Investigating Public Concerns About AI: A Comparative Topic Modeling Analysis of Tweets and News Articles

Elisa Degara **Francesca Dessalvi** **Clara Montemurro** **Tommaso Giacomello**

Abstract

We investigate public concerns about artificial intelligence (AI) by comparing their expression in traditional and social media. Using a curated set of English-language news articles and tweets from X (formerly Twitter), we isolated negatively framed content. For tweets, we manually annotated posts and fine-tuned a sentiment classifier (Twitter-RoBERTa) to extract negative opinions. We then applied three topic modeling methods (LDA, NMF, SVD) to both corpora. Based on coherence scores and interpretability, we selected NMF and manually refined its output into a taxonomy of concerns. Our findings reveal both shared and medium-specific anxieties: tweets emphasize speculative, emotional fears, while news articles focus on institutional and ethical risks. This analysis highlights how public narratives about AI are shaped by platform dynamics and communication norms.

1 Introduction

Artificial intelligence (AI) is increasingly influencing a wide range of social, economic, and technological domains. While public discourse reflects optimism about its potential, it also reveals persistent anxieties and criticisms about the consequences of this technology. Understanding these concerns is not only key to evaluating sentiment, but also to anticipating societal response, informing policy, guiding responsible communication against public misunderstanding, and identifying risks of misinformation/disinformation.

Existing research has examined AI-related sentiment and topics within individual media ecosystems, especially on platforms like Twitter ([8], [9]). However, few studies have systematically compared how concerns about AI are articulated across different types of media.

Are fears shaped more by emotional or institutional framings? Do traditional and social media amplify different risks?

We constructed two temporally aligned datasets: English-language tweets from X and news articles from LexisNexis, allowing for direct comparison of how similar developments are discussed across platforms.

We manually annotated a subset of tweets to fine-tune a Twitter-RoBERTa classifier, which we then used to extract negatively framed tweets from the full dataset.

We applied three topic modelling techniques — Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorisation (NMF) and Singular Value Decomposition (SVD) — to both corpora and evaluated their coherence. NMF was selected as the most interpretable and consistent.

We then refined the NMF topics into two taxonomies—one per medium—enabling a structured comparison of shared and medium-specific concerns.

2 Dataset creation

Collecting sufficient and relevant data was a significant challenge. To enable a meaningful comparative analysis, we aimed to ensure temporal and topical consistency across datasets. To achieve this, we had to make several methodological decisions, which are described in the following section and which inevitably influenced the results.

2.1 X data

We used a curated dataset that was available on Kaggle[4]. This dataset was obtained by web scraping more than 20,000 English-written tweets from July 2023 to November 2024, looking for specific word pairs (see Table 6 in the Appendix). The dataset included sentiment annotations (positive, negative, neutral), but we chose to re-annotate the tweets due to a difference in annotation criteria: we specifically aimed to capture the overall sentiment toward AI expressed by the tweet author, which differed from the original annotation task. We preprocessed the data by filtering for English, removing sensitive information, duplicates, hashtags, and non-ASCII characters, and lowercasing. To gain a preliminary understanding of the distribution of topics across the data, we extracted the most frequent bigrams. The results are shown in Table 7 in the Appendix.

2.2 News data

We collected a sample of news articles published during the same time period as the X posts. Using the *LexisNexis* [5] database, we retrieved 500 articles published between June 2023 and December 2024 containing the keywords *"AI"* or *"artificial intelligence"*, and applied a filter for "negative news." From these articles, we manually extracted 640 concern-focused sentences and stored them in a .CSV file.

2.3 Dataset Biases and Limitations

We relied on a pre-existing collection of tweets, which limited our control over topic diversity and coverage. Moreover, as the X data was obtained via web scraping, it lacks demographic information, which makes it difficult to assess the representativeness of the dataset in relation to the broader population. The news dataset is similarly constrained as it only includes English-language articles from the LexisNexis database, which may omit perspectives from less accessible sources.

We manually assigned sentiment labels and extracted concern-expressing sentences. This procedure might have introduced bias into the corpus, due to the inherently subjective nature of the task. Furthermore, download restrictions and time constraints limited the size of the news dataset.

3 Document-Level Sentiment Classifier

The X dataset required an additional sentiment classification step to isolate negative opinions. To address this, we fine-tuned two pre-trained models (DistilBERT, Twitter-RoBERTa-base) using a baseline of 2,000 manually annotated tweets and used the best performing one to annotate the rest of the dataset. Model performances at evaluation are shown in Table 1.

