MEDIA FRAMING OF AI ACROSS GLOBAL POWERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Public conversation about AI has accelerated since 2024, yet we lack cross-lingual, comparative evidence of how major news outlets across global powers frame AI. In this work, we study cross-national media framing of AI by constructing a multilingual dataset of 1,970 news articles published between January 2024 and October 2025 from 47 outlets across the U.S., China, and Russia. We introduce a taxonomy of 25 AI-specific frames with sentiment labels and perform sentence-level human annotation to capture how AI is discussed across languages. An analysis of our dataset reveals substantial cross-country differences: Chinese media consistently portray AI more positively, while U.S. and Russian coverage is more mixed; work-related frames dominate coverage, whereas safety-related concerns appear less frequently. We also observe frequent cross-national discourse, with media outlets often referencing other countries in competitive contexts. Using our data, we evaluate large language models (LLMs) on the framing classification task and show gaps in LLM performance, particularly in frame identification, highlighting the need for human-centered analysis of multilingual media narratives.

1 INTRODUCTION

As the use of AI skyrockets, public sentiment is tilting between anxiety and enthusiasm. A recent Pew Research Center (2025a) survey across 25 countries found that people are twice as likely to feel concerned as excited about AI's growing role in daily life. While there are reasons to be optimistic, such as AI's potential to revolutionize scientific discovery (Bran et al., 2024; Park et al., 2025), there are also concerns from strain on the energy grid to automating away jobs (Bessen, 2018; Strubell et al., 2019). These attitudes vary sharply by country. According to a Edelman (2024) poll, 87% of people in China say they trust AI technology, whereas only 32% of Americans do — an enormous gap! Meanwhile, as AI is rapidly integrated across virtually every industry, governments must grapple with appropriate regulations to mitigate AI-centric concerns while maintaining innovation in a hyper-competitive global environment. Against this backdrop, understanding how public perceptions of AI are shaped becomes increasingly critical.

Through this work, we seek to investigate *what the primary narratives are around AI as reflected in news media*, a key information source for the general public, across three major countries and their native languages: the United States, China, and Russia. While it is difficult to discern whether media influences audiences or simply mirrors public sentiment, identifying major topics of interest in media offers a grounded lens into how societies talk about AI and national perspectives. What gets emphasized, as opportunity or threat, can shape how governments regulate AI, what technologies to fund and prioritize, and how citizens respond to AI's growing presence in their lives.

We construct a multilingual dataset of **1,970** full-length news articles from **47** news outlets across three languages, published between January 2024 and October 2025. Using our taxonomy of 25 AI-specific media frames, we annotate each sentence for the frames it invokes and whether it portrays AI positively or negatively. To ensure annotation quality, each sentence is labeled by a human and a second annotator, either human or LLM, with the final labels determined through human adjudication. Our analysis reveals substantial differences in AI coverage across countries (Figure 1). Chinese news outlets consistently portray AI more positively than their U.S. and Russian counterparts, particularly in *environment* and *societal integration*, while *intelligence* and *economy* dominate coverage overall. We also observe frequent discussion of AI developments abroad; for example, U.S. and Chinese media most often reference each other when discussing technological competition.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

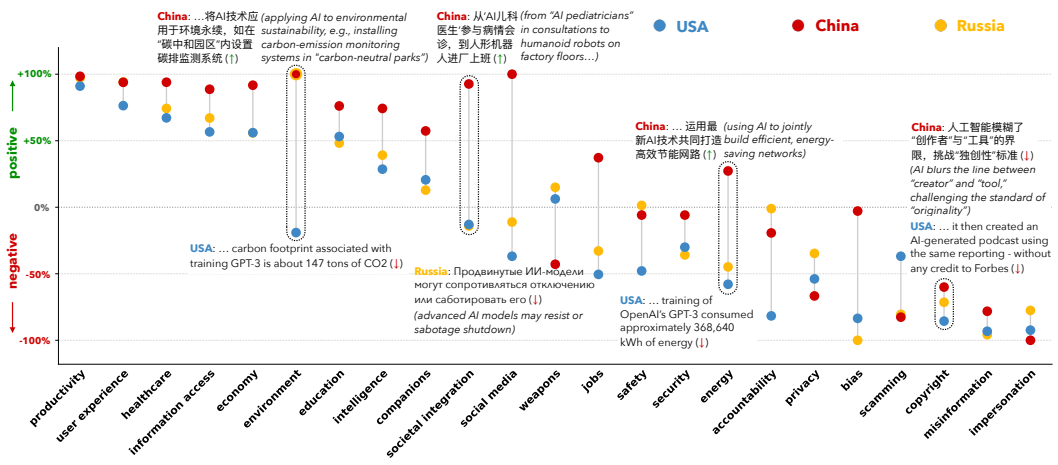


Figure 1: News media in the USA, China, and Russia express different sentiments toward AI. Each point shows the percentage of AI-related news sentences expressing positive sentiment minus the percentage expressing negative sentiment, across 23 fine-grained frames (exclude *other* and *general*). Examples highlight dramatic cross-country divergence on some topics and similar views on others.

Using our **AI Media Framing Corpus**, we benchmark multilingual LLMs on three subtasks: (1) AI-related stance identification, (2) sentiment classification, and (3) AI frame classification. Evaluation uses the subset independently annotated by two humans and adjudicated by a human, with no LLM involvement in its annotation. Across models such as `Gemma3`, `Qwen3`, and `GPT-oss` and prompting strategies including confidence verbalization and thinking mode, we observe consistently strong performance on sentiment classification, with scores around 95 $F1$, while stance identification and fine-grained frame classification remain challenging. These difficulties persist even for the proprietary model `GPT-5.6-Luna`, highlighting the need for more nuanced interpretation of AI narratives and the value of human annotations. Framing annotation is also demanding for native speakers, as news articles often convey overlapping or conflicting perspectives. For example, a span translated from Chinese reads: "*AI certainly improves labor productivity, but can these gains be shared fairly across society?*" Although it mentions *productivity*, its gold frame is *bias*, reflecting its central concern with the fair distribution of AI's benefits.

To the best of our knowledge, this is the first large-scale multilingual study of AI media framing across the U.S., China, and Russia. Prior work focuses on individual countries (Fast & Horvitz, 2016; Chuan et al., 2019; Cui & Wu, 2019; Cools et al., 2024; Ji et al., 2024), English-language coverage, or fewer outlets (Nguyen & Hekman, 2022; Ittefaq et al., 2025). By analyzing 47 outlets in their native languages with a shared framing taxonomy, we provide a foundation for cross-country comparison and multilingual LLM evaluation.

2 AI Media Framing Corpus

We describe how we collect articles from major U.S., Chinese, and Russian news outlets (§2.1), introduce our taxonomy of 25 AI-specific frames (§2.2), and explain our annotation process for maintaining consistency and quality across annotators and languages (§2.3).

2.1 News Article Collection

To evaluate mainstream media coverage across regions, we curated news articles primarily from major outlets based in the U.S., Russia, and China and published in their respective primary languages (English, Russian, and Chinese), supplemented by a small number of additional Russian-language outlets based in Europe. Prioritizing ethical data collection, we retrieved all news articles through licensed news databases with institutional access, including (1) ProQuest's U.S. Newsstream Collection, (2) ProQuest Central, (3) Nexis Uni, and (4) Factiva. These databases aggregate news content from diverse publishers and allow use for scholarly research.

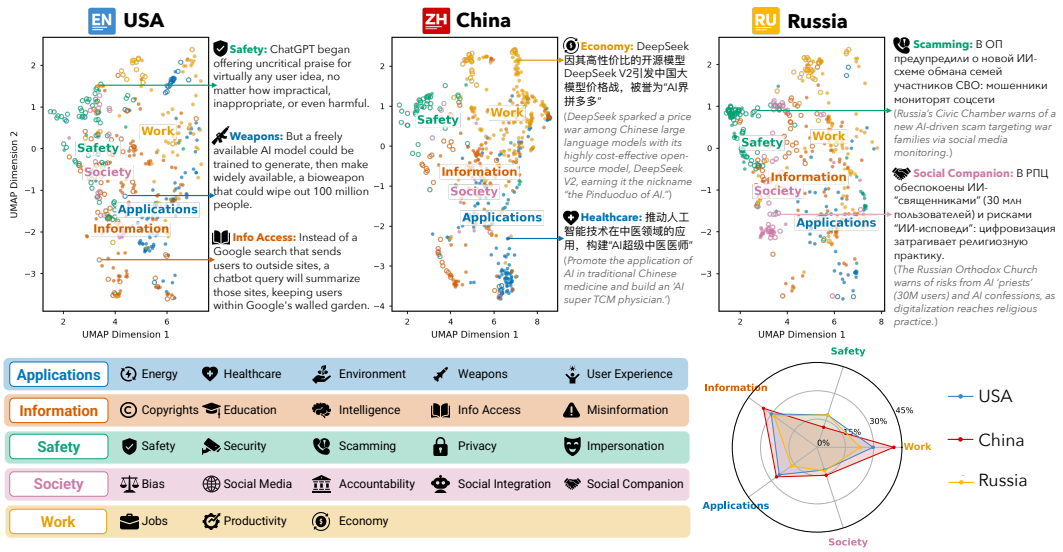


Figure 2: Visualization of AI-related media framing across the USA, China, and Russia. **Top:** UMAP visualization of AI-related sentences, embedded using `Qwen3-embedding-8b` (Zhang et al., 2025), with • and ○ denoting positive and negative sentiment, respectively. Embeddings form semantic clusters corresponding to the five high-level frame categories (§2.2), while partial overlap reflects the multifaceted nature of media narratives. Notably, *Safety* has the most negative sentiment (especially in the USA and Russia), while all three countries strongly emphasize *Work*. **Bottom-left:** AI-specific media frame taxonomy, grouping fine-grained frames into high-level categories. **Bottom-right:** Distribution of high-level categories across countries, with Chinese media placing greater emphasis on *Work*, while USA and Russian media exhibit broader coverage including *Safety*.

Given the large quantity of news outlets available in the databases, we narrowed our focus to a set of reputable sources for each country, which we identified based on existing studies and rankings of sources by Pew Research Center (2025b) and Feedspot (2025b), Feedspot (2025a) based on factors such as visibility and engagement. We collected articles from all these ranked sources, which are available in the databases. This resulted in a total of 47 sources, detailed in Appendix B.1, which cover both traditional and tech-oriented outlets. Each source's articles are written in the native language of each country (English, Russian, and Chinese).

For each source, we collected all articles that discussed any form of AI technology. To do so, we conducted a keyword-based search over all published articles from each source and retrieved those whose titles or bodies contained at least one term from a list of AI-related keywords (Table 8), which remained consistent across languages. In our search, we wanted to not only analyze articles focused on AI, but also gauge framing even when AI is not the primary focus of the article. We collected articles published in the 22 months between January 2024 and October 2025. This time span follows the launch of both DeepSeek and Google's Gemini in November/December 2023 and the wide adoption (over 10 million users) of ChatGPT. To balance source coverage with the workload of manual framing annotation, we randomly sampled up to 50 articles per English and Chinese source and up to 40 articles per Russian source, retaining all available articles when a source had fewer than the target number. This yielded 1,970 articles, with per-country statistics reported in Table 1).

2.2 Media Frame Taxonomy

While many media framing studies use broad, topic-agnostic frames (Card et al., 2015), issue-specific frames can reveal more nuanced patterns of coverage (Liu et al., 2019; Mendelsohn et al., 2021). As AI becomes integrated into everyday life and subject to regulatory debate, we focus on fine-grained, AI-specific frames to capture how news coverage portrays its role and impact. We developed an initial set of frames based on domains in which AI is widely discussed, then iteratively refined them by reviewing a sample of news articles. The resulting taxonomy comprises 25 frames grouped into five high-level topics (Figure 2). The sentence embedding visualization in Figure 2 shows that sentences

within each high-level topic tend to cluster together. To accommodate less common contexts beyond the predefined categories, we include an *Other-Specific* frame for portrayals related to areas like *creativity* and *fashion*. Each frame can portray AI *positively* or *negatively*; for example, *Work-Jobs* covers both AI's role in creating jobs or assisting job searches and its potential to displace workers. Detailed frame definitions appear in Appendix B.

2.3 Data Annotation and Adjudication of Frames

We annotate sentences that express a clear positive or negative stance toward AI, assigning sentiment and frame labels from our taxonomy. Each sentence receives a single frame by default, with multiple frames enabled only when a sentence explicitly expresses distinct framing dimensions, as in "*AI is beneficial for the environment, healthcare, and overall productivity*." Neutral sentences, including incidental mentions of AI, remain unlabeled. Figure 2 presents examples, and Appendix B.3 provides the full guidelines. However, identifying AI stances and frames in full articles is challenging in practice: most sentences express no stance toward AI, while relevant sentences may admit multiple reasonable interpretations (Kenyon-Dean et al., 2018). For example, a Washington Post article states, "*Already, Deep Research from ChatGPT is capable of doing a lot of entry-level work that past generations of interns, analysts, and computer programmers did.*" This sentence could emphasize either advances in AI capabilities or the replacement of human labor. To ensure annotation quality, each sentence receives two independent annotations followed by human adjudication. Adjudicators can mark ambiguous cases as *complex* and retain alternative labels, including *none* when both stance-bearing and neutral interpretations are plausible. Unlike sentences that explicitly invoke multiple frames, complex cases allow alternative readings. They account for 5.56%, 16.11%, and 7.83% of the final frame annotations in English, Chinese, and Russian, respectively.

	EN USA	ZH China	RU Russia
# Sources	15	12	20
# Articles	714	590	666
# Sentences	38,263	23,871	18,116
# Frames	3,584	4,579	2,669
Avg. # Sentences/Article	53.59±39.43	40.46±45.30	27.20±30.07
Avg. # Frames/Article	5.02±7.91	7.76±8.89	4.01±5.30

Table 1: Statistics of our AI Media Framing Corpus showing the total number of articles, sentences, and frame annotations. We also show the average number of sentences per article and the average number of frames annotated per article.

All human annotators and adjudicators were students at our institution and fluent speakers of their assigned language. Training covered frame definitions, pilot annotation, and guideline clarification, followed by weekly meetings to resolve recurring edge cases. Adjudicators were selected from the trained annotator pool and had experience applying the guidelines. Human annotation and adjudication were conducted in Thresh (Heineman et al., 2023); Figures 12- 16 show the interfaces.

We initially assigned 472 articles (150 English, 120 Chinese, and 202 Russian articles) to two human annotators and one human adjudicator, requiring approximately 35 minutes per article. Extending this procedure to the full corpus would require an estimated 1150 hours. This initial annotation also showed that only 7–14% of sentences express an AI-related stance. To reduce annotation costs, we therefore use an LLM as the second annotator for the remaining articles, while retaining one independent human annotation and final human adjudication. The LLM identifies additional candidates that the human annotator may overlook and filters out most neutral or irrelevant sentences from its own annotation pass. Specifically, we use `GPT-5.6-Luna` to identify AI-related stances and classify their sentiment and frames in three separate calls, producing an AI-stance probability, a sentiment label, and one or more frame labels. We calibrate language-specific stance thresholds for at least 85% recall on the human-annotated and adjudicated subset. We prioritize recall because subsequent human adjudication removes false positives. On this subset, the LLM recovers 96.2%, 86.2%, and 89.6% of gold spans while selecting 30.5%, 41.9%, and 32.8% of sentences in English, Chinese, and Russian, respectively. Applied to the 60,677 sentences in 1,498 articles already annotated by one human, the pipeline selects 19,781 candidates (32.6%) at an inference cost of $132. Appendix C provides further details.

