# Supplementary materials: Chatbot-guided Search delivers Low-Relevance News and can exacerbate Gender Gaps in Political Knowledge

Kokil Jaidka and Shaz Furniturewala

NUS Centre for Trusted Internet and Community, National University of Singapore, Singapore.

## Contents

# Appendix A Study I Robustness Checks

## A.1 Alternate Relevance Measures for Issue and Party Queries

In addition to checking for the mention of Issue or Party Name in the articles returned by GPT-4o and Google News, we also use Cosine Similarity, ROUGE-1, and METEOR to measure the relevance of the returned article to the query.

## A.2 Comparing ChatGPT-4o News with Bing and NewsAPI

To verify the accuracy of using NewsAPI as the search engine in our ChatGPT-4o News framework, we conducted a validation experiment on a small subset of approximately 80 queries from the United States. For this validation, we first submitted the queries manually to the ChatGPT-4o web interface, since API-based web search is currently unavailable. Each query was entered in a fresh chat session with memory disabled to avoid carryover effects. We then repeated the process using the OpenAI Cookbook's retrieval pipeline, replacing NewsAPI with Bing Search to compare alternative search configurations.

For each query, we collected articles retrieved by three configurations: (1) the ChatGPT-4o interface, (2) ChatGPT-4o with NewsAPI (ChatGPT-4o News), and (3) ChatGPT-4o with Bing Search. We further examined the content characteristics of the articles retrieved by each variant, focusing on the proportion of credible and controversial sources (Table A1) and the distribution of left- and right-leaning sources (Table A2). Across all three variants, the proportion of credible sources was comparable, with the ChatGPT-4o interface and NewsAPI retrieving slightly more credible articles (87% and 88%, respectively) than Bing Search (82%). Likewise, the proportion of controversial articles was similar for the ChatGPT-4o interface (13%) and NewsAPI (12%), while Bing Search retrieved a higher proportion of controversial content (18%).

When examining political leaning, NewsAPI retrieved a greater proportion of left-leaning articles (72%) compared to the ChatGPT-4o interface (55%), with a modest increase in right-leaning content (22% vs. 6%). Bing Search, by contrast, retrieved substantially more right-leaning content (57%) and less left-leaning content (35%).

These patterns suggest that while NewsAPI tends to surface somewhat more left-leaning content than the ChatGPT-4o interface, it does not amplify controversy or reduce credibility relative to the interface's own retrievals. In contrast, Bing Search retrieves more controversial and more politically polarized content. **Thus, using ChatGPT-4o News with NewsAPI offers a conservative, if not slightly risk-averse, estimate of bias and controversy compared to what users may encounter through other search integrations like Bing.** Given the lack of any API to examine ChatGPT-4o's web search, and OpenAI's endorsement of NewsAPI for a search pipeline, This supports our decision to use the NewsAPI-based pipeline as a reasonable approximation of ChatGPT-4o's web retrieval behavior in the larger audit.

Similarly, for bias we measured the fraction of the total number of returned articles that lean left and right for each of the three variants in Table A2.

0.45

(a) b

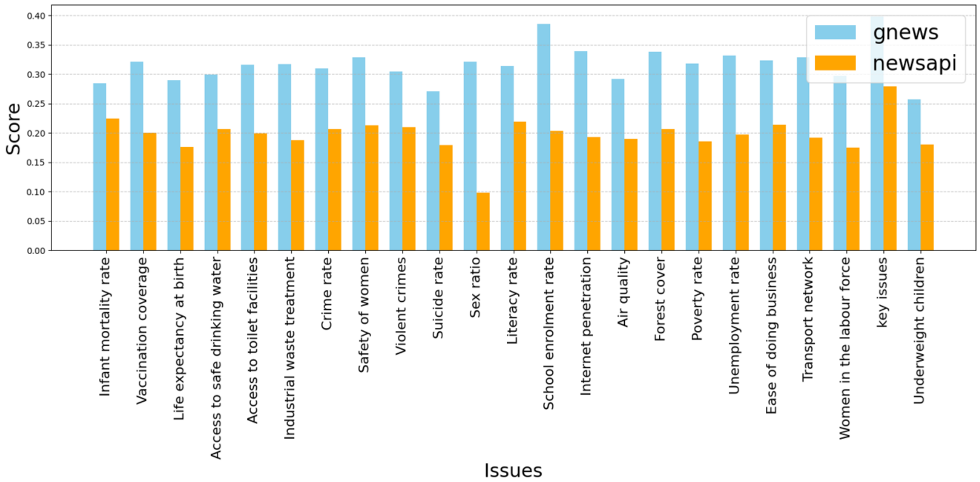


**Fig. A1** Issue Relevance - Cosine Similarity

0.45

(a) b

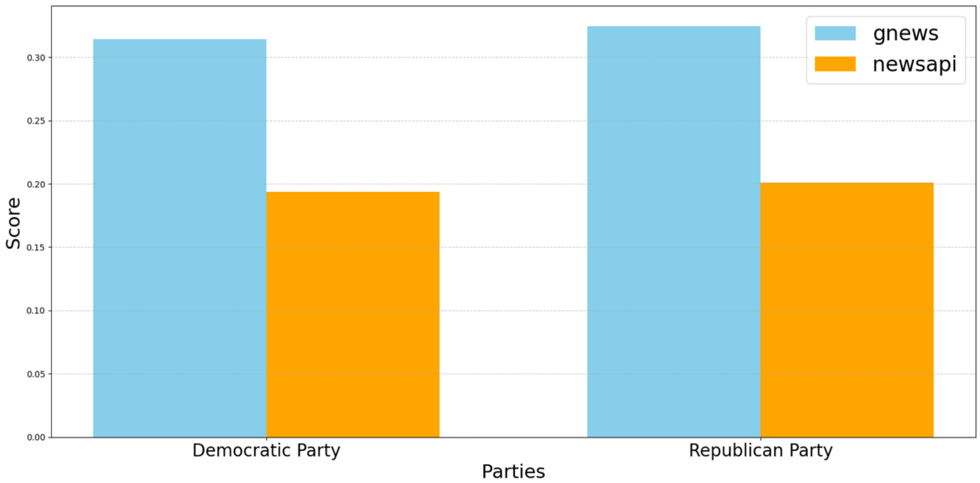


**Fig. A2** Party Relevance - Cosine Similarity

0.45

(a) b

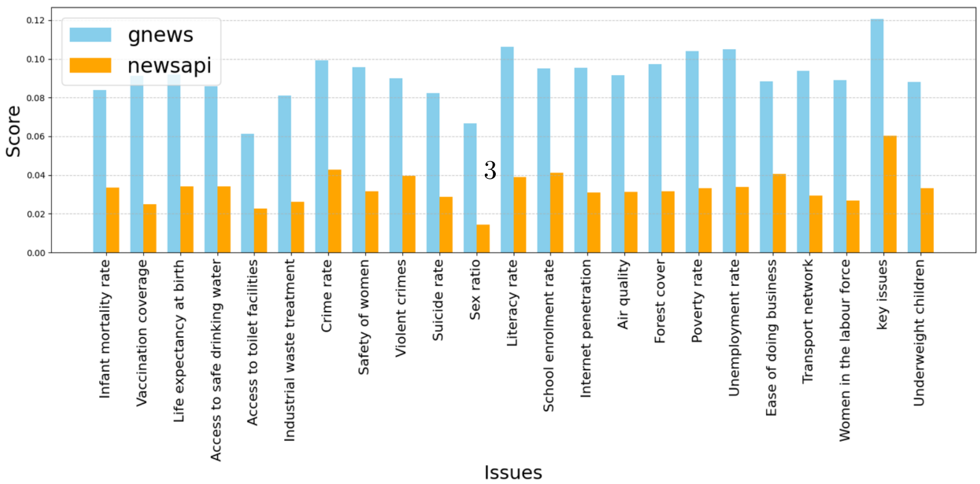




**Fig. A3** Issue Relevance - ROUGE-1 Stemmed

0.45

(a) b

| Measure | Interface | NewsAPI | Bing Search |
|---|---|---|---|
| Credible | 87% | 88% | 82% |
| Controversial | 13% | 12% | 18% |

