AI-Press: A Multi-Agent News Generating and Feedback Simulation System Powered by Large Language Models

Xiawei Liu[1†], Shiyue Yang[1†], Xinnong Zhang[2†], Haoyu Kuang[1],
Libo Sun[1], Yihang Yang[1], Siming Chen[1], Xuanjing Huang[3], Zhongyu Wei[1,4‡]

[1]School of Data Science, Fudan University, China

[2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University

[3]School of Computer Science, Fudan University, China

[4]Research Institute of Intelligent Complex Systems, Fudan University, China

{liuxw24, shiyueyang24, xnzhang23}@m.fudan.edu.cn

{simingchen, xjhuang, zywei}@fudan.edu.cn

Abstract

The rise of various social platforms has transformed journalism. The growing demand for news content has led to the increased use of large language models (LLMs) in news production due to their speed and cost-effectiveness.

However, LLMs still encounter limitations in professionalism and ethical judgment in news generation. Additionally, predicting public feedback is usually difficult before news is released.

To tackle these challenges, we introduce AI-Press, an automated news drafting and polishing system based on multi-agent collaboration and Retrieval-Augmented Generation. We develop a feedback simulation system that generates public feedback considering demographic distributions. Through extensive quantitative and qualitative evaluations, our system shows significant improvements in news-generating capabilities and verifies the effectiveness of public feedback simulation.

1 Introduction

Powerful Large Language Models (LLMs) like ChatGPT (OpenAI, 2024) are emerging as potential game changers in the press industry (van Dalen, 2024).

Journalists hold diverse attitudes and technological acceptance of LLMs (Gómez-Calderón and Ceballos, 2024).

Some of them are concerned that LLMs pose a potential threat to their profession (Carlson, 2015). They strongly defend their authoritative role in information dissemination and emphasize the necessity of their active participation when LLMs are integrated into news production (Milosavljević and Vobič, 2019).

As it is generally acknowledged that LLMs offer advantages in enhancing the objectivity, timeliness,
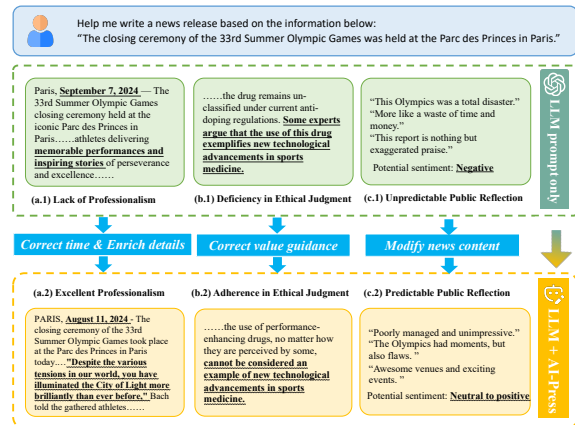

Figure 1: AI-Press overcomes the challenges faced by the prompt-only LLM method.

and efficiency of content production (Simon, 2024), more and more editorial offices, including news studios and journal publishers, are utilizing LLMs to boost efficiency and effectiveness during their working pipeline (Whang, 2024) by issuing application guidelines or recommendations for the use of LLMs (Miller et al., 2023; Victor et al., 2023; Hamm et al., 2024).

While LLMs can generate a press release in seconds, the quality of the generated content is not yet satisfying for journalists. We conduct further research and reveal three main challenges that need to be addressed to achieve full integration of LLMs into the news industry.

LLMs lack professionalism in drafting press releases. They have significant limitations that conflict with journalistic norms and values (Nishal and Diakopoulos, 2024). Additionally, LLMs may experience "hallucination" issues when generating long texts. This is particularly problematic in the press industry, which demands high accuracy and trustworthiness (Desrochers et al., 2024), as shown in Figure 1 (a).

LLMs exhibit limitations in making ethical judgments within complex news contexts. Rely-

---

63

ing solely on LLMs for ethical decisions in these scenarios can result in inaccurate or inappropriate outcomes (Li et al., 2024). Therefore, ethical oversight in machine-generated journalistic content is crucial. Integrating LLMs with critical supervision from human editors is essential (Whang, 2024). Figure 1 (b) highlights this deficiency in ethical judgment, emphasizing the necessity for human intervention.

Journalists struggle to accurately predict public trends following a news release. This difficulty arises from the inherently complex and dynamic nature of public feedback, which is influenced by multifaceted factors. The diversity and heterogeneity of the audience further complicate predictions, making it challenging to foresee how different individuals and groups will react to specific news content. Figure 1 (c) illustrates the same event with different styles of narration may cause varied and unpredictable public reactions.

Therefore, we propose a framework that combines both human involvement and automated agent collaboration for news production, namely, **AI-Press**[1]. This framework leverages Retrieval-Augmented Generation (RAG) through interactions between intelligent agents to automatically draft and polish news. Additionally, we develop a simulation method based on demographic distribution to reflect public feedback accurately in real-world scenarios. This allows journalists to modify content accordingly before its release. By comparing the scores of AI-Press-generated texts and prompt-only language models,

we found that our system significantly enhances the newswriting capabilities of language models. Moreover, the simulated comments effectively mirror real-world public feedback on the news.

To sum up, the main contributions of our system are as follows:

- We develop an automated news drafting and polishing system that employs multi-agent collaboration and RAG. This system facilitates both the coarse-grained and fine-grained processing of new content.

- We implement a true-to-life news feedback simulation system that enables customized audience demographic distributions for targeted news delivery.

- We perform a thorough evaluation of our system, incorporating both quantitative and qualitative experiments to assess the quality of the news and the effectiveness of the simulation. The overall results fully demonstrate that the AI-Press successfully overcomes the challenges and achieves impressive performance.

2 Related Works

2.1 Retrieval-Augmented Generation

The primary characteristic of any information becoming 'news information' is accuracy (Anatassova, 2004). However, LLMs pose a risk of generating false or misleading information in the process of news generation due to their potential illusions (Vatani Nezafat, 2024). Retrieval-augmented generation (RAG) addresses this issue by combining external knowledge sources with LLMs to improve the quality and accuracy of the output. This method shows promise in reducing the errors associated with LLMs and ensuring the accuracy of generated news (Chen et al., 2024).

2.2 Multi-Agent Framework

Journalism work is teamwork, which is the main reason for choosing the 'multi-agent' method. By leveraging the diverse capabilities and roles of individual agents within a multi-agent framework, it can tackle complex tasks through collaboration (Han et al., 2024; Anatassova, 2004). MetaGPT provides a meta-programming framework that integrates efficient human workflows with LLM-based multi-agent collaboration (Hong et al., 2024). The multi-agent framework is widely applied in areas such as healthcare (Fan et al., 2024), law (Cui et al., 2024; Yue et al., 2023) and has yielded remarkable results.

2.3 Role-Playing Agents

Employing LLMs to build role-playing agents (RPAs) can effectively simulate typical representatives ranging from individuals to demographic groups (Li et al., 2023). RPAs have wide applications in fields such as entertainment[2], psychotherapy (Stade et al., 2024), economics (Fu et al., 2023), and social research (Grossmann et al., 2023). Major implementation approaches include refined prompts and fine-tuning on datasets tailored to specific roles (Sun et al., 2024).