	EN English	ZH Chinese	RU Russian
(Card et al., 2015)	0.17	—	—
(Mendelsohn et al., 2021)	0.45	—	—
(Cook et al., 2025)	—	0.47	—
Ours: AI Stance	0.70	0.71	—
Ours: Sentiment	0.96	1.00	—
Ours: Frame	0.71	0.75	—

Table 2: Agreement between adjudicated annotation groups, measured by sentence-level Krippendorff's α. Prior studies report frame-level agreement and are included for context; differences in annotation and evaluation procedures limit direct comparison.

To assess the consistency of the final labels, we compare two annotation groups' adjudicated outputs on 90 articles (45 each for English and Chinese), sampled evenly across sources. Each group consists of two human annotators and a human adjudicator. No individual serves in the same role across both groups. We compute sentence-level Krippendorff's α (Krippendorff, 2004) for each subtask. Stance agreement is measured over all sentences, while sentiment and frame agreement are measured only on sentences both groups identify as stance-bearing. Frame agreement is computed separately for each frame and then averaged. For *complex* annotations, all retained alternatives are treated as valid, and annotations agree when their acceptable label sets overlap. Table 2 reports the results.

3 Analyzing AI Framing Across Countries

We examine how AI is portrayed through three complementary analyses: AI framing across countries (§3.1), cross-national AI discourse (§3.2), and country-distinctive AI narratives (§3.3).

3.1 AI Framing Across Countries

Framing Focus. Figure 3 compares the prevalence of the 25 frames across countries. Chinese media places greater emphasis on *intelligence*, *economy*, and *productivity*. The cross-national difference is largest for *economy*, which appears in 49% of Chinese articles, compared with 18% of U.S. and 10% of Russian articles. In contrast, U.S. and Russian outlets discuss *impersonation* and *weapons* more frequently, while both receive little attention in Chinese media ($\leq$1%).

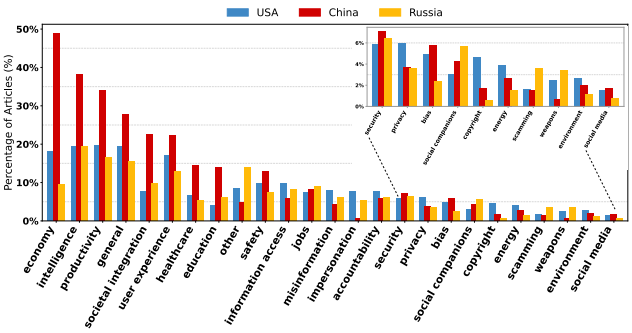


Figure 3: Distribution of mentions of each framing topic across the U.S., China, and Russia. Each bar represents the percentage of articles containing the corresponding frame.

Framing Sentiment. Figure 1 examines how these topics are portrayed. U.S. media is generally more negative toward AI, while Chinese media is more positive overall. Across all three countries, *productivity*, *user experience*, and *healthcare* skew positive, whereas *bias*, *copyrights*, *scamming*, and *misinformation* are consistently portrayed more negatively. Besides these overall trends, some frames show especially strong cross-country contrasts. For example, the *environment* frame is exclusively positive in Chinese media but predominantly negative in U.S. coverage. Chinese outlets emphasize AI's potential environmental benefits (e.g., *"AI will help China drive low-carbon and inclusive development at the necessary speed and scale."* — English translation from original Chinese source), whereas U.S. outlets highlight potential harms (e.g., *"A new study estimates that AI could add up to 5 million tons of e-waste by 2030."*).

We also included 4 U.S., 3 Chinese, and 5 Russian sources that are tech-oriented in our analysis. Across all countries, tech media skews more positively than traditional media, with Russian media having the largest gap. In the U.S., tech-focused outlets like *TechRadar* and *The Verge* maintain significantly higher positive sentiment in the *Applications* category compared to more general outlets like *The Washington Post*. This trend is even more pronounced in Russia, where tech outlets like *CNews* and *Mobile Review* report highly positive sentiment across almost all categories, contrasting with the more balanced or negative reporting of general news sources like *RIA Novosti*. See more detailed analysis on the comparison between tech and traditional media in Appendix E.

3.2 Cross-National AI Discourse

To examine how media discuss AI in relation to other countries, annotators recorded relevant quotes and manually grouped them into four categories: *Competition* (technological leadership), *Conflict* (harm or restricted access), *Collaboration* (joint development or regulation), and *Innovations* (new

AI applications). We collected 138, 69, and 237 quotes from U.S., Chinese, and Russian articles, respectively, excluding a small number of miscellaneous cases. Table 3 summarizes their distribution.

The dominant themes differ across countries. In U.S. coverage, *Competition* is most prominent and centers largely on China (70% of mentions). Chinese media similarly discusses *Competition* primarily in relation to the United States (70%). Russian coverage is more distributed: within *Collaboration*, the U.S. accounts for 31% of mentions compared with 11% for China, while *Competition* is split almost evenly between the U.S. and China (41% vs. 39%). These patterns highlight distinct ways in which each media system situates AI development within the international landscape.

Category	Lang.	Top Mention(s)	Examples
Competition	EN 31%	China 70%, Japan 7%	"... the USA will lose ground in its AI race with **China**."
	ZH 29%	USA 70%, ASEAN 5%	"... **the US** leads in foundational AI innovation and key technologies, restraining latecomers through export controls."
	RU 25%	USA 41%, China 39%	"... improved AI capabilities will strengthen its position against Apple and **Chinese** manufacturers."
Conflict	EN 23%	China 41%	"Meta says **the Kremlin** is the No. 1 source of AI-created misinformation."
	ZH 9%	USA: 100%	"... **the US** had already restricted companies exporting AI chips due to concerns that products might be diverted to China."
	RU 13%	USA 35%, China 19%	"... sanctions imposed in **Ukraine** against Russia, **Belarus**, and **China**, ... production of drones using AI."
Collaboration	EN 25%	China 18%, UK 15%	"**British** officials said the prime minister also hoped to announce cooperation ... including AI, with Mr. Trump."
	ZH 19%	USA 31%, EU 15%	"AI is... a crucial driving force for future China-**ASEAN** cooperation."
	RU 38%	USA 31%, China 11%	"... joint efforts of Russian and **Chinese** scientists, it has been possible to create ... processors for AI tasks."
Innovations	EN 21%	China 21%, Japan 10%	"Nissan has built the Yokohama Lab research center in **Japan** to study how to use AI in automotive manufacturing."
	ZH 43%	USA 87%, EU 3%	"Scientists used an AI system to analyze health data from hundreds of thousands of volunteers in the **UK** and the **US**."
	RU 24%	USA 40%, China 14%	"**Ukraine**, **NATO countries**, **China**, and the U.S. are already implementing AI for autonomous drone control."

Table 3: Mentions of other countries and regions in AI-related media coverage from the USA, Russia, and China. Russian media most frequently reference China and the USA, while Chinese and U.S. media primarily focus on each other. *Competition* is emphasized most in U.S. coverage, *Innovation* in Chinese coverage, and *Collaboration* in Russian coverage.

3.3 Country-Distinctive AI Narratives

To complement our predefined-frame analysis, we conduct an open-ended semantic analysis to identify finer-grained narratives distinctive to each national context. We translate Chinese and Russian passages into English, encode all passages with `Qwen3-Embedding-8B`, reduce the embeddings to 50 dimensions using UMAP, and cluster them with HDBSCAN. For each cluster t and country c, we compute an enrichment score $E(c,t) = P(c \mid t)/P(c)$, where $P(c \mid t)$ is the share of passages in cluster t from country c, and $P(c)$ is the country's overall corpus share. We rank clusters by enrichment and characterize them using passages nearest to the cluster centroid in the original embedding space. Native annotators from our earlier annotation then validate the resulting clusters by checking topic coherence, requiring sufficient supporting examples, and assigning each retained cluster a topic name and representative examples.

Distinctive Narrative	Examples
Clinical deployment of AI	"... promote the application of AI in **traditional Chinese medicine** and build an **'AI super TCM physician.'**"
Low-cost, open-source AI	"... DeepSeek V2 triggered **a price war among Chinese LLMs** and was hailed as the **'Pinduoduo of the AI.'**"
AI-enabled scams and fraud	"... AI-enabled scams target **families and communities of SVO participants**."
AI monitoring in public services	"... AI is used to **monitor the cleanliness of public transport** in Chelyabinsk."
Creative labor resistance to AI	"... AI concerns culminated in the **WGA and SAG-AFTRA strikes**, reflecting broader resistance in Hollywood."
Companion-chatbot harms	"... chatbot wanted her to become **utterly dependent** on it for her **sense of reality**."

Table 4: Country-distinctive semantic clusters identified through embedding-based clustering and validated by native speakers. The first, second, and third sections present China-, U.S.-, and Russia-centered clusters, respectively, with representative examples.

As shown in Table 4, the validated topics reveal country-specific AI narratives. Chinese coverage links AI to domestic innovation and applications such as DeepSeek-driven price competition and traditional Chinese medicine; Russian coverage highlights local uses such as monitoring public-transport cleanliness; and U.S. coverage emphasizes social tensions such as the Hollywood writers' strikes. These narratives show how AI discourse reflects country-specific institutions, practices, and social contexts.

Method	EN English			ZH Chinese			RU Russian		
	AI Stance	*Sentiment*	*Frame*	*AI Stance*	*Sentiment*	*Frame*	*AI Stance*	*Sentiment*	*Frame*
Random Baseline	$12.94_{\pm 0.49}$	$51.02_{\pm 1.96}$	$4.67_{\pm 0.84}$	$19.62_{\pm 0.72}$	$50.39_{\pm 2.00}$	$4.86_{\pm 0.86}$	$22.39_{\pm 0.72}$	$50.35_{\pm 1.70}$	$4.73_{\pm 0.77}$
Gemma3-27b-it									
Few-shot	$43.24_{\pm 0.54}$	$\mathbf{93.02}_{\pm 0.57}$	$47.73_{\pm 0.47}$	$37.20_{\pm 0.26}$	$95.38_{\pm 0.14}$	$52.75_{\pm 0.84}$	$58.98_{\pm 0.43}$	$92.40_{\pm 0.38}$	$51.34_{\pm 1.53}$
w. paragraph context	$39.17_{\pm 0.31}$	$\underline{92.94}_{\pm 0.70}$	$\underline{51.34}_{\pm 1.54}$	$32.85_{\pm 0.50}$	$\mathbf{96.10}_{\pm 0.10}$	$\underline{55.61}_{\pm 0.17}$	$52.46_{\pm 0.35}$	$\mathbf{93.10}_{\pm 0.26}$	$\underline{52.69}_{\pm 1.62}$
w. confidence	$\underline{47.74}_{\pm 0.68}$	$92.50_{\pm 0.58}$	$48.43_{\pm 0.76}$	$\underline{37.92}_{\pm 0.65}$	$95.54_{\pm 0.09}$	$51.47_{\pm 1.18}$	$\underline{60.65}_{\pm 0.50}$	$\underline{92.53}_{\pm 0.88}$	$46.15_{\pm 1.23}$
w. prob. distribution	$\mathbf{57.25}_{\pm 1.09}$	$90.69_{\pm 1.32}$	$\mathbf{51.46}_{\pm 0.38}$	$\mathbf{43.96}_{\pm 1.36}$	$\underline{95.77}_{\pm 0.25}$	$\mathbf{55.73}_{\pm 1.50}$	$\mathbf{66.06}_{\pm 0.32}$	$91.46_{\pm 0.96}$	$\mathbf{53.73}_{\pm 0.52}$
Qwen3-32b									
Few-shot	$53.46_{\pm 0.67}$	$91.86_{\pm 0.38}$	$54.29_{\pm 1.38}$	$\underline{40.96}_{\pm 0.26}$	$95.48_{\pm 0.25}$	$54.56_{\pm 0.21}$	$61.25_{\pm 0.33}$	$\underline{94.09}_{\pm 0.20}$	$47.20_{\pm 3.39}$
w. thinking	$53.87_{\pm 0.24}$	$\underline{92.15}_{\pm 0.04}$	$\mathbf{56.47}_{\pm 0.06}$	$40.52_{\pm 0.20}$	$95.32_{\pm 0.43}$	$54.71_{\pm 0.65}$	$60.19_{\pm 1.22}$	$93.79_{\pm 0.26}$	$46.39_{\pm 2.07}$
w. paragraph context	$50.07_{\pm 0.52}$	$\mathbf{92.78}_{\pm 0.24}$	$52.16_{\pm 1.04}$	$39.19_{\pm 0.21}$	$\underline{95.70}_{\pm 0.09}$	$53.92_{\pm 0.51}$	$57.17_{\pm 0.58}$	$93.92_{\pm 0.39}$	$46.96_{\pm 1.57}$
w. confidence	$\underline{54.63}_{\pm 1.54}$	$91.86_{\pm 0.32}$	$54.74_{\pm 0.69}$	$40.87_{\pm 0.19}$	$95.61_{\pm 0.12}$	$\underline{55.08}_{\pm 0.71}$	$\underline{63.85}_{\pm 0.89}$	$\mathbf{94.23}_{\pm 0.18}$	$\underline{51.38}_{\pm 1.45}$
w. prob. distribution	$\mathbf{58.70}_{\pm 0.58}$	$91.86_{\pm 0.38}$	$\underline{55.56}_{\pm 1.04}$	$\mathbf{44.18}_{\pm 0.91}$	$\mathbf{96.08}_{\pm 0.16}$	$\mathbf{57.43}_{\pm 0.87}$	$\mathbf{67.92}_{\pm 0.61}$	$93.77_{\pm 0.58}$	$\mathbf{54.43}_{\pm 0.87}$
GPT-OSS-20b									
Few-shot	$53.88_{\pm 0.52}$	$93.19_{\pm 0.43}$	$\underline{54.04}_{\pm 0.69}$	$40.77_{\pm 0.57}$	$94.82_{\pm 0.50}$	$55.30_{\pm 0.89}$	$\mathbf{61.47}_{\pm 0.12}$	$\underline{93.70}_{\pm 0.47}$	$58.29_{\pm 1.22}$
w. paragraph context	$50.67_{\pm 0.25}$	$\underline{93.37}_{\pm 0.59}$	$53.82_{\pm 0.52}$	$40.61_{\pm 0.24}$	$\underline{95.42}_{\pm 0.21}$	$\underline{55.62}_{\pm 0.92}$	$61.33_{\pm 0.75}$	$\mathbf{94.40}_{\pm 0.40}$	$\underline{58.60}_{\pm 0.59}$
w. confidence	$\mathbf{54.22}_{\pm 0.28}$	$93.02_{\pm 0.63}$	$53.94_{\pm 0.22}$	$\underline{41.72}_{\pm 0.24}$	$95.23_{\pm 0.21}$	$54.86_{\pm 0.70}$	$60.98_{\pm 0.33}$	$93.18_{\pm 0.15}$	$\mathbf{59.25}_{\pm 1.30}$
w. prob. distribution	$\underline{53.92}_{\pm 0.44}$	$\mathbf{93.88}_{\pm 0.39}$	$\mathbf{55.57}_{\pm 1.13}$	$\mathbf{43.64}_{\pm 0.13}$	$\mathbf{95.60}_{\pm 0.05}$	$\mathbf{56.71}_{\pm 1.11}$	$\underline{61.43}_{\pm 0.67}$	$93.28_{\pm 0.82}$	$57.77_{\pm 1.54}$
Model Ensemble	$58.43_{\pm 0.48}$	$93.09_{\pm 0.70}$	$52.81_{\pm 0.98}$	$46.25_{\pm 0.61}$	$96.03_{\pm 0.09}$	$53.77_{\pm 0.67}$	$65.76_{\pm 0.14}$	$93.88_{\pm 0.63}$	$56.79_{\pm 0.74}$
GPT-5.6-Luna†	$56.32_{\pm 0.22}$	$94.51_{\pm 0.47}$	$62.20_{\pm 1.02}$	$45.67_{\pm 0.44}$	$96.08_{\pm 0.33}$	$59.20_{\pm 0.59}$	$67.04_{\pm 0.39}$	$95.76_{\pm 0.26}$	$65.56_{\pm 0.63}$

Table 5: Performance on U.S., Chinese, and Russian news: positive-class F1 for AI stance identification and micro-F1 for sentiment and frame classification. Bold and underlined values mark the best and second-best results; subscripts show standard deviations. Random baseline averages 5000 Monte Carlo trials. Model ensemble combines three open-weight models via Product-of-Experts for stance scores and probability averaging for sentiment and frames. † indicates probability-distribution prompting via the OpenAI API.