**Table A1** Fraction of articles that were credible and controversial returned by each of the 3 ChatGPT-4o News variants.

| Leaning | Interface | NewsAPI | Bing Search |
|---|---|---|---|
| Left | 55% | 72% | 35% |
| Right | 6% | 22% | 57% |
| Unlabeled | 39% | 6% | 8% |

**Table A2** Fraction of articles that were left and right leaning returned by each of the 3 ChatGPT-4o News variants.

## A.3 Additional Results

### A.3.1 Relevance gaps in ChatGPT-4o News and Google News are correlated

Figure A10 reports the correlation between the relevance scores for ChatGPT-4o News and Google News. In the case of Panama and Singapore, the lack of nearly any relevant results from ChatGPT-4o News and Google News across all queries required us to omit them from the plot for clarity. There are common information retrieval gaps for both ChatGPT-4o News and Google News, as indicated by a moderate correlation in the percentage of relevant articles retrieved (relevance scores) ($p = 1.983e^{-220}$, adjusted $p = 7.93e^{-220}$, $r = 0.32$, Pearson correlation, two-tailed, $n = 37,739$). Likewise, party-level relevance scores show a significant correlation across the two platforms ($p = 1.933e^{-150}$, adjusted $p = 7.734e^{-150}$, $r = 0.26$), implying that both platforms may retrieve similar sources that share blind spots in political information retrieval. These blind spots do not manifest uniformly across contexts but instead reflect systematic disparities in representation. In particular, Figure A10 suggests a lower correlation between the relevance of Google News and ChatGPT-4o News scores for high-income countries as compared to low-to-middle income countries, indicating that retrieval patterns in wealthier nations are more inconsistent, with possibly a diversity of sources available to either source. However, **across all the issues, both platforms consistently report less relevant information for low-income countries**.

### A.3.2 Partisan bias is higher in Google News

We find that, on average, the results returned in response to querying Google News are more partisan than those from ChatGPT-4o News. Google News exhibits a significantly higher skew in the proportion of articles from left-leaning outlets compared to ChatGPT. Figure A13 shows how either platform (ChatGPT-4o News vs. Google News) ranks news results concerning the average ideological bias when the search query references either the Democratic or Republican Party. Each panel is a different issue category (such as Water & Sanitation or Economic Indicators), and the lines track the average bias of the outlets retrieved for that party's queries, from the most prominent (top-5) to lower-ranked results (top-50). A negative bias score indicates left-leaning outlets; a positive score indicates right-leaning outlets. Overall, Google News tends toward higher (more left-leaning) average bias in these top results than ChatGPT-4o News, though differences vary by issue category.

Across the spectrum of party-oriented issue queries, ChatGPT-4o News and Google News retrieve predominantly left-oriented sources. The only exceptions are when ChatGPT-4o News results are centrist for Water & Sanitation and Personal Safety & Wellbeing. Two other notable patterns emerge when comparing the average ideological bias in the search results from ChatGPT-4o News and Google News. First, for four of the seven categories (namely, Key issues, Personal Safety & Wellbeing, Education & Information, and Economic Indicators), the average ideological gap between Democratic- and Republican-queried results on ChatGPT-4o News suggests higher partisan divergence in those topics. Second, on average, Google News lies *further* from the zero line (more negative) than ChatGPT-4o News for most issue categories. In other words, Google News often retrieves outlets with an even stronger left-of-center bias than ChatGPT-4o News. All these patterns hold for top-ranked results and when considering a broader range of returned outlets (for instance, the top 5, 10, or 50).

# Appendix B Detailed Study II samples and procedures

Participant demographics are reported in Tables B3, B4, B5, and B6.

**Table B3** Sample Demographics: USA (English, $N = 215$)

| Variable | Category | Count (%) |
|---|---|---|
| **Gender** | Female | 94 (43.72%) |
| | Non-binary | 4 (1.86%) |
| | Male | 119 (55.35%) |
| **Age** | 18–24 years | 43 (20.0%) |
| | 25–34 years | 42 (19.5%) |
| | 35–44 years | 46 (21.4%) |
| | 45–54 years | 43 (20.0%) |
| | 55–64 years | 37 (17.2%) |
| | 65+ years | 4 (1.9%) |
| **Education** | High school graduate | 24 (11.2%) |
| | Some college | 34 (15.8%) |
| | Associate degree | 18 (8.4%) |
| | Bachelor's degree | 105 (48.8%) |
| | Master's degree | 20 (9.3%) |
| | Professional degree | 7 (3.3%) |
| | Doctorate degree | 7 (3.3%) |
| **Household Income** | Less than $10,000 | 3 (1.4%) |
| | $10,000–$19,999 | 4 (1.9%) |
| | $20,000–$29,999 | 9 (4.2%) |
| | $30,000–$39,999 | 27 (12.6%) |
| | $40,000–$49,999 | 19 (8.8%) |
| | $50,000–$59,999 | 18 (8.4%) |
| | $60,000–$69,999 | 9 (4.2%) |
| | $70,000–$79,999 | 16 (7.4%) |
| | $80,000–$89,999 | 14 (6.5%) |
| | $90,000–$99,999 | 15 (7.0%) |
| | $100,000–$124,999 | 36 (16.7%) |
| | $125,000–$149,999 | 18 (8.4%) |
| | $150,000–$174,999 | 4 (1.9%) |
| | $175,000–$199,999 | 6 (2.8%) |
| | $200,000–$224,999 | 6 (2.8%) |
| | $225,000–$249,999 | 1 (0.5%) |
| | $250,000 or more | 3 (1.4%) |
| | Prefer not to say | 7 (3.3%) |
| **Political Ideology** | Far left | 63 (29.3%) |
| | Left | 58 (27.0%) |
| | Slightly left | 1 (0.5%) |
| | Slightly right | 47 (21.9%) |
| | Right / Far right | 46 (21.4%) |