Before annotation began, we established clear guidelines to ensure consistency. Our core annotation question was: '*What is the overall sentiment towards AI expressed by the author of this tweet?*'. A team member labeled tweets as *Positive*, *Negative*, or *Neutral*, applying these guidelines consistently throughout the dataset.

3.1 Experiments

Baseline: To evaluate the performance of DistilBERT and Twitter-RoBERTa, we first implemented a simple baseline model for comparison. We extracted character-level n-grams (2–6), applying TF–IDF weighting, and training an L_2-regularized logistic regression classifier with balanced class weights. This model was trained on an 80/20 train-test split and evaluated using precision, recall, and F_1-score. As shown in Table 9 in the Appendix, the baseline model achieves an overall accuracy of 65.1% and a macro-averaged F_1 score of 64%.

DistilBERT Base: To improve upon our baseline model, we first attempted to fine-tune DistilBERT[11], a lightweight transformer that is not specifically trained on Twitter data. We trained the model for 8 epochs, with a learning rate of $1\mathrm{e}{-5}$, using early stopping with patience of 2. However, the model failed to outperform the baseline. It achieved a macro-averaged F_1 score of 63%. The complete performance scores can be seen in Table 10 in the Appendix.

Twitter-RoBERTa-base: Twitter-RoBERTa-base [1] is a RoBERTa-base model, pre-trained on approximately 58 million English tweets and fine-tuned for sentiment analysis using the TweetEval benchmark. We further fine-tuned this model on our labeled corpus using the Hugging Face ecosystem. The dataset was randomly split into 80% training, 10% validation, and 10% test subsets. Each tweet was tokenized using the model's tokenizer, with padding and truncation applied to ensure a consistent input length. Training was carried out using the AdamW optimizer and monitoring the macro-averaged F_1 score on the validation set to select the best-performing model. As shown in Table 8, the final model achieved an accuracy of 69.8% and a macro-averaged F_1 score of 69.4% on the test set, outperforming the baseline.

We then used the final Twitter RoBERTa model to label the remaining unlabeled tweets, and retained only those classified as negative for topic modeling.

4 Topic Modeling

We applied the same preprocessing and topic modeling pipeline to both the tweet corpus and the

Class	Baseline	DistilBERT	RoBERTA
Negative (*)	0.7323	0.6942	0.7031
Neutral (*)	0.5635	0.4390	0.6154
Positive (*)	0.6291	0.7603	0.7619
Accuracy	0.6512	0.6543	0.6975
Macro-avg F_1	0.6416	0.6312	0.6935

Table 1: Comparison of model performances on test set ((*): F1-score)

news-article corpus. Each document was lowercased, stripped of stop words and punctuation, lemmatized, and filtered to retain only content words (nouns, verbs, adjectives, and adverbs). To improve semantic grouping, we identified and merged frequent bigrams using Gensim's `Phrases` model [6], retaining those with normalized pointwise mutual information (NPMI) greater than 0.5 and occurring in more than two documents (Equation 2).

We then applied three topic modeling methods: Non-negative Matrix Factorization (**NMF**), Truncated Singular Value Decomposition (**SVD**), and Latent Dirichlet Allocation (**LDA**).

4.1 Dimensionality Reduction Models

We applied SVD and NMF to the TF–IDF matrix of each corpus. Before applying these methods, we filtered the vocabulary based on corpus-specific thresholds. For the news corpus, we removed terms appearing in more than 50% of documents or fewer than two documents. For the tweet corpus, we excluded terms appearing in more than 70% or fewer than 10% of documents, to account for the shorter and noisier nature of the content.

We set the number of components to *n_topics* = 10 for both NMF and SVD (See Section 5) and we extracted the 5 words with the highest weights from each component.

4.2 Latent Dirichlet Allocation

After applying the same preprocessing, we performed Latent Dirichlet Allocation (LDA) [3] to model latent topics in the. To tune the model, we conducted a grid search over key hyperparameters: the topic prior α, the word prior β, and the number of passes. We tested both fixed values (e.g., 0.1, 0.5) and automatic configurations (*symmetric*, *asymmetric*, *auto*) for α and β, and varied the number of passes in $\{1, 5, 10\}$. For each configuration, we trained models with a range of topic numbers between 3 and 15.