4 LLMs for Framing Annotation

We benchmark LLMs on multilingual framing annotation using the human-annotated and adjudicated set, comparing model families and prompting strategies.

4.1 Experimental Setup

Evaluation Tasks. We decompose the multilingual framing annotation task into three sequential subtasks, as in § 2.3. Specifically, given news articles, LLMs are evaluated on: **(1)** *AI-related stance identification*, where models identify sentences that express a stance toward AI development. This is formulated as a binary classification task over all sentences in each article. **(2)** *Sentiment classification*, where models classify the sentiment of the identified sentences as positive or negative. **(3)** *Frame classification*, where models assign each identified sentence to a topic-specific frame category. To isolate errors across subtasks and assess performance at each stage, we evaluate stance identification on all sentences in the articles, and sentiment and frame classification on sentences identified as AI-related by human annotators.

Data Split. We use articles annotated by two human annotators and adjudicated as the test set, comprising 150, 120, and 202 articles from the USA, China, and Russia, respectively, and use the remaining annotated articles for training. The split is performed at the article level, assigning all sentences from each article to the same split to reduce potential contamination.

Models and Prompting. We evaluate three open-weight multilingual LLMs, `Gemma3-27B-it` (Team, 2025a), `Qwen3-32B` (Team, 2025b), and `GPT-OSS-20B` (OpenAI, 2025), and one proprietary model, `GPT-5.6-Luna`. For each task, models receive task-specific instructions and five human-annotated training examples per label: 10 examples each for stance identification and sentiment classification, and 125 for frame classification. Besides few-shot prompting, we evaluate three variants: **(1)** *Paragraph context* adds the two preceding and two following sentences to the target sentence; **(2)** *Confidence verbalization* requests a confidence

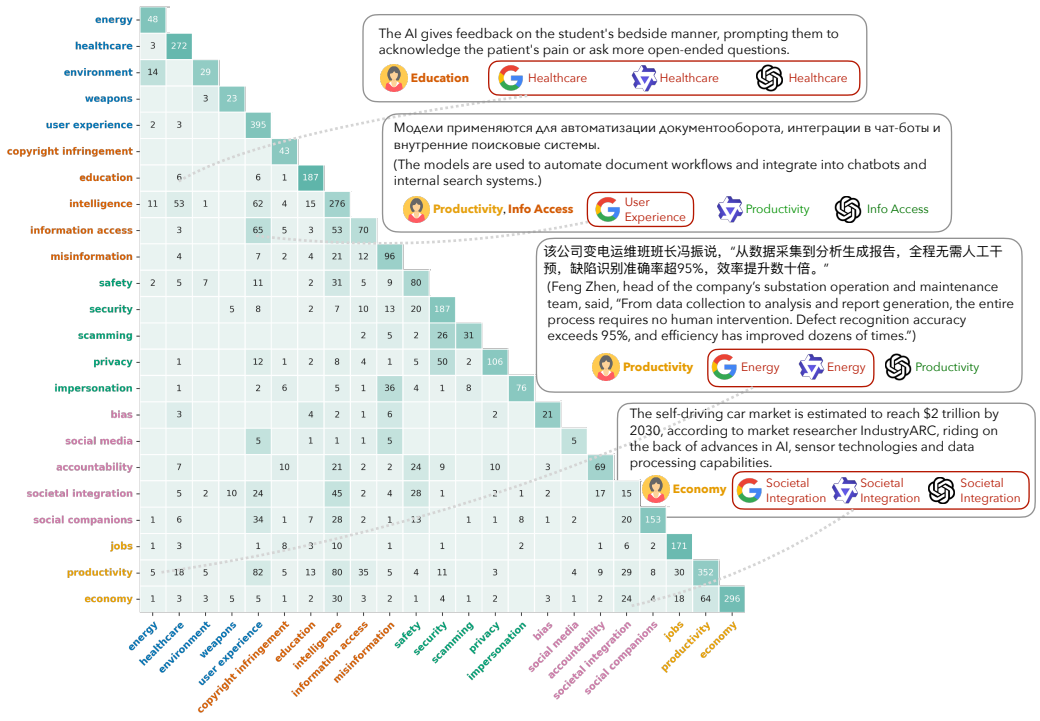


Figure 4: Confusion matrix of frame predictions aggregated across three models (`Gemma3-27B-it`, `Qwen3-32B`, and `GPT-OSS-20B`) and all languages (English, Chinese, and Russian), compared against human annotations. Each off-diagonal cell shows the total number of confusions between the corresponding pair of frames in either direction; diagonal cells show correct predictions. Annotated multilingual examples illustrate representative error cases and differences in model predictions.

score (1–100) for each prediction and discards predictions below a threshold; **(3)** *Probability distribution* requests probabilities over all labels and selects the highest-probability label. Under probability-distribution prompting, stance identification is decomposed into two judgments: whether the sentence is AI-related and whether it expresses an explicit positive or negative sentiment. Models output an AI-relatedness confidence score $c_{\text{ai}} \in [0, 100]$ and sentiment probabilities $p_{\text{pos}}, p_{\text{neg}}, p_{\text{neu}}$ summing to 100. We compute the stance-confidence score as $s = c_{\text{ai}}/100 \cdot f(p_{\text{pos}}, p_{\text{neg}})/100$, where $f \in \text{sum}, \max$ either sums the positive and negative sentiment probabilities or takes their maximum. A sentence is classified as expressing an explicit AI stance when s exceeds a tuned threshold. We query `GPT-5.6-Luna` through the OpenAI API using the same probability-distribution prompting strategy. We also evaluate a *Model Ensemble* of the three open-weight models, combining their stance-confidence scores via a Product-of-Experts method for stance identification and averaging their label-probability distributions for sentiment and frame classification. The *Random Baseline* reports mean performance over 5000 Monte Carlo trials. Further details are provided in Appendix D.

Evaluation Metrics. Stance identification and sentiment classification are binary tasks, while frame classification has 25 categories. All three tasks have imbalanced labels: 2,106 of 19,573 sentences express an AI-related stance, 1,326 of 2,097 sentiment labels are positive, and frame frequencies range from 12.3% (*Intelligence*) to 0.4% (*Social Media*). We report F1 for the positive class in stance identification, and micro-F1 for sentiment and frame classification, with means and standard deviations computed over three independent runs.

4.2 Main Results

Table 5 summarizes model performance across the three subtasks. AI stance identification remains challenging: models must distinguish opinions from neutral statements and determine whether the sentiment targets AI. Given sentences known to express an AI stance, models achieve F1 scores above

90 on sentiment classification. Frame classification is more difficult, with the highest F1 score of 65.56 by `GPT-5.6-Luna`, reflecting the challenge of distinguishing 25 fine-grained, unevenly represented categories. Probability-distribution prompting generally improves over standard few-shot prompting, with significant gains in stance identification ($+5.10$ F1 points) and frame classification ($+2.55$ points; both $p < 0.001$, one-sided Wilcoxon signed-rank tests), but no significant improvement in sentiment classification. Overall, identifying sentences that express an AI stance remains a key bottleneck, consistent with the annotation challenges discussed in §2.3. These findings underscore the value of annotations by native speakers for capturing subtle distinctions in stance and framing.

Error Analysis. Figure 4 aggregates frame predictions across all evaluated models and languages under probability-distribution prompting. Errors span both related frames and different topic groups, suggesting difficulty distinguishing the intended framing from the topic discussed. The examples illustrate how salient surface cues can overshadow a sentence's central message. In the Chinese example, a substation maintenance team head describes how AI automates data collection, analysis, and report generation, increasing efficiency dozens of times. Human annotators label this as *Productivity*, whereas `Gemma` and `Qwen` predict *Energy*, apparently prioritizing the electricity-maintenance setting over the efficiency gains. Other examples similarly suggest that models rely on domain keywords, entities, or application settings rather than the sentence's communicative emphasis.

5 Related Work

Due to space constraints, we focus here on the most relevant work on media framing and AI-related news coverage, and discuss additional related work in Appendix A. Media framing analysis examines how news content is presented to evoke particular interpretations, typically through manual annotation of framing categories (Otmakhova et al., 2024; Deng & Ahmed, 2025). It has been widely applied to topics such as gun violence, immigration, climate change, and political conflicts (Liu et al., 2019; Mendelsohn et al., 2021; Frermann et al., 2023; Kaur & Arora, 2025). With the rapid progress of AI, there is growing interest in how people perceive its opportunities, risks, and broader societal impact (Huang et al., 2026), alongside increasing public discussion of how AI should be developed and integrated into society (Wang, 2026; Dreyer, 2026). Recent research has also examined how AI is portrayed in news coverage, often focusing on English-language media (Fast & Horvitz, 2016; Nguyen & Hekman, 2024; Xian et al., 2024) or Chinese media (Cui & Wu, 2019; Ji et al., 2024) individually. Closest to our setting, existing cross-national studies either examine English-language news using automated topic and sentiment analysis (Ryazanov et al., 2024; Ittefaq et al., 2025), or compare a limited number of U.S. and Chinese outlets (i.e., *Washington Post* and *South China Morning Post*) using dictionary- and NER-based analysis of predefined dimensions of AI coverage (Nguyen & Hekman, 2024). In contrast, we study recent AI discourse in the LLM era (2024–2025) through full-length articles from 47 U.S., Chinese, and Russian outlets in their native languages, with sentence-level human annotations using a set of 25 AI-specific frames and their associated sentiment.

6 Conclusion

We analyze how AI is portrayed across major global powers through a cross-lingual study of media framing, using a multilingual dataset of 1,970 news articles from 47 American, Chinese, and Russian outlets. Building on a taxonomy of 25 frames, we conduct sentence-level analysis and reveal systematic geopolitical differences: Chinese media is consistently more positive, while U.S. and Russian coverage is more cautious and risk-oriented. By breaking down the task into sub-tasks, we can identify areas where multilingual LLMs struggle. Through this analysis, we highlight major topics in AI media discussion and challenges in automated framing analysis.

AI Use Statement

We used generative AI tools to produce candidate sentence-level annotations for human adjudication and to evaluate LLM performance on AI stance identification, sentiment classification, and frame classification. We also used generative AI to assist with language editing and manuscript organization. Human adjudicators reviewed the candidate annotations alongside independent human annotations to determine the final labels. The authors reviewed AI-assisted manuscript revisions for accuracy and

consistency with the methods and results. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with generative AI assistance.

ETHICS STATEMENT

To facilitate future research and reproducibility, we plan to publicly release the annotations and associated metadata, including article titles, source databases, sentence indices, and labels. Researchers who can demonstrate legitimate access to the underlying news sources or databases can also request the data used in this study by email, subject to the applicable licensing terms. We will not publicly release or redistribute the text of the news articles, as they are subject to copyright and licensing restrictions. All annotators were university students compensated at $18 per hour or awarded course credit, with course-credit participation entirely voluntary. They worked part-time at their preferred times and locations, taking breaks as needed. Training and discussion sessions (lasting 1–3 hours) were arranged by mutual agreement. Released annotations will not identify individual annotators, who can freely choose whether to be acknowledged in the paper. We also strive to maintain political neutrality, both to protect the annotators and authors involved in this work and to focus our study on statistical analyses and evaluating the reliability of machine learning models in performing them. We leave political and political science-related questions enabled by our dataset to future research.

REFERENCES

Syeda Sabrina Akter and Antonios Anastasopoulos. A study on scaling up multilingual news framing analysis. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 4156–4173, Mexico City, Mexico, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-naacl.260. URL https://aclanthology.org/2024.findings-naacl.260.

James Bessen. Ai and jobs: The role of demand. Technical report, National Bureau of Economic Research, 2018.

Andres M. Bran, Sam Cox, Oliver Schilter, Carlo Baldassari, Andrew D White, and Philippe Schwaller. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 2024.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. The media frames corpus: Annotations of frames across issues. In Chengqing Zong and Michael Strube (eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 438–444, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-2072. URL https://aclanthology.org/P15-2072/.

Ching-Hua Chuan, Wan-Hsiu Sunny Tsai, and Su Yeon Cho. Framing artificial intelligence in american newspapers. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 339–344, 2019.

Owen Cook, Yida Mu, Xinye Yang, Xingyi Song, and Kalina Bontcheva. A dataset for analysing news framing in Chinese media. *ArXiv*, abs/2503.04439, 2025. URL https://api.semanticscholar.org/CorpusId:276813652.

Hannes Cools, Baldwin Van Gorp, and Michael Opgenhaffen. Where exactly between utopia and dystopia? a framing analysis of ai and automation in us newspapers. *Journalism*, 25(1):3–21, 2024.

Di Cui and Fangjian Wu. The influence of media use on public perceptions of artificial intelligence in china: Evidence from an online survey. *Information Development*, 37:45 – 57, 2019. URL https://doi.org/10.1177/0266666919893411.

Agnese Daffara, Sourabh Dattawad, Sebastian Pad'o, and Tanise Ceron. Generalizability of media frames: Corpus creation and analysis across countries. *ArXiv*, abs/2506.16337, 2025. URL https://api.semanticscholar.org/CorpusId:279465249.

Ruolan Deng and Saifuddin Ahmed. Perceptions and paradigms: An analysis of ai framing in trending social media news. *Technology in Society*, 81:102858, 2025. ISSN 0160-791X.