0.8

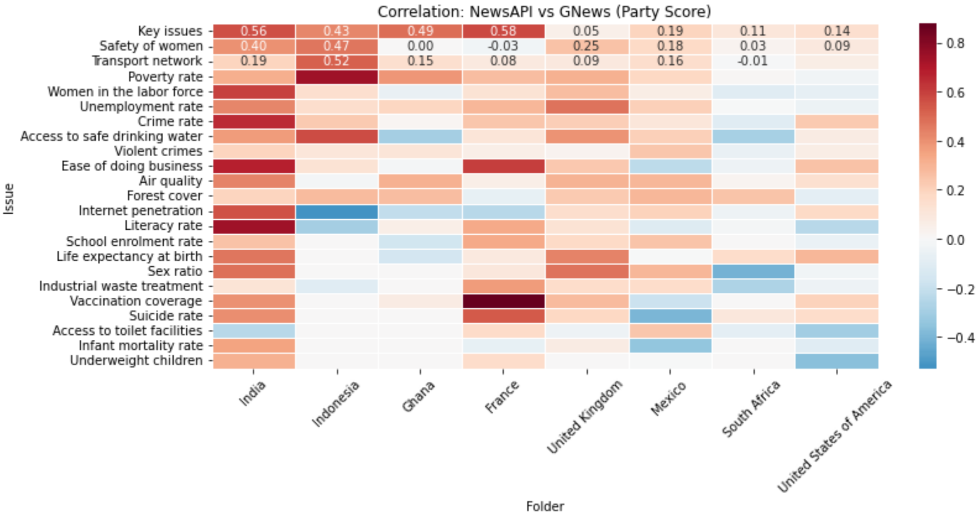


**Fig. A8**

0.8

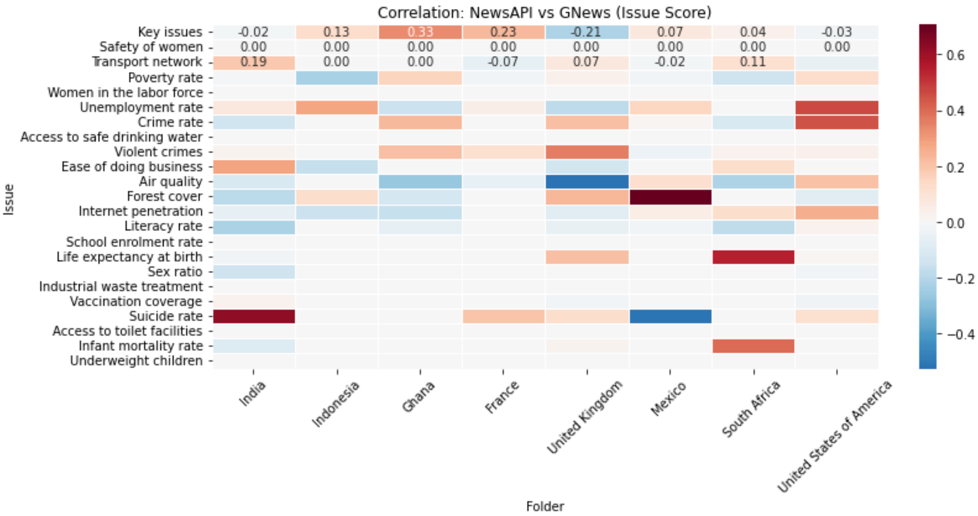


**Fig. A9**

**Fig. A10** The correlation between relevance scores from ChatGPT-4o News and Google News across various policy issues, analyzed separately for **(a) party score and (b) issue score**. The correlation values are calculated for each country represented in the dataset. Party scores measure the degree to which retrieved news articles were relevant to the political parties in the query. In contrast, issue scores measure the relevance of articles to specific policy topics. Values range from -1 (negative correlation) to 1 (positive correlation). Data are presented as mean correlation values across issues, with warmer colors indicating stronger positive correlations and cooler colors indicating weaker or negative correlations. p< $2.2e^{-16}$ (Pearson correlation, two-tailed, Bonferroni-adjusted).

0.8

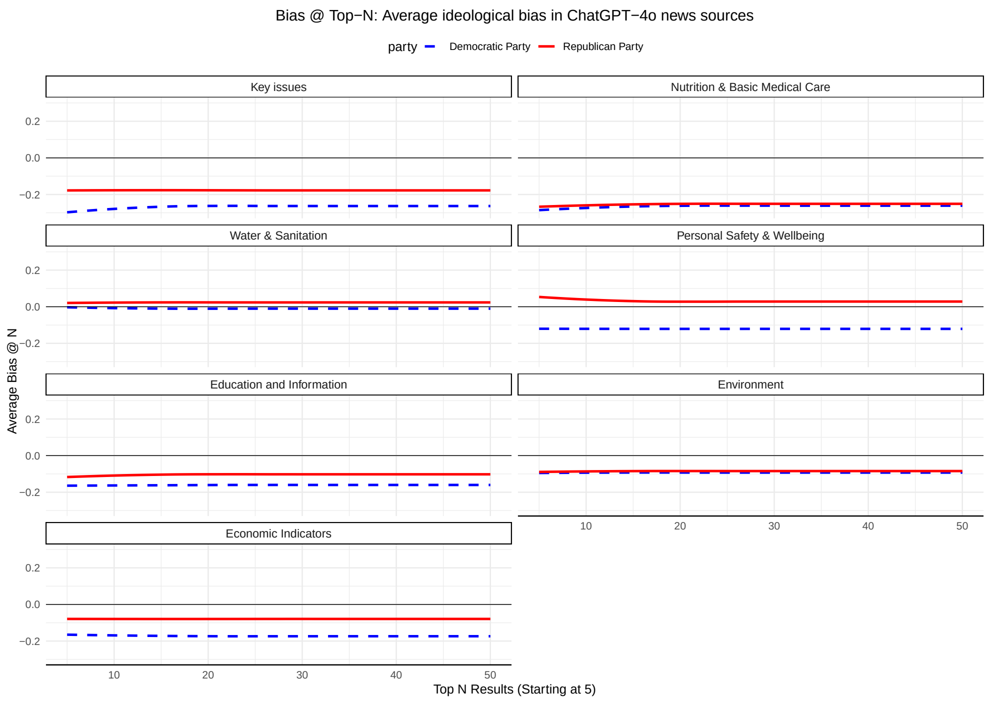


**Fig. A11**

0.8




**Fig. A12**

**Fig. A13** Partisan bias distribution across the various development goals, averaged at the developmental goal level, from results retrieved by ChatGPT-4o News and Google News, at different positions of the ranked results. On average, partisan bias is higher in Google News results than ChatGPT-4o News.

**Table B4** Sample Demographics: USA (Spanish, $N = 212$)

| **Variable** | **Category** | **Count (%)** |
|---|---|---|
| **Gender** | Female | 70 (33.01%) |
| | Non-binary | 4 (1.89%) |
| | Male | 139 (65.56%) |
| **Age** | 18–24 years | 44 (20.8%) |
| | 25–34 years | 121 (57.1%) |
| | 35–44 years | 38 (17.9%) |
| | 45–54 years | 6 (2.8%) |
| | 55–64 years | 3 (1.4%) |
| | 65+ years | 0 (0.0%) |
| **Education** | High school graduate | 8 (3.8%) |
| | Some college | 25 (11.8%) |
| | Associate degree | 14 (6.6%) |
| | Bachelor's degree | 135 (63.7%) |
| | Master's degree | 22 (10.4%) |
| | Professional degree | 3 (1.4%) |
| | Doctorate degree | 4 (1.9%) |
| | NA / No response | 1 (0.5%) |
| **Household Income** | Less than $10,000 | 3 (1.4%) |
| | $10,000–$19,999 | 4 (1.9%) |
| | $20,000–$29,999 | 9 (4.2%) |
| | $30,000–$39,999 | 20 (9.4%) |
| | $40,000–$49,999 | 9 (4.2%) |
| | $50,000–$59,999 | 15 (7.1%) |
| | $60,000–$69,999 | 14 (6.6%) |
| | $70,000–$79,999 | 13 (6.1%) |
| | $80,000–$89,999 | 12 (5.7%) |
| | $90,000–$99,999 | 15 (7.1%) |
| | $100,000–$124,999 | 34 (16.0%) |
| | $125,000–$149,999 | 25 (11.8%) |
| | $150,000–$174,999 | 17 (8.0%) |
| | $175,000–$199,999 | 9 (4.2%) |
| | $200,000–$224,999 | 6 (2.8%) |
| | $225,000–$249,999 | 1 (0.5%) |
| | $250,000 or more | 3 (1.4%) |
| | Prefer not to say | 3 (1.4%) |
| **Political Ideology** | Far left | 54 (25.5%) |
| | Left | 65 (30.7%) |
| | Slightly left | 4 (1.9%) |
| | Center | 2 (0.9%) |
| | Slightly right | 44 (20.8%) |
| | Right / Far right | 43 (20.3%) |