We defined the following score to select the best model and the optimal number of topics:

$$\text{Score} = \lambda C_v + (1 - \lambda)\text{UMass} \quad (1)$$

While C_v [10] captures semantic similarity among top topic words, UMass [7] reflects statistical co-occurrence patterns. We set $\lambda = 0.75$, to favor interpretability.

The C_v and UMass scores were separately normalized before being combined, as they are in different ranges. This allowed us to select topic models that were both semantically coherent and statistically grounded. We evaluated models on development sets from each corpus.

5 Results

The optimal LDA configuration uses α = 'symmetric', β = 'symmetric', and 5 passes. As shown in Figure 1a of the Appendix, the optimal number of topics is 10 for the tweet corpus and 13 for the news corpus. To allow for consistent comparison across models and datasets, we fixed the number of topics to *n_topics* $= 10$ for all methods. This value was chosen as the minimum of the two LDA-optimal values, to reduce topic redundancy, maintain coherence across the three models, and improve interpretability.

Each of the 10 topics extracted is represented by the five most representative words, as determined by the internal logic of each method. The top topics and associated keywords extracted by NMF are presented in Table 2, while the outputs for SVD and LDA are available in the Appendix (Tables 11, 12, and 13).

6 Discussion

6.1 Model Evaluation

From a qualitative perspective, the topics produced by NMF appear more interpretable and semantically coherent, especially when manually inspecting the topic-word sets.

To evaluate the models quantitatively, we used the combined coherence score introduced in Equation 1, ensuring a balanced trade-off between semantic interpretability (C_v) and statistical validity (UMass). As shown in Table 3, which reports the normalized coherence scores for a representative run, NMF outperforms both LDA and SVD across both news and tweet corpora.

However, it is important to note that topic modeling is inherently stochastic. Although NMF and SVD

News – NMF Topics	
Topic 1	people, technology, make, ai technology, information
Topic 2	job, replace, year, say, lose
Topic 3	company, openai, chatbot, chatgpt, lawsuit
Topic 4	bias, datum, train, woman, gender
Topic 5	use, artificial_intelligence, use artificial_intelligence, voice, use ai
Topic 6	image, create, video, abuse, real
Topic 7	human, risk, machine, pose, control
Topic 8	generate, ai generate, content, work, generate content
Topic 9	concern, raise, potential, ethical, privacy
Topic 10	tool, ai tool, political, fear, result
Tweets – NMF Topics	
Topic 1	bot, ai bot, problem, ai, bot problem
Topic 2	take_job, ai take_job, ai, away, take_job away
Topic 3	job, automation, problem, work, replace
Topic 4	use, war, use ai, ai, israel
Topic 5	danger, ai, danger ai, humanity, great
Topic 6	real, deepfake, problem, deep_fake, real problem
Topic 7	people, people job, work, need, problem people
Topic 8	just, try, say, ai, really
Topic 9	problem, education, ai, deepfake, need
Topic 10	human, think, ai, replace, solve

Table 2: NMF Topics from News and Twitter

Model	News	Tweets
LDA	0.5151	0.2474
NMF	0.9974	1.0000
SVD	0.2500	0.1316

Table 3: Normalized C_v Coherence Scores for Topic Models

are deterministic once initialized, their results can still vary depending on the random seed and initialization, particularly in the ordering and prominence of topics. LDA, being a fully probabilistic model, exhibits even more noticeable variability. To assess the robustness of our evaluation, we therefore conducted 200 repeated runs for each model using different seeds, collecting normalized coherence scores at each iteration.

Statistical Significance: To verify that the observed superiority of NMF was not due to random chance, we performed a one-sided Wilcoxon signed-rank test [2] (`alternative='greater'`) comparing the combined coherence scores of NMF to those of LDA and SVD across all runs. In both the news and tweet datasets, the test yielded p-values $< 10^{-4}$ in favor of NMF, indicating that the differences are statistically significant and not attributable to stochastic variation. Additionally, the test confirmed that the achieved performance gain is practically meaningful.

As a result, we selected NMF as the best-performing model, and used it for final topic refinement and downstream taxonomy construction.

6.2 Content-Based and Semantical Analysis

After identifying NMF as the most coherent method, we manually grouped its topics into broader, human-labeled categories of public concern. Overlapping topics were merged under unified, descriptive labels, and the final categories were ordered by their prevalence in each corpus.

