Automation Anxiety and AI Narratives Across News, Reddit, and Twitter (2017–2025)

Dheeraj Shetty
American University
ds5172a@american.edu

Abstract

Artificial intelligence (AI) and automation have sparked widespread public discourse about job displacement, inequality, and the future of work. This study analyzes 893,000 tweets, 885 Reddit comments, and 1,000 news articles (2017–2025) to examine how automation anxiety and AI-related narratives vary across platforms and communication genres. Using the CRISP-DM framework, the analysis integrates inductive methods topic modeling and word frequency patterns with deductive techniques including VADER sentiment analysis, chi-square statistical testing, and a theory-driven fear–opportunity dictionary. Findings show that Twitter and news media frame AI with greater optimism, while Reddit discussions contain elevated levels of fear and uncertainty. Topic modeling reveals shared themes of job loss, reskilling, technological innovation, and inequality. These results contribute to understanding how digital platforms shape emotional and thematic responses to technological change, offering a reproducible methodology for tracking automation anxiety over time.

Keywords: Automation Anxiety, Artificial Intelligence, Sentiment Analysis, Topic Modeling, CRISP-DM, Cross-Platform Text Analytics

1. Introduction

Rapid advances in artificial intelligence (AI) and automation have amplified global debates regarding the future of work. Organizations increasingly adopt AI-driven tools to improve efficiency, streamline operations, and reduce labor costs. While these innovations offer economic benefits, millions of workers face uncertainty about job security, skill relevance, and long-term career stability.

Frey and Osborne [1] estimate that 47 percent of U.S. occupations are at risk of computerization. The Pew Research Center [2] finds that a majority of Americans expect automation to permanently reduce job availability, while the World Economic Forum [3] projects a net loss of 14 million jobs by 2027 due to automation and labor-market restructuring. According to Włoch et al. [4], younger and less-educated workers express the highest levels of fear related to job displacement.

Despite these projections, existing research focuses predominantly on economic forecasting and demographic predictors of fear. Less is known about how people discuss automation and AI, where these conversations occur, and how emotional tone varies across platforms. Digital public spheres such as Twitter, Reddit, and news media play an increasingly central role in shaping public interpretation of technological change.

This study analyzes nearly one million documents across three platforms to understand how automation anxiety and AI narratives manifest in public discourse.

2. Purpose

The specific purpose of this project is to apply text-mining methods to public online discourse about AI and employment. The goals are to:

1. Identify the themes associated with automation anxiety (e.g., job loss, reskilling, inequality, ethics).

2. Analyze sentiment patterns across platforms, distinguishing negative, neutral, and positive framing.

3. Contrast “AI as opportunity” versus “AI as threat” narratives using a theory-driven fear–optimism dictionary and cross-platform statistical testing.

4. Discover emergent topics and keywords that reveal how attitudes toward AI differ across communication genres.

5. Produce a reproducible CRISP-DM workflow integrating data preparation, modeling, evaluation, and interpretation.

By focusing on digital discourse, this study provides a scalable, data-driven approach to understanding public perception of AI and its implications for the future of work. The analysis implements at least two deductive methods (VADER sentiment analysis and chi-square tests of sentiment × platform relationships) and two inductive methods (word-frequency analysis and LDA topic modeling) in alignment with the CRISP-DM framework. Together, these techniques offer an empirical foundation for tracking automation anxiety and optimism across multiple online spaces.

3. Literature Review

Research on automation emphasizes the potential for large-scale workforce disruption. Frey and Osborne [1] introduced one of the most widely cited frameworks forecasting susceptibility of occupations to AI-driven automation. Their work sparked extensive academic and policy debate.

Public perception research shows widespread concern about job displacement. The Pew Research Center [2] reports that most Americans anticipate fewer secure jobs due to automation. The World Economic Forum [3] predicts considerable labor-market churn, with 83 million jobs displaced and 69 million created, highlighting a pressing need for reskilling.