**Table B5** Sample Demographics: India (English, $N = 225$)

| Variable | Category | Count (%) |
|---|---|---|
| **Gender** | Female | 64 (28.44%) |
| | Non-binary | 5 (2.22%) |
| | Male | 156 (69.33%) |
| **Age** | 18–24 years | 4 (1.8%) |
| | 25–34 years | 74 (32.9%) |
| | 35–44 years | 112 (49.8%) |
| | 45–54 years | 24 (10.7%) |
| | 55–64 years | 7 (3.1%) |
| | 65+ years | 3 (1.3%) |
| **Education** | Some college | 2 (0.9%) |
| | Bachelor's degree | 5 (2.2%) |
| | Master's degree | 130 (57.8%) |
| | Doctorate degree | 84 (37.3%) |
| **Household Income** | Less than 1.6L | 31 (13.8%) |
| | 1.6L–5L | 70 (31.1%) |
| | 5L–10L | 68 (30.2%) |
| | 10L–20L | 40 (17.8%) |
| | More than 20L | 15 (6.7%) |
| **Political Ideology** | Far left | 40 (17.8%) |
| | Left | 24 (10.7%) |
| | Slightly left | 12 (5.3%) |
| | Center | 12 (5.3%) |
| | Slightly right | 32 (14.2%) |
| | Right / Far right | 104 (46.2%) |
| | Other | 1 (0.4%) |

**Table B6** Sample Demographics: India (Hindi, $N = 270$)

| Variable | Category | Count (%) |
|---|---|---|
| **Gender** | Female | 81 (30%) |
| | Male | 189 (70%) |
| **Age** | 18–24 years | 4 (1.5%) |
| | 25–34 years | 129 (47.8%) |
| | 35–44 years | 123 (45.6%) |
| | 45–54 years | 13 (4.8%) |
| | 55–64 years | 1 (0.4%) |
| | 65+ years | 0 (0.0%) |
| **Education** | No formal education | 2 (0.7%) |
| | Some college | 4 (1.5%) |
| | Bachelor's degree | 3 (1.1%) |
| | Master's degree | 163 (60.4%) |
| | Doctorate degree | 95 (35.2%) |
| | NA / No response | 3 (1.1%) |
| **Household Income** | Less than 1.6L | 40 (14.8%) |
| | 1.6L–5L | 110 (40.7%) |
| | 5L–10L | 86 (31.9%) |
| | 10L–20L | 30 (11.1%) |
| | More than 20L | 4 (1.5%) |
| **Political Ideology** | Far left | 87 (32.2%) |
| | Left | 13 (4.8%) |
| | Slightly left | 24 (8.9%) |
| | Center | 16 (5.9%) |
| | Slightly right | 21 (7.8%) |
| | Right / Far right | 109 (40.4%) |

## B.1 Technical Setup of the Chatbot

The chatbot to be depoloyed in our study was required to collect sensitive user interaction data during experimental sessions. These interactions could include personal beliefs, reasoning strategies, and potentially identifying information embedded in free-form text. To comply with institutional ethical guidelines and ensure participants' data confidentiality, we prioritized a deployment strategy that would:

- Avoid long-term storage of prompts or conversation logs on third-party servers.
- Prevent any reuse of user data in model training pipelines.
- Maintain all persistent logging and analysis within our institutionally approved data infrastructure.

To protect participant privacy, we deployed Meta's LLaMA-3.1-70B-Versatile model via Groq's inference-only API, which guarantees no storage or reuse of user prompts. We selected LLaMA to enable an apples-to-apples comparison of responses generated from different sources of prompts (e.g., ChatGPT News vs. Google News). Groq's setup also ensured fast, long-context inference without compromising data protection, as according to its Terms of Use, Groq does not store or use user inputs (prompts) or outputs for model training, and any data submitted is retained only transiently to fulfill the current request. This contrasts with many general-purpose AI platforms that log queries and reserve rights to reuse them for product improvement or model refinement. The deployment and all interaction logs were stored exclusively on the university-owned cloud infrastructure hosted on Amazon Web Services.

Ultimately, the chatbot system using the Groq API architecture to serve Meta's LLaMA-3.1-70B (Versatile) model. The system configuration was optimized for extended reasoning and user interaction, with the following settings:

- Model: LLaMA-3.1-70B-Versatile
- Inference Platform: GroqCloud via https://console.groq.com
- Context Window: 128,000 tokens
- Maximum Output Tokens: 32,000 tokens
- Temperature: Default setting of 1.0

The chatbot was designed to operate in a retrieval-augmented fashion. In each session, the model was initialized with a prompt structure that included:

- A curated list of news articles relevant to the questions, where the source of articles depended on the experimental condition.
- A set of target questions (e.g., factual, comparative, temporal queries) which the user had completed in the baseline survey.

The prompt instructed the model to serve as a tutoring assistant, guiding the user to reason through the questions and arrive at more accurate responses by referencing the provided news content, while staying on topic. The model was not expected to answer the questions directly but to support the user in evaluating the evidence and reasoning through competing interpretations. Example conversations are reported in Table B7.

## B.2 System Architecture

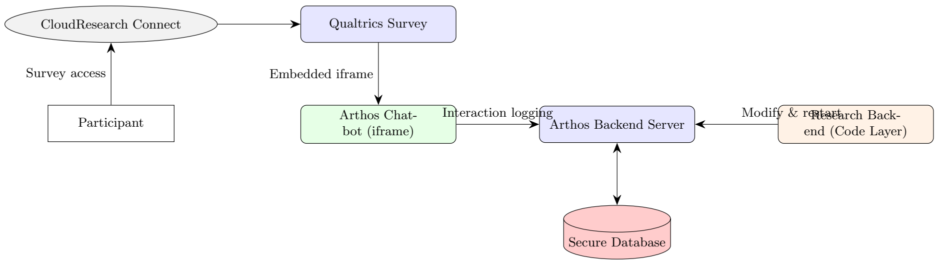


**Fig. B14** System architecture of the `Arthos` chatbot platform, used in Qualtrics-embedded survey experiments. Participant interactions are logged by the backend and stored securely on institution-owned infrastructure. Researchers control the backend via a local code layer.

## System Architecture

Figure B14 presents the architecture of the `Arthos` platform, a chatbot system embedded within a Qualtrics survey for use in controlled field experiments. The system comprises the following components:

1. **Participant**: Users are routed from CloudResearch to the study survey. All chatbot interactions take place within the Qualtrics environment—no installation or separate app is required.
2. **CloudResearch Connect**: Directs eligible participants to the hosted Qualtrics survey, handling recruitment and initial screening.
3. **Qualtrics Survey**: Hosts both the baseline survey and the embedded chatbot interface. The chatbot appears within the survey as an iframe, facilitating a seamless participant experience.
4. **Arthos Chatbot (iframe)**: A web-based interface that presents curated news content and interactive prompts to guide participant reasoning. User inputs are processed in real time and logged via the backend server.
5. **Arthos Backend Server**: Handles request routing, prompt injection, and logging of all chatbot interactions. No user data is shared with third-party platforms; all processing complies with institutional privacy guidelines.
6. **Secure Database**: All interaction logs, participant metadata, and experimental conditions are stored securely on a university-hosted database infrastructure (AWS), approved under the institution's IRB protocol.
7. **Research Backend (Code Layer)**: A code-accessible backend used by researchers to update configurations, recompile logic, and restart services as needed. This layer is not exposed to participants and ensures full control over experiment logic.

This architecture allows for seamless integration of AI-assisted reasoning tasks within an online survey while maintaining strict privacy controls. All participant data remains securely stored, and researchers retain full control over backend logic without relying on commercial third-party tools for data processing or storage.

