Orchestrating Large Language Models to Construct Interpretable Opinion Spaces for the Analysis and Emulation of Political Discourse

Ian Drumm, University of Salford, United Kingdom
Ser-Huang Poon, University of Manchester, United Kingdom

The Asian Conference on the Social Sciences 2026
Official Conference Proceedings

Abstract

This paper presents a transparent and reproducible pipeline for modelling online political discourse through clustered opinion spaces derived from real Reddit post–comment pairs. Addressing the challenge of mapping increasingly fragmented online ideologies, we use large-scale orchestration of LLMs for ordinal feature scoring and k-medoids clustering with Gower distance to construct interpretable archetypes that capture ideological, moral, and stylistic variation in public debate. These clusters act as filters within retrieval-augmented generation (RAG), enabling synthetic comments grounded in authentic discourse rather than arbitrary persona assumptions. Case studies on tariff debates and stock-market discussion illustrate the method's flexibility in identifying specific discursive tropes within polarized communities. The pipeline offers a scalable approach for studying online discourse, providing interpretable synthetic data and a controlled framework for examining how discursive styles contribute to digital polarization.

Keywords: large language models, opinion spaces, computational social science, retrieval-augmented generation, online political discourse

iafor
The International Academic Forum
www.iafor.org

Introduction

Online political discourse is increasingly fragmented, fast-moving, and difficult to interpret using conventional social media analytics alone. Keyword searches, sentiment analysis, and topic models can identify broad patterns, but they struggle with more nuanced features such as irony, moral framing, ideological implication, rhetorical stance, and the relationship between a post and its replies.

Recent developments in frontier large language models (LLMs) have shown they can have a seemingly uncanny ability to interpret language. Unlike conventional classifiers, LLMs can be prompted to assess multiple dimensions of discourse simultaneously, including dialogic meaning, implied stance, and communicative intent. Although they are not perfect, as this presentation will highlight, their reliability can be comparable to human coders when applied through carefully structured rubrics and reproducible scoring procedures. In this context, they offer a means of converting large volumes of unstructured social media interaction into interpretable feature spaces suitable for clustering, comparison, retrieval, and temporal analysis.

This presentation introduces Opinion-Space Modelling (OSM), an automated pipeline that uses batch-orchestrated LLMs to construct interpretable opinion spaces from thousands of social media post-comment pairs relating to a given topic. These pairs are scored at scale for ideological, moral, policy-related, values-based, and rhetorical features. Each post-comment pair is embedded into a vector database, making it amenable to vector search, while its associated scores are stored as metadata. This enables clustering within a high-dimensional feature space, effectively creating attitudinal groups that can be interpreted, compared, and tracked over time.

The presentation will discuss the development of the pipeline, the validity of the LLM-based scoring process, and two case studies: Reddit discourse on US tariff policy, and stock market sentiment during the 2025 tariff crisis.

Literature Review

Recent work in computational social science argues that large language models offer a new methodological opportunity, not simply as chatbots, but as scalable instruments for text annotation, discourse analysis, and synthetic social simulation. Ziems et al. (2024) describe LLMs as potentially transformative for computational social science because they can support annotation, measurement, and modelling of complex social constructs at scale. Thapa et al. (2025) similarly highlight the growing role of LLMs in social media analysis, including sentiment, stance, misinformation, discourse understanding, and content generation. However, this opportunity also raises methodological risks. LLM outputs can be inconsistent, biased, prompt-sensitive, and difficult to reproduce unless used within a transparent and auditable workflow. Törnberg (2024) and Abdurahman et al. (2025) stress the importance of clear coding schemes, validation against human judgement, and reproducible prompting strategies, particularly when scoring socially and politically sensitive data where models may infer too much from topic context rather than comment content.

The theoretical frameworks underpinning the scoring rubrics draw on well-established traditions in political psychology and social science. Ideological orientation is treated as a left-right summary variable (Jost et al., 2009). Authoritarian attitudes are operationalised through

