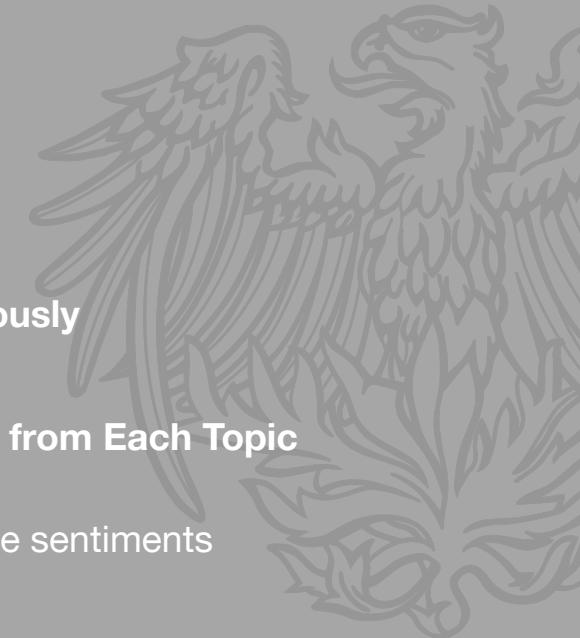
- Logistic regression trained on Yelp data with 255,717 records
- Use TF-IDF as word embeddings
- Test accuracy of 96.5%

Step 2: Applied Model on Articles in AI-Relevant Topics Identified Previously

Step 3: Filtered Out Articles with Extreme Positive/Negative Sentiments from Each Topic

- Filtering thresholds varied for different topics
- Used **LexRank Summarizer** to summarize articles with positive/negative sentiments

Step 4: Identified Specific Job Types/Industries Positively/Negatively Impacted by AI



Sentiment Analysis Cont.

Industry Candidates for AI Integration

+: positive
-: negative
?: not sure

Healthcare(+)

“Notably in healthcare, AI tools are identifying new antibiotics, helping people with brain damage...”

“...new capabilities within the AI Data Lab to scale predictive modeling to help clinicians in their efforts to anticipate and intervene before a person is in crisis...”

Sports Games(+)

“AI has reconfigured the sports experience for fans, both in stadiums and on any streaming device, by expanding fan engagement...”

Education(+)

“Students experience the direct advantages of low-code AI through a learning experience that is engaging, supportive, and responsive to their needs.”

“Educators can identify trends and intervene proactively to assist students who may be struggling.”

Art(+)

“...the upsides of AI-driven music: it could potentially streamline the creative process, allowing artists to focus more on the quality of their work...”

App Development(+)

“As OpenAI's innovations continue to advance, these organizations will experience streamlined app creation, increased innovation, cost reductions...”

Technology(?)

“How tech professionals can survive and thrive at work in the time of AI? ...working closely with the business to demonstrate his AI services...’We need to understand where the best place is to run the chatbot.’”

Government/Police(-)

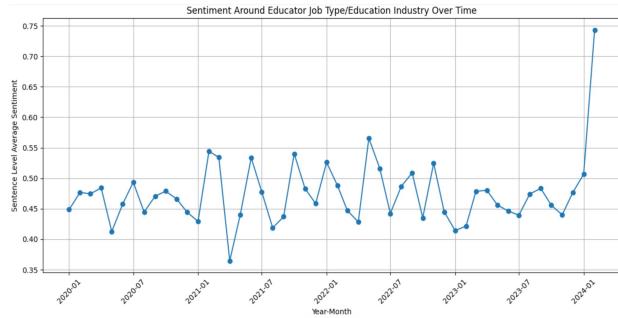
“Generative AI tools could be used as a ‘how to commit crime’ advisor...”

“Software that allows people to create a non-existent individual, with a fake accent, voice or face, poses legal complications.”

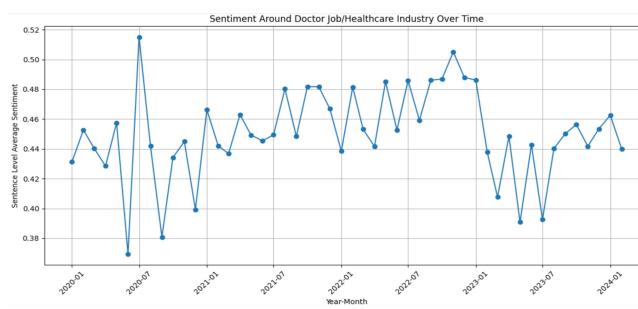
Sentiment Over Time Analysis 1

Job Type Sentiment(Sentence-Level) Impacted by AI

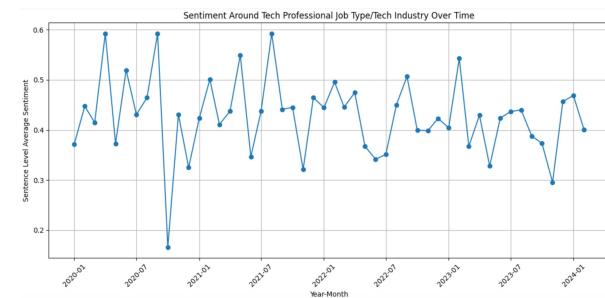
Methodology:
1 Target Sentence Sentiment
2 Mean Sentiment Aggregation