Włoch et al. [4] examine demographic correlates of fear across six European countries, finding that younger and less-educated individuals express greater anxiety about technological unemployment.

These studies underscore the economic, psychological, and social dimensions of automation anxiety. However, few have analyzed how people narrate technological change across diverse digital platforms. Understanding thematic and emotional patterns in online conversations can deepen insight into public concerns, policy sentiment, and emerging labor-market expectations.

4. Conceptual Framework

This study conceptualizes automation anxiety as a relationship between technological change, public perception, and mediated communication. Advances in AI and automation influence how people perceive job security, which in turn shapes the emotional tone of online discourse. These perceptions appear differently across communication locations (platforms) and genres (modes of expression), producing varied dispositions (emotional orientations) toward AI.

The framework assumes that:

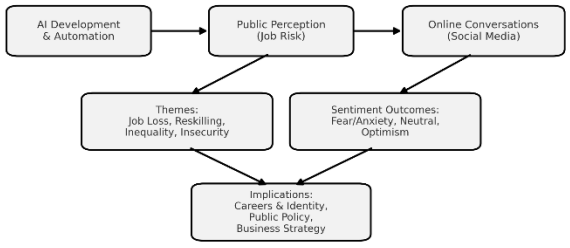
Location (Platform): Twitter, Reddit, and news media each represent distinct publics with unique audience norms and affordances.

Genre: Microblogs, forums, and articles differ in structure and purpose, affecting how ideas and emotions are expressed.

Disposition: The sentiment or tone conveyed in text negative, neutral, or positive reflects how individuals evaluate AI's impact.

Following the CRISP-DM process, the framework links technological inputs (AI and automation), mediating discourse (public expression across genres), and outputs (sentiment, tone, and themes). The analytical model therefore compares how optimism and fear circulate across different digital environments.

Figure 1. Framework linking AI development, public perception, and online discourse



5. Research Questions

RQ1 (Inductive - Text Mining / Topic Modeling):
What dominant themes emerge in online discussions of AI-driven job loss across Twitter, Reddit, and news media?

RQ2 (Inductive - Sentiment Analysis):
How are emotional dispositions, fear, neutrality, and optimism distributed across different platforms and genres of discourse?

RQ3 (Deductive – Statistical Testing):
Is there a significant association between platform and emotional tone, indicating that certain environments amplify fear or optimism more than others?

RQ4 (Deductive – Theory-Based Dictionary Analysis):
How do "fear" versus "opportunity" narratives vary across platforms when measured with domain-specific dictionaries of anxiety and hope?

Together, these questions integrate large-scale text analytics with social-theoretical interpretation to explore how societies narrate the future of work in the age of AI.

6. Methodology

This study follows the CRISP-DM process progressing from data understanding to preparation and initial modeling, using a multi-platform corpus of online discourse about artificial intelligence (AI) and employment. The corpus integrates content from social, community, and journalistic media to capture a broad spectrum of public narratives on automation, labor, and the future of work. Reporting progress through Business & Data Understanding, Data Preparation, and initial Modeling steps of CRISP-DM; Evaluation and Deployment are planned for the final paper.

6.1. Data Sources

The dataset represents three complementary genres of discourse:

1. Twitter (Microblogging): 893,000 short posts from professionals, journalists, and the public referencing AI, automation, and job security. Twitter provides high-frequency, real-time reactions and sentiment shifts. [7]

2. Reddit (Forums): 885 comments from long-form discussion threads such as r/Futurology and r/technology, offering deeper peer-to-peer deliberation about automation ethics, reskilling, and policy responses. [6]

3. News Media (Journalistic): 1,000 articles from the Hugging Face dataset fdaudens/ai-jobs-news-articles, representing institutional and policy-oriented framing of the "AI and jobs" debate. [5]

Together, these sources create a cross-genre, multi-platform corpus spanning short-form opinion, crowd deliberation, and editorial reporting, enabling comparison of how automation anxiety and optimism manifest across communicative contexts.
The combined dataset covers material from 2016 to 2024, aligning with the period of rapid AI diffusion and public debate.