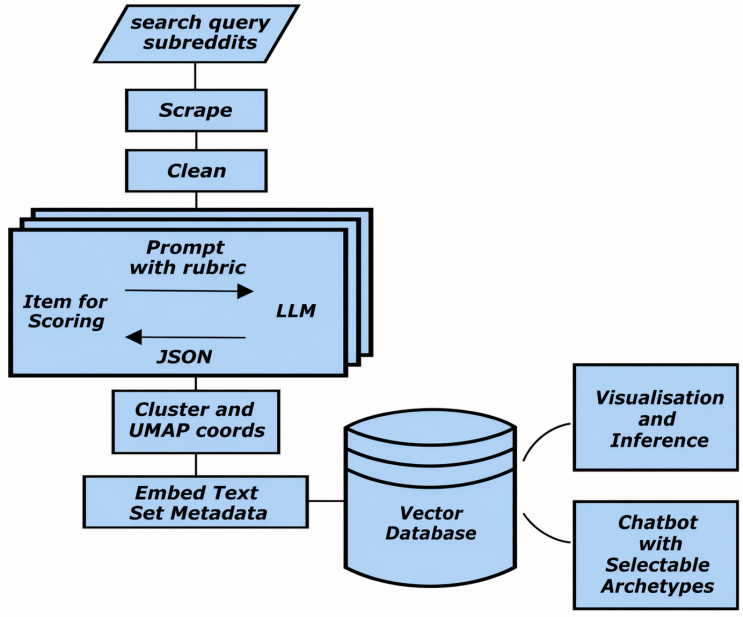
Altemeyer's (1981, 1996) (Dunwoody & Funke, 2016) three-component model of Right-Wing Authoritarianism, and intergroup hierarchy through the dominance and anti-egalitarianism dimensions of Social Dominance Orientation (Ho et al., 2015). Basic human values are drawn from Schwartz's (1992; Schwartz et al., 2012) ten-value circumplex, and intergroup dynamics from Social Identity Theory and Self-Categorisation Theory (Tajfel & Turner, 2000; Turner et al., 1987). Populist discourse is operationalised following ideational approaches that define populism through people-elite antagonism (Mudde & Kaltwasser, 2017). For the tariff case study, domain-specific features reflect the principal frames identified in the trade policy discourse literature, including reciprocity and fairness arguments, economic nationalism, manufacturing protection, and retaliation risk (Hiscox, 2006; Mutz, 2018). For the stock market case study, additional features draw on the financial sentiment literature (Shiller, 2015; Tetlock, 2007), research on retail investor communities and financial populism (Lyócsa et al., 2022), distributional critiques of financialised capitalism (Piketty, 2014), and the political economy of financial regulation (Barber & Odean, 2000; Helleiner, 2014).

The present work builds on Drumm and Tate (2026), which explored how social media post-comment pairs, LLM-based classification, vector databases, and retrieval-augmented generation (Lewis et al., 2020) can be combined to simulate responses shaped by attitudinal profile. The present work extends that framework into a more general methodology for analysing political discourse, modelling attitudinal groups, and generating cluster-conditioned synthetic responses grounded in real social media data.

Methodology

The OSM pipeline comprises six sequential stages: data collection, rubric-guided LLM scoring, embedding and vector database construction, clustering and dimensionality reduction, visualisation with LLM-powered cluster and axes interpretation, and retrieval-augmented generation. Each stage is implemented in Python and designed to be reproducible across different discourse domains and LLM backends.

Figure 1
The Opinion-Space Modelling (OSM) Pipeline



Data Collection

Social media data is collected using platform-appropriate libraries: PRAW for Reddit and atproto for Bluesky. Scraping is configured by specifying target subreddits or accounts, a search query, and a date range. Of the typically tens of thousands of post-comment pairs pulled, typically 3000 will be randomly sampled. Post-comment pairs with links, images, or unreadable text that may impair interpretation are filtered out.

Rubric Design and LLM Scoring

The analytical framework is specified as a structured JSON rubric. Each entry in the rubric defines a feature name, a scale (e.g. -1, 0, and 1), and a natural language description of what is being scored with scale-point definitions. An automatically generated output schema constrains the LLM to produce consistent, parseable JSON outputs of scores and justifications. Note while justifications frequently identify relevant textual evidence and serve a useful diagnostic function, particularly for understanding inter-rater disagreements, there is substantial evidence that LLM-generated explanations are often post-hoc rationalisations rather than faithful accounts of the model's scoring process (Turpin et al., 2023). The model may produce a plausible and textually grounded justification regardless of whether that reasoning causally determined the score. Justifications are therefore treated here as audit trails that support human review and rubric refinement, rather than as independent validation of scoring decisions.

For each post-comment pair, an LLM prompt is constructed automatically by combining the post and comment, instructions to score the comment and use the post as context, the rubric and the output schema. Scoring was performed using Gemini 3 Flash Preview via the Gemini backend, though the pipeline supports multiple backends including Ollama and OpenAI-compatible APIs. At scale, 5,000 post-comment pairs scored against 40 features produces 200,000 individual judgements, achievable in under two hours.