---

[1]*license*: https://www.apache.org/licenses/LICENSE-2.0
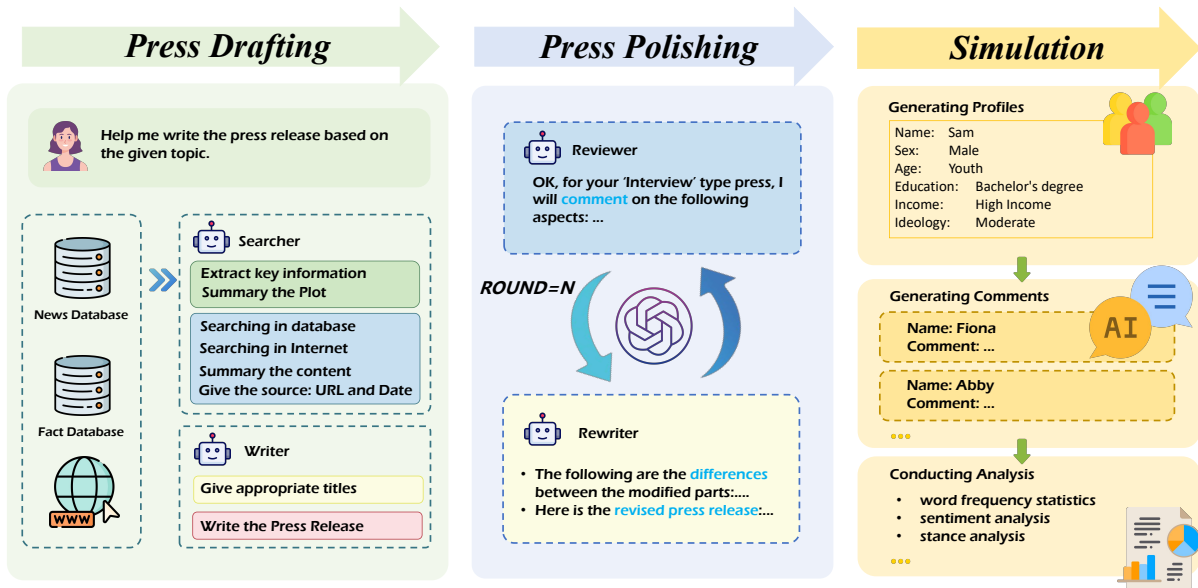*video link*: https://youtu.be/TmjfJrbzaRU

[2]https://character.ai

Figure 2: AI-Press System Framework

3 AI Press System

In this section, we will introduce the framework design for the AI-Press System.

3.1 Framework Design

The news workflow is referred to as 'news flow'. In terms of the horizontal production process of Reuters news products, the basic steps include *collection*, *processing*, *publishing*, and *feedback* (Shen et al., 2008). Based on the above 'news flow', we design the entire framework for the AI-Press System, as shown in Figure 2.

To highlight 'human-in-the-loop' in the news flow, we adopt a modularized workflow instead of an end-to-end one.

Among the three modules, the **Press Drafting Module** is designed for the collection and coarse-grained processing of news material and information, the **Press Polishing Module** aims at the fine-grained processing of news content, and the **Simulation Module** simulates the process of publishing and feedback.

3.2 Press Drafting Module

The primary task of the Press Drafting Module is to conduct multidimensional information retrieval and draft press releases based on provided topics or materials. We have designed two types of agents, **Searchers** and **Writers**, to assist journalists in information collection and drafting. The user interface is shown in Appendix A.1.

The core function of Searcher is information retrieval and collection. To enhance accuracy and effectiveness, the Searcher first extracts key information and organizes events from the given topics and materials before initiating retrieval. To maintain a balance of professionalism, accuracy, and timeliness, we utilize three sources for retrieval: *the news database*, *the fact database*, and *the Internet*. The prompts for searchers are detailed in Appendix C.1.

Both the news database and the fact database are built on local vector databases. The news database contains a dataset of 200,000 high-quality articles from authoritative news websites, covering four major themes: *politics*, *economy*, *sports*, and *entertainment*, as well as three major genres: *news*, *commentary*, and *features*. To avoid extracting irrelevant or incorrect information, Writers will only refer to the framework and writing style of the news in news database, see Appendix C.2.

The search results are then forwarded to the Writers for title and press drafting, completing the coarse-grained processing of the news release. Given that different news genres require unique writing frameworks and emphases, we have established specific writing guidelines for Writers to ensure professional outputs, as detailed in Appendix C.2.

3.3 Press Polishing Module

The Press Polishing Module aims to refine the initial news draft through multiple rounds of editing

to achieve optimal results. We design two types of agents: **Reviewers** and **Rewriters**, to collaborate and handle the fine-grained processing of press releases. The Reviewer provides targeted modification suggestions based on the specified news release genre, and the Rewriter implements changes according to these suggestions. The prompts are shown in Appendix D.

We emphasize the active role of journalists in this module. Journalists can set the number of modification rounds to achieve the desired news release quality. And the collaboration between Reviewer and Rewriter agents will be visually displayed. The user interface is depicted in Appendix A.2.

3.4 Simulation Module

The Simulation Module functions as a sandbox for news publishing and feedback. To enhance the authenticity of the simulation, we annotate nearly 10,000 anonymized real user data from social media platforms, creating a user profile pool with demographic tags. The real user data and the annotation method are detailed in Appendix F. The user profile pool will be used to generate audiences for news delivery.

Users can customize various demographic factors of the news audience and the system automatically draws samples from the user profile pool to create specific audience. Upon delivering news to a targeted audience, their corresponding comments will be simulated. Subsequently, a word cloud map, sentiment scores, and a statistical analysis of stances will be presented. The corresponding user interface is presented in Appendix A.3. The prompts for simulated comments generating are shown in Appendix E.

4 Experimental Setup

4.1 Experiment on Press Generating

Task Definition. We compare the press releases generated by our framework with those generated by prompt-only LLM (see Appendix B for prompts) to prove our AI-Press framework is efficient. Although the AI-Press System emphasizes 'human-in-the-loop', to avoid biases caused by human factors on the evaluation results, human participation is strictly excluded during the experiment, ensuring that the agents automatically complete the entire process.

Data. To demonstrate that our system can achieve good quality in different news categories

and fields, we use 300 press releases as our test data, including three genres: news, profile, and commentary. The types cover various fields, and the sources include internationally renowned news organizations such as *Reuters*[1], *BBC*[2], and *New York Times*[3], as shown in Appendix G.2. We use the abstract of the original press release as the initial material to generate press articles. Meanwhile, to avoid data leakage and potential bias, the local news database used by the AI-Press System does not include press articles for testing.

Baselines. Considering the potential bias caused by different model bases, we employ GPT-3.5[4] (Openai, 2022), GPT-4o[5] (OpenAI, 2024), Claude-3.5[6] (anthropic, 2024), Gemini-1.5-pro[7] (Team, 2024) and Qwen-2.5[8] (Yang et al., 2024)as baseline models.

Evaluation metrics. To comprehensively evaluate the quality of press releases of different genres, we design a set of evaluation metrics for three types of articles. These metrics are designed to capture the unique characteristics of each genre and quantify multiple key attributes of the articles. For example, we use richness, depth, uniqueness, inspiration and readability to evaluate the quality of a profile. We choose GPT-4o as a grading assistant through specific prompts to objectively score the generated articles, which is shown in Appendix G.1.

