

SCOPE: Source Credibility and Outlet Profiling Engine

Ilyas Nagiyev^{1*}, Miras Baisbay^{1*}, Daniil Orel^{2*}, Preslav Nakov²

¹Nazarbayev University, Astana, Kazakhstan

²MBZUAI, Abu Dhabi, United Arab Emirates

{ilyas.nagiyev, miras.baisbay}@nu.edu.kz
{daniil.orel, preslav.nakov}@mbzuai.ac.ae

Abstract

We present SCOPE, a multi-agent system for profiling news media outlets with respect to bias, factuality, and credibility. Given an outlet URL, SCOPE coordinates specialized agents for article acquisition, content analysis, external evidence search, scoring, and report generation. The system follows an MBFC-style methodology while making intermediate evidence and agent outputs inspectable through a structured demonstration interface. On a benchmark of 200 media outlets, SCOPE outperforms LLM-only, article-only, search-only, and media-background-check baselines on most metrics, achieving the lowest bias and factuality MAE, perfect failed-fact-check detection, and the lowest factual error rate. These results suggest that decomposed, evidence-grounded agent workflows can support more transparent and reliable media profiling. Our code is available at <https://github.com/mbaisbay/SCOPE>, and a live demonstration at <https://scope-demo.streamlit.app/>. Our video demonstration is available via the URL (clickable).

1 Introduction

The rapid expansion of online media has created an increasingly complex and fragmented information landscape, making it difficult for readers, researchers, journalists, and policy analysts to systematically evaluate the reliability, political leaning, and broader influence of news sources at scale.

While much work on misinformation focuses on verifying individual claims (Xie et al., 2025; Wang et al., 2025), detecting misleading articles, or evaluating the manipulation techniques applied by media outlets (Sajwani et al., 2024; Da San Martino et al., 2020b), other work has emphasized outlet-level profiling, such as estimating factuality, political ideology, framing, stance, and propagandistic content (Zhang et al., 2019).

Yet many real-world decisions require understanding the media outlet itself, including its topical focus, ownership, citation practices, audience positioning, and past behavior.

Media profiling is particularly valuable when claims are new, ambiguous, or hard to verify directly. In such cases, an outlet’s long-term patterns offer useful context for interpreting its content. Building reliable media profiles remains challenging because relevant evidence is distributed across many sources, including outlet articles, fact-checking platforms, social media accounts, public records, and independent third-party reports.

A single model may struggle to collect relevant evidence, evaluate its reliability, reconcile conflicting signals, and produce a transparent profile that users can inspect. Media profiling is not merely a classification task, but a multi-stage analytical process involving retrieval, extraction, comparison, reasoning, and explanation.

In this paper, we introduce SCOPE¹, a multi-agent media-profiling tool that distributes this workflow among specialized agents. Each agent performs a distinct function, including data acquisition, content analysis, evidence search, or report generation. By coordinating these agents, SCOPE combines complementary evidence streams to produce structured and explainable media profiles.

This design aims to improve coverage, reduce the burden on any single model, and make the profiling process more transparent through intermediate evidence and agent-level outputs.

2 System Description

In this section we describe how SCOPE works, which parts are included in the system, and the rationale behind our design choices.

¹Source Credibility and Outlet Profiling Engine

* Equal contribution.

2.1 Methodology

Media profiling requires a systematic methodology, since assessments of bias and reliability may otherwise become subjective. We follow the methodology proposed by Media Bias/Fact Check (MBFC)², an independent media-rating organization that evaluates news sources with respect to ideological bias, factual reporting, transparency, and credibility.

MBFC evaluates each outlet through a human review process that includes analysis of at least ten headlines and five full news articles. The review considers language, framing, tone, sourcing practices, balance of viewpoints, and consistency with verified external information. Additional factors, such as ownership, funding, authorship, and editorial transparency, are also considered when available.

The bias rating is computed as a weighted score on a left–right scale from -10 to $+10$. MBFC defines the overall bias score as

$$B = 0.35E + 0.35S + 0.15N + 0.15O,$$

where E denotes the outlet’s economic ideology, S its position on social values, N the balance of its straight news reporting, and O its editorial or opinion bias. Economic and social dimensions receive the largest weights because they capture the main ideological orientation of a media outlet. Straight news balance and editorial bias are included, since bias may also appear through story selection, framing, omission of opposing views, or emotionally loaded language. Negative values of B indicate left-leaning bias, positive values indicate right-leaning bias, and values close to zero correspond to least-biased reporting.

Factual reliability is computed from failed fact checks (C), sourcing quality (R), transparency (T), and one-sidedness, omission, or propaganda (M):

$$F = 0.40C + 0.25R + 0.25T + 0.10M.$$

Failed fact checks receive the highest weight as the strongest indicator of unreliability. Sourcing and transparency measure whether outlets provide credible evidence and disclose their ownership, funding, authorship, and editorial responsibility. The smaller weight for M reflects its overlap with bias while accounting for selective reporting and omission. Lower values of F indicate higher factual reliability.