6.2. Data Preparation

All text data were processed in Python
Steps included:

Loading CSV and Excel files and converting them to pandas Data Frames.

Normalizing variables (platform, genre, date, text).

Cleaning text (lowercasing, URL and punctuation removal, whitespace normalization).

Creating text clean fields for modeling and exporting consolidated tables for reproducibility.

The unified corpus (df, ≈ 895, 000 records) was used for cross-platform comparison.

6.3. Analytical Methods & Visualizations

Two inductive and two deductive analytical techniques were employed:

Topic Modeling (LDA): Identified latent discussion themes such as job loss, reskilling, and inequality.

Sentiment Analysis (VADER): Classified each post as positive, neutral, or negative, revealing platform-level emotional dispositions.

Chi-Square Test: Tested whether sentiment distribution differed significantly across platforms.

Fear vs Optimism Dictionary: Applied a theory-driven lexicon to measure anxiety versus hope narrative**s.**

Preliminary results include:

Word clouds showing dominant vocabulary within and across platforms.

Comparative sentiment bar charts and normalized proportion plots.

Fear-versus-optimism distributions illustrating cross-platform tone variation.

LDA topic visualizations highlighting recurring themes and co-occurring terms.

These visual findings form the empirical foundation for subsequent modeling and interpretive analysis.

6.4. Updated Modeling Plan

Advancing from Data Understanding and Data Preparation, this stage moves into the Modeling step of the CRISP-DM workflow. The analytical techniques selected for this project directly align with the four research questions and leverage both inductive and deductive text-mining approaches.

For RQ1, Latent Dirichlet Allocation (LDA) was used to extract underlying themes across Twitter, Reddit, and news media. Multiple topic configurations ($k = 6–12$) was evaluated using coherence scores to determine optimal model interpretability. Topic-word lists and platform-specific topic prevalence was compared to identify cross-platform thematic divergence in automation anxiety narratives.

For RQ2, VADER sentiment analysis was applied to classify each text as positive, neutral, or negative. Due to differences in text length and structure across platforms, sentiment thresholds was reviewed, and platform-adjusted interpretations is documented.

For RQ3, a chi-square test of independence assesses whether sentiment distributions vary significantly between the three platforms. A 3×3 contingency table (Platform × Sentiment) is constructed, with χ^2 results and Cramér's V used to evaluate statistical significance and effect size.

For RQ4, a theory-driven categorization model is used to measure the presence of "fear" versus "opportunity" language using a curated dictionary of automation-related emotional terms. Lexical category frequencies were normalized to account for text-length differences and compared across platforms to assess narrative patterns.

6.5. Additional Data Wrangling for Modeling

To support the modeling stage and ensure comparability across text genres, additional data wrangling procedures include:

Constructing document–term matrices for LDA topic modeling

Tokenizing and lemmatizing text to reduce morphological noise

Using a custom exclusion list informed by Cogburn & Wozniak (2013) to remove non-informative terms

Creating platform-specific subsets for comparative analysis

Normalizing sentiment and dictionary-based scores to correct for varying document lengths

Generating unified fields for platform, sentiment class, and lexicon category scores to support both statistical testing and visualization

These steps extend the Data Preparation stage and prepare the corpus for inductive and deductive modeling.