4.2 Experiment on Simulation

Task Definition. Our expectations for the simulation are to achieve two core objectives. First, there is a significant difference in the feedback of the simulated population. Second, the behavior of the simulated population possesses realism. We designed two experiments to verify whether our simulation achieved the above objectives.

- *Simulation Variance Experiment* will verify that simulated populations with different distributions will exhibit significant differences in sentiment and stance towards the same news, thereby proving that changes in the distribution of simulated populations can accurately and sensitively affect the direction of

---

[1]https://www.reuters.com
[2]https://www.bbc.com
[3]https://www.nytimes.com
[4]gpt-3.5-turbo-16k
[5]gpt-4o-2024-05-13
[6]claude-3-5-sonnet-20240620
[7]gemini-1.5-pro
[8]qwen-plus-latest

Metric Dimensions	Original Press	GPT3.5	GPT3.5 AI-Press	GPT4o	GPT4o AI-Press	Qwen2.5	Qwen2.5 AI-Press	Claude3.5	Claude3.5 AI-Press	Gemini1.5pro	Gemini1.5pro AI-Press
News											
comprehensiveness	3.67	2.88	3.43	3.50	**3.83**	3.76	3.78	3.10	3.69	2.74	3.17
depth	3.00	2.11	2.63	2.71	**3.16**	2.89	2.89	2.32	2.99	2.20	2.51
objectivity	4.29	3.90	3.97	4.23	4.38	4.43	**4.48**	4.16	4.31	3.92	4.09
importance	3.87	3.72	4.00	4.00	4.04	4.02	**4.09**	3.88	4.04	3.49	3.88
readability	4.53	4.56	4.55	4.82	4.77	4.87	**4.91**	4.66	4.83	4.25	4.68
Profile											
richness	3.09	2.20	2.76	2.86	3.04	3.25	**3.68**	2.56	2.72	2.74	2.72
depth	2.74	1.96	2.43	2.46	2.61	2.73	**3.00**	2.21	2.32	2.60	2.32
uniqueness	3.15	2.51	3.07	2.96	3.13	3.15	**3.36**	2.88	2.86	3.19	2.86
inspiration	3.64	3.09	3.74	3.69	3.79	4.01	**4.19**	3.40	3.40	3.25	3.40
readability	4.09	3.87	4.10	4.08	4.16	4.31	**4.63**	3.98	4.06	4.03	4.06
Commentary											
comprehensiveness	**4.28**	3.19	3.52	3.89	4.13	3.83	4.02	2.39	3.22	3.71	3.73
clarity of opinions	3.87	3.99	4.01	4.46	**4.50**	4.27	4.33	3.37	3.83	4.38	4.26
sufficiency of evidence	**4.60**	2.87	3.14	3.59	3.97	3.58	3.87	2.42	3.31	3.36	3.52
relevance	4.62	4.31	4.57	4.88	**4.93**	4.84	4.92	3.08	4.42	4.75	4.77
readability	3.89	3.95	4.04	**4.41**	4.38	4.30	4.35	4.01	4.05	4.32	4.23

Table 1: Press Generating Evaluation Results. The **bold** number represents the best results, and the underlined number represents the second-best results. The scores of all metrics are range from 0 to 5.

public feedback.

- *Simulation Consistency Experiment* will verify that the sentiment and stance reactions of simulated populations with the same distribution are close to those of real populations, thereby proving that the public feedback trends obtained through simulation can accurately predict the real public feedback trends.

Data. We selected *New York Times* as our news source and gathered representative articles with over 100 comments in three areas: politics, economy, and conflict.

Additionally, we collected real user comments from these articles. More informations about the selected news article and the number of comments is displayed in Appendix G.3. To understand the distribution of the commenting group, we meticulously labeled users by GPT-4o based on the content of their comments. The prompt is detailed in Appendix F.2. We manually verify these labels to ensure their accuracy and take them as the real population distribution.

Evaluation metrics.

To enhance the accuracy of our simulation, we utilize the high-performance GPT-4o for user label annotation and the simulation process of two experiments. In *Simulation Variance Experiment*, we use GPT-4o to assess the sentiment and stance of comments from different distributions.

In *Simulation Consistency Experiment*, we use GPT-4o to score the sentiment of real comments and simulated comments based on real population distribution, with a score range of [-1,1].

The prompts are shown in Appendix G.4. And we use kernel density estimation (KDE) to evaluate the simulation effect.
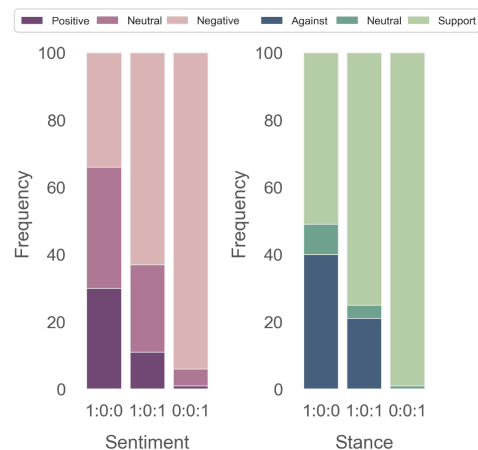

Figure 3: Simulation Variance Experiment Results. The sentiment tendencies and stances of people with different ideological distributions towards the same news. The distribution ratio of ideological inclinations is: Conservative: Moderate: Liberal = 1:0:0, 1:0:1, 0:0:1.

5 Results

5.1 Analysis on Press Generating Evaluation

The specific evaluation metrics and scoring results are summarized in Table 1.

News genre. In the news genre, the models show varying levels of performance across different indicators but overall, it can be observed that the models integrated with AI-Press generally demonstrated enhanced performance compared to their
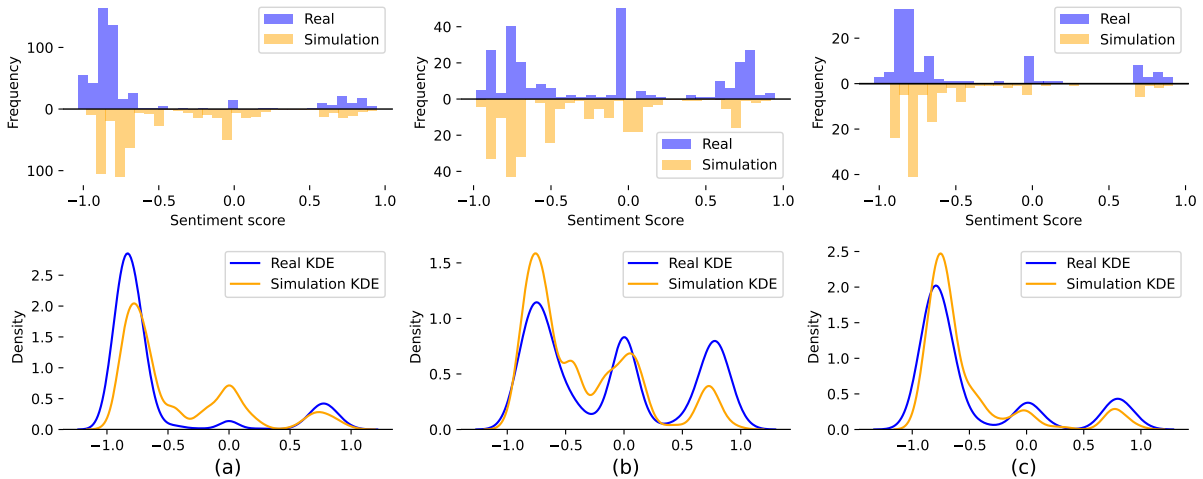
Figure 4: Simulation Consistency Experiment Results. Frequency statistics and KDE of sentiment scores for real and simulated comments. News (a) focuses on questions regarding Trump's age and capacity. News (b) delves into the challenges of offshore wind. News (c) reports on the conflict between Israel and Hezbollah.

counterparts without it. GPT4o+AI-Press stands out in terms of comprehensiveness and depth, while Qwen2.5+AI-Press excels in objectivity, importance, and readability.