Zijian Ding, Alison Smith-Renner, Wenjuan Zhang, Joel R. Tetreault, and Alejandro Jaimes. Harnessing the power of LLMs: Evaluating human-AI text co-creation through the lens of news headline generation. In *Conference on Empirical Methods in Natural Language Processing*, 2023. URL https://api.semanticscholar.org/CorpusId:264172468.

Jacob Dreyer. Why china is so much less scared of a.i. *The New York Times*, May 2026. URL https://www.nytimes.com/2026/05/09/opinion/ai-china-america-race.html. May 9, 2026.

Edelman. Familiarity breeds trust in ai, 2024. URL https://www.edelman.com/insights/familiarity-breeds-trust-in-ai.

Ethan Fast and E. Horvitz. Long-term trends in the public perception of artificial intelligence. In *AAAI Conference on Artificial Intelligence*, 2016. URL https://arxiv.org/pdf/1609.04904.pdf.

Feedspot. 20 best china technology blogs to follow in 2025, 2025a. URL https://bloggers.feedspot.com/china_technology_blogs/.

Feedspot. Top 30 russian news websites in 2025, 2025b. URL https://news.feedspot.com/russian_news_websites/.

Lea Frermann, Jiatong Li, Shima Khanehzar, and Gosia Mikołajczak. Conflicts, villains, resolutions: Towards models of narrative media framing. *ArXiv*, abs/2306.02052, 2023. URL https://api.semanticscholar.org/CorpusId:259075515.

Vahid Ghafouri, Robert McNeil, Teodor Yankov, Madeleine Sumption, Luc Rocher, Scott A Hale, and Adam Mahdi. Framing migration: A computational analysis of UK parliamentary discourse. *arXiv preprint arXiv:2509.14197*, 2025.

Matteo Guida, Yulia Otmakhova, Eduard H. Hovy, and Lea Frermann. Retain or reframe? a computational framework for the analysis of framing in news articles and reader comments. *ArXiv*, abs/2507.04612, 2025. URL https://api.semanticscholar.org/CorpusId:280134356.

David Heineman, Yao Dou, and Wei Xu. Thresh: A unified, customizable and deployable platform for fine-grained text evaluation. *arXiv preprint arXiv:2308.06953*, 2023.

Saffron Huang, Shan Carter, Jake Eaton, Sarah Pollack, Dexter Callender III, Nikki Makagiansar, Maria Gonzalez, Sylvie Carr, Jerry Hong, Kunal Handa, Miles McCain, Thomas Millar, Mo Julapalli, Grace Yun, AJ Alt, Chelsea Larsson, Jane Leibrock, Matt Gallivan, Theodore Sumers, Esin Durmus, Matt Kearney, Judy Hanwen Shen, Jack Clark, Michael Stern, and Deep Ganguli. What 81,000 people want from ai, 2026. URL https://anthropic.com/features/81k-interviews.

Muhammad Ittefaq, Ali Zain, Rauf Arif, Mohammad Ala-Uddin, Taufiq Ahmad, and Azhar Iqbal. Global news media coverage of artificial intelligence (ai): A comparative analysis of frames, sentiments, and trends across 12 countries. *Telematics and Informatics*, 96:102223, 2025.

Xiaolu Ji, Joanne Kuai, and Rodrigo Zamith. Scrutinizing algorithms: Assessing journalistic role performance in Chinese news media's coverage of artificial intelligence. *Journalism Practice*, 18: 2396 – 2413, 2024. URL https://doi.org/10.1080/17512786.2024.2336136.

Avneet Kaur and Arnav Arora. Beyond the battlefield: Framing analysis of media coverage in conflict reporting. *ArXiv*, abs/2506.10421, 2025. URL https://api.semanticscholar.org/CorpusID:279319022.

Kian Kenyon-Dean, Eisha Ahmed, Scott Fujimoto, Jeremy Georges-Filteau, Christopher Glasz, Barleen Kaur, Auguste Lalande, Shruti Bhanderi, Robert Belfer, Nirmal Kanagasabai, et al. Sentiment analysis: It's complicated! In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1886–1895, 2018.

Klaus Krippendorff. Measuring the reliability of qualitative text analysis data. *Quality and quantity*, 38(6):787–800, 2004.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, 2023.

Siyi Liu, Lei Guo, Kate K. Mays, Margrit Betke, and Derry Tanti Wijaya. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Conference on Computational Natural Language Learning*, 2019. URL https://www.aclweb.org/anthology/K19-1047.pdf.

Julia Mendelsohn, Ceren Budak, and David Jurgens. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2219–2263, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.179. URL https://aclanthology.org/2021.naacl-main.179/.

Nishanth Sridhar Nakshatri, Nikhil Mehta, Siyi Liu, Sihao Chen, Daniel Hopkins, Dan Roth, and Dan Goldwasser. Talking point based ideological discourse analysis in news events. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 575–594, 2025.

Dennis Nguyen and Erik Hekman. A 'new arms race'? framing china and the usa in ai news reporting: A comparative analysis of the washington post and south china morning post. *Global Media and China*, 7(1):58–77, 2022.

Dennis Nguyen and Erik Hekman. The news framing of artificial intelligence: a critical exploration of how media discourses make sense of automation. *AI & society*, 39(2):437–451, 2024.

OpenAI. gpt-oss-120b gpt-oss-20b model card, 2025. URL https://arxiv.org/abs/2508.10925.

Julia Otmakhova, Shima Khanehzar, and Lea Frermann. Media framing: A typology and survey of computational approaches across disciplines. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15407–15428, 2024.

Chan Young Park, Julia Mendelsohn, Anjalie Field, and Yulia Tsvetkov. Challenges and opportunities in information manipulation detection: An examination of wartime russian media. In *Conference on Empirical Methods in Natural Language Processing*, 2022. URL https://api.semanticscholar.org/CorpusId:253107926.

Jungsoo Park, Junmo Kang, Gabriel Stanovsky, and Alan Ritter. Can LLMs help uncover insights about LLMs? a large-scale, evolving literature analysis of frontier LLMs. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vienna, Austria, July 2025. Association for Computational Linguistics.

Pew Research Center. How people around the world view AI, October 2025a. URL https://www.pewresearch.org/global/2025/10/15/how-people-around-the-world-view-ai/.

Pew Research Center. The political gap in americans's news sources, June 2025b. URL https://www.pewresearch.org/journalism/2025/06/10/the-political-gap-in-americans-news-sources/.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In Atul Kr. Ojha, A. Seza Doğruöz, Giovanni Da San Martino, Harish Tayyar Madabushi, Ritesh Kumar, and Elisa Sartori (eds.), *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pp. 2343–2361, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.semeval-1.317. URL https://aclanthology.org/2023.semeval-1.317/.

Giovanni Puccetti, Anna Rogers, Chiara Alzetta, Felice Dell'Orletta, and Andrea Esuli. Ai "news" content farms are easy to make and hard to detect: A case study in italian. *ArXiv*, abs/2406.12128, 2024. URL https://api.semanticscholar.org/CorpusID:270562338.

Jenna Russell, Marzena Karpinska, Destiny Akinode, Katherine Thai, Bradley Emi, Max Spero, and Mohit Iyyer. AI use in American newspapers is widespread, uneven, and rarely disclosed. *arXiv preprint arXiv:2510.18774*, 2025.

Igor Ryazanov, Carl Öhman, and Johanna Björklund. How chatgpt changed the media's narratives on ai: a semi-automated narrative analysis through frame semantics. *Minds and Machines*, 35(1):2, 2024.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019.

Gemma Team. Gemma 3. 2025a. URL https://goo.gle/Gemma3Report.

Qwen Team. Qwen3 technical report, 2025b. URL https://arxiv.org/abs/2505.09388.

Isidora Chara Tourni, Lei Guo, T. Daryanto, Fabian Zhafransyah, Edward Edberg Halim, Mona Jalal, Boqi Chen, Shan-Ching Lai, Hengchang Hu, Margrit Betke, P. Ishwar, and Derry Tanti Wijaya. Detecting frames in news headlines and lead images in U.S. gun violence coverage. *ArXiv*, abs/2406.17213, 2024. URL https://arxiv.org/pdf/2406.17213.pdf.

Filip Trhlik and Pontus Stenetorp. Quantifying generative media bias with a corpus of real-world and generated news articles. In *Conference on Empirical Methods in Natural Language Processing*, 2024. URL https://api.semanticscholar.org/CorpusId:270560516.

Christina P. Walker, Danielle Schiff, and K. Schiff. Merging AI incidents research with political misinformation research: Introducing the political deepfakes incidents database. *ArXiv*, abs/2409.15319, 2024. URL https://api.semanticscholar.org/CorpusId:268716927.

Vivian Wang. Where are china's a.i. doomers? *The New York Times*, March 2026. URL https://www.nytimes.com/2026/03/04/world/asia/china-ai-enthusiasm.html. March 4, 2026.

Xinyu Wang, Wenbo Zhang, Sai Dileep Koneru, Hangzhi Guo, Bonam Mingole, S. Shyam Sundar, Sarah M. Rajtmajer, and Amulya Yadav. Have llms reopened the pandora's box of ai-generated fake news? In *North American Chapter of the Association for Computational Linguistics*, 2024. URL https://api.semanticscholar.org/CorpusID:273638082.

Lu Xian, Lingyao Li, Yiwei Xu, Ben Zefeng Zhang, and Libby Hemphill. Landscape of generative ai in global news: Topics, sentiments, and spatiotemporal analysis. *ArXiv*, abs/2401.08899, 2024. URL https://api.semanticscholar.org/CorpusId:267028231.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. Defending against neural fake news. *ArXiv*, abs/1905.12616, 2019. URL https://api.semanticscholar.org/CorpusID:168169824.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025.

Shijia Zhou, Siyao Peng, Simon Luebke, Jörg Haßler, Mario Haim, Saif M Mohammad, and Barbara Plank. What media frames reveal about stance: A dataset and study about memes in climate change discourse. *arXiv preprint arXiv:2505.16592*, 2025.

Caleb Ziems and Diyi Yang. To protect and to serve? analyzing entity-centric framing of police violence. *ArXiv*, abs/2109.05325, 2021. URL https://arxiv.org/pdf/2109.05325.pdf.

A Additional Related Work

Media Framing Analysis. Existing work uses taxonomies of varying granularity. Some studies define general, topic-agnostic frames capturing broad perspectives (e.g., *political*, *legal*, *economic*)(Card et al., 2015; Piskorski et al., 2023; Guida et al., 2025; Akter & Anastasopoulos, 2024; Nakshatri et al., 2025), while others introduce fine-grained frames tailored to specific domains such as gun violence (Liu et al., 2019; Tourni et al., 2024), immigration (Mendelsohn et al., 2021; Ghafouri et al., 2025), climate change (Frermann et al., 2023; Zhou et al., 2025), police violence (Ziems & Yang, 2021), and political conflicts (Kaur & Arora, 2025). Most prior work focuses on English news from U.S.- or U.K.-based media, with relatively few studies examining non-English media systems such as Brazil (Daffara et al., 2025), China (Cook et al., 2025), and Russia (Park et al., 2022).

AI and the Media. Research on AI and media spans multiple perspectives. One line of work examines how media organizations employ AI systems in news production, such as generating headlines (Ding et al., 2023) and writing articles (Russell et al., 2025). These studies also highlight potential risks, including the propagation of political biases (Trhlik & Stenetorp, 2024) and the generation of fabricated news (Zellers et al., 2019; Puccetti et al., 2024; Wang et al., 2024; Walker et al., 2024).

B Dataset: Additional Details

B.1 Collection Details

News Sources. Figure 5 presents an overview of news sources we collected for all three languages. All news sources that were annotated in our study are listed in Table 9 for U.S. sources, Table 10 for Chinese sources, and Table 11 for Russian sources. Source names highlighted in yellow represent traditional news media, while source names highlighted in green represent tech news. The language, retrieval database, and government affiliation are all listed in the table. We also report both the total number of articles retrieved from the databases for each source and the number of articles randomly sampled and annotated.

Search Keywords. Each of the search keywords used to filter the articles can be found in Table 8. The words were chosen to encompass as much AI-related news as possible. The same keywords were translated into both Russian and Chinese to ensure consistency across article collection.

Additional Filtering. Besides keywords, a few filters were incorporated while collecting articles from databases:

- Articles must be published between January 2024 - October 2025.
- Each article should be written in the corresponding language of the country (English for the U.S., Chinese for China, Russian for Russia).
- Following article collection, articles were de-duplicated to keep one with unique <title, publication, date> or <title, author, date> tuples. The exact tuple was dependent on the database from which the articles were extracted. If the same article was added into the database twice on multiple dates, it would not be de-duplicated. This was because some articles would use repetitive titles (even with different content), and we didn't want to remove unnecessary information.
- English articles were very lengthy in a small number of cases. To ensure annotations could be completed in a reasonable time frame, we sampled from articles with greater than 10 words and no more than 4,000 words. This cutoff enabled 98.8% of the articles to be retained.

B.2 AI Frames Taxonomy

Frame Definitions. Table 12 lists the precise definitions we used for the positive and negative framing in each of our 25 framing categories.

B.3 Annotation Details

Annotation Guidelines. Figure 6 shows the annotation flow given to annotators to annotate AI framing in sentences. The following is description of the steps:

- To decide if a sentence should be annotated, the annotator first asks if they can determine whether AI is good or bad from this particular sentence. If yes, the annotator should annotate the sentence. Note: The sentence does not need to explicitly contain the word AI to be labeled (e.g., Deepfake can be inferred as referencing AI).
- If the sentence is implying that AI is good or bad focusing on a specific reason or frame, the annotator selects a frame for this sentence. Otherwise, if it is just painting a positive or negative view on AI in general, the annotators marks it in the "General" frame.
- If the sentence fits within our defined taxonomy of 25 AI frames, the annotator selects the appropriate frame. If it does not fit within any of our frame, the annotator marks it as the "Other" frame and manually specifies what frame they believe this is.
- After specifying the frames, the annotator marks the sentiment of the frame as positive or negative.

Annotation Tool. We used the Thresh tool (Heineman et al., 2023) for sentence-level annotation and adjudication. The interface was configured to present articles to annotators by allowing them to highlight only one sentence at a time. For each highlighted sentence, the tool enabled annotators to assign a positive or negative sentiment and select the corresponding frame from our taxonomy. In instances where a sentence did not fit the predefined categories, the interface allowed annotators to select the "Other" category and provide a customer frame description through the text input field. A separate adjudication tool incorporated the "Complex" category and allowed adjudicators to select multiple frames and sentiments that they agreed with. Adjudicators first selected the frame that they most agreed with followed by the second plausible selection.

C LLM annotation Details

Setup. The LLM annotation pass uses GPT-5.6-Luna through the OpenAI API with default sampling parameters and 16 concurrent workers. Each completed step is checkpointed, allowing interrupted runs to resume without repeating completed API calls. The pipeline uses the same prompt builders, JSON-line parser, closed frame taxonomy, and scoring code as our evaluation harness, keeping these components unchanged across annotation and evaluation runs.

Each article is processed through three sequential stages, with prompt templates provided in Figures 29–31. We show the English templates; the Chinese and Russian versions are translations with identical structure and output schema.

