Selected Papers of #AoIR2025:
The 26th Annual Conference of the
Association of Internet Researchers
Niterói, Brazil / 15 – 18 Oct 2025

HUMANS AND MACHINES IN THE LOOP: RETHINKING LLMS FOR CONFLICT AND DISAGREEMENT IN CONTENT ANALYSIS OF SOCIAL COMPLEX PHENOMENA

Tariq Choucair
Digital Media Research Centre, Queensland University of Technology

Laura Vodden
Digital Media Research Centre, Queensland University of Technology

Ahrabhi Kathirgamalingam
Center for Advanced Internet Studies (CAIS)

Katharina Esau
Digital Media Research Centre, Queensland University of Technology

Axel Bruns
Digital Media Research Centre, Queensland University of Technology

Paul Pressman
ZeMKI / University of Bremen

Cornelius Puschmann
ZeMKI / University of Bremen

Fabio Giglietto
University of Urbino

Giada Marino
University of Urbino

Bruna Paroni
University of Urbino

Suggested Citation (APA): Choucair, T., Vodden, L., Kathirgamalingam, A., Esau, K., Bruns, A., Pressman, P., Puschmann, C., Giglietto, F., Marino, G., & Paroni, B. (2025, October). *Humans and machines in the loop: Rethinking LLMs for conflict and disagreement in content analysis of social complex phenomena*. Panel presented at AoIR2025: The 26th Annual Conference of the Association of Internet Researchers, Niterói, Brazil. Retrieved from http://spir.aoir.org

Panel Overview

The study of polarization, conflict, and ideological divergence has long challenged scholars across media, communication, and political science. Understanding these phenomena requires engaging with fundamental questions about how opinions are formed, reinforced, and contested in public discourse, and how language and discourses can have so many different meanings and interpretations. Traditional content analysis methods often try to capture this complexity - whether in large-scale media narratives, open-ended survey responses, or political discourse on social media. Computational approaches offer new possibilities, but also raise critical concerns about validity, interpretability, and methodological rigor (Baden et al., 2022; Boumans & Trilling, 2016).

Large Language Models (LLMs) have become increasingly central to communication research, with their capacity to process vast amounts of text, identify underlying patterns, and assist in qualitative coding (Chew et al., 2023; Alizadeh et al., 2025). Researchers have explored their use in a variety of tasks, from detecting polarization in open-ended survey responses to analyzing media frames and political discourse (DiGiuseppe & Flynn, 2025; Marino & Giglietto, 2024). However, the integration of LLMs into content analysis presents challenges: How well do they align with human interpretations? Can they enhance research beyond automation? And what role should they play in investigating contested or ambiguous meanings (Pilny et al., 2024; Gunes & Florczak, 2025)?

This panel moves beyond discussions of mere optimization and accuracy, instead critically and deeply discussing the use of LLMs to engage with conflict, disagreement, and interpretive diversity. Instead of seeing them as tools to impose consensus, we ask how researchers can use and interact with them to reveal tensions, challenge assumptions, and contribute to new methodologies (Dai et al., 2023; De Paoli, 2024). Across four studies, we assess LLMs mediating, amplifying, or reframing scholarly debates on the methods to analyse contentious political and social issues. Our discussion examines both the benefits and risks of these approaches, raising questions about the role of AI in media and communication methodologies.

Paper 1 present a computational approach to measure issue polarization from open-ended survey responses, leveraging Large Language Models (LLMs) to systematically code viewpoints on contentious topics such as climate change, trans rights, or political discourse on Ukraine. While traditional polarization research often relies on close-ended survey measures, this study explores how LLMs can introduce methodological complexity by computing nuanced, interpretive tensions in unstructured responses. The research highlights how LLMs not only enable large-scale automated coding but also challenge conventional measurement frameworks by accommodating ambiguity and ideological fluidity. Preliminary results underscore the potential of LLMs to expand analytical possibilities beyond traditional survey measures, reframing how scholars engage with disagreement and conflict in media and communications studies.

Paper 2 examines how LLMs can expand the scope of content analysis and assisting researchers in analyzing framing. The study applies Meta's Llama-3 model in a two-stage approach to study climate movements in Australian news coverage, using few-shot prompting to first extract frame elements - such as problem definitions, causes, and blame attributions - and then synthesizing these elements into coherent frames. While the human coders tended to construct more issue-specific and varied frames, LLM-generated outputs were generally broader in scope but more internally consistent across the dataset.

Paper 3 presents a literature review examining how LLMs are transforming content analysis workflows in social sciences. It identifies four key modes of LLM integration: scalable coders, human-assistive collaborators, autonomous decision-makers, and tools for semantic clustering. The study highlights the ongoing challenges of ensuring interpretability, reliability, and epistemic authority when LLMs are applied to human-generated texts.

Paper 4 applies LLMs to investigate the role of political narratives and user engagement on social media during Brazil's 2022 presidential election and the January 8, 2023, coup attempt. The study analyzes over 12 million social media posts, clustering content based on sentiment, audience reactions, and dissemination patterns to assess how different narratives are amplified. The research examines whether the framing of political content influences audience interaction and engagement levels.

Together, these studies push the boundaries of how LLMs are integrated into research on political communication, polarization, and media analysis. They provide an assessment of LLMs' methodological potential and risks - not just as tools for efficiency, but as channels to rethink how we engage with contested meanings in communication research.

References

Alizadeh, M., Kubli, M., Samei, Z., et al. (2025). Open-source LLMs for text annotation: A practical guide for model setting and fine-tuning. *Journal of Computational Social Science, 8(17).*

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures, 16(1), 1–18.* https://doi.org/10.1080/19312458.2021.2015574

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques. *Digital Journalism, 4(1), 8–23.*

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv.* http://arxiv.org/abs/2306.14924

Dai, S.-C., Xiong, A., & Ku, L.-W. (2023). LLM-in-the-loop: Leveraging large language models for thematic analysis. *arXiv.* https://doi.org/10.48550/arXiv.2310.15100

De Paoli, S. (2024). Using LLMs in qualitative research: An examination of interpretive and methodological challenges. *Qualitative Inquiry.*

DiGiuseppe, M., & Flynn, M. (2025). Scaling open-ended survey responses using LLM-paired comparisons.