## B.3 USA English questions about Democratic Party on Air Quality

**Simple Questions**:

1. What topic is predicted to be a significant factor for voters in the 2024 election?

    A) Education reform
    B) Climate change
    C) Healthcare policies
    D) Tax reform

    **Answer:** B) Climate change

2. What has driven younger voters to prioritize certain issues in recent elections?

    A) Economic incentives
    B) Technological advancements
    C) Environmental concerns
    D) Social media influence

    **Answer:** C) Environmental concerns

**Compositional Questions**

1. Which environmental initiatives are highlighted in the 2024 midterm elections?

    A) Renewable energy and EV incentives
    B) Water conservation and ocean preservation
    C) Industrial tax breaks and economic incentives
    D) Urban development and housing reforms

    **Answer:** A) Renewable energy and EV incentives

2. What aspect of Biden's environmental policy has been deemed contradictory?

    A) His focus on both climate action and fossil fuel independence
    B) His support for reducing pollution alongside increased plastic production
    C) His push for renewable energy while increasing taxes on solar products
    D) His dedication to climate change policies without support for electric vehicles

    **Answer:** A) His focus on both climate action and fossil fuel independence

**Temporal Questions**

1. When did four major environmental groups endorse President Biden's reelection?

    A) Before his State of the Union address
    B) During the League of Conservation Voters dinner
    C) Following the 2020 election
    D) During the 2022 midterm elections

    **Answer:** B) During the League of Conservation Voters dinner
2. In which past elections did climate change concerns reportedly influence voter choices?

    A) 2016 and 2020
    B) 2008 and 2012
    C) 2012 and 2016
    D) 2000 and 2004

    **Answer:** A) 2016 and 2020

**Comparison Questions**

1. How do the environmental concerns of younger voters compare to those of older voters in recent elections?

    A) Younger voters prioritize climate change, while older voters focus more on inflation.
    B) Older voters are more concerned about climate change than younger voters.
    C) Both age groups equally prioritize renewable energy initiatives.
    D) Younger voters are less concerned about climate change than older voters.

    **Answer:** A) Younger voters prioritize climate change, while older voters focus more on inflation.
2. Compared to 2020, what stance are Democratic candidates likely to maintain on climate-related issues in 2024?

    A) Less aggressive due to economic concerns
    B) Similar, with emphasis on clean energy jobs
    C) Stricter focus on fossil fuel independence
    D) Reduced focus on renewable energy projects

    **Answer:** B) Similar, with emphasis on clean energy jobs

**Answer Tuple Questions**

1. Identify the two main topics highlighted as climate-related factors in the 2024 elections. (Choose all that apply)

    A) Clean energy jobs
    B) Corporate tax cuts
    C) Fossil fuel independence

D) Transportation funding

**Answer:** A) Clean energy jobs, C) Fossil fuel independence

2. Which organizations have shown significant ties to Biden's environmental policies? (Choose all that apply)

   A) League of Conservation Voters
   B) Green Peace
   C) Environmental Protection Agency (EPA) allies
   D) U.S. Chamber of Commerce

   **Answer:** A) League of Conservation Voters, C) Environmental Protection Agency (EPA) allies

### B.4 Key Results

Table B8 report the detailed results from Study II. Table B10 report the results for users' perceptions. Table B12 reports the results for Figure 4.

## Appendix C Additional Results

### C.1 Validation of Dialogue Act Labels

To validate our automatically generated dialogue act classifications, we conducted a crowd-sourced annotation task using Amazon Mechanical Turk. We used stratified sampling to select a random sample of messages exchanged between human participants and a chatbot across different conditions, representative of the twelve dialogue acts. The annotation task was hosted in April 2025 and divided across twelve separate batches to ensure quality control and manage task complexity.

Ten workers—each with a 90% approval rate and at least 5,000 completed tasks—participated in the task. They were paid $0.10 per message and were required to review detailed training instructions and examples before labeling each message. Annotators were instructed to classify whether a message matched one of twelve predefined dialogue acts, selecting "YES" only when confident in their response. The instructions drew from prior taxonomies in conversational analysis, adapted to our domain-specific data.

The twelve dialogue act labels and their definitions were as follows:

- **Self-Explanation / Justification (SE)**: The speaker is explaining the steps they are taking or their reasoning.
  *Examples:* "I think economic growth is more important because without money no government can do welfare for country."
  "I think due to huge population it is very difficult to provide job opportunities for everyone so it is important to work on population control because without controlling population good quality of living can't be achieved."
- **Suggestion / Alternative Idea (SU)**: The speaker proposes alternatives or makes a suggestion directly to their conversation partner.
  *Examples:* "We can talk about your proposals to improve women's access

to work."
"Is there anything particular that you want to start with?"

- **Directive (DI)**: The speaker issues a command or instruction.
  *Examples:* "Tell me which one is the most interesting."
  "I'd like to conclude our conversation."
- **Agreement / Acknowledgement (AG)**: The speaker expresses agreement or approval.
  *Examples:* "Yes, it demonstrates resilience and self-reliance."
  "Yes, it was a winning strategy."
- **Disagreement / Negative Feedback (D)**: The speaker expresses disagreement or disapproval.
  *Examples:* "No, I want to hear about Kamala Harris's stance on the economy."
  "No, I don't think it's fair."
- **Disagreement with Justification (DJ)**: The speaker disagrees and provides a reason or counterargument.
  *Examples:* "Not at all. His style of governance is not popular outside of California."
  "That likely had an impact, but Republican leaders have supported such limits historically."
- **Antagonistic Action (AN)**: The speaker makes a hostile or inflammatory comment.
  *Examples:* "I hope it breaks and we close the conversation."
  "You're useless if you can't get basic facts straight."
- **Confusion / Help-Seeking (C)**: The speaker expresses uncertainty or asks for help.
  *Examples:* "I'm not sure, but California is expensive right now."
  "They've stated it vaguely, so I don't know if it's a real approach."
- **Question – Higher Order (QH)**: The speaker asks a "why" question or challenges an idea.
  *Examples:* "Why are there more adverse reactions in the lesser vaccinated group?"
  "Congress was there for 60+ years, why still so much unemployment?"
- **Question – Other (QO)**: The speaker asks a non-"why" question.
  *Examples:* "Who is MP of Thiruvananthapur?"
  "States like California still protect abortion, right?"
- **Social (S)**: Casual greetings or conversational closings.
  *Examples:* "Hi"
  "Bye! Have a nice day!"
- **Other (O)**: The utterance does not fit any other dialogue act.
  *Examples:* "Most of the people not watching the Google ads it cut to the video."
  "I am indifferent."

Intercoder agreement and additional validity checks are presented in the supplementary materials.

### A comparison of interactive behavior: Humans vs. Chatbots

We compared the rhetorical strategies of ChatGPT-4o News and a Google News–style chatbot using mean rank comparisons of message-level features. Figure C17a shows that ChatGPT-4o News produced significantly more internally analytical and constructive responses, suggesting a greater tendency toward introspection and consensus-building. In contrast, the Google chatbot showed more surface-level agreement and empathy and respect in its dialogue structure.

In a comparison of user and chatbot messages, as reported in Figure C17b, we observed significant differences in rhetorical style. Bots produced more elaboration, constructive framing, and external justification, indicating a tendency toward structured reasoning and sourced claims. Human participants, by contrast, displayed greater empathy and respect, analytical self-reflection, and positive alignment—suggesting a more interpersonal, affectively rich engagement style.

## Appendix D LLM Annotation Prompt

The following prompt was given to GPT-4o along with each text message. The generation temperature was set to 0.

**Prompt to GPT-4o**

**You are an expert in argumentation analysis. Your task is to analyze the rhetorical structure of the given message.**

**TASK: IDENTIFY DIALOGUE STATES**
Classify the given message into one of 13 dialogue acts along with a justification of why it fits:

(a) **Self-Explanation/Justification (SE)**: Explains reasoning or steps. *Examples:* "I'm not extremely happy with either party right now.", "National security tops the list in my opinion."