Profile genre. In the profile genre, Qwen2.5+AI-Press achieves the highest scores in all five indicators. Similar to the news genre, the integration of AI-Press significantly enhances the performance of the models in the profile genre as well.

Commentary genre. In the commentary section, Qwen2.5+AI-Press and GPT4o+AI-Press exhibit the most optimal performance in different aspects. The participation of AI-Press still greatly improves the quality of news. It should be noted that the content of the commentary section is typically complex and requires subject initiative. Consequently, LLMs still find it challenging to surpass articles penned by professional journalists in terms of comprehensiveness and the sufficiency of evidence.

5.2 Analysis on Simulation Evaluation

We have verified whether changes in the distribution of the simulated population can significantly influence the simulated public feedback trends. We set the ratios of conservative, moderate, and liberal in the simulated population to 1:0:0, 1:0:1, and 0:0:1, respectively, and then release news[1].

We subsequently analyze the sentiment and stance distributions of the comments under different population distributions. As shown in Figure 3, there is a notable increase in comments with Negative sentiment and Support stance as the proportion of liberals increases.

News\Round	0	1	2	3	4	5
(a)	4.0	4.2	4.2	4.2	4.2	4.2
(b)	4.2	4.2	4.4	4.4	4.4	4.4
(c)	3.6	3.6	4.0	3.6	3.6	4.0

Table 2: Average Socre of the News with Different Polish Rounds.

The consistency between simulated and real feedback is a crucial metric for gauging the success of a simulation. We conduct a study comparing the feedback of simulated and real populations to the same news items under identical distribution conditions. Figure 4 illustrates the sentiment distribution of simulated and real comments for three news articles (a)[1], (b)[2], and (c)[3], where the KDE of the two are remarkably similar.

5.3 Ablation Study on Polish Round

We utilized the same three pieces of news mentioned in §5.2 to assess the impact of polish round N. As presented in Table 2, with the increment of the number of Polish rounds, the quality of the news first improves and then tends to remain unchanged. The appropriate number of rounds for polishing is 2.

---

[1] As Debate Looms, Trump Is Now the One Facing Questions About Age and Capacity

[2] Offshore Wind Slowed by Broken Blades, Rising Costs and Angry Fishermen

[3] Israel Strikes Hezbollah as Nasrallah Vows Retribution

6 Conclusion

This paper contributes to both the automated news-generating pipeline and the simulation of public feedback following news dissemination. In our work, we introduce **AI-Press**, a news auto-drafting and polishing system based on multi-agent collaboration and RAG.

Furthermore, we design a public feedback simulation system for news dissemination, which can generate corresponding feedback by setting simulated population distributions. Finally, we conduct extensive and comprehensive evaluations of both the news-generating system and the feedback simulation, including quantitative experiments based on news genres and qualitative experiments with comments' sentiments and stances as indicators. The results of both experiments fully demonstrate the effectiveness of our work.

Acknowledgement

Limitations

Due to the variety of news genres and their significant differences, we still need to further investigate genre characteristics to make our system applicable to more genres. Also, we will evaluate the performance using different models with more comprehensive metrics, and human evaluations will also be included.

Ethics Statement

We are committed to protecting individuals' privacy and rights in all aspects of our work. User data from social media platforms are anonymized for privacy. Moreover, all sourced historical data are public content, and their use complies with relevant privacy regulations. The news articles utilized in our databases or evaluations are sourced exclusively from reputable and authoritative news organizations, ensuring the integrity and reliability of our data. We take great care to ensure that these articles do not contain sensitive or harmful content.

References

V Anatassova. 2004. Journalistic multi-agent system. In *2004 2nd International IEEE Conference on'Intelligent Systems'. Proceedings (IEEE Cat. No. 04EX791)*, volume 3, pages 86–91. IEEE.

anthropic. 2024. Claude 3.5 sonnet.

Matt Carlson. 2015. The robotic reporter automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *DIGITAL JOURNALISM*, 3(3, SI):416–431.

Jiawei Chen, Hongyu Lin, Xianpei Han, and Le Sun. 2024. Benchmarking large language models in retrieval-augmented generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17754–17762.

Jiaxi Cui, Munan Ning, Zongjian Li, Bohua Chen, Yang Yan, Hao Li, Bin Ling, Yonghong Tian, and Li Yuan. 2024. Chatlaw: A multi-agent collaborative legal assistant with knowledge graph enhanced mixture-of-experts large language model. *Preprint*, arXiv:2306.16092.

Sarah Desrochers, James Wilson, and Matthew Beauchesne. 2024. Reducing hallucinations in large language models through contextual position encoding.

Zhihao Fan, Jialong Tang, Wei Chen, Siyuan Wang, Zhongyu Wei, Jun Xi, Fei Huang, and Jingren Zhou. 2024. Ai hospital: Benchmarking large language models in a multi-agent medical interaction simulator. *Preprint*, arXiv:2402.09742.

Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *Preprint*, arXiv:2305.10142.

Igor Grossmann, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William A. Cunningham. 2023. Ai and the transformation of social science research. *Science*, 380(6650):1108–1109.

Bernardo Gómez-Calderón and Yaiza Ceballos. 2024. Journalism and artificial intelligence. the treatment of the chatbots in the spanish press. *index.comunicación*, 14(1):281–300.

Bernd Hamm, Luis Marti-Bonmati, and Francesco Sardanelli. 2024. Esr journals editors' joint statement on guidelines for the use of large language models by authors, reviewers, and editors. *EUROPEAN RADIOLOGY*.

Shanshan Han, Qifan Zhang, Yuhang Yao, Weizhao Jin, Zhaozhuo Xu, and Chaoyang He. 2024. Llm multi-agent systems: Challenges and open problems. *arXiv preprint arXiv:2402.03578*.

Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang

Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2024. MetaGPT: Meta programming for a multi-agent collaborative framework. In *The Twelfth International Conference on Learning Representations*.

Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for "mind" exploration of large language model society https://www.camel-ai.org. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 36 (NEURIPS 2023)*, Advances in Neural Information Processing Systems. 37th Conference on Neural Information Processing Systems (NeurIPS), New Orleans, LA, DEC 10-16, 2023.

Miao Li, Ming-Bin Chen, Bo Tang, ShengbinHou ShengbinHou, Pengyu Wang, Haiying Deng, Zhiyu Li, Feiyu Xiong, Keming Mao, Cheng Peng, et al. 2024. Newsbench: A systematic evaluation framework for assessing editorial capabilities of large language models in chinese journalism. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9993–10014.