²<https://mediabiasfactcheck.com/methodology/>

2.2 System and Components

While MBFC relies on human reviewers to apply its methodology, our goal is to approximate this process using LLM-based agents. The evaluation procedure is coordinated by an orchestrator that manages four groups of specialised agents (Figure 2). First, an *acquisition* agent scrapes and caches articles from the outlet. Second, a *content-analysis* group, consisting of nine agent-based analyzers, evaluates the articles according to the MBFC scoring dimensions. Third, an *evidence-search* agent retrieves information about the outlet, including its history, ownership, funding, transparency ranking, and proofs of failed fact-checks. Finally, a *report-generation* agent synthesises the intermediate outputs into an MBFC-style media profile.

Each agent is assigned a narrowly defined subtask and produces a validated, schema-constrained output. The orchestrator executes the agents in a fixed order, checks whether the collected evidence is sufficient, and then computes the bias, factuality, and credibility scores. This modular design separates evidence collection, content analysis, scoring, and report generation, making intermediate decisions inspectable and the final profile auditable.

Scraping. We collect up to 15 articles per outlet using a crawler (HTTP requests paired with BeautifulSoup parsing). Candidate URLs are discovered from RSS feeds, Substack archives, sitemaps, and homepages, then scored by a simple keyword-based heuristic that prioritises politics, economy, and world news; the top candidates are fetched with randomised rate-limiting to avoid blocking.

Analysis. After scraping, agents analyze editorial bias, transparency, one-sidedness, pseudoscience, fact-checking history, economic and social positioning, and sourcing quality. They examine the articles, outlet metadata, and evidence retrieved from external sources. Each agent returns a label, score, rationale, and supporting evidence. For example, the editorial-bias agent evaluates framing, headline tone, topic selection, and the treatment of political actors. The sourcing agent examines whether factual claims cite named sources, primary documents, expert quotations, or sites. The pseudoscience agent identifies unsupported scientific, medical, and conspiratorial claims using the MBFC taxonomy.³

³<https://mediabiasfactcheck.com/pseudoscience-dictionary/>



Figure 1: Propaganda techniques detected by SCOPE in articles, shown with the exact triggering spans and explanations. The system identifies Appeal to Fear/Prejudice, Loaded Language, Causal Oversimplification, Doubt, and Appeal to Authority — all consistent with MBFC’s Conspiracy-Pseudoscience classification of the outlet.

To estimate the economic and social components of the bias score, we use a two-axis framework. This follows the view that political ideology is not well captured by a single left–right continuum, but is better understood as having one economic and one social/cultural dimension (Eysenck, 1954; Treier and Hillygus, 2009; Jost et al., 2009).

The economic axis is classically defined by one’s stance toward equality, redistribution, and the role of the state (Bobbio, 1996), while the social axis captures progressive-versus-traditional positions on cultural questions such as gender, religion, immigration, and national identity (Norris and Inglehart, 2019). Both axes are scored from -10 to $+10$. On the economic axis, -10 corresponds to strongly state-interventionist positions and $+10$ to strongly free-market positions. On the social axis, -10 corresponds to progressive-liberal positions and $+10$ to traditional-conservative positions. Scores near zero indicate a centrist or mixed position on the corresponding axis.

For each scraped article, we first determine whether the text is relevant to either dimension. Articles that do not discuss economic policy, social values, civil liberties, or related political issues are excluded from the corresponding axis. Relevant articles are then scored on the applicable dimension. Using text to estimate ideological position follows the same general logic as work on party manifestos and media slant (Budge et al., 2001; Gentzkow and Shapiro, 2010; Groseclose and Milyo, 2005). The scoring rubrics are derived from the ideology bank covering 17 policy domains. The details are given in Appendix D. We additionally check the articles for loaded language and other propaganda and manipulation techniques, consistent with Da San Martino et al. (2020a) taxonomy. See examples on Figure 1.

Web-Search. In addition to scraped articles, SCOPE uses DuckDuckGo search to collect external evidence about each outlet, including ownership, funding, editorial history, and third-party fact-checking records from organizations such as Full Fact and other independent fact-checkers.

We explicitly exclude media-rating aggregators such as MBFC, AllSides⁴, NewsGuard⁵ to prevent gold-label leakage during profiling.

External knowledge base. We complement outlet-derived evidence with independent reference sources. The Tranco top-1M list (Le Pochat et al., 2019) estimates outlet reach using four traffic tiers: High for the top 10,000 domains, Medium for the top 100,000, Low for the top 1,000,000, and Minimal otherwise. Press-freedom ratings from the RSF Index (Reporters Without Borders, 2025) and Freedom House (Freedom House, 2024) produce a credibility adjustment of -1 for limited freedom and -2 for no freedom.

2.3 Demonstration interface

SCOPE provides a web application that presents each outlet profile in four tabs. The **Overview** tab (Figure 3) shows MBFC-style verdicts, numeric scores, outlet metadata, and a narrative summary with collapsible history, ownership, and analysis sections. The **Evidence** tab provides an audit trail containing article quotes, analyzer findings, propaganda spans with confidence scores, and fact-check results. The **Analyzer Trace** tab displays each agent’s structured output, while the **Compare** tab places two outlets side by side.