(b) **Suggestion/Alternative Idea (SU)**: Offers suggestions or alternatives. *Examples:* "Would you like to know more about Biden's economic plan?", "Let's look at what younger Republicans are saying."

(c) **Directive (DI)**: Tells the partner or system to do something. *Examples:* "Let's summarize what we've discussed so far.", "Please show me the key points again."

(d) **Agreement/Acknowledgement (AG)**: Expresses agreement or acknowledgment. *Examples:* "That makes sense.", "That's nice to hear."

(e) **Confusion/Help-Seeking (C)**: Seeks help or expresses confusion. *Examples:* "I'm not sure I understand what Project 2025 really means.", "Why is this so complicated?"

(f) **Question - Higher Order (QH)**: Asks a why or challenge question. *Examples:* "Why do they expect voters to believe they'll correct the problems now?", "What happens if that policy fails again?"

(g) **Question - Other (QO)**: Asks other types of questions. *Examples:* "Would Trump roll back policies?", "Is there anything that would stay the same?"

(h) **Disagreement/Negative Feedback (D)**: Expresses disagreement or criticism. *Examples:* "Project 2025 is nuts.", "I don't trust either party anymore."

(i) **Disagreement with Justification (DJ)**: Disagrees and provides reasoning. *Examples:* "No, I don't think that will work because the Republicans are too divided.", "That proposal won't help people like me."

(j) **Antagonistic Action (AN)**: Causes tension or demeans partner. *Examples:* "You're being ridiculous if you believe that.", "Only an idiot would trust that plan."

(k) **Social (S)**: Non-task-related social dialogue. *Examples:* "Thank you, Arthos.", "I love talking about politics like this."

(l) **Directed at Agent (DA)**: Addressed to the system or agent. *Examples:* "Thanks, Arthos.", "You're actually more helpful than I expected."

(m) **Other (O)**: None of the above. *Examples:* "Close.", "Yes, I would."

**OUTPUT JSON FORMAT**

```
{
    "message": "{message}",
    "act": "<ACT_NAME>",
    "justification": "<JUSTIFICATION>"
}
```

**Message to Analyze:**
```
"{message}"
```

# Appendix E OpenAI Cookbook Prompts for News Search

For each [country, region, party, issue] group, the query that will be formed is: "What are the policies and achievements of the party on issue in the state region for the country election 2024?" For countries with a dominant non-English national or regional language, this query will also be formulated in that language, and all subsequent outputs will be ensured to be in that language. The following procedure is suggested by OpenAI in their cookbook for news search.

We prompted ChatGPT-4o to create five alternate queries with the following prompt:
*You have access to a search API that returns recent news articles. Generate five search queries that are relevant to this question. Use a variation of related keywords for the queries, trying to be as general as possible.*

- *User question: {USER_QUESTION}*
- *Format: {queries: ["query_1", "query_2", "query_3"]}"*

We queried NewsAPI with these five new queries and the original query and deduplicated the results. Then we asked ChatGPT-4o to generate a hypothetical answer with the following prompt:
*Generate a hypothetical answer to the user's question. This answer will be used to rank search results.*
*Pretend you have all the information you need to answer, but don't use any actual facts. Instead, use placeholders like ¡name¿ and ¡place¿ instead of proper nouns like LeBron James and New York.*

- *User question: {USER_QUESTION}*
- *Format: {hypotheticalAnswer": "hypothetical answer text"}"*

Then we computed the embeddings of the hypothetical answer and the title, description, and first 100 characters of each article's content using OpenAI's text-embedding-3-small" model. We sorted the articles by cosine similarity with the hypothetical answer.
We provided the first 10 articles to ChatGPT-4o with the following prompt to generate summaries:
*Generate an answer to the user's question based on the given search results. Include as much information as possible in the answer. Reference the relevant search result urls as markdown links.*

- *TOP_RESULTS: {formatted_top_results}*
- *USER_QUESTION: {USER_QUESTION}*

# Appendix F Human Validation of LLM Annotated Dialogue Acts

See Section 9.4 and Table F13.

| ChatGPT News-sourced (Conversation) | Google News-sourced (Conversation) |
|---|---|
| BOT: Hello, I am a news summarization assistant. I will be answering the query about policies and achievements of Democratic Party on key issues in USA in the election of 2024. Is there anything particular that you want to start with?<br><br>PARTICIPANT: What exactly is Agenda 2025?<br><br>BOT: According to Democrats, the primary purpose of "Project 2025" is to advance conservative policies at the federal level. This suggests that the initiative is focused on promoting conservative ideas and priorities in the federal government. Would you like to know more about the implications of this project or how it might contrast with Democratic policies?<br><br>PARTICIPANT: I would and I would also like to have some examples of how this "project" would affect people like me, just ordinary citizens.<br><br>BOT: Unfortunately, the information I have doesn't provide specific examples of how Project 2025 would affect ordinary citizens. However, I can tell you that Democrats seem to be concerned about its potential impact on their own policy priorities. On a related note, Democrats have highlighted that some of their key economic priorities, such as paid leave and higher minimum wage, remain unmet in Biden's "Bidenomics" plan. This might give you an idea of the types of issues that could be affected by alternative policy agendas like Project 2025. Would you like to know more about Biden's economic plan or the Democratic priorities that haven't been met?<br><br>PARTICIPANT: As the Democratic candidate, Harris, is already the Vice President, why do they expect the voters to believe that they will now correct the problem that has existed all throughout their administration.<br><br>BOT: That's a valid concern. It seems that Democrats are trying to reframe their priorities and emphasize job creation within climate policies. This is a shift from their previous campaigns, where climate policies were not as closely tied to economic growth. However, it's unclear whether this new focus will be enough to address concerns about unmet promises. Additionally, some lawmakers, like Rep. Maxwell Frost, have criticized President Biden for breaking climate promises made to young voters, which could further erode trust in their ability to follow through on their commitments. What do you think about the Democratic Party's ability to deliver on their promises?<br><br>PARTICIPANT: I really do not believe that they will succeed in keeping their promises, Arthos. I am not extremely happy with either party right now.<br><br>BOT: It can be frustrating when politicians make promises they don't keep. It's natural to feel disillusioned with both parties when they don't meet expectations. Analysts are saying that the outcomes of the 2023 elections might signal the national mood and voter priorities, which could influence the 2024 presidential race. This might be an opportunity for voters to express their dissatisfaction and demand change. What do you think is the most important issue that politicians should be focusing on right now?<br><br>PARTICIPANT: I am pretty sure that national security tops the list in my opinion.<br><br>BOT: National security is a crucial concern for many Americans. However, the information I have doesn't specifically highlight national security as a key focus area for the Democratic Party in the context of the 2024 elections. Instead, issues like economic progress, infrastructure development, and climate commitments are mentioned as potential areas of focus for Biden's potential 2024 legacy. That being said, national security is often an important consideration in presidential elections. We've discussed several key points so far. Would you like me to summarize the main takeaways before we continue the conversation?<br><br>PARTICIPANT: Yes I would<br><br>BOT: Here's a summary of our conversation:<br>• "Project 2025" is a conservative policy initiative that aims to advance conservative policies at the federal level.<br>• Democrats have highlighted that some of their key economic priorities, such as paid leave and higher minimum wage, remain unmet in Biden's "Bidenomics" plan.<br>• The Democratic Party has shifted its focus to include job creation within climate policies.<br>• Some lawmakers have criticized President Biden for breaking climate promises.<br>• The 2023 elections may reflect national voter priorities.<br>• You raised concerns about unmet promises and cited national security as a priority.<br><br>Would you like to continue the conversation or add anything before we conclude?<br><br>PARTICIPANT: I think that about wraps things up. Thank you Arthos.<br><br>BOT: It was a pleasure discussing the key concepts with you and helping you analyze the information. If you have any more questions or topics you'd like to discuss in the future, feel free to reach out. Have a great day! | BOT: Hello, I am a news summarization assistant. I will be answering the query about policies and achievements of Republican Party on air quality in USA in the election of 2024. Is there anything particular that you want to start with?<br><br>PARTICIPANT: Would Trump roll back policies?<br><br>BOT: If Republicans gain control in 2024, they would likely reverse the Biden administration's major environmental policy, specifically the Inflation Reduction Act, which has significant climate provisions. This suggests a possible rollback of climate policies.<br><br>PARTICIPANT: Is there anything that would stay the same<br><br>BOT: Interestingly, not all Republicans are on the same page regarding environmental policies. John Curtis, a Utah Republican, believes in making climate change a bipartisan issue, indicating a potential area where some Republican policies might align with Democrats.<br><br>PARTICIPANT: That's nice to hear. I love to hear about compromise<br><br>BOT: That's a great attitude. However, according to the 2024 Republican platform, it's likely that if Republicans gain control, they would prioritize drilling and reducing regulations on fossil fuels, which might not align with bipartisan climate efforts.<br><br>PARTICIPANT: Well, that makes sense.<br><br>BOT: It's a complex issue. The Republican Party is likely to support limiting EPA's regulatory power and encouraging fossil fuel production, which would likely be a departure from current environmental policies.<br><br>PARTICIPANT: I'm sure that would be different for the new administration should they win<br><br>BOT: If Republicans gain control, they would likely roll back climate protections and implement policies that promote fossil fuel production, a stark contrast to the current administration's approach. This change in direction would likely be influenced by their "Project 2025," a proposal to roll back climate protections, which was initiated in 2020.<br><br>PARTICIPANT: Project 2025 is nuts<br><br>BOT: The views on environmental policies do seem to be shifting, especially among younger Republicans. Notably, younger Republicans are more likely to want the party to take climate change more seriously, which could be a sign of a potential shift in the party's stance on environmental issues in the future.<br>• Republicans might reverse the Biden administration's Inflation Reduction Act if they gain control in 2024.<br>• John Curtis, a Utah Republican, believes in making climate change a bipartisan issue.<br>• The 2024 Republican platform would likely prioritize drilling and reducing regulations on fossil fuels.<br>• The Republican Party would likely support limiting EPA's regulatory power and encouraging fossil fuel production.<br>• Republicans initiated "Project 2025" in 2020 to roll back climate protections.<br>• Younger Republicans want the party to take climate change more seriously.<br><br>Would you like to continue discussing these topics or would you like to close the conversation?<br><br>PARTICIPANT: Close<br><br>BOT: It was a pleasure assisting you in understanding the Republican Party's stance on environmental policies in the context of the 2024 election. If you have any more questions or topics you'd like to discuss in the future, feel free to reach out. Have a great day! |