News:

- AI and automation substituting human labor
- Algorithmic bias
- Abusive use of Generative AI
- Deceptive AI use in crime and scams
- General existential threat posed by AI
- Privacy and data security concerns

Tweets:

- Bot activity
- AI substituting human labor
- Fears surrounding the use of AI in warfare
- General existential threat posed by AI
- Misinformation and deepfakes
- AI's influence in thinking and education

Though concerns about AI overlap across news and social media, key differences reflect distinct communicative norms and audiences. On X, users emphasize bot activity and militarized AI, often framed emotionally or speculatively, likely due to direct personal exposure and the reactive nature of the platform. These themes are absent from news coverage, which focuses on structured, policy-relevant risks such as job displacement, algorithmic bias, and privacy, shaped by more balanced editorial standards.

News discourse is segmented and analytical, dedicating topics to areas like gender bias, child safety, or ethical regulation. Tweets, by contrast, often blur thematic boundaries, mixing fears of misinformation, existential threat, and automation, into broader expressions of unease. The informal, urgent language on social media ("bot problem," "can't think") contrasts with the abstract, formal register of traditional journalism ("replace," "ethical concern"). Geopolitical framing in tweets (e.g., China, Israel) further signals a more politicized and emotionally charged engagement with AI, shaped by the dynamics of user-driven discourse.

6.3 Topic Refinement

To improve interpretability, we refined the initial NMF output by manually reviewing the full topic-word distributions. From ten topics per corpus, we distilled six core concerns, merging overlapping or ambiguous topics and relabeling them into clearer, human-readable categories. This process yielded the final taxonomy shown in Table 5 and led to a measurable improvement in coherence scores (Table 4).

Stage	News	Tweets
Before Refinement	0.4861	0.6038
After Refinement	0.7480	0.8238

Table 4: C_v Coherence Scores Before and After Topic Refinement

News – Refined Topics	
Topic 1	`job, replace, lose work, future, worker`
Topic 2	`bias, woman, gender, candidate, female`
Topic 3	`generative ai, image, video, abuse, child`
Topic 4	`scam, phishing attack, criminal, voice, fake`
Topic 5	`risk, fear, control, human intelligence, negative consequence`
Topic 6	`privacy, ethical, concern, raise ethical, information`
Tweets – Refined Topics	
Topic 1	`bot, ai bot, bot problem, account bot, bot ai`
Topic 2	`job, automation, take_job away, work, people job`
Topic 3	`war, use ai, israel, propaganda, china`
Topic 4	`danger, danger ai, real, humanity, problem`
Topic 5	`make, generate, deep_fake, art, ai art`
Topic 6	`can't think, human, think, replace, problem`

Table 5: Refined NMF Topic Taxonomy for News and Tweets

7 Related work

Our study builds on recent work analyzing public sentiment and concerns around artificial intelligence (AI), especially through social media.

For model fine-tuning, we adopted the *TweetEval* benchmark by (Barbieri et al., 2020)[1], which provides pre-trained models and standardized sentiment analysis tasks for Twitter data. We used the *Twitter-RoBERTa-base* model from this benchmark to ensure domain-specific classification.

We were also inspired by *"Excitements and Concerns in the Post-ChatGPT Era"* [9], which uses sentiment analysis and topic modeling on large-scale Twitter data to track narratives around AI. Unlike our work, it focuses exclusively on social media, while we emphasize a comparative approach across media types and construct a refined taxonomy using interpretable NMF-based topics.

Finally, *"Public Perception of Generative AI on Twitter"* [8] analyzes discourse by demographic groups and use cases. While we do not examine demographic variation, our work shares its broader aim of understanding how different populations engage with AI-related issues.

8 Conclusion

We compared public concerns about AI in social and traditional media using sentiment classification and topic modeling. After evaluating three models, we selected NMF and refined its output into two taxonomies. News articles emphasized institutional concerns—privacy, bias, job automation—while tweets reflected personal fears, bot activity, and militarized AI. These contrasts underscore how public narratives about AI are shaped not only by events but also by the platforms through which they are discussed. Our work demonstrates the value of combining NLP techniques with qualitative analysis to uncover discourse patterns across platforms.