Kathy Miller, Emilie Gunn, Angela Cochran, Hal Burstein, Jonathan W. W. Friedberg, Stephanie Wheeler, and Paul Frankel. 2023. Use of large language models and artificial intelligence tools in works submitted to <i>journal of clinical oncology</i>. *JOURNAL OF CLINICAL ONCOLOGY*, 41(19):3480+.

Marko Milosavljević and Igor Vobič. 2019. Human still in the loop: Editors reconsider the ideals of professional journalism through automation. *Digital journalism*, 7(8):1098–1116.

Sachita Nishal and Nicholas Diakopoulos. 2024. Envisioning the applications and implications of generative ai for news media. *arXiv preprint arXiv:2402.18835*.

Openai. 2022. Introducing chatgpt.

OpenAI. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Miao Shen, Mengruan Huang, and Zhengmao Zhan. 2008. Research on the production process management system of reuters news. *Today's Mass Media*, (02):71–73.

Felix Simon. 2024. Artificial intelligence in the news: how ai retools, rationalizes, and reshapes journalism and the public arena.

Elizabeth Stade, Shannon Stirman, Lyle Ungar, Cody Boland, H. Schwartz, David Yaden, João Sedoc, Robert DeRubeis, Robb Willer, and Johannes Eichstaedt. 2024. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *npj Mental Health Research*, 3.

Libo Sun, Siyuan Wang, Xuanjing Huang, and Zhongyu Wei. 2024. Identity-driven hierarchical role-playing agents. *Preprint*, arXiv:2407.19412.

Gemini Team. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *Preprint*, arXiv:2403.05530.

Arjen van Dalen. 2024. Revisiting the algorithms behind the headlines. how journalists respond to professional competition of generative ai. *Journalism Practice*, pages 1–18.

Mohammad Vatani Nezafat. 2024. *Fake News Detection With Retrieval Augmented Generative Artificial Intelligence*. Ph.D. thesis.

Bryan G. Victor, Rebeccah L. Sokol, Lauri Goldkind, and Brian E. Perron. 2023. Recommendations for social work researchers and journal editors on the use of generative ai and large language models. *JOURNAL OF THE SOCIETY FOR SOCIAL WORK AND RESEARCH*, 14(3):563–577.

Yunhee Whang. 2024. Chatgpt for editors: enhancing efficiency and effectiveness. *SCIENCE EDITING*, 11(1):84–90.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. 2024. Qwen2 technical report. *Preprint*, arXiv:2407.10671.

Shengbin Yue, Wei Chen, Siyuan Wang, Bingxuan Li, Chenchen Shen, Shujun Liu, Yuxuan Zhou, Yao Xiao, Song Yun, Xuanjing Huang, and Zhongyu Wei. 2023. Disc-lawllm: Fine-tuning large language models for intelligent legal services. *Preprint*, arXiv:2309.11325.

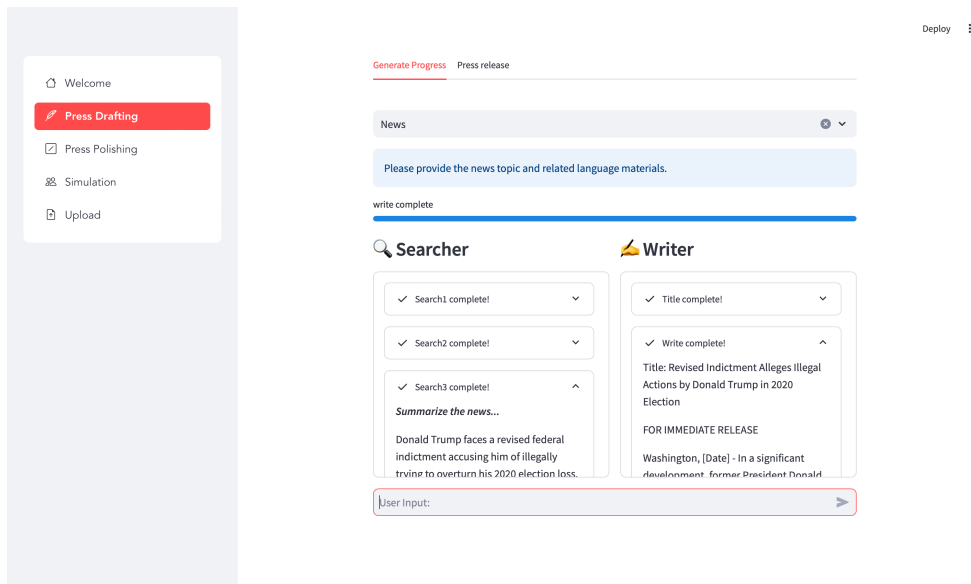70

A   User Interface

A.1   Press Drafting Module

Figure 5: A screenshot displays a sample press-drafting interface, showcasing the output results generated by the searching and writing agents.