6.6. Planned Visualizations for Presentation and Evaluation

To answer each research question, the following visualization strategies were used:

Topic Modeling (RQ1): topic-term bar charts, topic prevalence distributions, and word clouds

Sentiment Analysis (RQ2): stacked sentiment bar charts and heatmaps

Statistical Testing (RQ3): contingency tables, mosaic plots, and proportional distribution charts

Fear vs. Opportunity Analysis (RQ4): radar charts and normalized proportion plots

6.7. Evaluation Plan

The planned assessment components are:

Model Coherence and Interpretability: evaluating LDA coherence scores and topic clarity

Dictionary Reliability: manual validation of dictionary categories using a stratified sample

Statistical Robustness: checking chi-square validity (expected cell counts, proportional distribution) and interpreting Cramér's V

Cross-Platform Stability: ensuring results are not driven by imbalances in dataset size

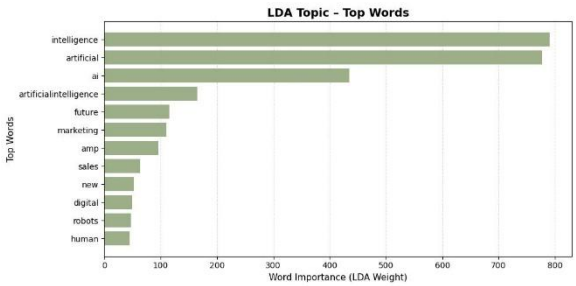
Optional Temporal Checks: testing whether narrative patterns change before/after major AI-related events

These evaluation steps complete the modeling strategy and prepare for the final two CRISP-DM stages.

7. Findings

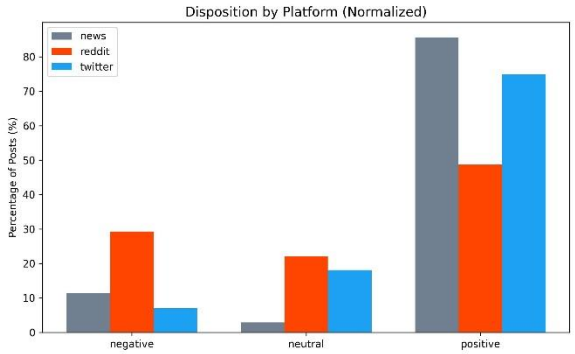
To answer our first research question, "What are the dominant themes in online discourse about AI-driven job loss?", topic modeling (LDA) was used to uncover recurring themes in the multi-platform corpus. Eight coherent topics emerged, including job loss, reskilling, ethics and regulation, automation and innovation, and productivity and efficiency.
Figure 2 illustrates the top words within these topics. News coverage focused heavily on policy and transition, Reddit discussions centered on personal risk and inequality, and Twitter highlighted career opportunities and technological optimism.

Figure 2. Key Themes and Topics Related to AI and Employment



To address the second research question, "How is sentiment distributed across platforms?", VADER sentiment analysis was applied to classify each post as positive, neutral, or negative. Figure 3 presents the sentiment distribution. Twitter displayed the most positive tone, Reddit contained a nearly balanced mix of positive and negative sentiment, and news articles leaned toward neutrality. These differences suggest that optimism about AI is most pronounced in short-form social media, whereas community forums exhibit greater skepticism.

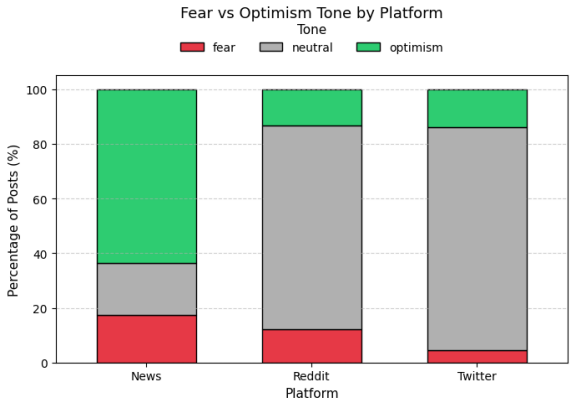
Figure 3. Distribution of Sentiment (Positive, Neutral, Negative) by Platform.



