Analysis

This analysis involved extracting data from Reddit over two years, manually labeling posts into specific categories, training machine learning models, and performing exploratory data analysis (EDA) and topic modeling to glean insights into AI and employment discussions.

# Data Collection and Preprocessing

## Data Extraction

Due to limitations with Reddit’s API, we utilized the Reddit data archive, which is accessible via torrents. The data spans 2005 to 2023, but we focused on posts from July 2022 to July 2024. The data was downloaded using academic torrents, processed using scripts from the [PushshiftDumps repository](https://github.com/Watchful1/PushshiftDumps), and filtered within ten subreddits and a query with key terms.

**Subreddits:** anti-work, AskReddit, careerguidance, changemyview, Economics, Futurology, jobs, NoStupidQuestions, Showerthoughts, technology

**Query:**

(("ai " OR " ai" OR "artificial intelligence" OR chatgpt) AND (job OR jobs OR work OR career OR employment OR profession OR worker OR workers OR employee OR replace OR replaced OR replaces OR replacement OR affected OR affect OR affecting OR disappear OR disappearing OR disappeared OR fired OR hiring OR hire OR lose OR lost OR losing OR eliminate OR eliminates OR eliminating OR redundant OR safe OR obsolete))

For two years, we gathered **7,616,585** Reddit posts within the ten subreddits. After that, we performed a double validation process. First, we used the query above to filter only relevant Reddit posts, and second, we used Llama 3.1 7B to automatically validate whether they were talking about AI taking over jobs or affecting the job market. The number of Reddit posts filtered with the query and labeled by the LLM was around **18,159**.

**Data Labeling**

To narrow down the topic of the post, we assigned it to 3 levels (below). We manually labeled **555** Reddit posts validated by the LLM using multilabel classification. We selected the posts from 10 to the highest score.

1. **C1 Work**: Posts discussing task-oriented thoughts, such as new summarization models.
2. **C2 Worker**: Posts about how a worker might have suffered an impact on their work environment.
3. **C3 Workforce**: Posts indicating that a significant number of workers in a specific sector or in general are affected.

**Label Distribution:**

|  |  |
| --- | --- |
| **Category** | **Number of Posts** |
| C1 Work | 67 |
| C2 Worker | 192 |
| C3 Workforce | 80 |

# Methodology

## Machine Learning Models

We trained two machine learning models to predict the categories of unlabeled data:

1. Random Forest Classifier
2. Support Vector Machine (SVM)

## Data Preprocessing:

* Text data was cleaned by converting it to lowercase and removing URLs, punctuation, numbers, and stop words.
* Lemmatization was applied to reduce words to their base forms.
* The cleaned text was transformed into numerical features using TF-IDF vectorization.

## Model Training:

* The models were trained using the manually labeled data through multilabel classification (the categories are not mutually exclusive)
* The data was split into training and validation sets using an 80/20 split.
* We used the following evaluation metrics: Precision, Recall, F1-Score, and Support (number of occurrences of each class in the validation set)

**Model Performance:**

**Random Forest:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **Precision** | **Recall** | **F1-Score** | **Support** |
| C1 Work | 1.00 | 0.10 | 0.18 | 10 |
| C2 Worker | 0.61 | 0.33 | 0.43 | 33 |
| C3 Workforce | 0.00 | 0.00 | 0.00 | 16 |

**SVM:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **Precision** | **Recall** | **F1-Score** | **Support** |
| C1 Work | 0.00 | 0.00 | 0.00 | 10 |
| C2 Worker | 0.69 | 0.27 | 0.39 | 33 |
| C3 Workforce | 0.50 | 0.12 | 0.20 | 16 |

**Predicted Posts per Category and Model:**

|  |  |  |
| --- | --- | --- |
| **Category** | **Random Forest** | **SVM** |
| C1 Work | 0 | 10 |
| C2 Worker | 920 | 942 |
| C3 Workforce | 7 | 61 |

**Model Selection:**

* Both models struggled with predicting C1 Work and C3 Workforce due to low support and class imbalance.
* The Random Forest model had a higher F1-Score for C2 Worker, so it was selected for further analysis.

## Exploratory Data Analysis (EDA)

EDA was conducted on the manually labeled and predicted data (classified by the Machine Learning model into Category 2, about the Worker) to understand trends and patterns.

1. **Time Series Analysis**: Plots showing the number of monthly Reddit posts were created to identify trends over time.

A graph with blue lines

Description automatically generated

A graph with blue lines

Description automatically generated

A graph of a graph

Description automatically generated

A graph with a line

Description automatically generated

A graph of different colored lines

Description automatically generated

1. **Sentiment Analysis**: The average sentiment per month was calculated using TextBlob’s sentiment analyzer to assess the emotional tone of the discussions.