⁴<https://www.allsides.com/unbiased-balanced-news>

⁵<https://www.newsguardtech.com/>

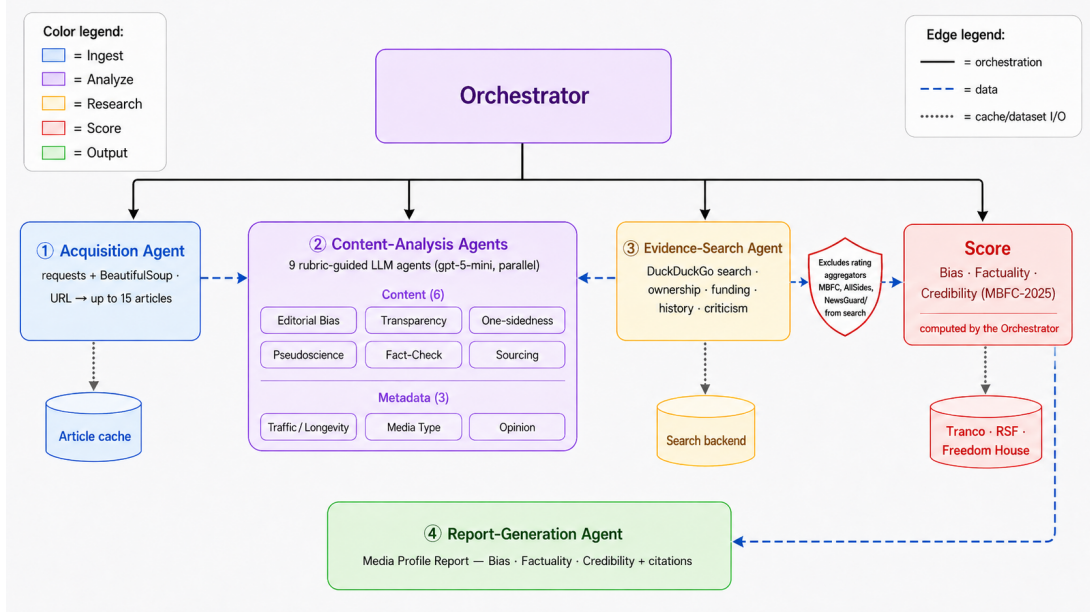


Figure 2: Schematic pipeline of SCOPE system. Given an outlet URL, four agent groups-acquisition, content analysis, evidence search, and report generation-are coordinated by a deterministic orchestrator to produce a structured, evidence-grounded profile.

3 Experiments and Results

To demonstrate the usability of our system, we run several comparative experiments. We ablate SCOPE by comparing it against three LLM baselines. The first baseline asks the LLM to make the assessment using only its internal knowledge, without any external context. The second provides the LLM with 15 articles from the website. The third gives the LLM access to a web search tool, while excluding MBFC results to avoid label leakage.

We further compare SCOPE with Media Background Checks (MBC) by [Schlichtkrull \(2024\)](#), a retrieval-augmented system for generating website background checks. MBC issues seven web-search queries, retrieves 30 documents per query, and uses a DeBERTa-based question-answering model ([He et al., 2021](#)) to extract key facts such as founders and prior reports of misinformation. We use the original implementation, replacing its unsupported search backend with the same search tool used in our experiments alongside the LLM to answer the questions instead of the DeBERTa-based question-answering model.

3.1 Data

All experiments are conducted on 200 media outlets sampled from the MBFC catalogue, selected by round-robin sampling across seven bias categories and six factuality levels for broad label coverage.

We use each outlet’s MBFC bias, factuality, and credibility labels as gold annotations, and the free-text MBFC fields (summary, history, ownership, analysis) as reference text for evaluating generated reports. All systems are evaluated on the same fixed set of collected articles.

3.2 Setup and metrics

We evaluate political leaning and factuality using MAE, following GREENER ([Panayotov et al., 2022](#)), [Manzoor et al. \(2026\)](#), and related work. MAE reflects ordinal distance: confusing Left with Left-Center is penalized less than confusing it with Right. Bias is ordered from Extreme-Left, Left, Left-Center, Least Biased, Right-Center, Right, to Extreme-Right; factuality ranges from Very Low to Very High through Low, Mixed, Mostly Factual, and High. We assess reports with FACTScore ([Min et al., 2023](#)): precision measures claim support, recall measures MBFC-reference coverage, and error rate is one minus precision. FC-Detection measures how often the system finds an outlet’s failed fact checks.

To assess the content overlap of the generated reports, we also compute METEOR ([Banerjee and Lavie, 2005](#)) and ROUGE-L ([Lin, 2004](#)), using the MBFC text as the reference. All experiments use GPT-5 mini, the Pareto-optimal model in the cost-quality comparison of Appendix A.