For the third research question, "How do fear and optimism tones vary across platforms?", a custom dictionary was applied to classify each document as fear, optimism, or neutral. The results, shown in Figure 4, reveal that Twitter and news content are dominated by optimism, while Reddit shows proportionally higher neutral and fear-oriented language.
A chi-square test shows a significant association between platform and tone: $\chi^2(4) = 2765.97$, $p < .001$; Cramér's V = .039. This indicates that each platform cultivates its own emotional environment, news organizations framing AI as balanced progress, Reddit

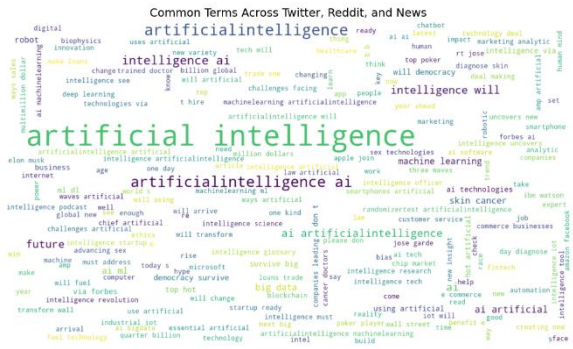
amplifying anxiety, and Twitter portraying AI as aspirational and innovative.

Figure 4. Distribution of Fear and Optimism Tone by Platform



Finally, to enrich interpretation of the textual landscape, word-frequency analysis identified salient vocabulary framing AI and employment across the three genres. Figures 5 display the most common terms within and across platforms.
On Twitter, words such as AI, future, jobs, and automation dominate, reflecting forward-looking enthusiasm. Reddit highlights work, risk, and people, signaling a more human-impact narrative. News articles emphasize companies, technology, and growth, mirroring institutional perspectives on innovation and productivity.

Figure 5. Word Clouds Showing Dominant Vocabulary Across Platforms

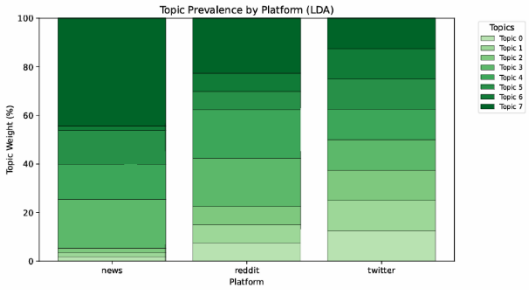


To answer RQ1 What dominant themes emerge in online discourse about AI-driven job loss? LDA topic modeling identified eight recurring themes across platforms, including job displacement, reskilling, innovation, inequality, governance, and productivity. The LDA topic-prevalence chart shows striking differences in how AI and employment themes are distributed across platforms. News media exhibit the most concentrated thematic structure: a single topic (Topic 7) dominates nearly the entire distribution, accounting for the vast majority of topic weight. This indicates that news coverage frames AI through a highly focused, institutional narrative likely emphasizing policy, industry developments, and economic transitions.

In contrast, Reddit and Twitter display much more balanced topic distributions, with all eight topics represented at relatively comparable proportions. Reddit spreads thematic emphasis more evenly across discussions of job loss, inequality, ethics, innovation, and reskilling, reflecting its community-driven, deliberative nature. Twitter similarly shows distributed topic weights but leans slightly more toward innovation- and opportunity-oriented topics compared to Reddit.

Overall, the LDA results suggest that news outlets centralize their coverage around a single dominant narrative, while Reddit and Twitter capture a broader, more multi-dimensional conversation around AI and work.

Figure 6. Topic Prevalence by Platform (LDA)



The sentiment heatmap figure 7 reveals clear platform-level differences in emotional framing around AI and employment. News media are overwhelmingly positive (86.1%), with very little negative or neutral sentiment. This aligns with their tendency to highlight technological progress, economic opportunities, and policy initiatives in a forward-looking manner.