Gunes, S., & Florczak, S. (2025). Evaluating LLM-driven multiclass classification of congressional bills and policy documents under different levels of human oversight. *AI & Society.*

Marino, G., & Giglietto, F. (2024). Integrating large language models in political discourse studies on social media: Challenges of validating an LLM-in-the-loop pipeline. *Sociologica, 18(2), 87–107.* https://doi.org/10.6092/issn.1971-8853/19524

Pilny, A., McAninch, K., Slone, A., & Moore, K. (2024). From manual to machine: Assessing the efficacy of large language models in content analysis. *Communication Research Reports, 41(2), 61–70.*

Paper 1

AN ISSUE POLARIZATION SCALE FROM OPEN TEXT RESPONSES

Paul Pressman
ZeMKI / University of Bremen

Cornelius Puschmann
ZeMKI / University of Bremen

Introduction and State of Research

Issue polarization and its connection with media consumption is a phenomenon that has attracted substantial scholarly attention over the course of the last two decades, as well as drawing a high degree of public interest (Brüggemann & Meyer, 2023; Mason, 2015). In Western societies, there is a widespread belief that both individual attitudes and public discourse have become more extreme and adversarial over time, and research in political science, sociology, and media and communication research has aimed to test this assumption, to theorize its causes, and to measure its effects (Kubin & von Sikorski, 2021). Much is contested in the polarization literature, most of all whether societies have become more polarized over time, whether and under which conditions polarization should be considered destructive and, perhaps even more fundamentally, whether polarization is a *state* or a *process* that should be understood to be applicable to individual attitudes, public opinion – or both (Esau et al., 2024). In particular, the role of (social) media in accentuating extreme stances and thus acting as a catalyst of polarized public opinion is controversial in political communication (Barberá, 2020; Bruns et al., 2024). Attitudinal polarization on key issues is also contested, with some arguments for no substantial change in attitudes and their relative extremity over time (Wojcieszak et al., 2021).

The debate extends to methodological questions about how polarization should be measured and conceptualized. Researchers disagree about whether to focus on issue-specific attitudes, ideological consistency across issues, or affective responses to political outgroups. Furthermore, cross-national comparative studies have yielded inconsistent findings, suggesting that polarization dynamics may be context-dependent rather than universal across democratic societies. On a methodological level, there is the challenge of accurately measuring issue polarization, with a variety of approaches having been previously applied, most notably scaling models (Asker & Dinas, 2019; Lowe et al., 2011; Olechowska, 2022).

An intriguing option is not only to apply techniques for measuring issue polarization automatically to news articles, social media and party manifestos, but also to extend them to open survey responses. However, traditional approaches to studying polarization often rely on survey instruments that assume clear-cut ideological positions, but emerging computational methods, such as large language models (LLMs), offer an opportunity to move beyond rigid self-assessments on close-answer scales. At the same time, analyzing open-ended responses presents unique challenges, as polarization in public communication – such as social media & journalistic articles – operates with distinct intrinsic motives, compared to the expression of opinions in surveys (Cavari & Freedman, 2018; Geer, 1988; Reja et al., 2003).

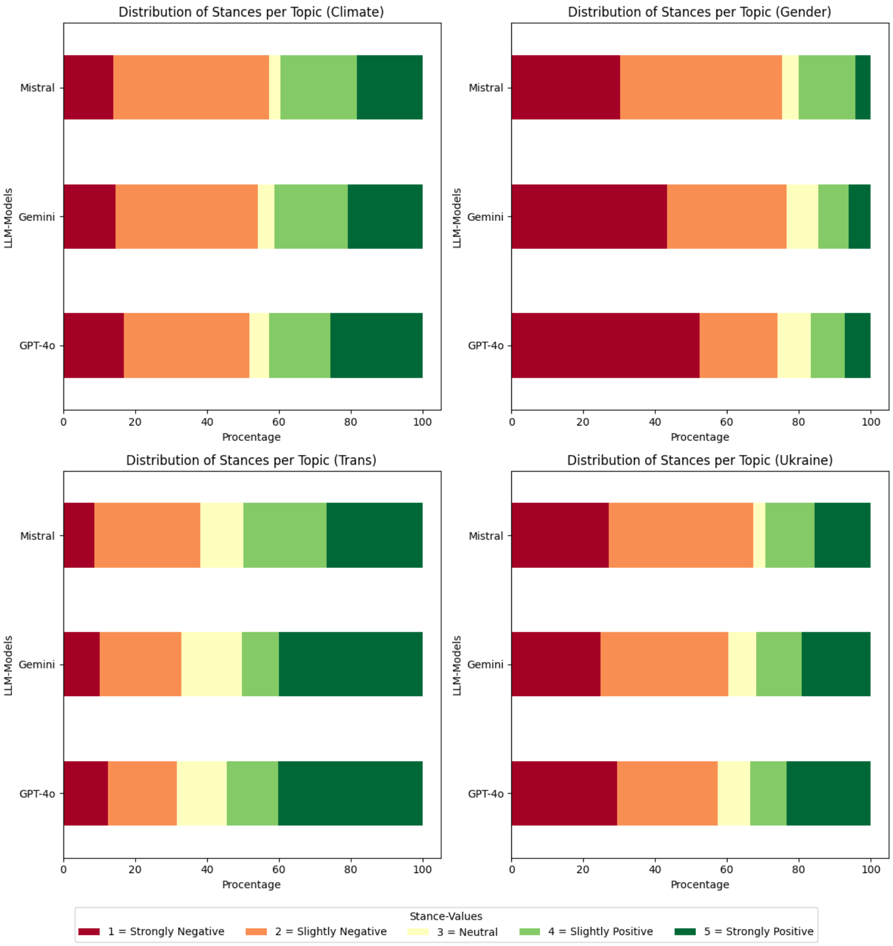
We contribute to this effort by developing a standardized measurement scale to assess the degree of extreme viewpoints in open survey responses collected on a set of high-salience contentious topics, including the war on Ukraine, climate change, inflation, trans rights, and gender-sensitive language. Unlike traditional closed-ended survey methods, we are able to capture nuanced viewpoints by relying on unstructured responses, providing richer data for analysis. To systematically categorize these viewpoints, we employ several Large Language Models (Mistral, Gemini, GPT-4o) to code these viewpoints systematically, as well as benchmarking and developing an automated assessment framework that can be applied at scale. This computational approach allows for consistent application of coding criteria across large datasets while minimizing human resource constraints (DiGiuseppe & Flynn, 2025;).

Preliminary Results

Our analysis develops and applies a systematic computational pipeline to categorize polarized opinions in answers to open-eded survey questions about potentially controversial topics, leveraging 30.919 open-ended survey responses. Using a combination of human-coded annotations and closely monitored Few-Shot Prompting strategies, we evaluate the effectiveness of automated stance detection. In a secondary step of analysis, we will compare these annotation results with the self-assessed closed survery responses provided by the same participants.