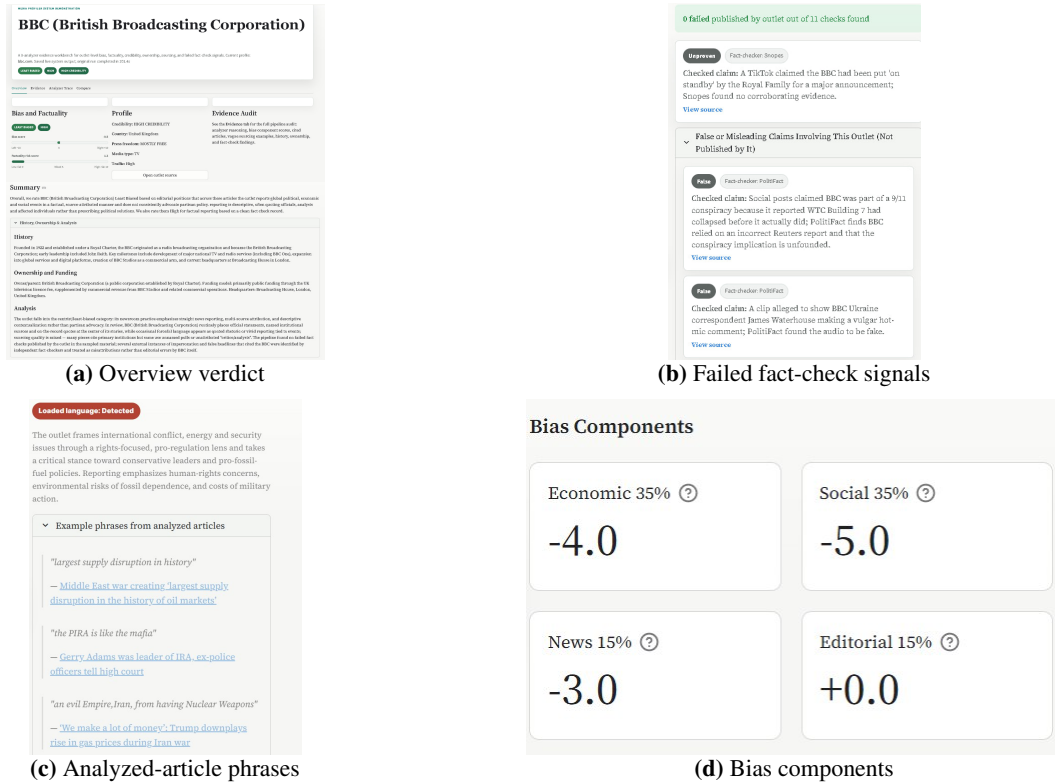


Figure 3: BBC (a,b) and The Guardian (c,d) profiles in SCOPE demonstration interface, showing the overview verdict, supporting failed fact-check signals, analyzed-article phrases, and bias-component scores.

3.3 Main results

Figure 4 summarises the comparison across the approaches.

SCOPE is the strongest overall, achieving the lowest bias MAE (0.65 vs. 1.17 for the LLM-only baseline), the lowest factuality MAE (0.67), perfect fact-check detection (1.00), the best ROUGE-L (0.16), the highest FACTScore precision (0.48), and the lowest error rate (0.52).

It also always finds if there are failed fact-checks by the outlet (as, for example, for *NaturalNews.com*, as shown in the Appendix C)

The LLM with its web-search tool raises fact recall (0.36) and METEOR (0.26), but also produces the second highest error rate (0.62), showing that live web text increases coverage at the cost of unsupported assertions. Adding offline articles improves bias calibration over the LLM-only baseline (0.89 vs. 1.17) but slightly hurts factuality, consistent with article text being more informative about an outlet’s stance than about its factual track record. On the ordinal axes, SCOPE predicts the exact bias class for nearly 50% of outlets and lands within one ordinal class for about 90%; on factuality it is exact for nearly 42% and within one class for about 90% (Figure 5; full breakdown in Appendix B).

The full pipeline costs only \$0.054 per outlet—comparable to the LLM-only baseline (\$0.050) and roughly half the cost of Search (\$0.101).

4 Conclusion

We presented SCOPE, a multi-agent system that transforms a single outlet URL into a structured, evidence-grounded profile of political bias, factual reliability, and credibility. SCOPE divides media profiling among specialised agents for acquisition, content analysis, evidence search, and report generation. Their typed, schema-constrained outputs are visible in the interface, making each verdict traceable to its evidence rather than an opaque label.

On a 200-outlet MBFC benchmark, SCOPE achieves the lowest bias and factuality MAE and perfect failed-fact-check detection among the evaluated baselines. It also keeps 90% of predictions within one ordinal class on both axes, at about \$0.05 per outlet.

Qualitative analysis shows that SCOPE’s evidence remains useful even when its final label is incorrect, especially for conspiracy and pseudoscience outlets that do not fit a left–right scale. We release the system, live demo, and leakage-controlled evaluation code.