Scoring Validation

Inter-rater reliability was assessed by comparing LLM scores against those produced by a human annotator on a stratified sample of comments. The LLM-versus-human disagreement rate was below 9.7%, comparing favourably with a human-versus-human baseline of below 10.3% on the same material, indicating that LLM scoring is operating within normal human annotator variance. Qualitative analysis of disagreement cases reveals consistent failure patterns: the LLM occasionally attributes features to the broader topic of a post rather than to the comment itself, relies on surface syntactic cues rather than pragmatic meaning, or infers values without sufficient textual evidence. These failure modes inform rubric refinement and are consistent with known limitations reported in the LLM annotation literature.

Embedding and Vector Database Construction

Once scored, each post-comment pair is embedded using Gemini Embedding 001 and stored in a ChromaDB vector database. The embedding captures semantic content, while all rubric scores are stored as metadata fields alongside each post-comment pair. This separation of semantic representation from structured scoring is intentional: it allows the database to support both similarity-based retrieval and metadata-filtered querying independently or in combination.

Clustering and Dimensionality Reduction

The optimum number of clusters was determined empirically by computing the Silhouette Score and Davies-Bouldin Index across a range of candidate values, selecting the solution that maximised the former and minimised the latter. Clustering was performed using Gower distance K-Medoids, which accommodates the mixed feature types present in the rubric (continuous scores and nominal categories) without requiring normalisation assumptions appropriate only to purely numerical data. Medoid selection ensures that each cluster is represented by an actual post-comment pair from the corpus rather than an abstract centroid, supporting both interpretability and downstream retrieval.

The high-dimensional feature space was projected to two dimensions using UMAP for visualisation. Cluster and axis interpretation was then performed using Google Gemini 3 Pro. Rather than passing raw comment text directly to the model, a structured statistical context was first computed for each cluster from the rubric feature scores, comprising central tendency patterns (median position relative to the global distribution), consistency and consensus patterns, polarisation in feature distributions, and the most representative sample comments selected as those closest to the cluster centroid in score space. Pearson correlations between each rubric feature and the two UMAP coordinate axes were also computed and included. This statistical summary was assembled into a structured prompt instructing the model to characterise each cluster as a coherent attitudinal archetype, interpret the UMAP axes as opinion-space spectra based on the feature correlations, and identify the features most instrumental in driving cluster separation. The model functioned as an interpretive layer over pre-computed statistical descriptions rather than performing any analysis itself, and its outputs were treated as summaries subject to researcher review rather than as authoritative findings.

Retrieval-Augmented Generation

The vector database and cluster structure together support a RAG-based comment generation system. A user submits a simulated post and selects one or more cluster archetypes. The system performs a maximum marginal relevance similarity search filtered to the selected cluster metadata, retrieving k post-comment pairs as grounding context. These are passed to an LLM alongside the user's post and an instruction to generate a comment consistent with the attitudinal profile of the selected archetype. This allows synthetic discourse to be generated from specific, empirically grounded regions of the opinion space rather than from demographic assumptions or unconstrained generation.

Results

Two case studies are presented to demonstrate the OSM pipeline across different but related discourse domains: Reddit discussion of US tariff policy, and Reddit discussion of stock market conditions during the same period. Both datasets were collected from politically diverse subreddits and scored using variants of the same ideology rubric adapted to their respective domains.

Case Study 1: US Tariff Discourse

Data were collected using the query “tariffs” across five subreddits (r/conservative, r/liberal, r/libertarian, r/news, and r/worldnews) for the period 1 March to 31 August 2025, yielding

48,105 unique top-level comments, of which 45,416 remained after cleaning; a stratified sample of 3,000 was scored using Google Gemini 3 Flash.

The theoretical frameworks underpinning the scoring rubrics draw on well-established traditions in political psychology and social science. Ideological orientation is treated as a left-right summary variable (Jost et al., 2009). Authoritarian attitudes are operationalised through the three-component model of Right-Wing Authoritarianism (Altemeyer, 1981, 1996; Dunwoody & Funke, 2016), while intergroup hierarchy is captured through the dominance and anti-egalitarianism dimensions of Social Dominance Orientation (Ho et al., 2015).