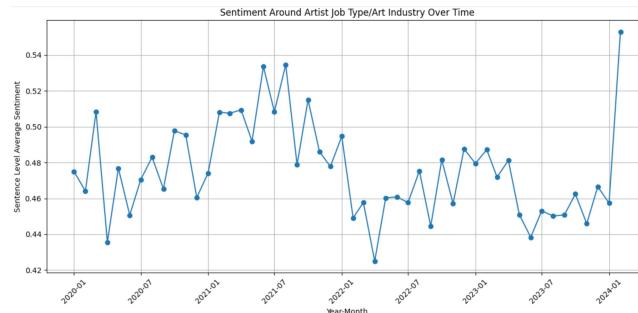
Educators:
Recent Positive



Doctors:
Fluctuations/
Unsure



Technicians:
Fluctuations/
Unsure



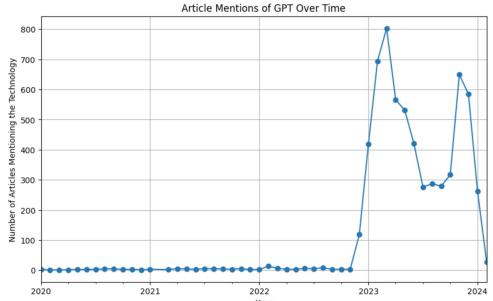
Artists:
Recent Positive

Note: Doctor fluctuated sentiments don't perfectly align with the positive sentiments in healthcare industry identified earlier. Possible reason is that previous article sampling didn't capture the overall sentiment picture or AI has different impacts on doctors and the general healthcare/clinical industry.

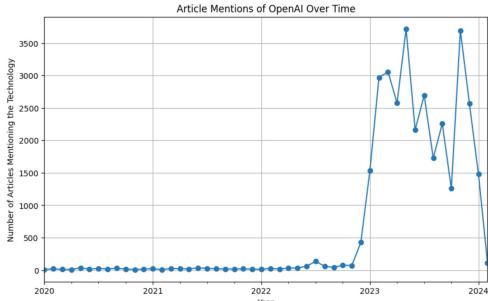
Sentiment Over Time Analysis 2

Timeline for Introduction of New AI Technologies

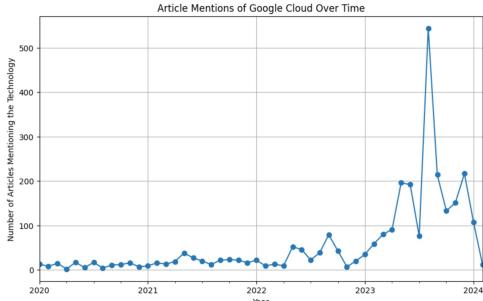
Methodology:
1 Moderate Cleaning
2 NER from Spacy Package & Total Counts
3 Article Mentions Aggregation



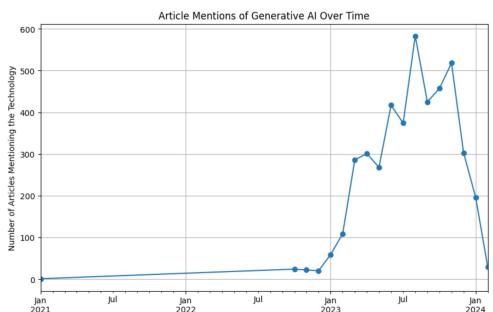
GPT | 2022 Winter | 13742 Counts



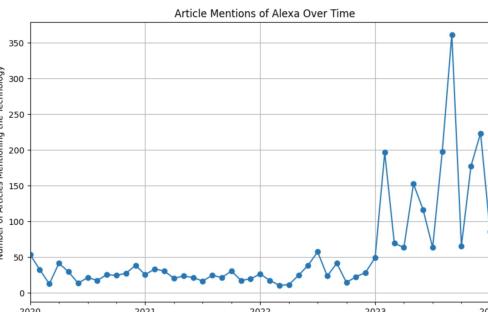
OpenAI | 2022 Winter | 120768 Counts



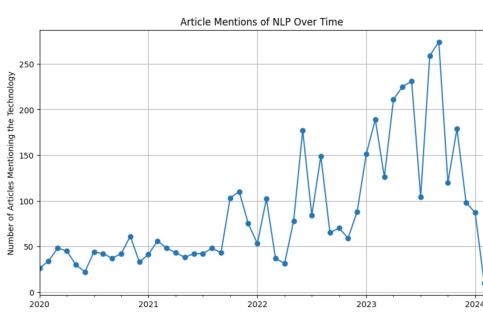
Google Cloud | 2023 Spring | 7942 Counts