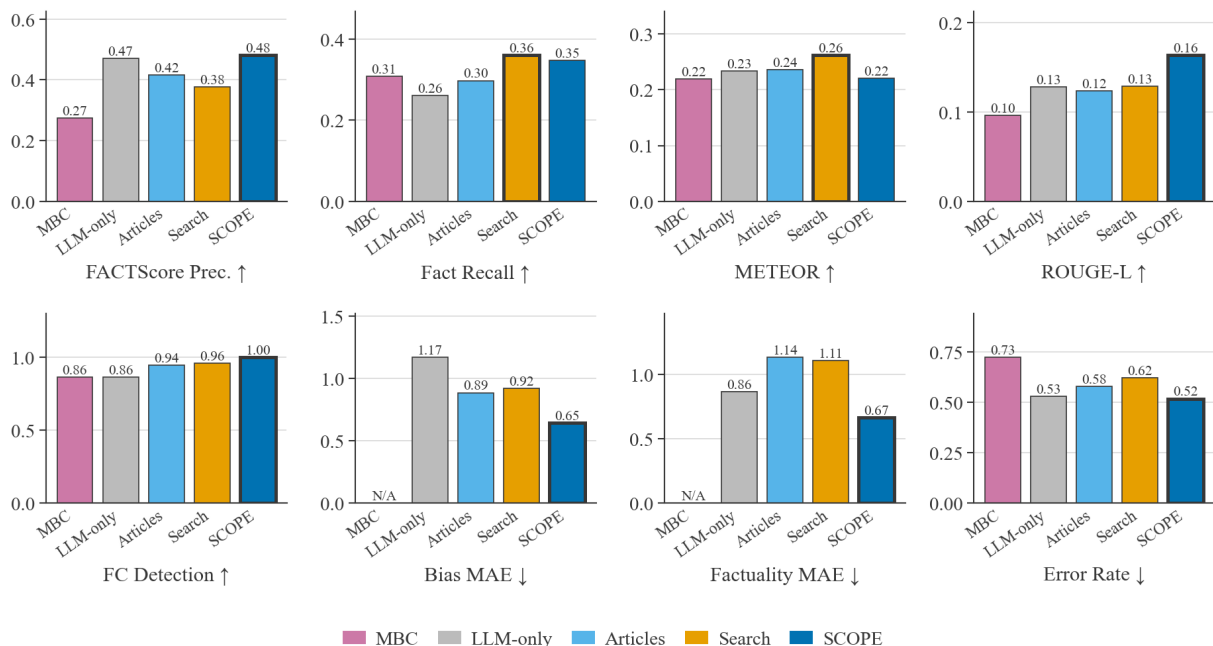


Figure 4: Ablation results across eight metrics (↑ higher is better; ↓ lower is better). The full *SCOPE* wins on six of eight metrics, including both ordinal MAEs, failed fact-check detection, and error rate. *Search* leads only on recall-oriented metrics (fact recall, METEOR). We provide no Bias and Factuality evaluation for MBC, since it does not provide a direct label and gives a description only.

In future work we aim to extend coverage beyond English-language and U.S.-centric outlets, and to generalise the agent scaffolding to adjacent entities such as politicians, pollsters, and government bodies.

Limitations

Although *SCOPE* performs well in our evaluation, several limitations remain and should be considered when interpreting the results.

First, our evaluation focuses on U.S.-centric, English-language outlets because labeled fact-checking and media-profiling data are scarce for other languages and regions. *SCOPE* itself is not limited to English and can work with other languages. However, the lack of suitable labels prevented us from evaluating its multilingual performance. The reported results therefore may not fully generalize to non-Western, non-English, or low-visibility outlets.

Second, while we explicitly filter web-search results to exclude MBFC, AllSides, NewsGuard, and similar media-rating platforms, indirect leakage is still possible, since other sites may reproduce or cite their ratings. We manually inspected the retrieved web-search traces in our experiments and did not find evidence of such leakage, but this risk cannot be ruled out completely.

Third, the sourcing analyzer may miss hyper-link metadata. The scraper converts each article to plain text, retaining the displayed anchor text but removing the target URL and any indication that the text was clickable. A citation embedded behind a phrase may therefore be invisible to the analyzer unless the source name or URL also appears in the text. In manual testing, this sometimes led the system to report missing evidence even when the article linked to a relevant source. We will address this in the next release.

Finally, conspiracy and pseudoscience outlets can be difficult to place on a left-right political scale. *SCOPE* usually captures their low factuality and credibility, but it may occasionally assign the wrong political direction when the outlet’s content is anti-establishment, conspiratorial, or ideologically mixed.

Ethics and Broader Impact

SCOPE scores politically sensitive attributes of news outlets, and such scores can be misused to dismiss or amplify sources. Every score is backed by inspectable intermediate evidence and withheld when the evidence-sufficiency gate is not met. The system is intended as an auditable assistive tool for researchers, journalists, and analysts—not as an arbiter of truth.