A graph of blue lines

Description automatically generatedA graph of a graph

Description automatically generated with medium confidenceA graph with blue lines

Description automatically generated

A graph with a line

Description automatically generated

A graph of different colored lines

Description automatically generated

# Topic Modeling

Topic modeling was performed using Latent Dirichlet Allocation (LDA) on four datasets:

1. Manually Labeled Data for All 3 Categories:
   1. C1 Work
   2. C2 Worker
   3. C3 Workforce
2. Predicted Data for C2 Worker

**Table 1: Top Words for C1 Work Topics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| ai (0.008) | ai (0.024) | job (0.013) | ai (0.011) | ai (0.028) |
| could (0.005) | job (0.012) | human (0.011) | ’ (0.011) | job (0.017) |
| people (0.003) | potential (0.006) | ai (0.011) | human (0.008) | people (0.010) |
| use (0.003) | article (0.006) | company (0.007) | travel (0.008) | ’ (0.009) |
| going (0.003) | human (0.006) | laundry (0.006) | like (0.008) | new (0.007) |

**A graph with a bar chart and a bar chart with a bar chart and a bar chart with a bar chart and a bar chart with a bar chart and a bar chart with a bar chart and

Description automatically generated**

*C1 Work LDA*

**Table 2: Top Words for C2 Worker Topics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| ai (0.025) | ai (0.027) | ai (0.021) | would (0.013) | happens (0.011) |
| art (0.025) | job (0.025) | job (0.017) | copyright (0.006) | ai (0.007) |
| artist (0.017) | people (0.016) | ’ (0.010) | bootcamps (0.005) | art (0.006) |
| ’ (0.014) | like (0.009) | career (0.008) | won't (0.005) | would (0.005) |
| like (0.009) | human (0.009) | I'm (0.007) | robot (0.004) | I'm (0.005) |

*A screenshot of a graph

Description automatically generated*

*C2 Worker LDA*

**C3 Workforce**

**Table 3: Top Words for C3 Workforce Topics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| ai (0.024) | ai (0.027) | ai (0.020) | agi (0.011) | people (0.015) |
| job (0.021) | job (0.017) | job (0.015) | ai (0.008) | work (0.009) |
| ’ (0.016) | people (0.013) | art (0.013) | translation (0.007) | ai (0.009) |
| would (0.014) | human (0.007) | people (0.012) | general (0.006) | collar (0.009) |
| people (0.009) | even (0.007) | human (0.008) | translator (0.006) | job (0.007) |

*A graph with text and numbers

Description automatically generated with medium confidence*

*C3 Workforce LDA*

**Table 4: Top Words for Topics 1 to 5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** | **Topic 4** | **Topic 5** |
| long (0.033) | code (0.024) | scientist (0.046) | andor (0.029) | pace (0.059) |
| suffer (0.021) | ethical (0.012) | chinese (0.013) | planning (0.027) | sooner (0.016) |
| term (0.017) | vast (0.012) | jumping (0.004) | viable (0.008) | eliminate (0.010) |
| replacing (0.016) | intelligence (0.011) | escape (0.002) | success (0.005) | administrative (0.001) |
| fathom (0.015) | tool (0.010) | monitoring (0.001) | progression (0.002) | warning (0.001) |

**Table 5: Top Words for Topics 6 to 10**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Topic 6** | **Topic 7** | **Topic 8** | **Topic 9** | **Topic 10** |
| problem (0.011) | service (0.017) | ’ (0.273) | suffering (0.041) | ai (0.051) |
| position (0.010) | government (0.015) | “ (0.023) | ipad (0.016) | job (0.023) |
| created (0.010) | unemployed (0.013) | ” (0.022) | ideally (0.014) | people (0.017) |
| collar (0.010) | population (0.013) | stop (0.016) | device (0.011) | like (0.012) |
| massive (0.009) | physical (0.012) | dentist (0.013) | random (0.008) | human (0.012) |

*A graph of a diagram

Description automatically generated with medium confidence*

*Predicted C2 Worker LDA*

**Findings from C1 Work**

The C1 Work category delves into discussions focused on task-oriented aspects of AI, particularly how new technologies and models influence specific work-related activities.