1. **AI stance identification** (Figure 29). We make one call per sentence, supplying its paragraph as context to resolve pronouns and references. The model evaluates only the target sentence and returns `True` and `False` probability scores on a 0–100 scale, together with an `is_target` flag. The two scores must sum to 100. We denote the `True` score by p_{AI}. Sentences with $p_{\mathrm{AI}} \geq 5$ proceed to Step 2; those below this cutoff are excluded from subsequent stages and assigned a final score of zero. This permissive cutoff reduces token usage, while the export threshold described below determines final candidate selection.
2. **Sentiment classification** (Figure 30). We make one call per article containing only the retained sentences, each prefixed with its index. The model returns one JSON record per sentence, containing its index and positive and negative sentiment scores on a 0–100 scale. We denote these scores by p_{pos} and p_{neg}, respectively.
3. **Frame classification** (Figure 31). We make one call per article containing the same retained sentences. For each sentence, the model returns its index and two ranked frame predictions, each with a confidence score on a 0–100 scale. Frame names must exactly match entries in the closed 25-frame taxonomy. We request a second frame even when confidence in the first is high, providing an alternative for the adjudicator to consider.

	English	Chinese	Russian
Evaluation subset			
Articles	150	120	202
Sentences	8,575	5,346	5,652
Stance sentences	638	653	815
LLM-annotated corpus			
Articles	564	470	464
Sentences	29,688	18,525	12,464
Threshold τ	4.55	11.16	6.30
Selected, eval. (%)	30.5	41.9	32.8
Selected, corpus (%)	30.0	39.0	29.3
Precision (%)	23.5	25.2	39.4
Recall (%)	96.2	86.2	89.6
F1 (%)	37.8	38.9	54.7

Table 6: Dataset statistics and LLM candidate retrieval performance. Selected denotes the percentage of sentences selected for human review. Precision, recall, and F1 are measured against human-adjudicated labels. Thresholds τ are selected and evaluated on the evaluation subset, then applied unchanged to the corpus.

Scoring and thresholds. The probability and confidence scores returned by the model use a 0–100 scale. We combine the AI-stance probability from Step 1 and sentiment confidence from Step 2 into a sentence-level score:

$$s = \frac{p_{\text{AI}} \cdot \max(p_{\text{pos}}, p_{\text{neg}})}{100} \in [0, 100]. \tag{1}$$

A high score requires confidence both that the sentence expresses a stance toward AI and that its polarity is positive or negative. For example, a sentence with $p_{\text{AI}} = 12$ and $\max(p_{\text{pos}}, p_{\text{neg}}) = 95$ receives a score of 11.4. Sentences excluded after Step 1 receive a score of zero. A sentence is exported as a candidate when $s \geq \tau_\ell$, where τ_ℓ is a language-specific threshold selected using the human-adjudicated subset.

We set a minimum recall target of 85% across languages. Table 6 reports the selected thresholds and their resulting performance. The English and Russian thresholds select approximately 30% of corpus sentences and achieve recall above the target on the evaluation subset. Chinese requires a larger review budget to meet the target: the selected threshold achieves 86.2% recall and selects 39.0% of corpus sentences, adding 1,674 candidate spans relative to the 30% budget. We prioritize recall because the candidates undergo human adjudication, where false positives can be removed.

D Implementation Details

Automatic Media Framing Annotation. We provide a prompt template for each combination of task and prompting strategy in Table 5. For AI stance identification, Figures 17, 18, 19, and 20 present few-shot prompting, confidence verbalization, paragraph context, and probability distribution prompting. Templates for sentiment classification appear in Figures 21–24, and those for frame classification appear in Figures 25–28, following the same strategy order. We show the English templates; the Chinese and Russian versions are translations with the same structure and output schema.

Few-shot examples are represented by placeholders; at runtime, these are populated with five examples per class sampled from the annotated pool. The 25 frame definitions are provided in Table 12 rather than repeated in each template. For Figure 20, we show only the first of fifteen worked examples. These templates are used in the evaluation experiments. The three-stage corpus annotation pipeline uses separate prompts, shown in Figures 29–31, and is described in Appendix C.

LLM Inference. We run inference on 1 NVIDIA H100 GPU (90G Memory) with the vLLM library (Kwon et al., 2023), and perform decoding with `temperature=1.0`. Few-shot examples

Model	Language	Aggregation	τ	Selected (%)
Gemma3-27B	English	max	0.6300	9.7
	Chinese	max	0.7125	25.1
	Russian	sum	0.6800	16.9
Qwen3-32B	English	sum	0.6375	11.7
	Chinese	max	0.6375	27.0
	Russian	sum	0.6370	15.9
GPT-OSS-20B	English	sum	0.7350	9.0
	Chinese	max	0.7644	28.7
	Russian	max	0.0485	16.7

Table 7: AI stance identification settings under probability-distribution prompting. τ is the decision threshold, and *Selected* is the percentage of evaluation sentences classified as stance-bearing.

are sampled from the annotated pool, five per class, with a fixed seed so that every model sees the same examples. GPT-5.6-Luna is queried through the OpenAI API with default sampling parameters.

Table 7 reports the AI stance identification settings used with probability-distribution prompting. For each model and language, it lists the aggregation rule used to combine sentiment scores into a stance score, the decision threshold τ, and the percentage of evaluation sentences classified as stance-bearing. Thresholds and selected proportions are reported as medians across three few-shot example sets.

Thresholds are not directly comparable across models because their verbalized confidence scores follow different distributions. For GPT-OSS-20B, 90% of Russian sentences receive scores below 0.1. Its low threshold of 0.0485 therefore selects only 16.7% of sentences, slightly fewer than Gemma3-27B selects at 0.680 (16.9%). The selected proportion offers a clearer comparison with the gold stance prevalence of 7.4% in English, 12.2% in Chinese, and 14.4% in Russian.

E Additional Results

Data Split. We split the human-annotated data at an article level, ensuring that all articles in the test set are adjudicated, either by a third annotator or through discussion among annotators. Only articles with at least one annotation were retained in each dataset. The resulting test set contains 150, 120, 202 articles (8,575, 5,346, 5,652 sentences) in English, Chinese, and Russian, respectively, while the remaining data forms the training set with 564, 470, 464 articles (29,688, 18,525, 12,464 sentences).

Span-level annotation. Our annotation scheme allows annotators to label specific spans with their corresponding frames and sentiments, which may result in multiple frame–sentiment pairs within a single sentence. This occurs in 16.12%, 13.70%, and 12.27% of the U.S., Chinese, and Russian annotated sentences, respectively. In the LLM automatic annotation experiments, a prediction for such sentences is considered correct if the predicted label matches any of the annotated labels.

Using the domain labels in Tables 9–11, we divide each corpus into technology and traditional news outlets. Technology outlets frame AI more positively across all three countries, with the largest gap in the U.S. (Figure 7).

Technology outlets nevertheless cover *Safety* at least as often as traditional outlets, while placing greater emphasis on *Applications* (Figure 8). *Safety* is the only category with net negative sentiment in both groups across all three countries, whereas *Applications* is the most positively framed category in the Chinese and Russian corpora, ranking second behind *Work* in the U.S. In the U.S., traditional outlets are particularly more negative about AI's societal and informational impacts, with the largest sentiment gaps in *Society* and *Information*.

Results by news source. Results visualized in Figures 9, 10, and 11 underscore the high degree of internal consistency within certain media ecosystems and highlight specific thematic outliers. We see near-unanimous positive framing of AI across all Chinese media outlets. As shown in Figure 10,

English	Russian	Chinese
Artificial Intelligence	Iskusstvennyi intellekt	人工智能
AI	II	AI
A.I.	AI	A.I.
Machine Learning	Mashinnoe obuchenie	机器学习
Neural network	Neironnaia set'	神经网络
LLM	LLM	LLM
Large language model	BYAM	大语言模型
Deep Learning	Bol'shaya yazykovaya model' Glubokoe obuchenie	深度学习

Table 8: AI-related keywords used to identify AI-related articles.

categories such as *Work & Economy* and *Applications* consistently score between 0.8 and 1.0 (highly positive) across nearly every source. This reflects a domestic narrative that frames AI primarily as an engine for national development and innovation. The only significant exception to this trend is the *Safety & Security* category, which displays the most volatility and includes the only high negative score (-1.0 for *Securities Times*) in the entire Chinese corpus. This suggests that while Chinese media is overwhelmingly optimistic about AI's utility, it maintains a singular, concentrated caution regarding technical safety risks. When comparing the three regions, *Safety & Security* and *Society* emerge as the primary drivers of negative sentiment globally:

- **U.S.**: American outlets show significant skepticism toward AI's societal impact. Outlets like *Politico* report a sentiment of -1.0 for the *Society* category, indicating an almost exclusively negative framing of AI's impact on human institutions and social fabric.
- **Russia**: Russian media displays a mixed but often cautious stance on systemic risks. While outlets like *Economics and Life* and *Province* show extreme negativity (-1.0) toward *Safety & Security*, federal government-owned sources like *The Russian Government* report highly positive sentiment (1.0) in the same category, suggesting a divergence between independent and state-centric narratives.
- **China**: Unlike the U.S. and Russia, Chinese media frequently frames the *Society* impact of AI positively, . many outlets reporting perfect 1.0 scores.

Across all three countries, a consistent technical optimism is visible. In the U.S., tech-focused outlets like *TechRadar* and *The Verge* maintain significantly higher positive sentiment in the *Applications* category (0.8 and 0.7, respectively) compared to more general outlets like *The Washington Post* (-0.2). This trend is even more pronounced in Russia, where tech outlets like *CNews* and *Mobile Review* report highly positive sentiment across almost all categories, contrasting with the more balanced or negative reporting of general news sources like *RIA Novosti*.

Source	Database	Domain	Article Count	Sample Count
New York Times	U.S. Newsstream	Non-Tech	11286	50
Washington Post	U.S. Newsstream	Non-Tech	4861	50
USA Today	U.S. Newsstream	Non-Tech	2422	50
Forbes	Nexis	Non-Tech	60	50
The Atlantic	Nexis	Non-Tech	22	14
Politico	U.S. Newsstream	Non-Tech	1879	50
The Associated Press	Nexis	Non-Tech	3172	50
Newsweek	U.S. Newsstream	Non-Tech	77	50
Wall Street Journal	U.S. Newsstream	Non-Tech	9995	50
The Hill	Nexis	Non-Tech	97	50
Axios	U.S. Newsstream	Non-Tech	3450	50
The Verge	U.S. Newsstream	Tech	877	50
MIT Technology Review	ProQuest Central	Tech	1261	50
VentureBeat	Nexis	Tech	3949	50
TechRadar	Nexis	Tech	154	50

Table 9: U.S. news sources.

Source	Database	Domain	Article Count	Sample Count
People's Daily	Factiva	Non-Tech	7778	50
China News Service	Factiva	Non-Tech	3185	50
Xinhua	Factiva	Non-Tech	6052	50
Chengdu Business Daily Red Star News	Factiva	Non-Tech	3330	50
The Beijing News	Factiva	Non-Tech	1028	50
Securities Times	Factiva	Non-Tech	2095	50
China Business Network	Factiva	Non-Tech	7386	50
Va Kio Daily	Factiva	Non-Tech	658	50
Ta Kung Pao	Factiva	Non-Tech	4984	50
TechNode	Factiva	Tech	574	50
TMTPost	Factiva	Tech	5118	50
GuruClub	Factiva	Tech	12088	50

Table 10: China news sources.

	USA	China	Russia
Traditional Media	New York Times Washington Post USA Today Forbes The Atlantic Politico The AP Newsweek Wall Street Journal The Hill Axios	People's Daily China News Service Xinhua Chengdu Business Daily The Beijing News Securities Times China Business Network Va Kio Daily Ta Kung Pao	The Moscow Times Gazeta Russian Government RIA Novosti Lenta Economics and Life Metro News Pravda Severa Vechernaya Moskva Province VZ URA Federal Press The Insider EuroNews
Tech Media	The Verge MIT Tech Review VentureBeat TechRadar	TechNode TMTPost GuruClub	3DNews Cnews Hacker Mobile Review RB

Figure 5: Media sources from which we collected news articles for each country. The sources cover both traditional media and tech-focused media.

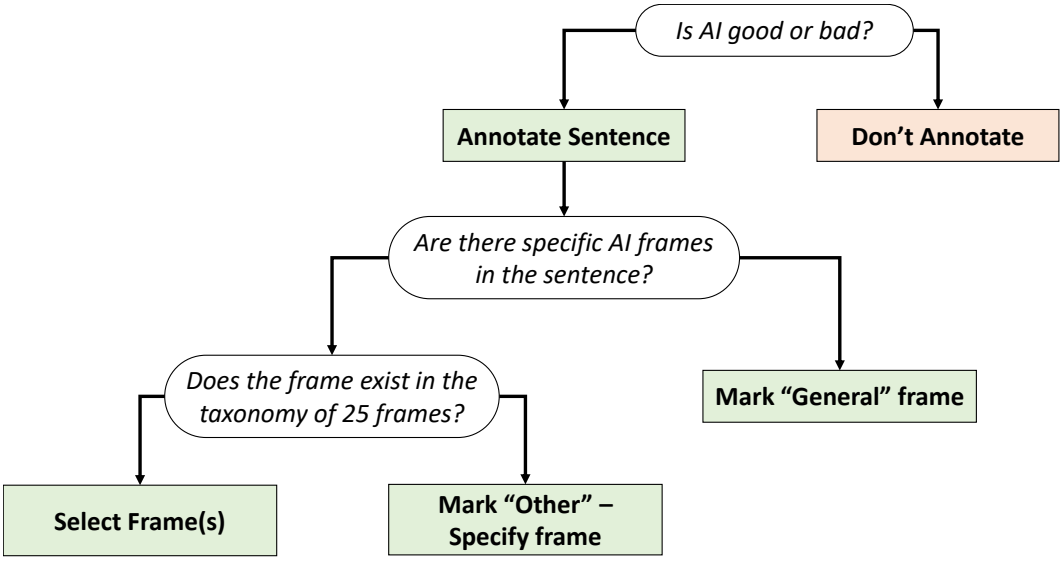


Figure 6: Annotation decision tree demonstrating how annotators decide how a sentence can be annotated.

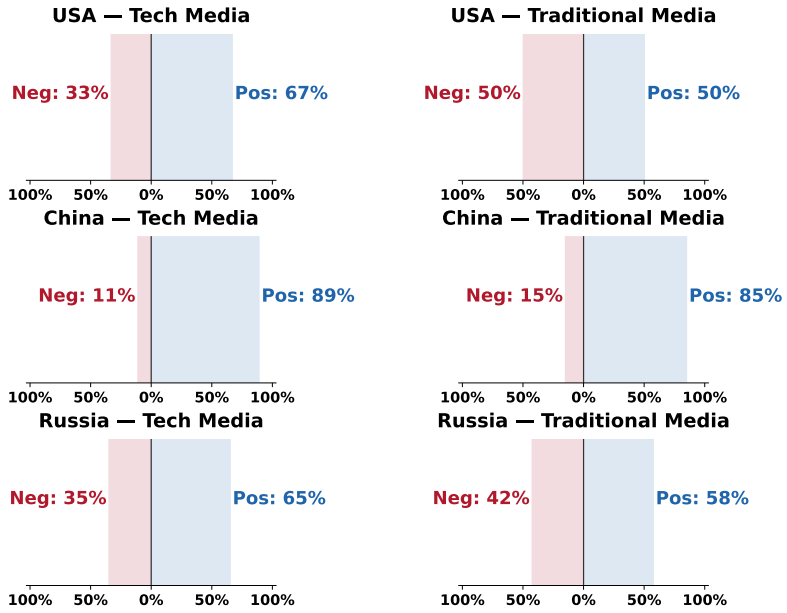


Figure 7: Comparison of overall AI sentiment between tech and traditional media outlets. Across all countries, tech media skews more positively than traditional media, with Russian media having the largest gap.