9 Limitations and future directions

In addition to the data collection limitations and potential biases previously discussed, the absence of demographic metadata limits analysis of population-specific concerns. Inferring user categories based on their susceptibility to particular topics could enable more targeted follow-up research or interventions. Additionally, tweets in our dataset were analyzed in isolation, preventing us from studying conversational context or responses. This limits the ability to construct emotion or influence networks and identify key opinion leaders. Finally, while topic modeling offers a high-level view of emerging themes, the manual interpretation of topics may overlook subtle or overlapping meanings.

References

[1] Francesco Barbieri et al. "TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 1644–1650. DOI: 10.18653/v1/2020.findings-emnlp.148. URL: https://aclanthology.org/2020.findings-emnlp.148.

[2] Ana Durango and Craig Refugio. *An Empirical Study on Wilcoxon Signed Rank Test*. Dec. 2018. DOI: 10.13140/RG.2.2.13996.51840.

[3] Matthew Hoffman, Francis Bach, and David Blei. "Online Learning for Latent Dirichlet Allocation". In: *Advances in Neural Information Processing Systems*. Ed. by J. Lafferty et al. Vol. 23. Curran Associates, Inc., 2010. URL: https://proceedings.neurips.cc/paper_files/paper/2010/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf.

[4] *Kaggle dataset*. 2025. URL: https://www.kaggle.com/datasets/mizanurrahmanrafi/sentiment-classification-of-ai-related-tweets.

[5] LexisNexis. *LexisNexis Academic Database*. 2025. URL: https://www.lexisnexis.com.

[6] Tomas Mikolov et al. *Distributed Representations of Words and Phrases and their Compositionality*. 2013. arXiv: 1310.4546 [cs.CL]. URL: https://arxiv.org/abs/1310.4546.

[7] David Mimno et al. "Optimizing Semantic Coherence in Topic Models". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Ed. by Regina Barzilay and Mark Johnson. Edinburgh, Scotland, UK.: Association for Computational Linguistics, July 2011, pp. 262–272. URL: https://aclanthology.org/D11-1024/.

[8] Kunihiro Miyazaki et al. "Public perception of generative AI on Twitter: an empirical study based on occupation and usage". In: *EPJ Data Science* 13.1 (2024), p. 2. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-023-00445-y. URL: https://doi.org/10.1140/epjds/s13688-023-00445-y.

[9] Weihong Qi et al. *Excitements and Concerns in the Post-ChatGPT Era: Deciphering Public Perception of AI through Social Media Analysis*. 2023. arXiv: 2307.05809 [cs.SI]. URL: https://arxiv.org/abs/2307.05809.

[10] Michael Röder, Andreas Both, and Alexander Hinneburg. "Exploring the Space of Topic Coherence Measures". In: *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining* (Feb. 2015), pp. 399–408. DOI: 10.1145/2684822.2685324.

[11] Victor Sanh et al. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *ArXiv* abs/1910.01108 (2019).

Appendix

.1 Additional Dataset creation results

Word Pairs	
AI Danger	AI Ethics
AI Problems	AI Creating Problems on Jobs
AI Threat	AI Used in War
AI Bot Problems	AI Problems for Students

Table 6: Word Pairs used for web scraping in Kaggle dataset

Bigram	**Count**
ai taking	2060
taking jobs	1742
using ai	1698
ai bots	767
ai bot	709
problem solving	696
problem ai	676
taking job	530
ai war	469
bot problem	466

Table 7: Top 10 Bigrams in Tweets

.2 Additional fine-tuning results

Class	F_1–Score
Negative (0)	0.7031
Neutral (1)	0.6154
Positive (2)	0.7619
Accuracy	0.6975
Macro-average F_1	0.6935

Table 8: 'Twitter RoBERTa' model performance on the test set.

Class	Precision	Recall	F_1–Score
Negative (0)	0.7099	0.7561	0.7323
Neutral (1)	0.6296	0.5100	0.5635
Positive (2)	0.5982	0.6634	0.6291
Accuracy		0.6512	
Macro Average	0.6459	0.6432	0.6416
Weighted Average	0.6503	0.6512	0.6480

Table 9: Baseline model performance on the test set.