Ensuring methodological rigor, we implement a comprehensive validation strategy, though a full validation is still pending due to the preliminary nature of this abstract. As a benchmark for assessing the reliability and validity of the LLM-based coding approach, human parallel coding is performed by three independent coders. Additionally, we evaluate the correspondence between our open response-based scale and traditional standardized survey responses on identical issues, enabling a direct comparison between methodologies. While our preliminary results provide valuable insights, they remain provisional until the validation process is completed.

Figure 1: *Distribution of Stances detected by different LLM models, per topic*



Note. Stance distribution based on a subset of n = 7.915 open-ended survey responses across 4 survey waves.

The preliminary results indicate that responses to high-salience issues exhibit varying degrees of polarization, with notable differences across topics. Our LLM-assisted analysis demonstrates that while some topics, such as gender-sensitive language or Trans Rights, display larger opinion clusterings across all models, others, like the war on Ukraine, show a more diffuse distribution of viewpoints. Thus, our findings underscore the potential of LLM-based approaches in systematically identifying and measuring issue polarization in open-ended survey data. By benchmarking our computational models against human-coded annotations, we ensure reliability while maintaining scalability. This methodological advancement allows for a more granular analysis of public opinion dynamics, moving beyond binary ideological classifications to capture the fluidity and contextual nature of polarization.

To further validate our findings and refine our methodology, we plan a systematic comparison between open-ended and close-ended survey responses. This comparison serves two key purposes: (1) validation, ensuring that automated stance detection aligns with established polarization measures, and (2) investigating whether there are systematic differences in how respondents articulate their opinions depending on the response format. Specifically, we will examine whether individuals express more nuanced or extreme positions in open-ended responses compared to the more structured but potentially restrictive format of closed-ended survey questions. By analyzing these patterns, we aim to uncover potential biases in existing survey methodologies and explore whether the scalable implementation of automated content analysis strategies on open-ended responses could provide deeper insight into latent polarization that may be overlooked in conventional polling.

This study contributed to the broader discourse on the development of valid and reliable tools to measure contentious concepts of the Social Sciences, such as polarization. While still under refinement, this classifier provides a scalable tool for analyzing such latent polarization dynamics. This project aims to produce a validated measurement instrument that researchers can apply to assess issue polarization more effectively. The results will contribute to ongoing debates about whether and how polarization manifests. The methodological advancements offered by this project have potential applications for researchers, policymakers, and media organizations seeking to understand and address the complex dynamics of issue polarization in democratic societies.

References

Kubin, E., & von Sikorski, C. (2021). The Role of (Social) Media in Political Polarization: A Systematic Review. *Annals of the International Communication Association*, 45(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Geer, J. G. (1988). What do open-ended questions measure?. *Public Opinion Quarterly*, *52*(3), 365-367.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics*, *19*(1), 159-177.

DiGiuseppe, M., & Flynn, M. (2025). Scaling Open-ended Survey Responses Using LLM-Paired Comparisons.

Brüggemann, M., & Meyer, H. (2023). When debates break apart: Discursive polarization as a multi-dimensional divergence emerging in and through communication. Communication Theory, 33(2–3), 132–142. https://doi.org/10.1093/ct/qtad012

Mason, L. (2015). “I Disrespectfully Agree”: The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science*, 59(1), 128–145. https://doi.org/10.1111/ajps.12089

Esau, K., Choucair, T., Vilkins, S., Svegaard, S. F., Bruns, A., O'Connor-Farfan, K. S., & Lubicz-Zaorski, C. (2024). Destructive polarization in digital communication contexts: a critical review and conceptual framework. *Information, Communication & Society*, 1-22.

Bruns, A., dos Santos Choucair, T., Esau, K., Svegaard, S. F., & Vilkins, S. (2024). Polarization in online spaces: Distinguishing forms of polarized politics. In *The Routledge Handbook of Political Campaigning* (pp. 45-57). Routledge.

Wojcieszak, M., de Leeuw, S., Menchen-Trevino, E., Lee, S., Huang-Isherwood, K. M., & Weeks, B. (2023). No Polarization From Partisan News: Over-Time Evidence From Trace Data. *The International Journal of Press/Politics*, 28(3), 601–626. https://doi.org/10.1177/19401612211047194

Asker, D., & Dinas, E. (2019). Thinking Fast and Furious: Emotional Intensity and Opinion Polarization in Online Media. *Public Opinion Quarterly*, *83*(3), 487–509. https://doi.org/10.1093/poq/nfz042

Lowe, W., Benoit, K., Mikhaylov, S., & Laver, M. (2011). Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly*, *36*(1), 123–155. https://doi.org/10.1111/j.1939-9162.2010.00006.x

Olechowska, P. (2022). Divisions of Polish Media and Journalists as an Example of Polarization and Politicization. *Journalism Practice*, *16*(10), 2125–2146. https://doi.org/10.1080/17512786.2021.1884991

Cavari, A., & Freedman, G. (2018). Polarized mass or polarized few? Assessing the parallel rise of survey nonresponse and measures of polarization. *The Journal of Politics*, *80*(2), 719-725.

Barberá, P. (2020). Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, 34-55.

Paper 2

EXTENDING OUR CAPABILITIES: LLM-ASSISTED FRAME ANALYSIS OF AUSTRALIAN CLIMATE MOVEMENT NEWS COVERAGE

Laura Vodden
Digital Media Research Centre, Queensland University of Technology

Katharina Esau
Digital Media Research Centre, Queensland University of Technology

Axel Bruns
Digital Media Research Centre, Queensland University of Technology

Frame analysis is a content analysis method that seeks to understand how narratives are shaped and presented to audiences, via their framing in communication. Manual frame analysis is a time- and labour-intensive task, which limits the scope of empirical research (Kuang et al., 2024; Walter & Ophir, 2019). These constraints along with the increasing volume of text-based data have prompted researchers to employ computational methods to analyse data at scale (Kermani et al., 2023; Kroon et al., 2023). Computational methods offer the advantage of processing large volumes of media content, allowing for broader and more comprehensive analyses (Matthes & Kohring, 2008), but currently do not sufficiently capture media frames (Ali & Hassan, 2022), which is a core concept in political, media and communication studies.