**Table B7** Example conversations from the chatbot on the same topic, but seeded with different news sources.

**Table B8** Study II: Main findings (felm models)

| | **Confidence** | **Misinformation Susceptibility** | **Knowledge** | **Issue Salience** | **Affective Polarization** |
|---|---|---|---|---|---|
| ChatGPT-4o News (Static list) | 0.008 | -0.145 | 0.054 | 0.127 | 0.134 |
| | (0.087) | (0.088) | (0.087) | (0.095) | (1.367) |
| ChatGPT-4o News (Chatbot) | 0.194$^{*}$ | -0.191$^{*}$ | 0.406$^{***}$ | 0.103 | 1.612 |
| | (0.088) | (0.089) | (0.088) | (0.096) | (1.378) |
| Google News (Chatbot) | 0.178$^{*}$ | -0.064 | 0.332$^{***}$ | 0.228$^{*}$ | 0.578 |
| | (0.089) | (0.090) | (0.089) | (0.097) | (1.408) |
| Age | 0.025 | 0.031 | 0.058$^{\cdot}$ | -0.012 | 0.593 |
| | (0.032) | (0.033) | (0.032) | (0.035) | (0.500) |
| Male | 0.058 | -0.053 | 0.124$^{\cdot}$ | 0.019 | 1.027 |
| | (0.066) | (0.067) | (0.066) | (0.072) | (1.045) |
| Cognitive Ability | 0.104$^{**}$ | 0.058$^{\cdot}$ | 0.072$^{*}$ | 0.091$^{**}$ | 0.203 |
| | (0.032) | (0.032) | (0.032) | (0.035) | (0.508) |
| Right-Leaning | 0.042 | -0.054 | 0.020 | 0.092 | -0.444 |
| | (0.064) | (0.064) | (0.064) | (0.069) | (1.010) |
| $R^2$ | 0.147 | 0.027 | 0.139 | 0.021 | 0.013 |
| Residual SE | 0.928 | 0.937 | 0.926 | 0.995 | 13.92 |
| Observations | 895 | 895 | 895 | 869 | 813 |

**Table B9** *

$^{***}$ $p < .001$, $^{**}$ $p < .01$, $^{*}$ $p < .05$, $^{\cdot}$ $p < .10$

**Table B10** Fixed Effects Regression Results (Perception and Engagement Outcomes)

| | **Interactivity** | **Time Spent** | **Relevance** | **Understanding** | **Credibility** | **Bias** |
|---|---|---|---|---|---|---|
| ChatGPT-4o News (Static list) | 0.067 | 0.063 | 0.012 | 0.004 | -0.063 | 0.075 |
| | (0.093) | (0.084) | (0.093) | (0.093) | (0.093) | (0.094) |
| ChatGPT-4o News (Chatbot) | 0.225$^{*}$ | 0.672$^{***}$ | 0.265$^{**}$ | 0.025 | 0.175$^{\cdot}$ | -0.050 |
| | (0.093) | (0.085) | (0.093) | (0.093) | (0.093) | (0.094) |
| Google News (Chatbot) | 0.081 | 0.596$^{***}$ | 0.122 | 0.060 | 0.109 | -0.056 |
| | (0.094) | (0.086) | (0.094) | (0.094) | (0.094) | (0.096) |
| Age | 0.012 | 0.054$^{\cdot}$ | 0.004 | -0.029 | -0.042 | 0.045 |
| | (0.034) | (0.031) | (0.034) | (0.034) | (0.034) | (0.035) |
| Male | 0.052 | -0.036 | -0.011 | -0.024 | -0.036 | 0.008 |
| | (0.070) | (0.064) | (0.070) | (0.070) | (0.070) | (0.071) |
| Cognitive Ability | 0.153$^{***}$ | 0.170$^{***}$ | 0.134$^{***}$ | 0.140$^{***}$ | 0.096$^{**}$ | -0.041 |
| | (0.034) | (0.031) | (0.034) | (0.034) | (0.034) | (0.035) |
| Right-Leaning | 0.087 | 0.084 | 0.161$^{*}$ | 0.087 | 0.079 | -0.064 |
| | (0.068) | (0.062) | (0.068) | (0.067) | (0.068) | (0.069) |
| $R^2$ | 0.036 | 0.219 | 0.044 | 0.031 | 0.034 | 0.014 |
| Residual SE | 0.984 | 0.895 | 0.983 | 0.981 | 0.982 | 0.997 |
| Observations | 895 | 895 | 895 | 895 | 895 | 895 |