Class	F_1–Score
Negative (0)	0.6942
Neutral (1)	0.4390
Positive (2)	0.7603
Accuracy	0.6543
Macro-average F_1	0.6312

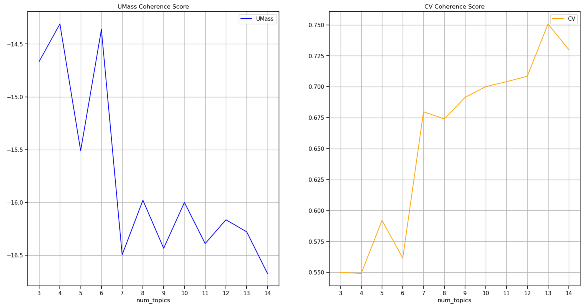
Table 10: 'DistilBERT' model performance on the test set.

.3 NPMI Calculation

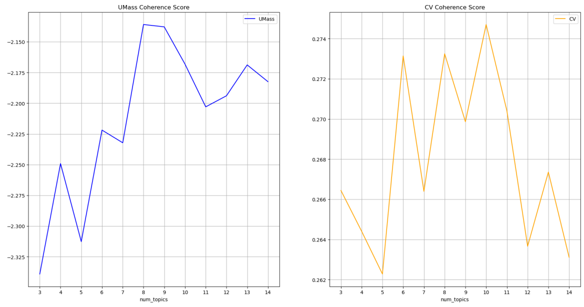
The Normalized Pointwise Mutual Information (NPMI) between two terms x and y is computed as:

$$\mathrm{NPMI}(x,y) = \frac{\log\big(p(x)\,p(y)\big)}{\log p(x,y)} - 1 \in [-1,1], \tag{2}$$

A value of 1 indicates perfect co-occurrence, 0 indicates independence, and –1 indicates complete mutual exclusion.



(a) News corpus



(b) Tweet corpus

Figure 1: Coherence scores across different topic counts. Left of each subplot: UMass coherence. Right: CV coherence. These metrics were used to guide the selection of topic count, as discussed in Section 4.2.

.4 Additional topic modelling results

Topic 1	`combine, realistic, violence, success, target`
Topic 2	`gender, such, artificial_intelligence, openai, use`
Topic 3	`regulate, u.s., fake, also, too`
Topic 4	`keep, benign, exploit, request, designer`
Topic 5	`scale, research, world, draw, ghost`
Topic 6	`investor, race, computer, time, argue`
Topic 7	`fraud, abuse, fuel, app, war`
Topic 8	`ready, much, depression, website, medium`
Topic 9	`date, million, times, york, microsoft`
Topic 10	`bring, quickly, low, publish, effect`
Topic 11	`article, disinformation, thing, falsely, face`
Topic 12	`increase, say, adopt, pose, be`
Topic 13	`screenwriter, celebrate, copyright_infringement, previously, industrial`

Table 11: LDA Topics from News Articles

Topic 1	bot, problem, have, ai, go
Topic 2	automation, job, solve, problem, need
Topic 3	even, get, take, good, bot
Topic 4	real, only, see, problem, human
Topic 5	know, work, do, danger, have
Topic 6	people, time, now, use, ai
Topic 7	take_job, take, job, think, do
Topic 8	use, war, ai, education, now
Topic 9	more, be, human, use, war
Topic 10	make, say, so, just, thing

Table 12: LDA Topics from Tweets

Topic 1 – News	use, artificial_intelligence, technology, say, people
Topic 1 – Tweets	ai, problem, job, bot, use
Topic 2 – News	job, human, concern, risk, technology
Topic 2 – Tweets	take_job, ai take_job, job, away, take_job away
Topic 3 – News	company, datum, work, train, openai
Topic 3 – Tweets	job, automation, problem, people, work
Topic 4 – News	bias, woman, datum, generate, content
Topic 4 – Tweets	bot, take_job, ai bot, problem, bot problem
Topic 5 – News	job, use, artificial_intelligence, tool, use artificial_intelligence
Topic 5 – Tweets	danger, danger ai, ai, real, think
Topic 6 – News	job, image, woman, year, say
Topic 6 – Tweets	job, ai, bot, ai bot, human
Topic 7 – News	human, risk, artificial_intelligence, machine, chatbot
Topic 7 – Tweets	people, bot, people job, ai bot, just
Topic 8 – News	generate, work, ai generate, human, say
Topic 8 – Tweets	just, make, job, art, say
Topic 9 – News	tool, generative, generative ai, ai tool, increase
Topic 9 – Tweets	education, ai, problem ai, people, ai take_job
Topic 10 – News	tool, say, ai tool, need, student
Topic 10 – Tweets	human, make, think, bad, art

Table 13: SVD Topics from News and Twitter