References

- Satanjeev Banerjee and Alon Lavie. 2005. [METEOR: An automatic metric for MT evaluation with improved correlation with human judgments](#). In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.
- Norberto Bobbio. 1996. *Left and Right: The Significance of a Political Distinction*. University of Chicago Press, Chicago, IL.
- Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tanenbaum. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford University Press, Oxford.
- Edmund Burke. 1790. *Reflections on the Revolution in France*. J. Dodsley, London.
- Judith Butler. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge, New York.
- David Card and Alan B. Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton University Press, Princeton, NJ.
- Joseph H. Carens. 2013. *The Ethics of Immigration*. Oxford University Press, New York.
- Ta-Nehisi Coates. 2014. [The case for reparations](#). The Atlantic.
- Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020a. [SemEval-2020 task 11: Detection of propaganda techniques in news articles](#). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1377–1414, Barcelona (online). International Committee for Computational Linguistics.
- Giovanni Da San Martino, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barrón-Cedeño, and Preslav Nakov. 2020b. [Prta: A system to support the analysis of propaganda techniques in the news](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 287–293, Online. Association for Computational Linguistics.
- Angela Y. Davis. 2003. *Are Prisons Obsolete?* Seven Stories Press, New York.
- Gøsta Esping-Andersen. 1990. *The Three Worlds of Welfare Capitalism*. Princeton University Press, Princeton, NJ.
- Hans J. Eysenck. 1954. *The Psychology of Politics*. Routledge and Kegan Paul, London.
- Freedom House. 2024. [Freedom in the World 2024](#). Accessed 2025.
- Richard B. Freeman and James L. Medoff. 1984. *What Do Unions Do?* Basic Books, New York.
- Milton Friedman. 1955. The role of government in education. In Robert A. Solo, editor, *Economics and the Public Interest*, pages 123–144. Rutgers University Press, New Brunswick, NJ.
- Milton Friedman. 1962. *Capitalism and Freedom*. University of Chicago Press, Chicago, IL.
- Matthew Gentzkow and Jesse M. Shapiro. 2010. [What drives media slant? evidence from U.S. daily newspapers](#). *Econometrica*, 78(1):35–71.
- Tim Groseclose and Jeffrey Milyo. 2005. [A measure of media bias](#). *The Quarterly Journal of Economics*, 120(4):1191–1237.
- Jens Hainmueller and Daniel J. Hopkins. 2014. [Public attitudes toward immigration](#). *Annual Review of Political Science*, 17:225–249.
- Friedrich A. Hayek. 1944. *The Road to Serfdom*. Routledge, London.
- Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. [DeBERTa: Decoding-enhanced BERT with disentangled attention](#). In *International Conference on Learning Representations*.
- John T. Jost, Christopher M. Federico, and Jaime L. Napier. 2009. [Political ideology: Its structure, functions, and elective affinities](#). *Annual Review of Psychology*, 60:307–337.
- Ibram X. Kendi. 2019. *How to Be an Antiracist*. One World, New York.
- Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. [Tranco: A research-oriented top sites ranking hardened against manipulation](#). In *Proceedings of the 26th Annual Network and Distributed System Security Symposium*.
- Chin-Yew Lin. 2004. [ROUGE: A package for automatic evaluation of summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Bjørn Lomborg. 2001. *The Skeptical Environmentalist: Measuring the Real State of the World*. Cambridge University Press, Cambridge.
- Muhammad Arslan Manzoor, Dilshod Azizov, Daniil Orel, Umer Siddique, Zain Muhammad Mujahid, Yufang Hou, and Preslav Nakov. 2026. [A multi-view media profiling suite: Resources, evaluation, and analysis](#). In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 25265–25286, San Diego, California, United States. Association for Computational Linguistics.