Generative AI | 2023 Spring | 9481 Counts

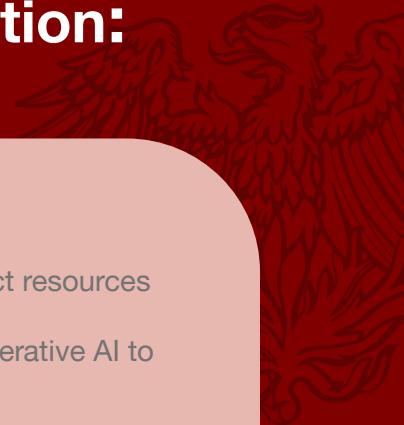


Alexa | 2023 Spring | 6275 Counts



NLP | 2021 Winter | 7483 Counts

Recommendations & Insights from Text Summarization: what can we do to accelerate the transformative AI development?



Companies:

1. Incorporate AI to **build and scale their applications** & improve enterprise operations

- “...AI Center...provide businesses with the industry expertise, technical knowledge, and product resources to build and scale applications...improve enterprise search, enhance customer contact center experiences...” | “... AI...support code generation and completion for customers—bringing generative AI to software engineering teams to enhance every stage of the development lifecycle...”

2. Use AI for any form of **content creation**

- “...AI includes more sophisticated and nuanced content generation...creative process, and ongoing collaboration between AI systems and human creators...”

Academic Institutions:

1. Replace human instructors with AI technologies to **support self learning**

- “AI also helps people who want to self-learn using the latest technology aids available to them...” | “AI chatbots can provide instant feedback and assistance, offering a 24/7 support system...”

Governments:

1. Utilize trusted **AI-enabled Decision Intelligence solutions**

- “Global investments aimed at accelerating..government agency’s ability to use trusted AI-enabled Decision Intelligence solutions...create over 170 new jobs in London AI Innovation Centre team...”

Targeted Entity Sentiment Identification 1

Types of Companies Planning to Invest in AI Technologies

Methodology:

- 1 Target Sentence Sentiment Around Different Company Types/People/Locations
- 2 Filter Out Positive Sentiment for Visuals



Automotive Companies:

Tesla



Social Media & Online Service Companies:

Facebook, Instagram,
LinkedIn, Google



Technology Consumer Electronics Companies:

Apple, Microsoft, Oracle, Samsung



Sam Altman:
(OpenAI CEO)
Companies in Health,
Sport, Tech, Education

Types of Companies Planning to Invest in AI Technologies

Successful Stories & Insights from Text Summarization



Automotive Companies

- “Tesla recall million vehicles China over autopilot safety controls...Samsung partners with Tesla and Hyundai offer deeper smart home and controls...expand SmartThings into the areas home energy and vehicle home automation...”
- “Tesla robot Optimus can now ...”
- “Tesla and Twitter CEO ...reportedly assembling team artificial intelligence experts build alternative Chat GPT and other similar ideas may also develop...”

Social Media & Online Service Companies

- “Dr. Amrie Grammer Google Reuters webinar focused machine learning approaches select the right...”
- “Launched News Writing Artificial Intelligence Write your Twitter experience...”
- “Altman posted Friday formerly Twitter loved time openai...”
- “Tech giant Google has seen great potential for artificial intelligence technologies South ...”

Technology Consumer Electronics Companies

- “...Smart Interviewer powered the world largest source ...and advanced Natural Language Processing ...enabling organizations interrupt hiring bias scale get the right talent fast and give every candidate experience they love...”
- “...which use intelligent light detection combine Oppo advanced algorithms enable users quickly capture professional quality videos with ease...”

Success Around Sam Altman

- “New global university and study program search platform backed artificial intelligence and machine learning...help students find their ideal university and study program...”
- “Appearing Microsoft event ...Sam Altman said the the new Bing experience looks fantastic and based part learnings from its GPT line large language models...”
- “Sam Altman ...join Microsoft help lead new advanced artificial intelligence team...”

Targeted Entity Sentiment Identification 2

Types of Applications Cannot Currently be Transformed by AI

Methodology:

- 1 Target Sentence Sentiment Around Applications
- 2 Filter Out Negative Sentiment for Visuals
- 3 Text Summarization for Insights

Environmental Issue Domain



- Average Sentence-Level Sentiment < 0.5
 - Percentage of Negative Sentiment 56%

Misuse of AI exacerbates environmental issues: “ environmentalist group says Amazon Microsoft and Google have been undermining their own climate change pledges ...looked for new technology get more oil and gas out the ground...”

Ethical concerns & biases in AI recommendations: “unclear about important issues such environmental impact bias and the ethics OpenAI ...”

Airline Domain



- Average Sentence-Level Sentiment < 0.5
 - Percentage of Negative Sentiment 79%

The work of airport customer service staff is **various and highly dependent** on individual customer situations; the work of captains and flight attendants is **complex and requires high professional knowledge**. The cost of inappropriate service/decisions is high in airline domain. AI technology needs further **domain-specific development and expertise** in order to transform airline domain.