1. **Topic 1**This topic centers on the potential applications of AI in various contexts. The frequent mention of words like "could," "use," "going," "interview," "technology," and "manager" suggests conversations about how AI might be integrated into managerial tasks and interviews. People are contemplating the future trajectory of AI technology and its practical implications in the workplace.
2. **Topic 2**Here, the discussion shifts towards the potential impact of AI on jobs and human labor. With words like "job," "potential," "article," "human," "work," and "use," the topic likely includes debates spurred by articles about AI's capabilities. There's a focus on whether AI will replace or augment human workers and how it's being utilized in different industries.
3. **Topic 3**This topic appears to address specific jobs or tasks susceptible to AI automation. Words such as "company," "laundry," "software," and "writing" indicate discussions about industries like software development and laundry services. People are exploring how companies are adopting AI to automate these roles and the consequences for employees who currently perform these tasks.
4. **Topic 4**The emphasis here is on AI's role in areas like travel and language processing. The presence of words like "travel," "llm" (large language model), "idea," and "question" suggests speculative conversations about how AI could revolutionize travel experiences or language-related services. There's a sense of curiosity and exploration of innovative possibilities enabled by AI advancements.
5. **Topic 5**This topic captures people's reflections on new developments in AI and their effects on jobs and work practices. Words like "new," "think," "like," "way," and "even" indicate a general contemplation about the evolving landscape due to AI innovations. Discussions may involve opinions on recent AI technologies and their potential to transform traditional work methods

**Findings from C2 Worker**

The C2 Worker category focuses on the individual worker's perspective, examining how AI impacts personal job security, career paths, and day-to-day work experiences.

* 1. **Topic 1**This topic is heavily centered on the intersection of AI and the art world. Words like "art," "artist," "artwork," and "human" suggest discussions about AI-generated art and the challenges it poses to human artists. There's an exploration of creativity, originality, and the value of human input in artistic endeavors when AI can produce art.

1. **Topic 2**Here, the emphasis is on personal concerns regarding AI replacing human jobs. The frequent use of words like "job," "people," "human," "work," "new," and "I'm" indicates that individuals are reflecting on how AI advancements are affecting their employment and considering new types of work emerging from technological progress.
2. **Topic 3**

This topic delves into career-related anxieties and considerations. Words like "career," "work," "year," "article," and "time" suggest that workers are discussing the influence of AI on their career trajectories. There may be conversations about the need to adapt, retrain, or make timely decisions in response to AI developments.

1. **Topic 4**

Legal and financial concerns are prominent in this topic. The presence of words like "copyright," "bootcamps," "income," "industry," and "robot" points to discussions about the implications of AI on intellectual property rights, income stability, and the necessity of education programs to help workers adapt to technological changes.

1. **Topic 5**

This topic reflects uncertainty and questions about the future impact of AI on specific fields, particularly in the arts. Words like "happens," "would," "field," "content," and "question" indicate that individuals are pondering the fate of their professions and expressing apprehension about how AI might alter their work landscapes.

**Findings from** **C3 Workforce**

The C3 Workforce category examines the broader implications of AI on large groups of workers and entire industries, emphasizing systemic changes rather than individual experiences.

1. **Topic 1**

This topic discusses the need for adaptation as AI reshapes job markets and company operations. Words like "would," "people," "human," "work," "company," "new," and "need" suggest conversations about how organizations and the workforce must evolve. There's likely a focus on acquiring new skills and the restructuring required to stay relevant in an AI-driven economy.

1. **Topic 2**

The focus here is on the displacement of human labor by AI and the societal challenges that arise. With words like "even," "like," "work," "would," and "make," discussions may involve debates about the extent to which AI can or should replace human workers, the ethical considerations, and the potential consequences for employment rates.

1. **Topic 3**

This topic addresses automation's impact on jobs, including in creative sectors such as art. Words like "art," "think," "automation," and "going" indicate reflections on the direction of technological advancement and its effects on employment opportunities. There's an acknowledgment of the accelerating pace of automation across various industries.

1. **Topic 4**

The emergence of Artificial General Intelligence (AGI) is the central theme here. Words like "agi," "translation," "general," "translator," "intelligence," and "artificial" suggest discussions about AI reaching a level of general cognitive abilities. Concerns may revolve around the future of professions like translation and the broader implications of AGI on the job market.

1. **Topic 5**

This topic focuses on the impact of AI on white-collar jobs. Words like "collar," "white," "work," "day," and "lot" imply that everyday professional work is changing due to AI technologies. There's a contemplation of how AI affects routine tasks and the potential need for workers in professional settings to adapt to new tools and processes.

**Findings from Predicted C2 - Worker**

1. **Topic 1: Long-Term Impacts of AI on the Workforce**

This topic revolves around discussions about the long-term effects of AI on employment. Words like "long," "suffer," "term," "replacing," and "workforce" suggest that individuals are contemplating how AI might replace human labor over extended periods. There's a concern about the potential suffering or hardship that could result from AI-driven job displacement. The mention of "fathom" and "it'll" indicates the difficulty in fully understanding or accepting these future changes.