- Karl Marx and Friedrich Engels. 1848. *The Communist Manifesto*. Workers' Educational Association.
- Aaron M. McCright and Riley E. Dunlap. 2011. [The politicization of climate change and polarization in the American public's views of global warming, 2001–2010](#). *The Sociological Quarterly*, 52(2):155–194.
- John Stuart Mill. 1859. *On Liberty*. John W. Parker and Son, London.
- Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2023. [FACTScore: Fine-grained atomic evaluation of factual precision in long form text generation](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12076–12100, Singapore. Association for Computational Linguistics.
- Charles Murray. 1984. *Losing Ground: American Social Policy, 1950–1980*. Basic Books, New York.
- Pippa Norris and Ronald Inglehart. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press, Cambridge.
- Panayot Panayotov, Utsav Shukla, Husrev Taha Sencar, Mohamed Nabeel, and Preslav Nakov. 2022. [GREENER: Graph neural networks for news media profiling](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7470–7480, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Thomas Piketty. 2014. *Capital in the Twenty-First Century*. Harvard University Press, Cambridge, MA.
- John Rawls. 1971. *A Theory of Justice*. Harvard University Press, Cambridge, MA.
- Reporters Without Borders. 2025. [2025 World Press Freedom Index](#). Accessed 2025.
- David Ricardo. 1817. *On the Principles of Political Economy and Taxation*. John Murray, London.
- Dani Rodrik. 2011. *The Globalization Paradox: Democracy and the Future of the World Economy*. W. W. Norton, New York.
- Emmanuel Saez and Gabriel Zucman. 2019. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. W. W. Norton, New York.
- Ahmed Sajwani, Alaa El Setohy, Ali Mekky, Diana Turmakhan, Lara Hassan, Mohamed El Zeftawy, Omar El Herraoui, Osama Mohammed Afzal, Qisheng Liao, Tarek Mahmoud, Zain Muhammad Mujahid, Muhammad Umar Salman, Muhammad Arslan Manzoor, Massa Baali, Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2024. [FRAPPE: FRAMing, Persuasion, and Propaganda explorer](#). In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 207–213, St. Julians, Malta. Association for Computational Linguistics.
- Michael Sejr Schlichtkrull. 2024. [Generating media background checks for automated source critical reasoning](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 4927–4947, Miami, Florida, USA. Association for Computational Linguistics.
- Adam Smith. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. W. Strahan and T. Cadell, London.
- Christina Hoff Sommers. 1994. *Who Stole Feminism? How Women Have Betrayed Women*. Simon & Schuster, New York.
- Thomas Sowell. 1987. *A Conflict of Visions: Ideological Origins of Political Struggles*. William Morrow, New York.
- George J. Stigler. 1971. [The theory of economic regulation](#). *The Bell Journal of Economics and Management Science*, 2(1):3–21.
- Joseph E. Stiglitz. 2010. *Freefall: America, Free Markets, and the Sinking of the World Economy*. W. W. Norton, New York.
- Shawn Treier and D. Sunshine Hillygus. 2009. [The nature of political ideology in the contemporary electorate](#). *Public Opinion Quarterly*, 73(4):679–703.
- Yuxia Wang, Minghan Wang, Hasan Iqbal, Georgi N. Georgiev, Jiahui Geng, Iryna Gurevych, and Preslav Nakov. 2025. [OpenFactCheck: Building, benchmarking customized fact-checking systems and evaluating the factuality of claims and LLMs](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 11399–11421, Abu Dhabi, UAE. Association for Computational Linguistics.
- Tim Wu. 2018. *The Curse of Bigness: Antitrust in the New Gilded Age*. Columbia Global Reports, New York.
- Zhuohan Xie, Rui Xing, Yuxia Wang, Jiahui Geng, Hasan Iqbal, Dhruv Sahnan, Iryna Gurevych, and Preslav Nakov. 2025. [FIRE: Fact-checking with iterative retrieval and verification](#). In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 2901–2914, Albuquerque, New Mexico. Association for Computational Linguistics.
- Yifan Zhang, Giovanni Da San Martino, Alberto Barrón-Cedeño, Salvatore Romeo, Jisun An, Haewoon Kwak, Todor Staykovski, Israa Jaradat, Georgi Karadzhov, Ramy Baly, Kareem Darwish, James Glass, and Preslav Nakov. 2019. [Tanbih: Get to know what you are reading](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 223–228, Hong Kong, China. Association for Computational Linguistics.

A Model Selection

We selected the model by evaluating five candidates on the same set of outlets (≈ 50), while keeping the token cap fixed, so that differences in quality-per-dollar reflect model pricing rather than usage variation.

Quality was measured as an equal-weight, min-max-normalised composite of the task metrics. We divided this quality score by cost to obtain the quality-per-dollar metric reported in Table 1.

By this metric, GPT-5 mini is Pareto-optimal: none of the evaluated models is both cheaper and higher quality. It achieves a score of 12.1, about $3\times$ higher than premium models such as GPT-5 and GPT-5.2. This comparison is based on the LLM-only setting.

Model	\$/outlet	Quality	Quality/\$
GPT-5 nano	0.0110	0.086	7.8
GPT-5 mini	0.0507	0.614	12.1
GPT-4.1	0.2201	0.858	3.9
GPT-5	0.2491	0.980	3.9
GPT-5.2	0.3483	0.576	1.7

Table 1: Cost-quality trade-off in LLM-only mode. Quality is an equal-weight, min-max-normalised composite over five task metrics. The recommended operating point is shown in **bold**.

B Ordinal Accuracy

Figure 5 gives the per-class breakdown, for both the bias and factuality axes, behind the aggregate within-one-class figures reported in Section 3.3.

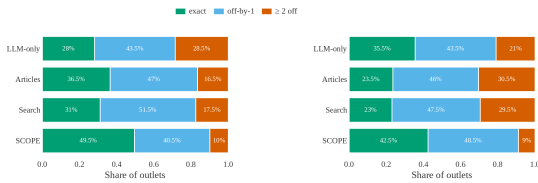


Figure 5: Ordinal accuracy. It shows the share of outlets whose predicted class is exact, off by one, or ≥ 2 classes from the gold label, for bias (left) and factuality (right).

It shows that *SCOPE* keeps roughly 90% of predictions within one class for both factuality and bias.

C Failed Fact-Check Detection Example

Figure 6 shows the “Signals at a Glance” panel *SCOPE* produces for *NaturalNews.com*.