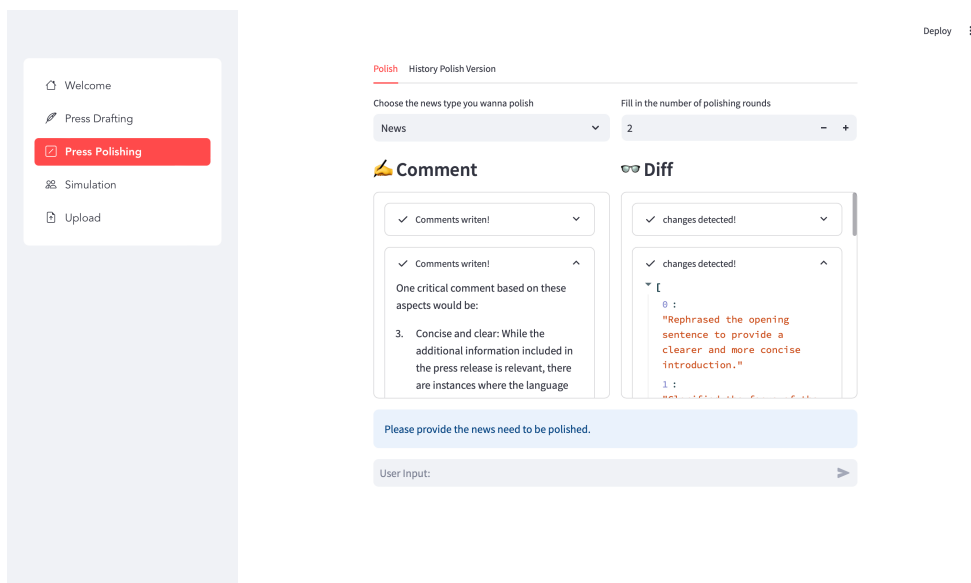## A.2   Press Polishing Module

Figure 6: A screenshot presents a sample press-polishing interface, illustrating the output results produced by the reviewing and rewriting agents.
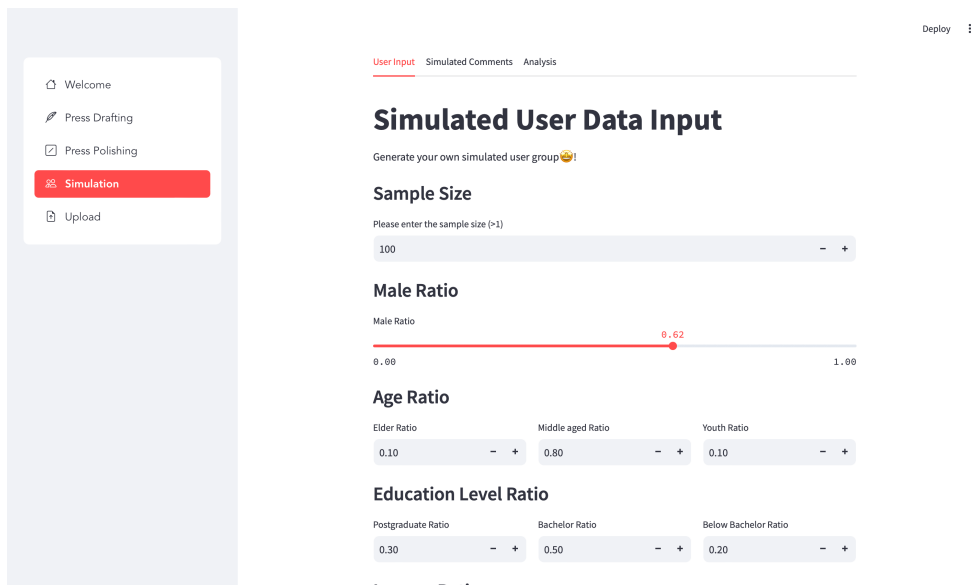
A.3   Simulation Module

Figure 7: A screenshot illustrates the user interface for generating a specific population distribution, utilizing customized demographic indicators.
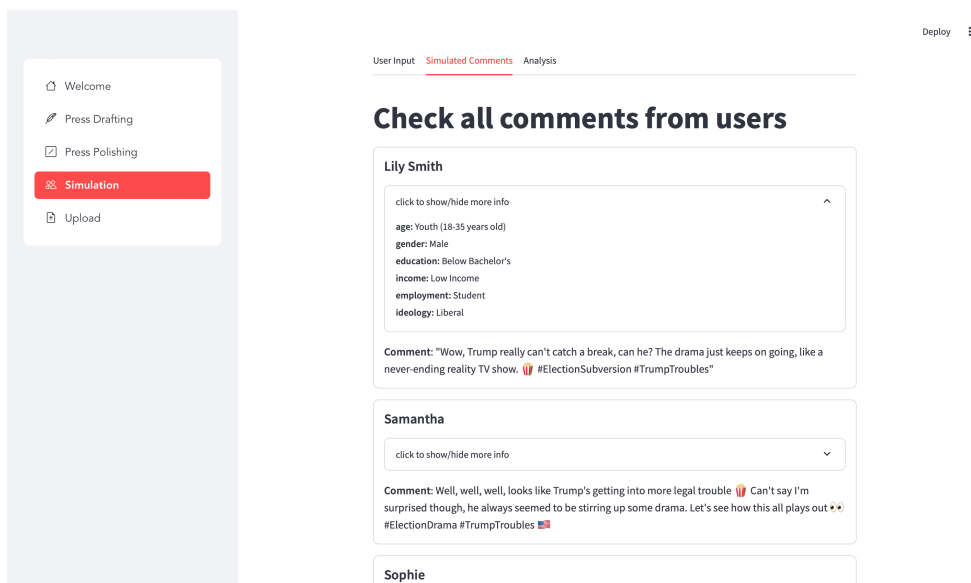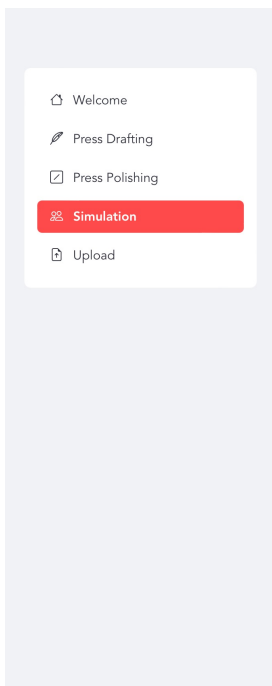

Figure 8: A screenshot presents the generated simulated comments. The specific demographic indicators of the simulated commenters are also displayed.
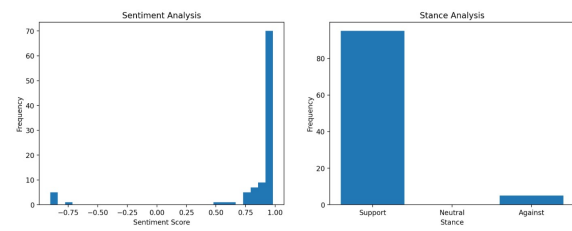
Figure 9: The screenshot shows the analysis interface of simulated feedback, which displays the word cloud of comments and the frequency statistics of sentiments and stances.

B Prompts for LLMs

Genres	Prompts
News	Write a news release with a genre NEWS based on the following content. News reports on current events with timeliness and objectivity. content:{content}.
Profile	Write a news release with a genre of PROFILE based on the following content. Profile journalism offers in-depth looks at people or topics. content:{content}.
Commentary	Write a news release with a genre of COMMENTARY based on the following content. Commentary provides analysis and opinion on current events. content:{content}.

Table 3: Prompts for LLMs. {content} is news corpus.

C Prompts for Agents on Press Drafting Module

C.1 Prompts for Searchers

Searchers	Prompts
Searcher1	{content} Based on the above corpus, extract the core elements of the event, including time, place, key people, etc. Here are the detailed instructions: Identify the exact date and timeframe of the event. Identify the exact location of the incident, including the city, region, or even a specific place. Identify the key people involved, such as the dominant person, the victim, or the relevant authority.
Searcher2	{content} Sort out the passage of time timeline and key plot points based on the provided news corpus. Describe the course of events and record important steps and twists in time sequential order. Extract key episodes and details, e.g. flashpoints, important decisions, or actions. For example, at YYYY-MM-DD, some events happened, etc.

Table 4: Prompts for Searchers. {content} is news corpus, the same as the ones in Table 3.

C.2 Prompts for Writers

Writers	Prompts
Title	{content} Based on the content of UserRequirement, extract the core elements, process, key plot of the event, and the collected background information and impact. Propose 3-5 headlines for the news. Please return the result based on the following JSON structure: [{{"title": str}}].
Content of News	{content} Select the most suitable title from the Title. A good news headline should be accurate, concise, and attractive. Complete the writing of the press release and present a complete, professional, excellent, and directly publishable press release. Requirement: - Refer to the language style, article structure, and narrative techniques of [News Database], maintain objectivity and neutrality, and present the occurrence, development, and results of events in a clear structure. Use concise and clear language, avoiding lengthy and complex sentences as well as obscure vocabulary. - You can refer to the factual basis that may be used in the [Fact Database] (or not) to correct the misinformation in the press release. If there are references, please indicate the source at the end of the news. - Based on the information from UserRequirement and the [Internet Surfer] as the main basis and theme for your writing, use all facts as a benchmark and write according to the format of news reports, including titles, introductions, main body, and endings.
Content of Profile	{content} Select the most suitable title from the Title. A good profile title for a profile should highlight its characteristics and charm. Complete the writing of the character profile and present a complete, vivid, and in-depth close-up of the character. Requirement: - Using information from UserRequirement and the [Internet Surfer] as the main materials, the characters' personalities, achievements, and stories are presented through descriptions of their appearance, personality, behavior, and language. By using detailed descriptions and scene reproduction, readers can feel the true existence and emotional world of the characters. - Referring to the language style, article structure, and narrative techniques of [News Database]. - Referring to factual evidence from the [Fact Database] that can be used to enrich character images, such as their experiences, achievements, and contributions. Language expression should be delicate and emotional, using appropriate adjectives and adverbs to enhance the character's infectiousness and affinity.
Content of Commentary	{content} Select the most suitable title from the Title. A good comment title should introduce the event directly with a viewpoint or appeal. Complete the writing of news commentary and present a complete, professional, and in-depth news commentary manuscript. Requirement: - Based on information from UserRequirement and the [Internet Surfer], conduct an in-depth analysis of the background, causes, and impact of the event. - Propose unique perspectives and analyses. Use logical reasoning and evidence to support viewpoints, avoiding subjective speculation and emotional expression. - Referring to the language style, article structure, and narrative techniques of [News Database], it is appropriate to cite factual evidence from [Fact Database] to support viewpoints and enhance the credibility and persuasiveness of comments. - Language expression should be persuasive and infectious, using vivid vocabulary and vivid metaphors to attract readers' attention and evoke resonance.