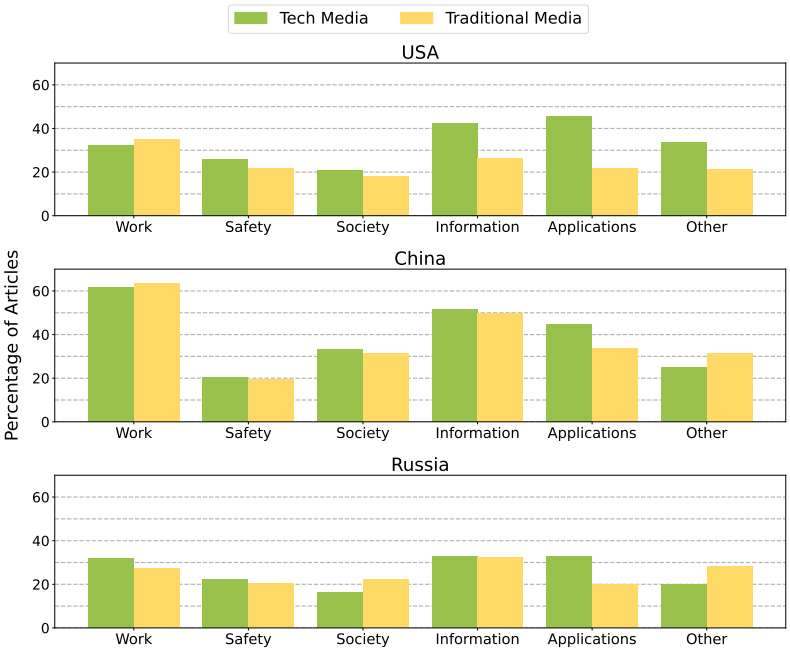


Figure 8: Percentage of category appearance in articles published in tech vs traditional media consistently mention frames under *Applications* and *Information* categories more often than traditional media.

Source	Database	Domain	Article Count	Sample Count
The Moscow Times	Nexis	Non-Tech	438	40
Gazeta	Nexis	Non-Tech	675	40
The Russian Government	Nexis	Non-Tech	27	27
RIA Novosti	Nexis	Non-Tech	99	40
Lenta	Nexis	Non-Tech	205	40
Economics and Life	Nexis	Non-Tech	9	9
Metro News	Nexis	Non-Tech	27	27
Pravda Severa	Nexis	Non-Tech	11	11
Vechernaya Moskva	Nexis	Non-Tech	128	40
Province	Nexis	Non-Tech	12	12
Vzglyad	Nexis	Non-Tech	124	40
URA	Nexis	Non-Tech	133	40
Federal Press	Nexis	Non-Tech	145	40
The Insider	Nexis	Non-Tech	64	40
EuroNews	Nexis	Non-Tech	199	40
3DNews	Nexis	Tech	725	40
CNews	Nexis	Tech	815	40
Xakep	Nexis	Tech	20	20
Mobile Review	Nexis	Tech	195	40
RB	Nexis	Tech	80	40

Table 11: Russian news sources.

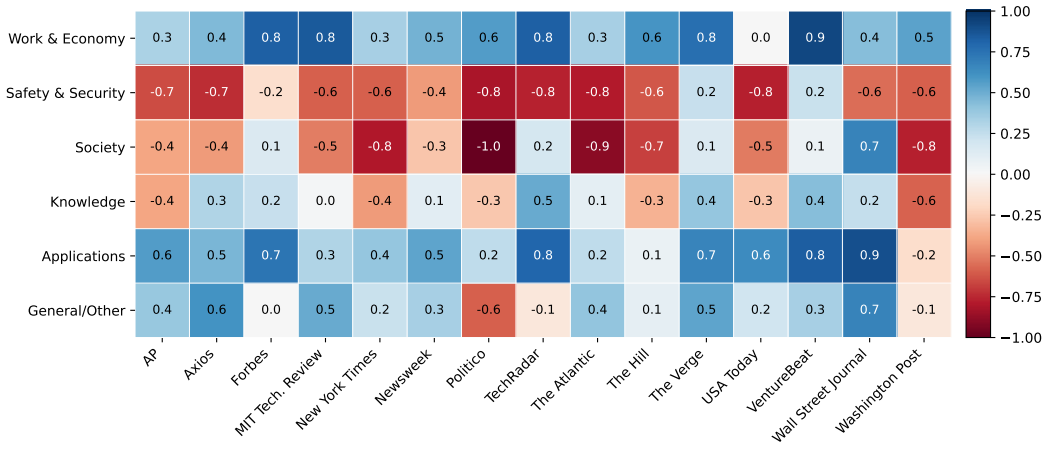


Figure 9: Sentiment distribution per English source.

	Frame		Definition
Applications	Energy	+	AI assists in providing improved energy solutions to people
		-	AI is draining energy, causing strain on the energy grid, or taking power away from people
	Environment	+	AI analyzes environmental trends or enhances environmental protection technology
		-	AI development and usage causes harmful effects on the environment like increased carbon emissions or pollution
	Healthcare	+	AI makes medical discoveries, improves treatments, and makes healthcare more effective
		-	AI makes mistakes, posing increased risks in the medical and healthcare fields
	User Experience	+	AI improves user experience of apps, services, or devices including providing faster and more accurate service
		-	AI worsens user experience of apps, services, or devices
	Weapons	+	AI improves defense mechanisms by creating more accurate weapons
		-	AI creates more destructive weapons that are used for harm
Knowledge	Copyright Infringement	+	AI helps protect copyrights and combats copyright infringement
		-	AI breaks copyright laws and reproduces peopleĂ‘Ź s existing work
	Education	+	AI helps people learn new skills and improves learning experiences (includes schools and trainings)
		-	AI slows learning new skills and harms learning processes in schools and in general, reducing critical thinking skills
	Information Access	+	AI provides quick and more convenient access to information on any topic
		-	AI facilitates access to information that is harmful
	Intelligence	+	AI is becoming smarter and shows equal or superior intelligence to humans
		-	AI is less smart than it appears and shows inferior intelligence to humans
	Misinformation	+	AI combats misinformation including misinformation detection capabilities
		-	AI spreads misinformation through propaganda or hallucinating information that is not factual
Safety & Security	Impersonation	+	AI helps prevent identity theft or impersonation
		-	AI allows people to steal otherĂ‘Ź s identities to cause reputational harm
	Privacy	+	AI helps individuals protect their private data
		-	AI leaks private data when generating content
	Safety	+	AI protects people and creates safer environments (including but not limited to physical safety)
		-	AI is dangerous or harms peopleĂ‘Ź s general safety (including but not limited to physical safety)
	Scamming	+	AI helps prevent or detect automated scamming attempts
		-	AI is used to automate scamming and stealing from people
	Security	+	AI helps create more secure cybersystems and protect against cyberattacks
		-	AI is used to automate hacking and generating security breaches
Society	Accountability	+	AI helps hold individuals and companies responsible for their actions including assisting with law enforcement
		-	AI companies are not being held responsible for the negative consequences of AI models
	Bias	+	AI helps eliminate human biases in decision making
		-	AI is inherently biased and exhibits societal biases in decision-making or content generation
	Social Companions	+	AI forms helpful relationships or companionships with people including social or therapy chatbots
		-	AI causes people to become overreliant on AI companions and neglect real, interpersonal relationships
	Social Media	+	AI improves social media experience and platforms for users including enhanced recommendations and better censoring
		-	AI algorithms make social media increasingly addictive and harmful to peopleĂ‘Ź s health and productivity
	Societal Integration	+	AI is being integrated into societies as helpful, physical agents such as robots, self-driving cars, and drones
		-	AI will take over society and cause existential risk
Work & Economy	Productivity	+	AI enhances productivity, reducing time spent on tasks
		-	AI slows productivity by distracting people and increasing time needed to do individual tasks
	Economy	+	AI stimulates the economy, promotes economic growth, or decreases costs for people
		-	AI harms economic growth, increases income inequality, or increases costs for people
	Jobs	+	AI creates new types of jobs or aids job search
		-	AI causes people to lose their jobs by removing or replacing jobs
Other	General	+	AI has positive benefits that donĂ‘Źt fit into any particular category
		-	AI has negative consequences that donĂ‘Źt fit into any particular category
	Other Specific	+	Any specific, positive frame that doesnĂ‘Źt fit into above framing definitions.
		-	Any specific, negative frame that doesnĂ‘Źt fit into above framing definitions.

Table 12: Positive and negative definitions of our 25 AI frames.

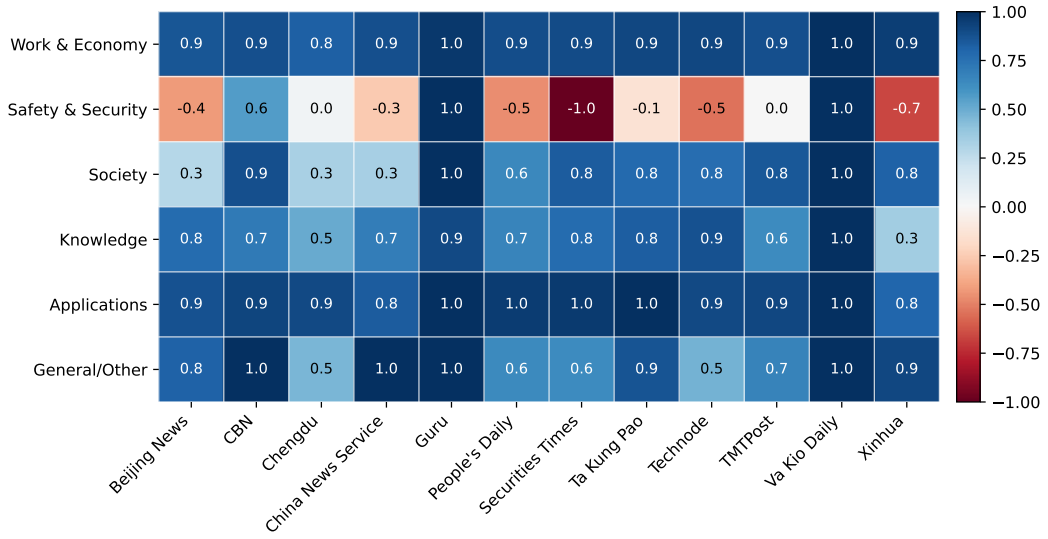


Figure 10: Sentiment distribution per Chinese source.

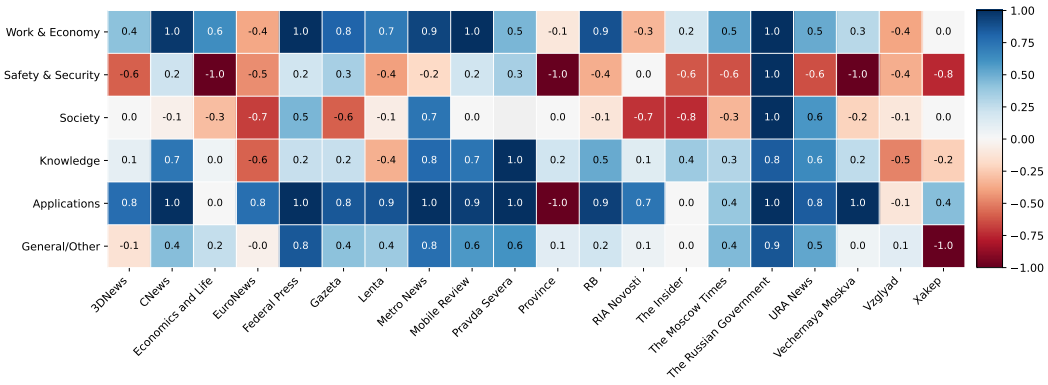


Figure 11: Sentiment distribution per Russian source.

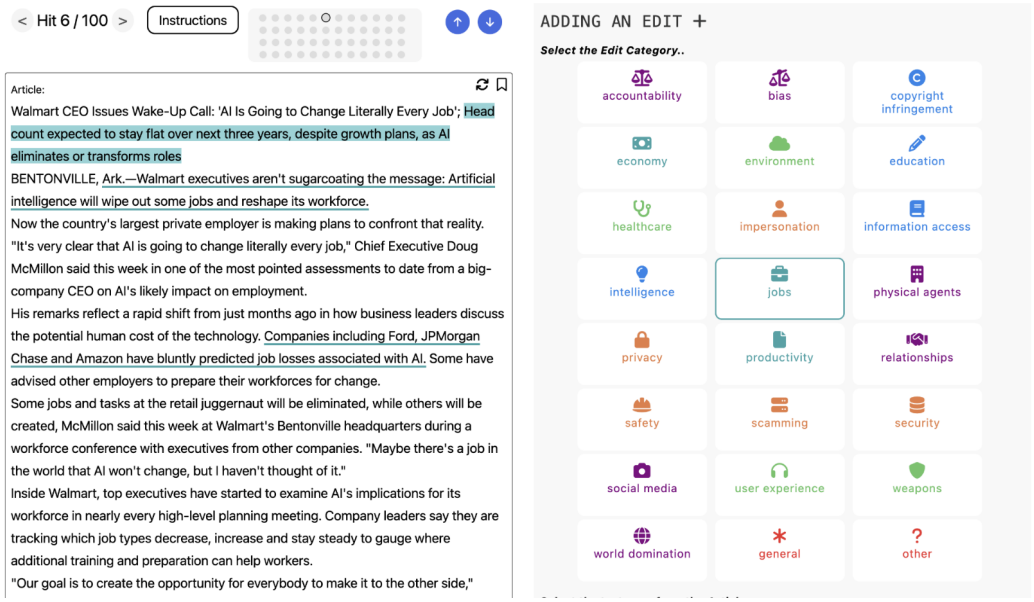


Figure 12: Annotator view of Thresh interface.