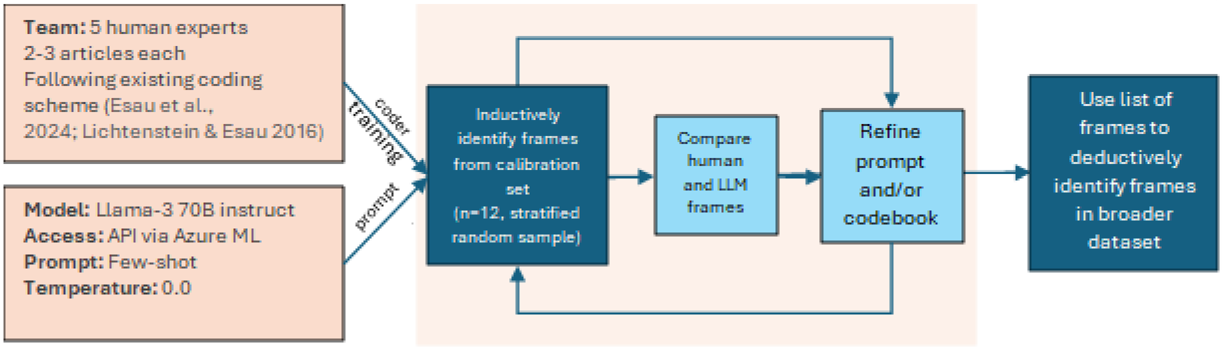
Large Language Models (LLMs) have the potential to bridge this gap and expand the scope of content analysis in general and frame analysis specifically. Several authors (e.g. Alizadeh et al., 2025) have identified the potential value of LLMs in frame analysis, but presently there is no established methodology available for applying LLMs to analyse framing in the news. Our pilot study develops a methodology to incorporate LLMs to aid researchers in analysing the framing of climate movements in news coverage. We find that LLMs can be used to inductively build frames from news content, and, additionally, enhance manual approaches to frame analysis.

We selected 12 mainstream and alternative Australian media outlets with national reach. Articles were sourced from ProQuest between January 2019 and October 2023, containing at least one of the following key terms: “Fridays for Future” OR “School Strike 4 Climate” OR “School Strike for Climate” OR “Extinction Rebellion”. This search yielded a corpus of 3,183 articles. From this corpus, we randomly selected 12 articles (one per media outlet), ensuring coverage across different time periods, for manual coding, while setting the remainder aside for LLM-assisted coding.

Five human coders applied an established qualitative frame analysis approach to the sampled articles, using Entman’s (1993) frame elements and an existing coding scheme from Lichtenstein and Esau (2016). Coders analysed each statement in the articles to identify the speaker (journalist, cited person, or organisation) and recorded any problem definitions, causes, blame attributions, proposed solutions, and solution addressees. The coders then discussed and synthesised their findings to produce a coherent list of the frames present in the articles and used by particular speakers. These synthesised frames were further refined into a distinct set of frame descriptions.

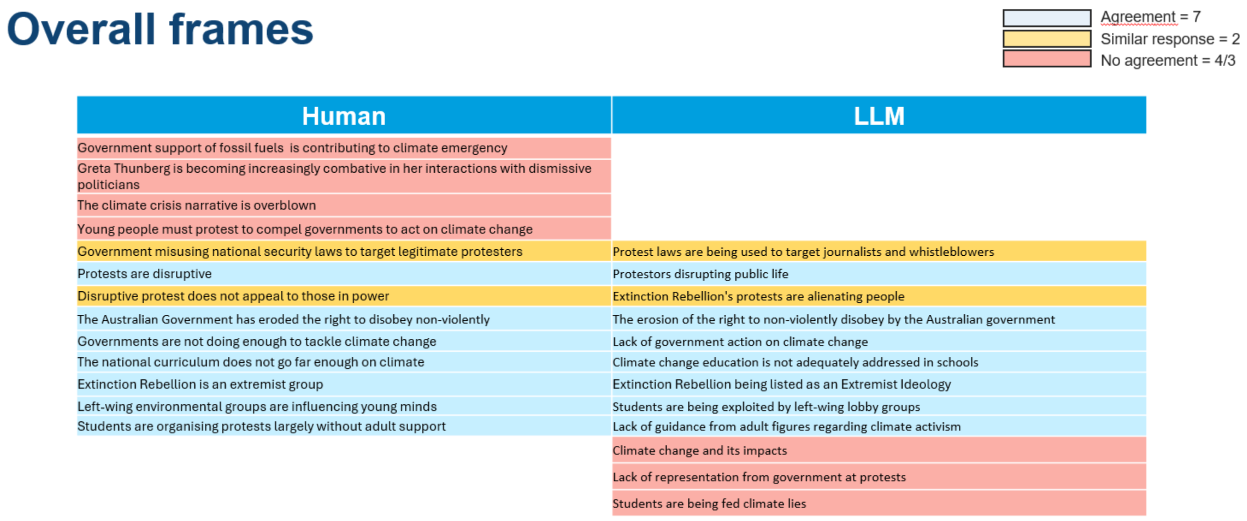
We then applied Meta's Llama-3 language model to replicate this analysis. Following a framework for LLM-assisted content analysis developed by Chew et al. (2023), we iteratively refined the prompt, ran the model, qualitatively compared its results to our manually identified frames, and adjusted the prompt as needed (see Figure 1).

Figure 1: Flowchart demonstrating the process of LLM-assisted frame analysis


Team: 5 human experts
2-3 articles each
Following existing coding scheme (Esau et al., 2024; Lichtenstein & Esau 2016)
Model: Llama-3 70B instruct
Access: API via Azure ML
Prompt: Few-shot
Temperature: 0.0
coder training
prompt
Inductively identify frames from calibration set (n=12, stratified random sample)
Compare human and LLM frames
Refine prompt and/or codebook
Use list of frames to deductively identify frames in broader dataset

In keeping with the process undertaken by the human coders, we divided the process into two stages, requiring two separate prompts. We structured the first prompt to extract the frame elements from the news text by outlining the overall task and providing a summary of frame analysis and a small number of examples (few-shot prompting), then breaking the task further into subtasks (e.g., identifying all problem definitions and corresponding causes and blame attributions from the text). The second prompt contained instructions to synthesise the output of the first prompt, to arrive at a list of frames. We instructed the LLM to withhold responses when uncertain and to provide supporting evidence by extracting relevant excerpts directly from the news articles, fostering traceability throughout the process. Finally, we specified a standardised output format—JSON—to facilitate structured analysis. We verified the existence of each LLM-generated frame within the relevant news article, and compared the frames against those produced by the human codes.