1. **Topic 2: Ethical Considerations and AI Tools**

This topic focuses on ethical issues surrounding AI and its applications. Words such as "code," "ethical," "intelligence," "tool," and "development" highlight discussions about responsible AI development and the importance of ethics in coding and deploying AI technologies. The term "vast" suggests the expansive nature of AI's capabilities, and "potential" points to the opportunities and risks involved.

1. **Topic 3: Scientists and International Perspectives**

This topic appears to center on scientists and possibly international collaborations or competitions in AI. The prominent words are "scientist" and "Chinese," indicating discussions about scientists from different countries and their contributions or roles in AI development. Other words like "jumping," "escape," and "monitoring" could suggest topics related to movement or shifts in scientific communities or concerns about surveillance.

1. **Topic 4: Career Planning and Success in the AI Era**

This topic focuses on personal career trajectories in the context of AI advancements. Words like "planning," "viable," "success," and "progression" suggest that individuals are considering how to plan their careers to remain relevant and successful. The mention of "and/or" might indicate weighing different options or paths. There's an emphasis on embracing change and seeking opportunities in the evolving job market.

1. **Topic 5: Accelerating Pace of Automation**

This topic addresses the rapid acceleration of AI and automation. The word "pace" stands out significantly, indicating a focus on the speed at which changes are occurring. Words like "sooner," "eliminate," and "warning" suggest that automation may eliminate certain jobs faster than anticipated, raising concerns and prompting calls for caution. There may be discussions about administrative roles being affected.

1. **Topic 6: Employment Challenges and Economic Policies**

This topic delves into problems related to employment and economic policies in the face of AI advancements. Words like "problem," "position," "created," "collar," "massive," "policy," and "job" indicate discussions about the challenges workers face, particularly in blue-collar positions. There's mention of "massive" changes and the need for policies to address job displacement.

1. **Topic 7: Government Services and Automation**

This topic focuses on the government's role in addressing unemployment and the impact of automation on services. Words like "service," "government," "unemployed," "population," "physical," "system," and "automated" suggest discussions about how government services can support unemployed populations and how automation affects public services. "UBI" (Universal Basic Income) appears to be a point of discussion as a potential solution.

1. **Topic 8: Communication and Expression Challenges**

This topic seems to highlight issues related to communication and expression. The high frequency of punctuation marks like "’," "“," and "”" suggests textual elements or perhaps quoting and dialogues in discussions. Words like "stop," "voice," "concerned," "affected," and "negative" indicate conversations about being heard, expressing concerns, or negative impacts on communication.

1. **Topic 9: Technological Devices and User Experience**

This topic revolves around technology devices and user experiences. Words like "suffering," "ipad," "ideally," "device," "ipados," "macos," "remove," and "cyber" suggest discussions about issues with devices, perhaps frustrations with technology or software updates. "Suffering" in this context may refer to user dissatisfaction or challenges with adapting to new technologies.

1. **Topic 10: General Discussions on AI and Employment**

This topic captures general conversations about AI's impact on jobs and the future. Words like "ai," "job," "people," "like," "human," "I'm," "art," "work," "future," and "think" indicate broad discussions about how AI affects employment, human roles, and thoughts about the future. There's a mix of personal reflections and general observations.

# Summary

**EDA Findings**

* Trends Over Time: There was an observable increase in discussions on AI and employment following the release of ChatGPT.
* Sentiment Analysis: The average sentiment fluctuated over time, indicating varying emotional responses to AI developments.
* Word Frequencies: Common words included “ai,” “job,” “people,” “work,” and “human,” highlighting the central themes of employment and human-AI interaction.

**Model Performance**

* Random Forest Model: Demonstrated better performance for the C2 Worker category but failed to effectively predict C1 Work and C3 Workforce.
* SVM Model: Showed limited effectiveness across all categories.
* Challenges: The models struggled due to class imbalance and the complexity of the multilabel classification task.

**Topic Modeling Insights**

Across all datasets, several common themes emerge:

1. **AI's Impact on Employment:** There's a pervasive concern about how AI will affect jobs, both at the individual and industry levels. Workers are contemplating the potential for job displacement and the need to adapt to new technologies.
2. **Human vs. AI Creativity:** Particularly in the art sector, there's significant discourse on the differences between human-created and AI-generated content. The value of human creativity is being examined in light of AI's capabilities.
3. **Future Uncertainty:** Many discussions express uncertainty about the future. Questions about how AI will reshape careers, industries, and society are prevalent, reflecting both anxiety and curiosity.
4. **Ethical and Societal Considerations:** Ethical issues, such as the implications of AGI and the societal impact of widespread automation, are important topics. There's an awareness of the need for responsible AI development and consideration of its effects on humanity.
5. **Need for Adaptation:** Workers recognize the necessity of adapting to the changing landscape, whether through upskilling, retraining, or shifting to roles less susceptible to automation.