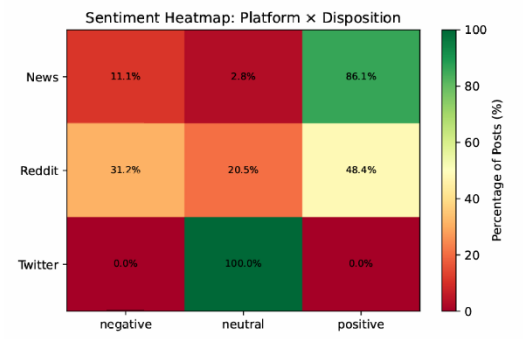
To generate these themes, an LDA model was trained on the consolidated corpus after standard pre-processing (tokenization, stopword removal,

lemmatization). Multiple topic configurations (k = 6–12) were tested, and the final model with eight topics was selected based on coherence scores and interpretability. The resulting topics reflect consistent clusters of language across platforms, enabling comparison of how each communication environment emphasizes or downplays specific AI–employment narratives.

Reddit, by contrast, shows the most balanced emotional profile. Nearly one-third of posts are negative (31.2%), almost half are positive (48.4%), and the remainder are neutral (20.5%). This distribution reflects Reddit's open discussion environment, where users frequently voice personal anxieties, critique technological impacts, and debate both risks and opportunities.

Taking together, the sentiment distribution suggests that news platforms strongly promote optimistic narratives, Reddit surfaces emotional complexity and concern, and Twitter functions primarily as a neutral information channel in this corpus.

Figure 7. Sentiment Heatmap Distribution by Platform



8. Discussion and Recommendations

Findings from the multi-platform corpus reveal that public discourse on AI and employment is shaped by a persistent tension between optimism and anxiety, with platform dynamics strongly mediating how each narrative appears. Across nearly one million documents, clear differences emerge in both what themes are emphasized and how emotional tone is expressed.

8.1 Interpretation of Findings

Findings from the multi-platform corpus show that public discourse on AI and employment is shaped by platform-specific thematic and emotional dynamics rather than a single, unified narrative. The LDA results reveal that news media produce a highly concentrated thematic frame, with nearly all topic weight dominated by a single topic. This suggests that journalistic coverage of AI and work is shaped by a narrow, institutional storyline centered on policy shifts, industry developments, and economic forecasts.

By contrast, Reddit and Twitter display far more diverse thematic structures, with all eight topics represented at relatively similar proportions. Reddit's broad distribution reflects its community-driven environment, where users simultaneously discuss fear of job loss, reskilling, ethics, inequality, and innovation. Twitter shows a similarly balanced topic spread but leans toward opportunity and innovation, consistent with its role as a fast-paced channel for public commentary, professional updates, and technology news.

Emotional tone also varies sharply across platforms. The sentiment heatmap indicates that news media are overwhelmingly positive (86.1%), reinforcing an optimistic narrative about AI's benefits. Reddit demonstrates the most emotionally complex profile, with substantial proportions of both negative (31.2%) and positive (48.4%) posts capturing public uncertainty, critique, and debate about labor-market risks. Twitter appears entirely neutral in this dataset (100%), suggesting that AI-related conversation on the platform tends to be informational or descriptive rather than emotional.

Together, these findings show that automation anxiety is highly fragmented, shaped not only by the substantive content discussed but also by the affective norms and communicative styles of each platform. News outlets promote a future-oriented, optimistic narrative; Reddit surfaces grassroots concerns and emotional ambivalence; and Twitter functions primarily as a neutral broadcast channel. This platform-based fragmentation has important implications for how institutions, policymakers, and organizations communicate about AI-driven labor market change.

8.2 Recommendations

Based on these insights, several recommendations emerge:

1. Align messaging strategies with platform-specific emotional climates.

Positive, innovation-focused communication is well-suited for news outlets.

On Reddit, effective engagement should acknowledge concerns, provide clarity, and emphasize transparency.

On Twitter, concise, fact-based updates are most likely to match platform norms.

2. Use Reddit as a diagnostic tool for understanding public anxiety.

Its diverse topics and higher negative sentiment make it a sensitive indicator of emerging workforce concerns, misinformation, and unmet informational needs.