While the human experts identified 13 distinct frames in total, the LLM identified 12. Of these, nine common frames were identified by both the humans and the LLMs. The human coders identified four frames that the LLM did not, and the LLM identified three frames that the human coders did not. The LLM was generally more succinct in its construction of frames than the human experts, but the LLM struggled to identify frames that were specific to a particular issue, and as such the LLM-generated frames were more general in their scope than those identified by the human experts. The human experts did not all construct frames consistently; the style of language and depth of the frames varied widely between human coders, and further work on this project will see our human researchers benefit from a second round of coder training, to address possible inconsistencies in defining a frame and possibly allow for improvement of LLM-generated results as a result of refining the overall approach to this frame analysis.



Overall frames

🟦 Agreement = 7
🟨 Similar response = 2
🟥 No agreement = 4/3

Human	LLM
Government support of fossil fuels is contributing to climate emergency	
Greta Thunberg is becoming increasingly combative in her interactions with dismissive politicians	
The climate crisis narrative is overblown	
Young people must protest to compel governments to act on climate change	
Government misusing national security laws to target legitimate protesters	Protest laws are being used to target journalists and whistleblowers
Protests are disruptive	Protestors disrupting public life
Disruptive protest does not appeal to those in power	Extinction Rebellion's protests are alienating people
The Australian Government has eroded the right to disobey non-violently	The erosion of the right to non-violently disobey by the Australian government
Governments are not doing enough to tackle climate change	Lack of government action on climate change
The national curriculum does not go far enough on climate	Climate change education is not adequately addressed in schools
Extinction Rebellion is an extremist group	Extinction Rebellion being listed as an Extremist Ideology
Left-wing environmental groups are influencing young minds	Students are being exploited by left-wing lobby groups
Students are organising protests largely without adult support	Lack of guidance from adult figures regarding climate activism
	Climate change and its impacts
	Lack of representation from government at protests
	Students are being fed climate lies

Table 1: Frames identified by human coders, and by the LLM, showing where responses overlapped and diverged.

Despite these teething issues, our results indicate the future utility of LLMs in inductively generating frames from news articles. The next step is to refine our master list of frames using the LLM to analyse a much larger corpus using these frames as a guide—offering benefits in terms of increased scale and reduced time that would otherwise be spent on manual coding. We will also conduct further systematic comparisons of frame detection performance between different LLM systems and models, and between zero- and few-shot prompting approaches.

While this study deals with climate change, we also plan to extend this methodology to other topics that are part of our ongoing research project (e.g., transgender communities rights, indigenous rights, abortion). Regardless of topic, LLMs may identify frames within a text that researchers, due to their own biases and preconceptions, or the limitations of coder training, may overlook or misinterpret, taking the use of LLMs in research beyond simply extracting information from content, and towards improving human-led research by causing us to question our own approaches to doing research.

References

Ali, M., & Hassan, N. (2022). A survey of computational framing analysis approaches. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 9335–9348). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.633

Alizadeh, M., Kubli, M., Samei, Z. et al. (2025). Open-Source LLMs for Text Annotation: A Practical Guide for Model Setting and Fine-Tuning. *Journal of Computational Social Science*, *8*(17). https://doi.org/10.1007/s42001-024-00345-9

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. arXiv. http://arxiv.org/abs/2306.14924

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Kermani, H., Makou, A. B., Tafreshi, A., Ghodsi, A. M., Atashzar, A., & Nojoumi, A. (2023). Computational vs. qualitative: Analyzing different approaches in identifying networked frames during the Covid-19 crisis. International Journal of Social Research Methodology, 0(0), 1–15. https://doi.org/10.1080/13645579.2023.2186566

Kroon, A., Welbers, K., Trilling, D., & Van Atteveldt, W. (2023). Advancing automated content analysis for a new era of media effects research: The key role of transfer learning. Communication Methods and Measures, 1–21. https://doi.org/10.1080/19312458.2023.2261372

Kuang, X., Liu, J., Zhang, H., & Schweighofer, S. (2024). Towards Algorithmic Framing Analysis: Expanding the Scope by Using LLMs. Research Square. https://doi.org/10.21203/rs.3.rs-5032053/v1

Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. Journal of Communication, 58(2), 258–279. https://doi.org/10.1111/j.1460-2466.2008.00384.x

Walter, D., & Ophir, Y. (2019). News frame analysis: An inductive mixed-method computational approach. Communication Methods and Measures, 13(4), 248–266. https://doi.org/10.1080/19312458.2019.1639145

Paper 3

HUMAN-AI DYNAMICS: STATE OF THE ART REVIEW AND FUTURE DIRECTIONS FOR LARGE LANGUAGE MODELS USE IN CONTENT ANALYSIS

Tariq Choucair
Digital Media Research Centre, Queensland University of Technology

Ahrabhi Kathirgamalingam
Center for Advanced Internet Studies (CAIS)

Axel Bruns
Digital Media Research Centre, Queensland University of Technology

Laura Vodden
Digital Media Research Centre, Queensland University of Technology

It has been only three years since Baden et al. (2022) presented a comprehensive review on the use of computational methods for text analysis, covering 854 research articles from leading journals that applied quantitative text analysis, and survey responses from the authors of such studies. They identified three key challenges stemming from the perspectives of social scientists regarding computational text analysis methods: a stronger focus on technological advances than on ensuring validity; a misalignment between what computational methods analyse and the more complex measurements that social scientists need; and an overreliance on English-language tools and models. These key challenges add to those already discussed in previous literature (Boumans & Trilling, 2016; Domahidi et al., 2019): the slow adoption of such methods by social science scholars; a lack of awareness or literacy of computational advancements; limited interdisciplinary collaboration with computer scientists; and difficulties in accessing and ethically using large-scale data.

Since then, however, extremely rapid advancements in and widespread adoption of large language models (LLMs) and generative AI (GenAI) technologies and tools have dramatically shifted the landscape. Not only have these challenges and possibilities changed, but the very nature of our interaction with methods seems to have transformed – now, the boundaries between human-led and computationally assisted analysis are increasingly blurred. What this means for content analysis in the investigation of social phenomena – and how this transformation has unfolded – is the focus of this study.

Research Questions and Methodology

We focus on answering the following research questions:

RQ1: What types of content analysis tasks are LLMs being applied to in social science research?