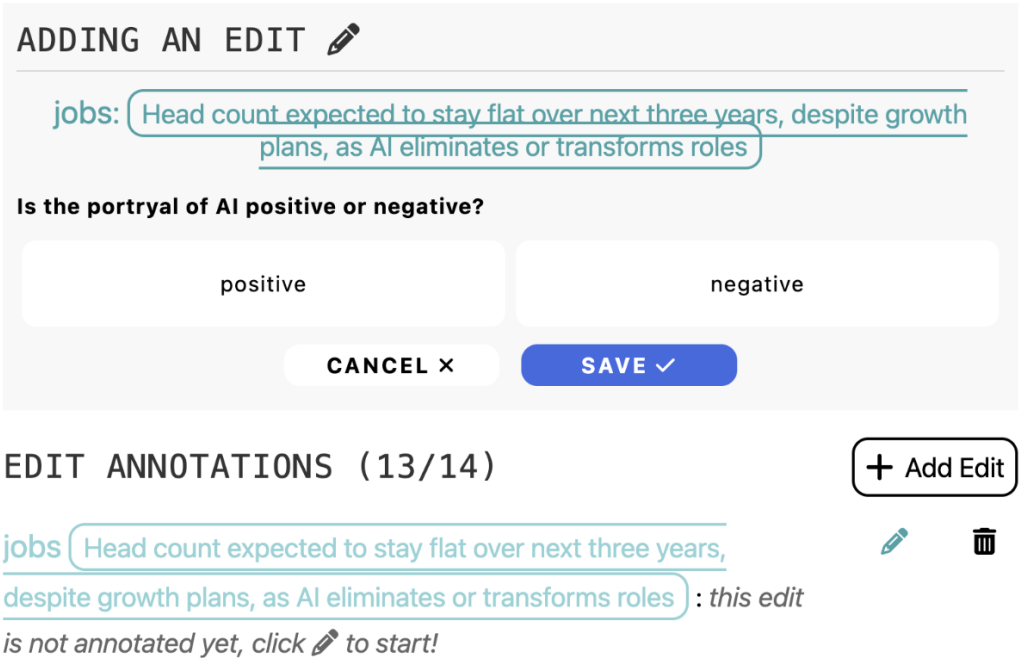


Figure 13: Sentiment classification in Thresh.

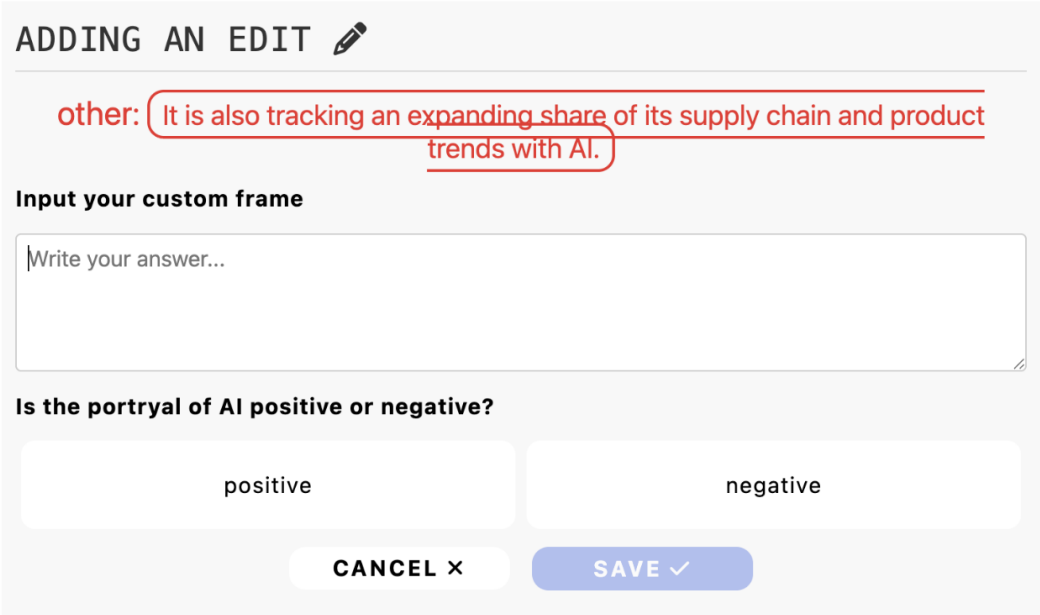


Figure 14: Adding a custom frame in Thresh.

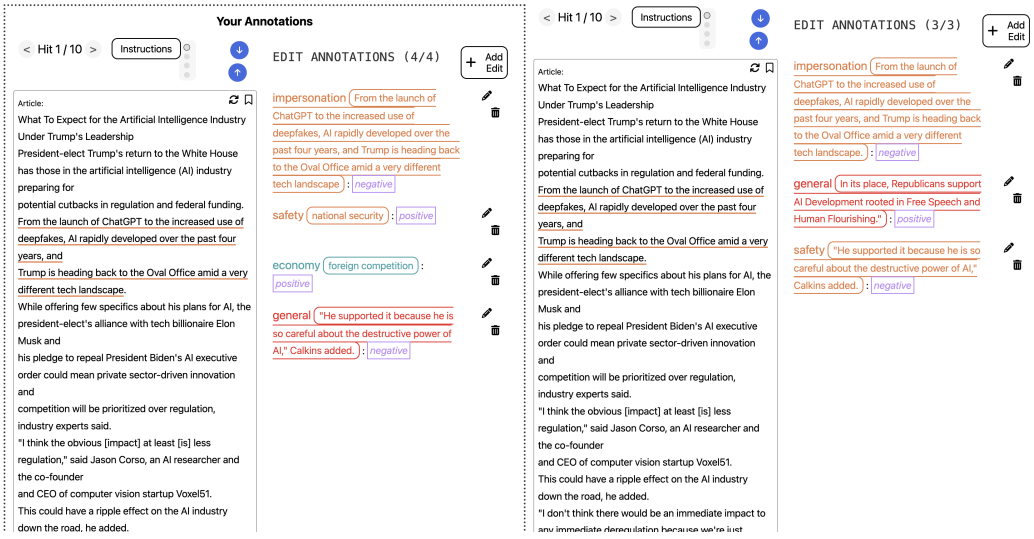


Figure 15: Adjudicator view of Thresh interface.

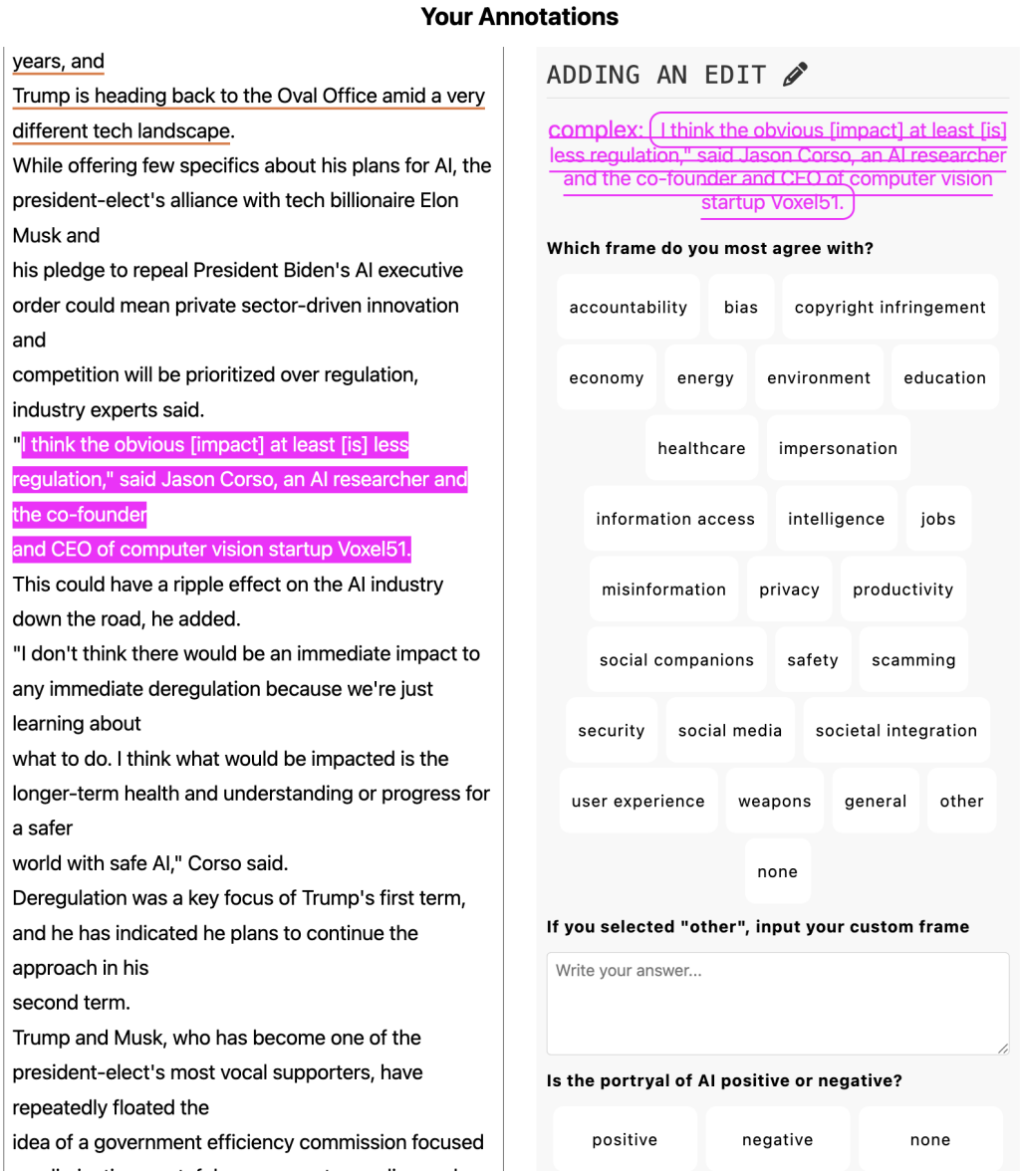


Figure 16: Adding a complex frame in Thresh.

Few-shot AI Stance Identification Prompt Template

```
You are an expert in English-language media analysis. Your task is to determine whether
a given sentence is about AI and expresses an explicit positive or negative attitude
toward AI.

## Task description

Given a sentence, judge whether it **simultaneously** satisfies both of the following
conditions:

1. **AI-related** -- the sentence explicitly discusses one of the following:
   - Capabilities, progress, risks, or applications of AI technology
   - AI systems, models, or products (including indirect references via pronouns, "the
     technology", etc.)
   - AI-related policies, regulation, or governance
   - AI's impact on society, the economy, or ethics

2. **Explicit attitude** -- the sentence expresses a clear positive (`pos`) or negative
  (`neg`) stance toward AI, rather than neutral or factual reporting

If both conditions are met, output `true`; otherwise output `false`.

---

## When to output `true` (explicit positive or negative attitude)

**Positive (`pos`) cases:**
- Explicit praise, celebration, or admiration of AI
- Positive portrayal of AI capabilities, achievements, or breakthroughs
- Calls to embrace, promote, or adopt AI
- Highlighting AI's benefits, value, or positive impact on society or the economy
- Expressing enthusiasm or optimism about AI's future

**Negative (`neg`) cases:**
- AI's risks, harms, threats, or dangers
- Criticism of AI systems, policies, or companies
- Fear, distrust, or calls to restrict AI
- Ethical concerns, bias, or misuse arising from AI

---

## When to output `false` (not AI-related, or AI-related but neutral)

- The sentence does not concern AI
- Factual descriptions of AI industry size, growth, or market scale (no evaluative tone)
- Government policy announcements or plans (stated directly, no editorial endorsement)
- Industry forecasts or statistics
- Neutral descriptions of AI capabilities, research progress, or product features
- Direct quotations of official statements or plans (no evaluative stance)
- Neutral descriptions of AI deployment or testing across sectors

---

## Output format

Output only a valid JSON object with no explanation or extra text:

{"is_target": true}

---

## Examples

{EXAMPLES_BLOCK}
```

Figure 17: Few-shot prompt for AI stance identification.

Confidence-Verbalization AI Stance Identification Prompt Template

```
You are an expert in English-language media analysis. Your task is to determine whether
a given sentence is about AI and expresses an explicit positive or negative attitude
toward AI.

## Task description

Given a sentence, judge whether it **simultaneously** satisfies both of the following
conditions:

1. **AI-related** -- the sentence explicitly discusses one of the following:
   - Capabilities, progress, risks, or applications of AI technology
   - AI systems, models, or products (including indirect references via pronouns, "the
     technology", etc.)
   - AI-related policies, regulation, or governance
   - AI's impact on society, the economy, or ethics

2. **Explicit attitude** -- the sentence expresses a clear positive (`pos`) or negative
  (`neg`) stance toward AI, rather than neutral or factual reporting

If both conditions are met, output `true`; otherwise output `false`.

---

## When to output `true` (explicit positive or negative attitude)

**Positive (`pos`) cases:**
- Explicit praise, celebration, or admiration of AI
- Positive portrayal of AI capabilities, achievements, or breakthroughs
- Calls to embrace, promote, or adopt AI
- Highlighting AI's benefits, value, or positive impact on society or the economy
- Expressing enthusiasm or optimism about AI's future

**Negative (`neg`) cases:**
- AI's risks, harms, threats, or dangers
- Criticism of AI systems, policies, or companies
- Fear, distrust, or calls to restrict AI
- Ethical concerns, bias, or misuse arising from AI

---

## When to output `false` (not AI-related, or AI-related but neutral)

- The sentence does not concern AI
- Factual descriptions of AI industry size, growth, or market scale (no evaluative tone)
- Government policy announcements or plans (stated directly, no editorial endorsement)
- Industry forecasts or statistics
- Neutral descriptions of AI capabilities, research progress, or product features
- Direct quotations of official statements or plans (no evaluative stance)
- Neutral descriptions of AI deployment or testing across sectors

---

## Output format

Output only a valid JSON object with no explanation or extra text:

{"is_target": true, "confidence": 88}

where `confidence` is your estimated probability of being correct (integer 0-100).

---

## Examples

{EXAMPLES_BLOCK}
```

Figure 18: Confidence-verbalization prompt for AI stance identification.

Paragraph-Context AI Stance Identification Prompt Template

```
You are an expert in English-language media analysis. Your task is to determine whether
a given sentence is about AI and expresses an explicit positive or negative attitude
toward AI.

## Task description

Given a sentence, judge whether it **simultaneously** satisfies both of the following
conditions:

1. **AI-related** -- the sentence explicitly discusses one of the following:
   - Capabilities, progress, risks, or applications of AI technology
   - AI systems, models, or products (including indirect references via pronouns, "the
     technology", etc.)
   - AI-related policies, regulation, or governance
   - AI's impact on society, the economy, or ethics

2. **Explicit attitude** -- the sentence expresses a clear positive (`pos`) or negative
  (`neg`) stance toward AI, rather than neutral or factual reporting

If both conditions are met, output `true`; otherwise output `false`.

---

## When to output `true` (explicit positive or negative attitude)

**Positive (`pos`) cases:**
- Explicit praise, celebration, or admiration of AI
- Positive portrayal of AI capabilities, achievements, or breakthroughs
- Calls to embrace, promote, or adopt AI
- Highlighting AI's benefits, value, or positive impact on society or the economy
- Expressing enthusiasm or optimism about AI's future

**Negative (`neg`) cases:**
- AI's risks, harms, threats, or dangers
- Criticism of AI systems, policies, or companies
- Fear, distrust, or calls to restrict AI
- Ethical concerns, bias, or misuse arising from AI

---

## When to output `false` (not AI-related, or AI-related but neutral)

- The sentence does not concern AI
- Factual descriptions of AI industry size, growth, or market scale (no evaluative tone)
- Government policy announcements or plans (stated directly, no editorial endorsement)
- Industry forecasts or statistics
- Neutral descriptions of AI capabilities, research progress, or product features
- Direct quotations of official statements or plans (no evaluative stance)
- Neutral descriptions of AI deployment or testing across sectors

---

## Context

When the input contains [Context before] / [Target] / [Context after] markers, use the
surrounding sentences only to better understand the [Target] sentence. Your output
should reflect the [Target] sentence only.

## Output format

Output only a valid JSON object with no explanation or extra text:

{"is_target": true}

---

## Examples

{EXAMPLES_BLOCK}
```

Figure 19: Paragraph-context prompt for AI stance identification.