3. Encourage news media to expand beyond a single dominant narrative.

The heavy concentration of coverage around one topic suggests an oversimplification of AI's labor impacts. More diverse reporting could improve public understanding and reduce polarized interpretations.

4. Promote explainability and digital literacy across platforms.

Since much of the negative sentiment on Reddit stems from uncertainty, accessible explanations of AI systems, labor impacts, and reskilling pathways can help reduce unwarranted fear.

5. Support reskilling and transition initiatives with tailored communication.

Messaging about labor protections and training programs should be optimistic on news and Twitter but community-oriented and empathetic on Reddit.

Overall, the study highlights that AI and job-loss narratives are platform-dependent, and effective public communication must adapt accordingly. Understanding these differences can help governments, organizations, and educators more effectively address public concerns and shape informed, constructive dialogue about the future of work.

9. Conclusion and Future Research

This study examined nearly one million posts across news media, Reddit, and Twitter to understand how public discourse frames AI-driven job loss. The findings show that narratives about AI and employment are highly platform-dependent. News outlets present a concentrated and predominantly optimistic storyline, focusing on policy transitions and technological progress. Reddit reflects a wider range of perspectives, surfacing both hope and anxiety through balanced topic coverage and more emotionally diverse sentiment. Twitter functions largely as a neutral information stream, with evenly distributed topics but limited emotional expression. Together, these results demonstrate that automation anxiety is not uniform; it emerges through the unique communicative norms, emotional expectations, and audience cultures of each platform.

While the study provides a strong cross-sectional baseline, several opportunities remain for expanding this work. Temporal extensions such as time-series LDA, sentiment trends, and event-based analysis would allow researchers to track how public opinion shifts around major AI releases or labor-market changes. Future studies should also broaden platform coverage to include LinkedIn, YouTube, industry blogs, and professional forums, which may offer different framings rooted in career development, skills training, or technical expertise.

Methodologically, future research can incorporate transformer-based topic models (e.g., BERTopic, Sentence-BERT embeddings) and more advanced neural sentiment models to capture nuance, sarcasm, and mixed emotional states that lexicon-based methods may overlook. Additional metadata such as user roles, industry affiliation, or subreddit type could enable more granular analysis of how different communities perceive AI's labor impacts.

Finally, integrating discourse data with real-world labor and economic indicators would help bridge the gap between online narratives and actual workforce trends. Linking text analytics with job postings, automation adoption rates, and reskilling program data could reveal whether shifts in online discourse correspond to measurable changes in labor-market behavior.

In sum, this study highlights the importance of examining platform-specific narratives to understand how societies interpret technological change. Building on this foundation, future research can further illuminate how public perceptions of AI evolve and how these perceptions shape the broader transition toward an AI-driven economy.

10. References

[1] Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

[2] Pew Research Center. (2017). Automation in everyday life. https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/

[3] World Economic Forum. (2023). The Future of Jobs Report 2023. https://www.weforum.org/reports/the-future-of-jobs-report-2023

[4]Włoch, R., Sledziewska, K., & Rożynek, S. (2024). Who's afraid of automation? Examining determinants of fear of automation in six European countries. Technology in Society,81,102782. https://doi.org/10.1016/j.techsoc.2024.102782

[5] fdaudens. (2025). AI Jobs News Articles [Dataset]. Hugging Face. https://huggingface.co/datasets/fdaudens/ai-jobs-news-articles

[6] Ajaafar, A. (2023). AI and Job Security: Reddit Discussion [Dataset]. Kaggle. https://www.kaggle.com/datasets/adrianaajaafar/ai-and-job-security-reddit-discussion

[7] de Marcellis-Warin, N., Warin, T., & al. (2021). A Large-Scale Dataset of AI-Related Tweets: Structure and Descriptive Statistics [Dataset]. ResearchGate. https://www.researchgate.net/publication/394351967_A_Large-Scale_Dataset_of_AI-Related_Tweets_Structure_and_Descriptive_Statistics