RQ2: How are LLMs being integrated in content analysis workflows (e.g., schema development, inter-coder reliability tests)?
RQ3: What levels of human oversight and intervention are typically implemented when using LLMs for content analysis, and how does this impact the validity and reliability of the findings?
RQ4: What interdisciplinary approaches and methodological innovations are emerging in the realm of LLM-driven content analysis of human-generated text for social science research?
RQ5: How are human-AI interactions changing the work of researchers in LLM-assisted content analysis, and what are the implications for interpretability, agency, and epistemic authority in social science research?

We performed a systematic literature search across major academic databases (Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and arXiv).

Our query was designed to capture a broad range of relevant studies, incorporating various terminologies associated with LLMs and content analysis. The search string included general terms related to LLMs (e.g., “large language models”, “generative AI”), specific processes (e.g., “fine-tuning,”), and specific model names (e.g., “GPT”, “BERT”, “Gemini”). To ensure relevance, we also included terms related to content analysis methodologies (e.g., “discourse analysis”, “thematic analysis”). The query structure required retrieved articles to contain at least one term from the LLM category AND at least one term from the content analysis category, ensuring a focus on the intersection of these fields. To test the inclusiveness of the query, the authors, well versed in the topic, listed from their knowledge ten key publications that should definitely be included. All ten appeared in the search, validating the query. The total number of publications found, for which we retrieved their key information (e.g., title, authors, abstract) was 33,910:

Database	**Results**
Web of Science	3,174
Scopus	7,714
IEEE	2,593
ACM	18,450
arXiv	1,979
Total	**33,910**

We then screened the title, keywords, and abstracts to filter for relevant publications. Articles were classified as relevant or irrelevant for the purpose of this study. Relevant articles were those in which content, narrative, discourse, or any form of textual analysis was performed using, either partially or fully, Large Language Models or any other form of Generative AI on textual corpora composed of human-generated text (including, but not limited to, news articles, interviews, social media posts and comments, and video transcriptions) to interpret, analyse, or investigate a social phenomenon in any way.

To facilitate this classification, as we analysed samples of the data we listed key reasons for the exclusion of irrelevant publications. These were, for example, studies that manually analysed LLM-generated text without using LLMs as tools for analysis (e.g., manual content analysis of LLM outputs), and studies examining user perceptions of LLMs through interviews or surveys rather than applying LLMs for textual analysis (e.g., manual discourse analysis of opinions on AI).

The validation of the relevant/irrelevant classification is still in progress through an intercoder reliability test, but an initial classification performed over a random sample of the material (1%, 339 publications) has yielded 6% relevant and 94% irrelevant publications, which applied to the entire dataset would mean a total of approximately 2,034 relevant publications.

Preliminary Results

As a work in progress, the coding scheme for this literature review is still under development and validation. We began its development by conducting an in-depth close reading of key relevant publications. From this we extracted four initial categories of distinct interactions between researchers and LLMs for content analysis (and related) tasks:

(a) LLMs as Scalable Coders for Deductive Content Analysis

In a similar path (or mode of interaction) to previous automatic content analysis methods, many researchers have been using LLMs to perform deductive content analysis at scale. Chew et al. (2023), for instance, build upon traditional frameworks for content analysis. Their process involves prompt engineering to instruct models in applying a codebook, comparing inter-rater reliability between LLMs and human coders through traditional measures. Similarly, Tai et al. (2024) employ an LLM coding workflow in which the same input text is analysed multiple times across multiple interactions with an LLM, mirroring the variability in human coder decisions.

(b) LLMs as Assistants in Human-Led Hybrid Workflows

Rather than replacing human coders, LLMs are also being used as iterative “collaborators” in qualitative coding workflows, assisting researchers with thematic identification, codebook refinement, and interpretations. Dai et al. (2023) propose an LLM-in-the-loop framework, which integrates human coders and LLMs in a four-step thematic analysis pipeline. The framework introduces a machine coder alongside a human coder in a mutual feedback loop to collaboratively identify themes. De Paoli (2024) takes a similar approach in the domain of inductive thematic analysis, applying six phases of using LLMs to analyse interviews. The study finds that LLMs are particularly effective in generating preliminary theme structures, but struggle at the later stages of analysis.

(c) LLMs as Primary Decision-Makers

Some studies are testing the feasibility of LLMs as standalone decision-makers, allowing models to categorise, annotate, and classify textual data without, or with minimal, human intervention. Pilny et al. (2024) systematically evaluate the comparative efficacy of multiple LLMs in classifying relational uncertainty within text. Their methodology follows a quantitative evaluation framework, wherein the LLMs are assigned classification tasks without any example-based prompts (zero-shot learning). The study showed that while LLMs perform comparably to human coders on straightforward classification tasks, they show greater variance when tasked with context-dependent interpretations. This result aligns with Gunes and Florczak (2025), who examine LLMs' role in multiclass classification of congressional bills and policy documents, employing three different human-involvement scenarios: minimal, moderate, and major human interference. The study assesses model performance, revealing that instruction-tuned LLMs can reach high agreement with human coders in the highest-supervision condition, though they perform significantly worse when used autonomously. These findings emphasise that while LLMs can serve as autonomous decision-makers in structured classification tasks, they require human oversight when handling more socially complex tasks.

(d) LLMs as Semantic Categorisation and Clustering Tools

LLMs are also applied in semantic clustering and content categorization. Marino and Giglietto (2024) propose an LLM-integrated research pipeline for analysing political discourse on social media, using LLM-based embeddings to classify, cluster, and label political content. Their approach combines fine-tuned classification models with unsupervised clustering algorithms. Applied to Facebook posts from the Italian general elections (2018 and 2022), their framework classifies political links, groups similar narratives, and generates descriptive labels, though human researchers must intervene to verify category consistency.

We note that these patterns are preliminary and based on the analysis of a sample of articles from our full dataset. The final paper will present these findings in full detail and identify further gaps and opportunities for the principled and meaningful use of LLMs in content analysis. In doing so we provide further impetus to a rapidly developing methodological conversation in our field.