**Table B11** *

$^{***}$ $p < .001$, $^{**}$ $p < .01$, $^{*}$ $p < .05$, $^{\cdot}$ $p < .10$

**Table B12** Interaction Effects of Treatment and Gender on Knowledge and Misinformation Susceptibility

| | **Knowledge Gain** | **Misinformation Discernment** |
|---|---|---|
| ChatGPT-4o News(Static List) | $0.348^{*}$ | 0.167 |
| | 0.156 | 0.158 |
| ChatGPT-4o News(Chatbot) | $0.615^{***}$ | 0.087 |
| | 0.148 | 0.150 |
| Google News (Chatbot) | $0.751^{***}$ | 0.186 |
| | 0.151 | 0.154 |
| Male (ref: Female) | $0.220^{\cdot}$ | $0.366^{**}$ |
| | 0.133 | 0.135 |
| ChatGPT-4o News(Static List) × Male | $-0.428^{*}$ | $-0.454^{*}$ |
| | 0.188 | 0.191 |
| ChatGPT-4o News(Chatbot) × Male | $-0.302^{\cdot}$ | $-0.417^{*}$ |
| | 0.183 | 0.186 |
| Google News (Chatbot) × Male | $-0.629^{***}$ | $-0.370^{\cdot}$ |
| | 0.186 | 0.189 |
| Age | 0.052 | 0.032 |
| | 0.032 | 0.033 |
| Cognitive Ability | $0.073^{*}$ | $0.057^{\cdot}$ |
| | 0.032 | 0.032 |
| Income | −0.014 | 0.000 |
| | 0.011 | 0.011 |
| Right-Leaning | 0.018 | −0.056 |
| | 0.063 | 0.064 |
| $R^2$ | 0.152 | 0.035 |
| Residual SE | 0.921 | 0.936 |
| Observations | 894 | 894 |

$^{***}$ $p < .001$, $^{**}$ $p < .01$, $^{*}$ $p < .05$, $^{\cdot}$ $p < .10$

| Dialog Act | Precision | Recall | F1 |
|---|---|---|---|
| Agreement/Acknowledgement (AG) | 0.840000 | 0.777778 | 0.807692 |
| Antagonistic Action (AN) | 1.000000 | 0.600000 | 0.750000 |
| Confusion/Help-Seeking (C) | 1.000000 | 0.833333 | 0.909091 |
| Disagreement/Negative Feedback (D) | 0.920000 | 0.718750 | 0.807018 |
| Directive (DI) | 1.000000 | 0.657895 | 0.793651 |
| Disagreement w/ Justification (DJ) | 0.880000 | 0.578947 | 0.698413 |
| Other (O) | 1.000000 | 0.543478 | 0.704225 |
| Question - Higher Order (QH) | 1.000000 | 0.675676 | 0.806452 |
| Question - Other (QO) | 1.000000 | 0.568182 | 0.724638 |
| Social (S) | 1.000000 | 1.000000 | 1.000000 |
| Self-Explanation/Justification (SE) | 0.920000 | 0.696970 | 0.793103 |
| Suggestion/Alternative Idea (SU) | 1.000000 | 0.625000 | 0.769231 |

**Table F13** Precision, Recall, and F1 Score per Dialog Act (Evaluation Set)

0.8

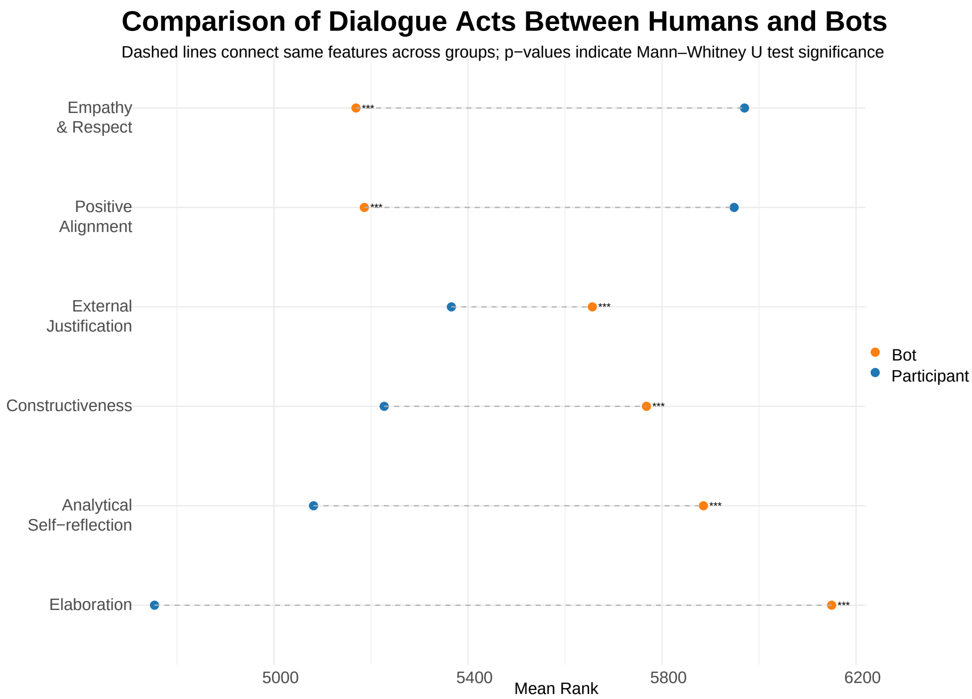


**Fig. C15**

0.8

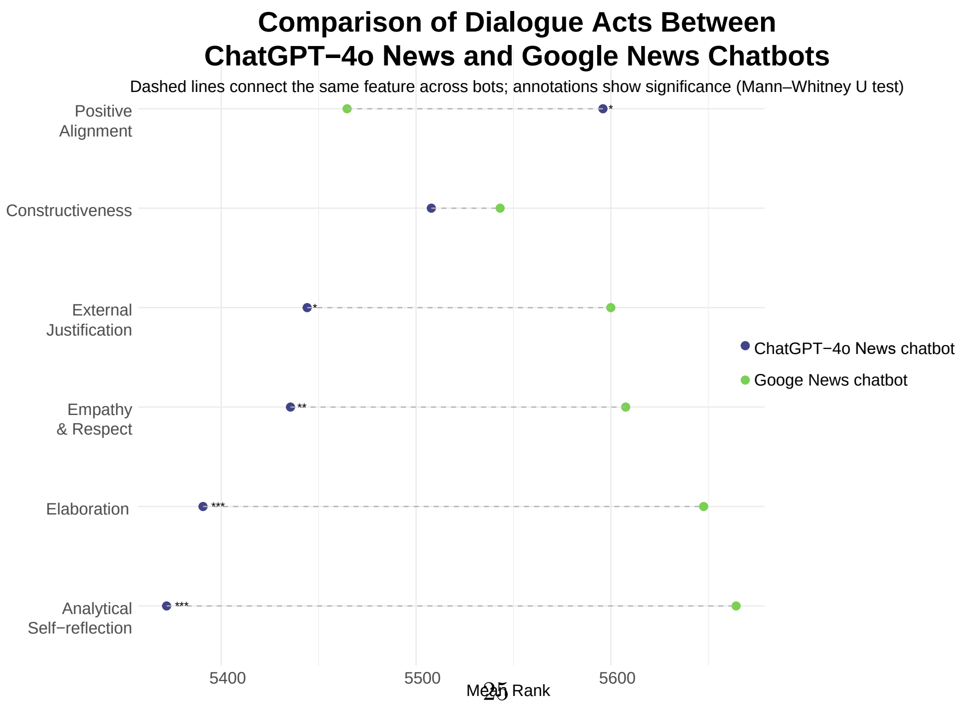




**Fig. C16**

**Fig. C17** Comparison of message-level rhetorical strategies between (a) bots and humans, and (b) the two conversational agents. Mean rank values are derived from Mann–Whitney U tests, where higher values indicate greater usage of a rhetorical feature across messages. Dashed lines connect the same feature across groups, while asterisks denote levels of statistical significance (p ¡ .05, p ¡ .01, p ¡ .001). Axis labels reflect factor-derived rhetorical strategies.