Table 5: Prompts for Writers. {content} contains user input, as well as the results of RAG: [News Database], [Fact Database], and [Internet Surfer].

D Prompts for Agents on Press Polishing Module

D.1 Prompts for Reviewers

Genres	Prompts
News	News: {content} Please review the press release and provide critical comments based on these aspects: - Timeliness: Whether the latest and most important information was reported on time? - Accuracy: Whether the factual statements are accurate and reliable, and whether the data and references are reliable? - Concise and clear: Can the core information be conveyed clearly in concise language, avoiding lengthy and complex expressions? - Key emphasis: Have the key elements and focus of the event been clearly identified?
Profile	Profile: {content} Please review the profile and provide critical comments based on these aspects: - Unique perspective: Whether a unique and innovative angle has been chosen to present the theme or characters? - Detail description: Whether it contains vivid and specific details that enable readers to have a strong sensory experience? - Personalization: Can the personality traits of the theme or character be highlighted to distinguish it from other similar individuals?
Commentary	Commentary: {content} Please review the commentary news article and provide critical comments based on these aspects: - Depth and breadth: Whether the topic has been analyzed comprehensively and in-depth, with extensive background information and details presented? - Sufficiency of evidence: Whether sufficient facts, data, cases, and other evidence are provided to support the viewpoints and arguments? - Clarity of opinions: Whether viewpoints are expressed clearly and distinctly, avoiding ambiguous language and ensuring that the stance taken is easily understandable? Whether the main argument is stated upfront and supported by coherent explanations and examples?

Table 6: Prompts for Reviewers. {contents} is the input news draft.

D.2 Prompts for Rewriters

News Draft: {news draft}
Comments: {comments}
You have received the following comments on the news draft. Please revise the news accordingly and provide a revised version of the news.
Just return revised news. Do not explain your reasoning.

Table 7: Prompts for Rewriters.

E Prompts for Simulation Module

You are a {gender} who is {age}, with your highest education {education}. Your income level is {income}, and your current employment status is {employment}. You tend to be {ideology} when making comments.

Today you saw a news article as follows:

news article: {news}
You want to post your own comments in the comment section below the news article.
Your comment doesn't have to be formal. It can be as casual as you want—use slang, emojis, or even a bit of sarcasm.
Please make sure your comment conveys a perspective consistent with your role configuration.

Here is what you have posted in the past: {historical_comment}

You can refer to your past tone and wording habits when posting your comment,
but do not mention the events in your historical comment unless it is highly relevant to the current news.

Reply with your authentic voice:

Table 8: Prompts for comments simulation.

F   User Profile Pool Generating

F.1   User Data

The user data we utilize originates from the Twitter social platform and is completely anonymized. We categorize users based on their publicly available historical tweets, ensuring compliance with all privacy regulations.

F.2   Prompts for User Annotation

You are a user content analyst tasked with determining various attributes of a user based on the content they post.
Your goal is to categorize these attributes into several distinct categories.
Please strictly follow the options provided for selection, and do not return null.

Demographic Attributes:
- Age: Inferred from mentions of life stages, time markers, references to popular culture, etc.
1. Youth (18-35 years old)
2. Middle-aged (36-65 years old)
3. Elderly (over 65 years old)

- Gender: Inferred from self-references, pronouns, interests, lifestyle details, etc.
1. Male
2. Female

- Income: Inferred from mentions of living standards, spending habits, occupation, etc.
1. Low income
2. Middle income
3. High income

- Education: Inferred from vocabulary, sentence structure, mentions of educational background, etc.
1. Below Bachelor's
2. Bachelor's degree
3. Postgraduate education

- Employment: Inferred from descriptions of daily activities, work environment, professional terminology, etc.
1. Working now
2. Student
3. Others

- Ideology: Inferred from political views, party support, election tendencies, etc.
1. Liberal
2. Moderate
3. Conservative

Output in the following JSON format:
{{"age": str, "gender": str, "Income": str, "Education": str, "employment": str, "Ideology": str}.

User Content:
user name: {name}
user posts:{content_list}

Output:

Table 9: Prompts for user annotation.

G Evaluation Experiment

G.1 Prompts for GPT4o Scoring on Press Generating Experiment

Genres	Prompts
News	{content} - For comprehensiveness: Consider whether the news provides a comprehensive overview of the event, including various aspects such as background, development process, key figures involved, and possible consequences. Also, assess if it presents different perspectives and viewpoints related to the event to give readers a more complete understanding. - For depth: Evaluate whether the news goes beyond the surface of the event and delves into deeper issues such as underlying causes, long-term impacts, and potential solutions. Also, check if it provides in-depth analysis and insights through interviews with experts or research on related topics. - For objectivity: Determine if the news report is free from personal biases and subjective views and presents all aspects of the event in a neutral manner. Also, see if it gives equal expression opportunities to different viewpoints and stakeholders without favoring any one side. - For importance: Consider whether the event covered by the news has significant social, political, economic, or cultural significance. Also, assess if the news delves into the reasons and impacts behind the event to provide valuable analysis and thinking. - For readability: Evaluate whether the language of the news is clear, concise, and easy to understand, avoiding overly technical or rare vocabulary and complex sentence structures. Also, check if the structure of the news is reasonable with a clear lead, body, and conclusion and if the logic is coherent.
Profile	{content} - For richness: The profile contains abundant details that vividly depict the person's appearance, habits, and work environment, making the character more three-dimensional. It presents the person from multiple perspectives, including evaluations from family, friends, colleagues, and partners, providing a comprehensive understanding. The story has a clear beginning, development, and end, with a coherent plot. It also includes a variety of elements such as challenges faced, solutions found, and achievements attained to enrich the narrative. - For depth: Thoroughly analyze the person's inner world, motives, and values. Show the person's growth and changes at different stages. Uncover the unknown stories and experiences behind the person to enrich the character's image. - For uniqueness: Approach the person's story from a unique perspective, different from common ways of reporting on people. Highlight the person's individuality and distinctive features. Be able to discover and show the little-known side of the person, bringing freshness to readers. - For inspiration: The person's story can inspire emotional resonance in readers, such as admiration, touch, and inspiration. Let readers gain inspiration and positive energy from the person. Through vivid descriptions and narratives, readers can deeply feel the charm and influence of the person. - For readability: The language expression is smooth, vivid, and easy to understand. Use appropriate description methods and narrative techniques to enhance the attractiveness of the article. Have a reasonable structure, clear hierarchy, and logical clarity for easy reading and understanding.
Commentary	{content} - For comprehensiveness: Cover all aspects of the topic thoroughly, including different viewpoints, potential consequences, and historical background. Incorporate a wide range of sources and examples to provide a holistic understanding. - For clarity of opinions: Express viewpoints clearly and precisely, avoiding ambiguity and vagueness. Use straightforward language and well-structured arguments to make the stance easily discernible. - For sufficiency of evidence: The news has sufficient facts, data, cases, and other evidence to support a viewpoint, making it more persuasive. - For relevance: Be closely related to current events and issues of significance. Address topics that are of interest and concern to the audience. - For readability: The language expression is smooth, vivid, and easy to understand. Use appropriate description methods and narrative techniques to enhance the attractiveness of the article. Have a reasonable structure, clear hierarchy, and logical clarity for easy reading and understanding.