References

Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Zahedivafa, M., Bermeo, J. D., Korobeynikova, M., & Gilardi, F. (2024). Open-source LLMs for text annotation: A practical guide for model setting and fine-tuning. *Journal of Computational Social Science*, *8*(1), 17. https://doi.org/10.1007/s42001-024-00345-9

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda. *Communication Methods and Measures*, *16*(1), 1–18. https://doi.org/10.1080/19312458.2021.2015574

Boumans, J. W., & Trilling, D. (2016). Taking Stock of the Toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, *4*(1), 8–23. https://doi.org/10.1080/21670811.2015.1096598

Carius, A. C., & Teixeira, A. J. (2024). Artificial Intelligence and content analysis: The large language models (LLMs) and the automatized categorization. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-01988-y

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). *LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding* (arXiv:2306.14924). arXiv. https://doi.org/10.48550/arXiv.2306.14924

Dai, S.-C., Xiong, A., & Ku, L.-W. (2023). *LLM-in-the-loop: Leveraging Large Language Model for Thematic Analysis* (arXiv:2310.15100). arXiv. https://doi.org/10.48550/arXiv.2310.15100

De Paoli, S. (2024). Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. *Social Science Computer Review*, *42*(4), 997–1019. https://doi.org/10.1177/08944393231220483

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Computational Communication Science | Outlining the Way Ahead in Computational Communication Science: An Introduction to the IJoC Special Section on “Computational Methods for Communication Science: Toward a Strategic Roadmap”. *International Journal of Communication*, *13*(0), Article 0.

Gunes, E., & Florczak, C. K. (2025). Replacing or enhancing the human coder? Multiclass classification of policy documents with large language models. *Journal of Computational Social Science*, *8*(2), 31. https://doi.org/10.1007/s42001-025-00362-2

Marino, G., & Giglietto, F. (2024). Integrating Large Language Models in Political Discourse Studies on Social Media: Challenges of Validating an LLMs-in-the-loop Pipeline. *Sociologica*, *18*(2), Article 2. https://doi.org/10.6092/issn.1971-8853/19524

Pilny, A., McAninch, K., Slone, A., & Moore, K. (2024). From manual to machine: Assessing the efficacy of large language models in content analysis. *Communication Research Reports*, *41*(2), 61–70. https://doi.org/10.1080/08824096.2024.2327547

Tai, R. H., Bentley, L. R., Xia, X., Sitt, J. M., Fankhauser, S. C., Chicas-Mosier, A. M., & Monteith, B. G. (2024). An Examination of the Use of Large Language Models to Aid Analysis of Textual Data. *International Journal of Qualitative Methods*, *23*, 16094069241231168. https://doi.org/10.1177/16094069241231168

Törnberg, P. (2023). *How to use LLMs for Text Analysis* (arXiv:2307.13106). arXiv. https://doi.org/10.48550/arXiv.2307.13106

Paper 4

USING LLMs FOR INVESTIGATING THE RELATIONS BETWEEN POLITICAL NARRATIVES AND USERS' INTERACTIONS ON SOCIAL MEDIA IN THE BRAZILIAN CONTEXT

Fabio Giglietto
University of Urbino

Giada Marino
University of Urbino

Bruna Paroni
University of Urbino

Introduction

The 2022 presidential elections in Brazil were marked by intense polarization and the widespread dissemination of misinformation online, both during and after the campaign period (Bastos & Recuero, 2023). This polarization peaked on January 8, 2023, with an attempted coup.

In light of this scenario, our study addresses a growing concern: the rise of destructive political polarization (Esau et al., 2024). Destructive polarization goes beyond ideological disagreement, posing a serious threat to democracy. It involves a communication breakdown between individuals with opposing political views, not merely as a lack of interaction but as a dysfunctional form of communication driven by negative emotions and antagonism toward anyone outside one’s political community.

Emotions are central to this form of political polarization. Research highlights that anger and mistrust toward those outside one’s ideological group often intensify group boundaries (Barnes, 2022; Recuero et al., 2021; Sandvoss, 2020). Social media platforms, particularly Facebook, play a crucial role in this dynamic.

Facebook’s affordances—such as reactions, comments, and shares—enable researchers to measure these negative emotions (Anwar & Giglietto, 2024). Prior studies indicate that comments often reflect intentions to engage in discussion, which can sometimes escalate into incivility (Giglietto et al., 2019; McCosker, 2014). Meanwhile, 'angry' reactions frequently signal frustration or anger directed at opposing content (Anwar & Giglietto, 2024; Eberl et al., 2020; Muraoka et al., 2021).

Our ongoing research builds on these insights by analyzing 'angry' and 'love' reactions alongside comments and shares to track polarized engagement. Angry reactions and comments may indicate negative sentiments toward the 'other,' whereas shares and 'love' reactions typically signify support and alignment with the content (Eberl et al., 2020).

To guide our analysis, we propose the following research question:

RQ. What is the role played by emotions in the amplification of a partisan political post?

Method

As part of the European research project Vera.ai, researchers developed an innovative news alert system to identify links shared in a coordinated manner by actors known for spreading problematic information on Facebook (Giglietto et al., 2023). Analysis of these flagged links revealed repeated and regular dissemination of content related to Brazilian political pages and groups.

We created a graph of Facebook accounts that coordinately shared these links, filtered the top 10% of edges by weight, and used a modularity algorithm to identify communities. This process identified 58 coordinated Facebook pages and groups supporting Bolsonaro.

Over 12 million posts shared by this network between January 1, 2021, and December 31, 2023, were collected. Quantitative analysis measured interactions (comments, shares, love, and angry reactions) to map engagement trends. Time series analysis of love/angry and share/comment ratios was conducted to assess levels of emotional polarization over time and its fluctuations in response to political events. Seven periods of high volatility, mostly in 2021 and 2023, were identified, resulting in the selection of 1,161,126 posts.

We conducted a qualitative analysis using a grounded approach, applying it to a sample of 200 posts with the highest sum of comments, shares, loves, and angry reactions from each period of instability, totaling 1,400 posts. A stratified sample of these posts was categorized for fine-tuning a Large Language Model, resulting in 760 classified posts. The analysis scheme includes three groups of categories: the target, the sentiment, and the typology of the post (see Table 1). We will use this categorized sample as a gold standard to fine-tune an OpenAI gpt-4o-mini model and calculate a regression model to assess if the narrative of the post impacts its amplification.

Table 1. Codebook and Classification Frequencies.

Group of Categories	Category	Frequencies
Target (multiple choices)	Bolsonaro, his family, allies and supporters	422
	The Supreme Federal Court and other public institutions	198
	Armed forces / Military Police	78
	Media Mainstream	116
	Lula, his family, allies and supporters	220
	Other	112
Sentiment	Positive	193

	Negative	507
	Neutral	60
Content Typology	News	547
	User Generated Post	213

Preliminary Findings and Next Steps

We focused on three years of content shared by accounts primarily supporting Bolsonaro (N=58). The angry/love and comment/share ratios range between 1 and -1. Values close to 1 indicate dominance of love reactions and shares, while values near -1 reflect a predominance of angry reactions and comments.