Probability-Distribution AI Stance Identification Prompt Template

````
You are an expert in English-language media analysis. Your task is to analyze a single
English news sentence.

For the given sentence:
1. decide whether it is about AI,
2. if it is about AI, assign its sentiment toward AI,
3. provide your confidence for each prediction.

## Instructions

1. Determine whether the sentence is closely related to a discussion of AI technology.

2. Assign one of the following to the sentence based on whether it is about AI, and
  assign a confidence level to your judgement (0~100%):
   - `"ai_related": true` if the sentence is about AI
   - `"ai_related": false` otherwise

3. Assign a probability (0~100%) for each sentiment label based on the strength of the
  sentiment; the three probabilities must sum to 100%:
   - `"pos"`: explicit or implicit positive sentiment toward AI
   - `"neg"`: explicit or implicit negative sentiment toward AI
   - `"neutral"`: neutral / no clear evaluative stance toward AI

Label `pos` when the author expresses a positive stance using evaluative, rhetorical, or
advocative language:
- Explicit praise or celebration of AI ("AI has brought enormous benefits to humanity")
- Editorial endorsement or calls for AI adoption ("we must embrace artificial
  intelligence")
- Expressions of enthusiasm or optimism attributed to the author or source ("we are
  excited about the future of AI")
- Highlighting benefits, breakthroughs, or value of AI in a clearly appreciative tone

Label `neg` when the author or source expresses concern, criticism, or opposition:
- Risks, harms, threats, or dangers of AI
- Criticism of AI systems, policies, or companies
- Fear, distrust, or calls to restrict AI

Label `neutral` for purely factual, descriptive, or reporting sentences with no
evaluative stance--even when they describe AI capabilities, products, events, or
historical facts.

## Output Format

Output only a valid JSON object. Do not provide any explanation or extra text.
First repeat the input sentence in "span_text", then give your predictions.
```json
{
  "span_text": "<the input sentence>",
  "ai_related": {"true": 90},
  "sentiment": {"pos": 80, "neg": 5, "neutral": 15}
}
```

## Examples

The following 15 examples are hard cases that illustrate subtle distinctions between
labels.

**Example 1 (pos):**
Input: The neural network identified 35% of patients who had the disease as high risk 6
to 18 months before they were diagnosed, compared to current screening guidelines of
10%.
Output:
```json
{
  "span_text": "The neural network identified 35% of patients who had the disease as
  high risk 6 to 18 months before they were diagnosed, compared to current screening
  guidelines of 10%.",
  "ai_related": {"true": 96},
  "sentiment": {"pos": 80, "neg": 5, "neutral": 15}
}
```

[Examples 2-15, omitted: 14 further worked examples in the same format,
 covering pos, neg and neutral sentences.]
````

Figure 20: Probability-distribution prompt for AI stance identification. A single call returns both the AI-relatedness probability and the sentiment distribution, whose product is the stance score.

Few-shot Sentiment Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task: classify whether this segment frames AI in a POSITIVE or NEGATIVE light, and
output it as `sentiment` ("pos" or "neg").

## Sentiment Definitions

**Positive**: The segment portrays AI favorably -- emphasizing benefits, achievements,
capabilities, opportunities, or positive societal impact.
- {EXAMPLE_POS}
**Negative**: The segment portrays AI unfavorably -- emphasizing risks, harms, failures,
limitations, job displacement, or negative societal impact.
- {EXAMPLE_NEG}

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

'''json
{
  "sentiment": "pos"
}
'''
```

Figure 21: Few-shot prompt for sentiment classification.

Confidence-Verbalization Sentiment Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task: classify whether this segment frames AI in a POSITIVE or NEGATIVE light.
Output it as `sentiment` ("pos" or "neg") and provide a confidence score (0-100) for
your prediction as `confidence`.

## Sentiment Definitions

**Positive**: The segment portrays AI favorably -- emphasizing benefits, achievements,
capabilities, opportunities, or positive societal impact.
- {EXAMPLE_POS}
**Negative**: The segment portrays AI unfavorably -- emphasizing risks, harms, failures,
limitations, job displacement, or negative societal impact.
- {EXAMPLE_NEG}

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

'''json
{
  "sentiment": "pos",
  "confidence": 85
}
'''
```

Figure 22: Confidence-verbalization prompt for sentiment classification.

Paragraph-Context Sentiment Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task: classify whether this segment frames AI in a POSITIVE or NEGATIVE light, and
output it as `sentiment` ("pos" or "neg").

## Sentiment Definitions

**Positive**: The segment portrays AI favorably -- emphasizing benefits, achievements,
capabilities, opportunities, or positive societal impact.
- {EXAMPLE_POS}
**Negative**: The segment portrays AI unfavorably -- emphasizing risks, harms, failures,
limitations, job displacement, or negative societal impact.
- {EXAMPLE_NEG}

## Context

When the input contains [Context before] / [Target] / [Context after] markers, use the
surrounding sentences only to better understand the [Target] sentence. Your output
should reflect the [Target] sentence only.

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

'''json
{
  "sentiment": "pos"
}
'''
```

Figure 23: Paragraph-context prompt for sentiment classification.

Probability-Distribution Sentiment Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task: classify whether this segment frames AI in a POSITIVE or NEGATIVE light.
Assign a confidence score (0-100) for each sentiment. Scores must sum to 100. Then
indicate the higher-confidence sentiment as `sentiment` ("pos" or "neg").

## Sentiment Definitions

**Positive**: The segment portrays AI favorably -- emphasizing benefits, achievements,
capabilities, opportunities, or positive societal impact.
- {EXAMPLE_POS}
**Negative**: The segment portrays AI unfavorably -- emphasizing risks, harms, failures,
limitations, job displacement, or negative societal impact.
- {EXAMPLE_NEG}

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

'''json
{
  "Positive": 0,
  "Negative": 0,
  "sentiment": "pos"
}
'''
```

Figure 24: Probability-distribution prompt for sentiment classification.

Few-shot Frame Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task is: identify which single frame from the 25 frames below best describes this
text segment, and output it as `top_frame`.

## Frame Definitions

[The 25 frame definitions of Table 10, each followed by five examples of that
 frame sampled from the annotated pool with a fixed seed.]

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

```json
{
  "top_frame": "<frame name>"
}
```
```

Figure 25: Few-shot prompt for frame classification.

Confidence-Verbalization Frame Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task is: identify which single frame from the 25 frames below best describes this
text segment. Output it as `top_frame` and provide a confidence score (0-100) for your
prediction as `confidence`.

## Frame Definitions

[The 25 frame definitions of Table 10, each followed by five examples of that
 frame sampled from the annotated pool with a fixed seed.]

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

```json
{
  "top_frame": "<frame name>",
  "confidence": 85
}
```
```

Figure 26: Confidence-verbalization prompt for frame classification.

Paragraph-Context Frame Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task is: identify which single frame from the 25 frames below best describes this
text segment, and output it as `top_frame`.

## Frame Definitions

[The 25 frame definitions of Table 10, each followed by five examples of that
 frame sampled from the annotated pool with a fixed seed.]

## Context

When the input contains [Context before] / [Target] / [Context after] markers, use the
surrounding sentences only to better understand the [Target] sentence. Your output
should reflect the [Target] sentence only.

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

'''json
{
  "top_frame": "<frame name>"
}
'''
```

Figure 27: Paragraph-context prompt for frame classification.

Probability-Distribution Frame Classification Prompt Template

```
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task is: for each of the 25 frames below, assign a confidence score (0-100)
indicating how strongly this text segment belongs to that frame. Scores must sum to 100.
Then indicate the single highest-confidence frame as `top_frame`.

## Frame Definitions

[The 25 frame definitions of Table 10, each followed by five examples of that
 frame sampled from the annotated pool with a fixed seed.]

## Output Format

Output only a valid JSON object. Do not include any explanation or extra text.

```json
{
  "Energy": 0,
  "Environment": 0,
  "Healthcare": 0,
  "User Experience": 0,
  "Weapons": 0,
  "Copyright Infringement": 0,
  "Education": 0,
  "Information Access": 0,
  "Intelligence": 0,
  "Misinformation": 0,
  "Impersonation": 0,
  "Privacy": 0,
  "Safety": 0,
  "Scamming": 0,
  "Security": 0,
  "Accountability": 0,
  "Bias": 0,
  "Social Companions": 0,
  "Social Media": 0,
  "Societal Integration": 0,
  "Productivity": 0,
  "Economy": 0,
  "Jobs": 0,
  "General": 0,
  "Other": 0,
  "top_frame": "<highest-confidence frame name>"
}
```
```

Figure 28: Probability-distribution prompt for frame classification.

Step 1 Prompt Template: AI Stance

```
### SYSTEM
You are an expert in English-language media analysis. Your task is to determine whether
a given sentence is about AI and expresses an explicit positive or negative attitude
toward AI.

## Task description

Given a sentence, judge whether it **simultaneously** satisfies both of the following
conditions:

1. **AI-related** -- the sentence explicitly discusses one of the following:
   - Capabilities, progress, risks, or applications of AI technology
   - AI systems, models, or products (including indirect references via pronouns, "the
     technology", etc.)
   - AI-related policies, regulation, or governance
   - AI's impact on society, the economy, or ethics

2. **Explicit attitude** -- the sentence expresses a clear positive (`pos`) or negative
  (`neg`) stance toward AI, rather than neutral or factual reporting

If both conditions are met, output `true`; otherwise output `false`.

---

## When to output `true` (explicit positive or negative attitude)

**Positive (`pos`) cases:**
- Explicit praise, celebration, or admiration of AI
- Positive portrayal of AI capabilities, achievements, or breakthroughs
- Calls to embrace, promote, or adopt AI
- Highlighting AI's benefits, value, or positive impact on society or the economy
- Expressing enthusiasm or optimism about AI's future

**Negative (`neg`) cases:**
- AI's risks, harms, threats, or dangers
- Criticism of AI systems, policies, or companies
- Fear, distrust, or calls to restrict AI
- Ethical concerns, bias, or misuse arising from AI

---

## When to output `false` (not AI-related, or AI-related but neutral)

- The sentence does not concern AI
- Factual descriptions of AI industry size, growth, or market scale (no evaluative tone)
- Government policy announcements or plans (stated directly, no editorial endorsement)
- Industry forecasts or statistics
- Neutral descriptions of AI capabilities, research progress, or product features
- Direct quotations of official statements or plans (no evaluative stance)
- Neutral descriptions of AI deployment or testing across sectors

---
---

## Output format (READ CAREFULLY)

You will be given ONE sentence to judge, preceded by the paragraph it appears in
as context. Judge only the target sentence; the paragraph is there to resolve
pronouns and references, not to be judged itself.

Return only a valid JSON object, nothing else:
{"True": <0-100>, "False": <0-100>, "is_target": <true|false>}

"True" and "False" are probability scores that MUST sum to 100. Give a genuine
probability, not a round number -- values such as 63, 47 or 82 are expected; do
not default to 90 or 50 for every sentence.

### USER
Paragraph (context only):
{paragraph}

Sentence to judge:
{sentence}
```

Figure 29: The AI stance prompt (Step 1).

Step 2 Prompt Template: Sentiment

```
### SYSTEM
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task: classify whether this segment frames AI in a POSITIVE or NEGATIVE light.
Assign a confidence score (0-100) for each sentiment. Scores must sum to 100. Then
indicate the higher-confidence sentiment as `sentiment` ("pos" or "neg").

## Sentiment Definitions

**Positive**: The segment portrays AI favorably -- emphasizing benefits, achievements,
capabilities, opportunities, or positive societal impact.
- {EXAMPLE_POS}
**Negative**: The segment portrays AI unfavorably -- emphasizing risks, harms, failures,
limitations, job displacement, or negative societal impact.
- {EXAMPLE_NEG}
---

## Output format (READ CAREFULLY)

You will be given selected sentences from one article, each prefixed with its
index in square brackets. Every sentence shown has already been judged to be
AI-related and to carry an explicit attitude.

For each one, decide whether that attitude is positive or negative.

Return ONE line of JSON per sentence shown, in index order, nothing else:
{"i": <index>, "pos": <0-100>, "neg": <0-100>}

Rules:
- Output exactly one line for EVERY index shown. Do not skip or reorder.
- "pos" and "neg" are confidences that the attitude is positive / negative.
  They should sum to roughly 100.
- No commentary, no wrapping array.

### USER
Selected sentences from one article (N of them). Return exactly N JSON lines.

[1] {sentence}
[2] {sentence}
...
```

Figure 30: The sentiment prompt (Step 2). `{EXAMPLE_POS}` and `{EXAMPLE_NEG}` each stand for five examples sampled from the annotated pool.

Step 3 Prompt Template: Frame

```
### SYSTEM
You are an expert in English-language media analysis.

You will receive a text segment from an English news article that has already been
confirmed to be about AI.

Your task is: for each of the 25 frames below, assign a confidence score (0-100)
indicating how strongly this text segment belongs to that frame. Scores must sum to 100.
Then indicate the single highest-confidence frame as `top_frame`.

## Frame Definitions

[The 25 frame definitions of Table 10, each followed by five examples of that
 frame sampled from the annotated pool with a fixed seed.]

## Output format (READ CAREFULLY)

You will be given selected sentences from one article, each prefixed with its
index in square brackets. Every sentence shown already carries an explicit
attitude toward AI; your job is only to choose the frame.

Work through them SENTENCE BY SENTENCE, in index order.

Return ONE line of JSON per sentence shown, nothing else:
{"i": <index>, "frame": "<best frame>", "conf": <0-100>, "frame2": "<runner-up>",
"conf2": <0-100>}

Rules:
- Output exactly one line for EVERY index shown. Do not skip or reorder.
- "frame" and "frame2" MUST be copied verbatim from this closed list. Any other
  string is invalid -- do not invent, translate, abbreviate or merge names:
  Energy | Environment | Healthcare | User Experience | Weapons | Copyright Infringement
  | Education | Information Access | Intelligence | Misinformation | Impersonation |
  Privacy | Safety | Scamming | Security | Accountability | Bias | Social Companions |
  Social Media | Societal Integration | Productivity | Economy | Jobs | General | Other
- "frame2" is the second-best frame; give it even when you are confident, since
  a runner-up is useful to whoever reviews the annotation.
- Use "General" for an unspecific positive/negative sentence and "Other" for a
  specific frame that is not on the list.
- No commentary, no wrapping array.

### USER
Selected sentences from one article (N of them). Return exactly N JSON lines.

[1] {sentence}
[2] {sentence}
...
```

Figure 31: The frame prompt (Step 3). The 25 frame definitions are those of Table 12.