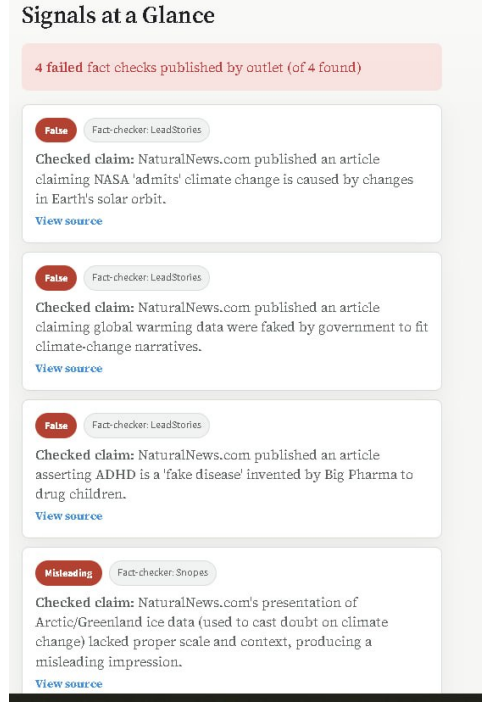


Figure 6: “Signals at a Glance” panel for *NaturalNews.com*: *SCOPE* locates four failed fact-checks published by external fact-checkers (LeadStories, Snopes).

D The Ideology Item Bank

The economic and social bias axes are anchored by an offline ideology item bank.

The bank contains reference-backed indicators organised into 17 policy domains: 7 economic domains and 10 social domains.

Each domain contains nine possible positions: four left/progressive stances, one centrist stance, and four right/conservative stances.

The bank’s purpose is to make the scoring scale explicit: each item describes an observable editorial stance and maps it to a point on the corresponding -10 to $+10$ axis, and in the released version is annotated with a primary or scholarly source supporting that mapping.

Table 2 reports the 17 domains with a representative source for each pole; where no single canonical source defines a pole, we instead cite literature showing that the issue loads on the economic or social/cultural dimension of ideology (Jost et al., 2009; Treier and Hillygus, 2009; Budge et al., 2001).

Concretely, the prompt embeds the two axis scales and domain summaries as reference points, fixing the scale’s meaning before the model assigns each outlet’s economic and social scores in one step.

Policy domain	Left / progressive pole (grounding)	Right / conservative pole (grounding)
<i>Economic axis</i> (−10 state-interventionist → +10 free-market)		
Government ownership & nationalisation	Public ownership of the means of production (Marx and Engels, 1848)	Private enterprise, the “invisible hand” (Smith, 1776)
Taxation & redistribution	Progressive and wealth taxation (Piketty, 2014; Saez and Zucman, 2019)	Low, flat taxation; limited state (Friedman, 1962; Hayek, 1944)
Labour, unions & minimum wage	Collective bargaining, wage floors (Freeman and Medoff, 1984; Card and Krueger, 1995)	Market-set wages, labour deregulation (Friedman, 1962)
Corporate regulation & antitrust	Aggressive antitrust enforcement (Wu, 2018)	Deregulation; regulatory-capture critique (Stigler, 1971)
Trade & globalisation	Managed / fair trade, worker protection (Rodrik, 2011)	Free trade, comparative advantage (Ricardo, 1817)
Social-welfare programmes	Social-democratic welfare state (Esping-Andersen, 1990)	Welfare-dependency critique (Murray, 1984)
Financial-sector regulation	Post-crisis bank regulation (Stiglitz, 2010)	Market discipline, free banking (Friedman, 1962)
<i>Social axis</i> (−10 progressive-liberal → +10 traditional-conservative)		
LGBTQ+ rights	Equal rights, anti-discrimination (Butler, 1990)	Traditional family / religious objection (Burke, 1790)
Abortion & reproductive rights	Bodily autonomy as individual liberty (Mill, 1859)	Traditional morality, foetal personhood (Burke, 1790)
Immigration (social / cultural)	Permissive borders, inclusion (Carens, 2013; Hainmueller and Hopkins, 2014)	Restriction; cultural-cohesion priority (Norris and Inglehart, 2019)
Climate change & environment	Urgent, human-caused crisis (McCright and Dunlap, 2011)	Cost-sceptic, anti-regulation (Lomborg, 2001)
Racial justice & equity	Anti-racism, structural remedy (Kendi, 2019; Coates, 2014)	Colour-blind individualism (Sowell, 1987)
Gender equality	Feminist / structural equality (Butler, 1990)	Equity-feminism critique (Sommers, 1994)
Gun rights & control	Stronger firearm regulation (Jost et al., 2009)	Robust individual right (Sowell, 1987)
Criminal justice	Decarceration, reform / abolition (Davis, 2003)	Deterrence, tough-on-crime (Sowell, 1987)
Religion & secularism	Strict church–state separation (Mill, 1859)	Religion in public life (Burke, 1790)
Education policy	Universal public provision (Rawls, 1971)	School choice and vouchers (Friedman, 1955)

Table 2: The 17 policy domains of the ideology item bank, with a representative source grounding each pole. Every domain expands into a 4-left / 4-right / 1-centre ladder of observable editorial stances, each mapped to a calibrated point on the −10 to +10 economic or social axis and individually cited in the released bank. These two scales and the 17 domains are distilled into the single scoring prompt described in the main text; the items themselves are not executed per outlet.