In 2023, following the attempted coup, the share/comment ratio and love/angry ratio shifted significantly, with comments and angry reactions becoming more prominent compared to previous years (see Figure 1).

Volatility analysis identified periods of instability where these ratios diverged significantly from previous patterns. While the ratios remained stable throughout 2022, seven periods of instability were concentrated in 2021 and 2023.

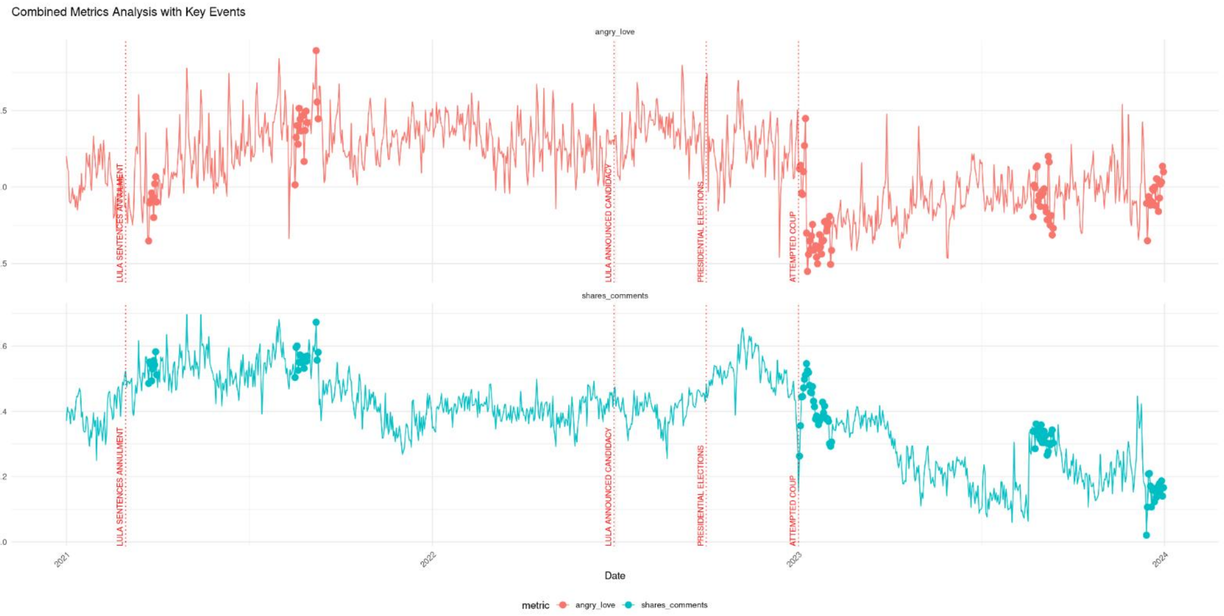


Figure 1 – Time series analysis of angry/love and comment/share ratios.

After having clustered narratives by target, sentiment, and content typology, we will estimate a regression model predicting total interactions as an amplification proxy and using narratives and emotional reactions (comments, shares, angry, and love) as independent variables while controlling for timeframe, the posting account, and the type of content.

The presentation will describe the study findings in light of how LLMs enhance our understanding of the relations between political narratives and Facebook users' reactions regarding emotions or desire to amplify or criticize content, highlighting the methodological approach insights.

References

Anwar, S., & Giglietto, F. (2024). Facebook reactions in the context of politics and social issues: A systematic literature review. *Frontiers in Sociology*, 9. https://doi.org/10.3389/fsoc.2024.1379265

Barnes, R. (2022). Loving to hate: Fandom fueling polarized behavior. In *Fandom and Polarisation in Online Political Discussion* (pp. 61–86). Springer International Publishing.

Bastos, M., & Recuero, R. (2023). The insurrectionist playbook: Jair Bolsonaro and the National Congress of Brazil. *Social Media + Society*, 9(4). https://doi.org/10.1177/20563051231211881

Eberl, J.-M., Tolochko, P., Jost, P., Heidenreich, T., & Boomgaarden, H. G. (2020). What's in a post? How sentiment and issue salience affect users' emotional reactions on Facebook. *Journal of Information Technology & Politics*, 17(1), 48–65. https://doi.org/10.1080/19331681.2019.1710318

Esau, K., Choucair, T., Vilkins, S., Svegaard, S. F. K., Bruns, A., O'Connor-Farfan, K. S., & Lubicz-Zaorski, C. (2024). Destructive polarization in digital communication contexts: A critical review and conceptual framework. *Information, Communication and Society*, 1–22. https://doi.org/10.1080/1369118x.2024.2413127

Giglietto, F., Marino, G., Mincigrucci, R., & Stanziano, A. (2023). A workflow to detect, monitor, and update lists of coordinated social media accounts across time: The case of the 2022 Italian election. *Social Media + Society*. https://doi.org/10.1177/20563051231196866

Giglietto, F., Valeriani, A., Righetti, N., & Marino, G. (2019). Diverging patterns of interaction around news on social media: Insularity and partisanship during the 2018 Italian election campaign. *Information, Communication and Society*, 22(11), 1610–1629. https://doi.org/10.1080/1369118X.2019.1629692

Marino, G., & Giglietto, F. (2024). Integrating large language models in political discourse studies on social media: Challenges of validating an LLMs-in-the-loop pipeline. *Sociologica*, 18(2), 87–107. https://doi.org/10.6092/issn.1971-8853/19524

McCosker, A. (2014). Trolling as provocation: YouTube's agonistic publics. *Convergence*, 20(2), 201–217. http://journals.sagepub.com/doi/abs/10.1177/1354856513501413

Muraoka, T., Montgomery, J., Lucas, C., & Tavits, M. (2021). Love and anger in global party politics: Facebook reactions to political party posts in 79 democracies. *Journal of Quantitative Description: Digital Media*, 1. https://journalqd.org/article/view/2568

Recuero, R., Soares, F., & Zago, G. (2021). Polarization, hyperpartisanship, and echo chambers: How the disinformation about COVID-19 circulates on Twitter. *Contracampo – Brazilian Journal of Communication*, 40(1), 1–16.

Sandvoss, C. (2020). 6. The politics of against: Political participation, anti-fandom, and populism. In *Anti-Fandom* (pp. 125–146). New York University Press. https://doi.org/10.18574/nyu/9781479866625.003.0009