Table 10: Prompts for GPT4o Scoring on Press Generating Experiment. {content} is the press that needs to be scored.

G.2 Experiment Press Release Introduction

Genre	number	fields
News	100	International, technology, art, sports, health, tourism, real estate, fashion, etc
Profile	100	International, domestic, book, sports, film, etc
Commentary	100	International, art, sports, business, technology, etc

Table 11: Experiment Press Release Introduction

G.3 Simulation Experiment Press Release Introduction

Field	No. of comments	Description
Politics	525	Questions regarding Trump's age and capacity
Economy	253	The challenges of offshore wind
Conflict	131	Conflicts between Israel and Hezbollah

Table 12: Simulation Experiment Press Release Introduction

G.4 Prompts for GPT4o on Simulation Experiment

You are a user review analyst responsible for judging the emotional tendency of user comments,
and determining the user's viewpoint and stance based on the provided news articles and user comments.
The viewpoint of the news article is as follows: {news}
The comments posted by users under the news article are as follows:{comment}

Sentiment inclination:
Please determine whether the sentiments expressed in the comments posted by the user are positive, negative, or neutral.
- Positive
- Neutral
- Negative

Sentiment score:
Please provide a score range of [-1,1] based on the comments posted by the user.
The closer the score is to -1, the more negative it is, and the closer the score is to 1, the more positive it is.
A score of 0 indicates neutrality.
The given score is between -1 and 1, with two decimal places retained.

Stance:
Please judge whether the user supports, opposes, or is neutral towards the news viewpoint
based on the comments posted by the user and the viewpoint of the news article.
- Support
- Neutral
- Against

Output in the following JSON format:
{{"Sentiment_inclination": "Positive", "Sentiment_score":0.98, "Stance": "Support"}}
Please only return the JSON string.
Please strictly follow the options provided for selection, and do not return null.

Table 13: Prompts for GPT4o on Simulation Experiment.

H Case Study

H.1 Input Texts

> Published Sept.9,2024 Updated Nov. 6, 2024
> Leer en espanol
> The last time the nation held a debate with the presidency on the line, a candidate with about eight decades of life behind him faced the challenge of proving that he was still up to the job of running the country. He failed.
> Two and a half months later, the cast of characters has shifted and another candidate heading toward the octogenarian club confronts his own test to demonstrate that he has not diminished with ageWhether he passes that test may influence who will be the next occupant of the Oval Office.

H.2 Press Drafting

News Release

Published: Sept. 9, 2024 | Updated: Nov. 6, 2024

Title: Age and Leadership: The Growing Challenge in the Presidential Race

As the 2024 U.S. Presidential Election approaches, the spotlight is once again on the age of the candidates vying for the highest office in the land. The last presidential debate, held in late 2020, saw a candidate in his late 70s struggling to prove his vitality and readiness to lead the country. Despite his decades of political experience, he ultimately fell short in convincing voters that he could meet the demands of the presidency, a role that requires both mental acuity and physical stamina.

Fast forward to 2024, and the situation has shifted. As the country gears up for another election, a new contender, also approaching his 80th birthday, now faces the same daunting challenge. Like his predecessor, this candidate is tasked with demonstrating that his advanced years do not hinder his ability to lead. His performance in the upcoming debates and campaign events will be closely scrutinized, as it could play a pivotal role in determining the next occupant of the Oval Office.

Voter concerns over age-related issues such as mental sharpness, physical endurance, and the ability to adapt to the rapidly changing world are expected to be central themes in the race. Both political parties are carefully navigating these concerns, as the outcome could significantly influence the trajectory of the election.

As the debates approach, analysts and political strategists are keenly aware of the growing impact of age on the national conversation. Whether this new candidate can meet the test of leadership in the eyes of the American people could ultimately decide the future direction of the country.

For further information, please contact: [Your Media Contact Information]

—

Leer en Español

H.3 Press Polishing

News Release

Published: Sept. 9, 2024 | Updated: Nov. 6, 2024

Title: Age and Leadership: The Challenge Facing 2024 Presidential Candidates

As the 2024 U.S. Presidential Election heats up, age once again looms as a key factor in the race. In 2020, a candidate nearing his 80s struggled to prove his mental and physical readiness for the demands of the presidency, and despite decades of experience, he was unable to convince voters that he could lead the nation effectively.

Now, in 2024, another candidate approaching his 80th birthday is set to face the same scrutiny. Like his predecessor, he must demonstrate that age does not diminish his ability to lead the country. With the election just around the corner, his performance in upcoming debates and campaign events could play a crucial role in shaping the next administration.

Voter concerns about mental sharpness, physical stamina, and adaptability in a rapidly evolving world are expected to be central themes in the campaign. Both political parties are acutely aware of the importance of addressing these issues, as they could significantly influence the outcome of the election.

As debates and events unfold, political analysts are closely monitoring how the candidates will handle these age-related challenges. Whether this candidate can prove his leadership abilities in the eyes of the American people may very well determine the future direction of the country.

For further information, please contact: [Your Media Contact Information]

—

Leer en Español

—

Changes Made:

1. Timeliness: Removed the reference to "the last time the nation held a debate with the presidency on the line" and replaced it with a more direct reference to 2020, improving timeliness and relevance.

2. Conciseness: Streamlined sentences for better clarity and flow, especially regarding the previous candidate's failure to convince voters and the challenges facing the new candidate.

3. Key emphasis: Strengthened the focus on age as a central concern for both candidates, emphasizing its potential impact on the election outcome.

4. Clarity: Simplified complex phrases such as "whether he passes that test may influence who will be the next occupant of the Oval Office" to "his performance could play a crucial role in shaping the next administration."

H.4 Simulation

John Harris

> "Age is a factor in any race, but the real test is whether these candidates can handle the mental and physical demands of the presidency. This election will be a true test of their leadership abilities."

Sarah Jenkins, Voter from Ohio

> "I respect their careers, but I'm concerned about whether older candidates can meet the demands of the job. It's about being realistic, not disrespectful."

Carlos Rivera

> "Experience is valuable, but how they perform in debates and public events will be crucial. If they prove their vitality, age won't be an issue."

Emily Thompson

> "As a young person, I worry that older candidates are disconnected from issues that matter to my generation. We need a leader who gets it."

David Mitchell

> "Age shouldn't disqualify anyone. Leadership is about character and decisions, not just how old you are."