Basic human values were operationalised across ten dimensions following Schwartz (1992; Schwartz et al., 2012). Intergroup dynamics were captured via ingroup and outgroup reference, evaluation, threat, and boundary policing, grounded in Social Identity and Self-Categorisation theories (Tajfel & Turner, 1979; Turner et al., 1987). Populist discourse was represented through anti-elite sentiment and people-elite antagonism (Mudde & Kaltwasser, 2017). Domain-specific features covered the principal frames in the US tariff debate, including reciprocity and fairness, economic nationalism, manufacturing protection, consumer price harm, supply chain disruption, retaliation risk, trade deal leverage, national security justification, threat framing, and revenue justification (Hiscox, 2006; Mutz, 2018).

Cluster evaluation metrics identified five as the optimum number of clusters. The resulting opinion space was projected on a 2D UMAP as shown in Figure 2. Table 1 shows cluster meanings and Table 2 shows the axes meanings interpreted by Gemini 3 Pro.

Figure 2

2D UMAP Projection of the Opinion Space Pertaining to Tariffs With Axes and Cluster Interpretations

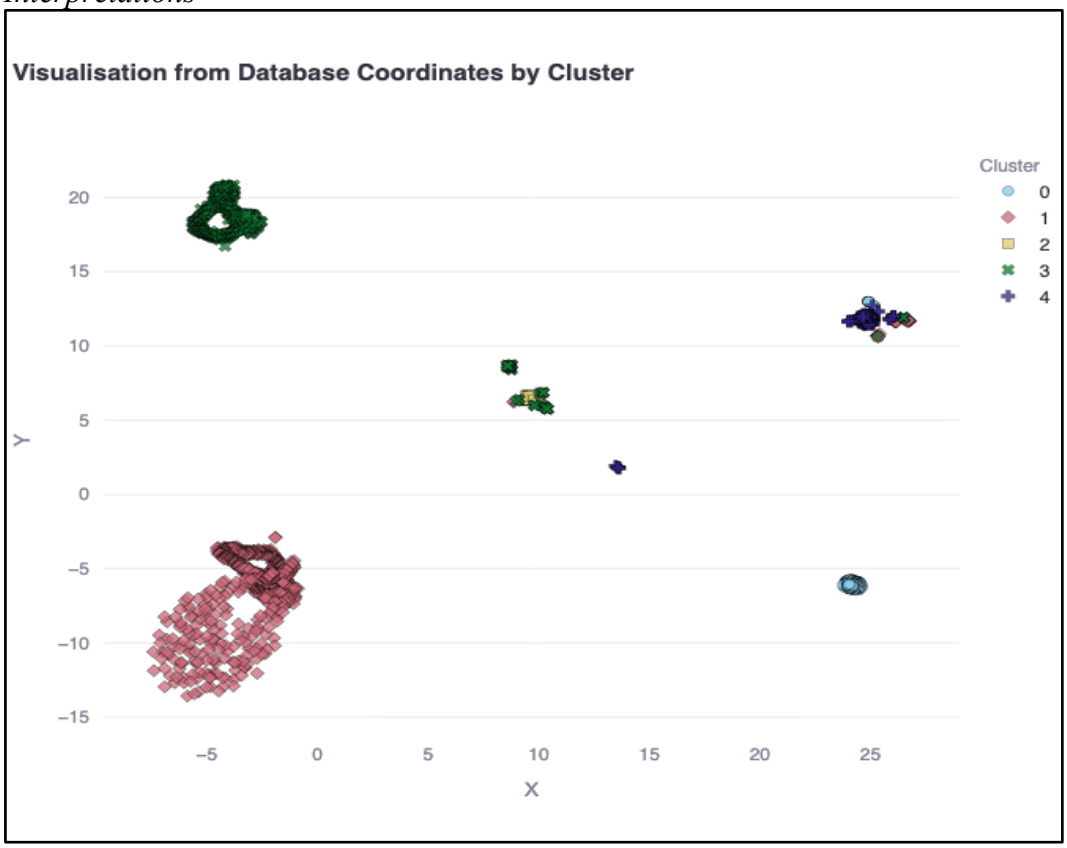


Table 1
LLM Interpretation of Clusters Pertaining to Tariffs

Cluster	Label	%	Distinguishing Characteristics
0	Consumer Price Critics	9.4%	Opponents focused on inflation and the domestic cost of living.
1	Blunt Policy Opponents	41.5%	Large group expressing non-specific, often dismissive opposition to tariffs.
2	Nationalist Proponents	2.8%	Vocal supporters framing tariffs as essential for economic sovereignty.
3	Skeptical Observers	35.9%	Neutral or ambivalent commenters focused on general political exhaustion.
4	Trade War Realists	10.4%	Opponents primarily concerned with foreign retaliation and global isolation.

Table 2
LLM Interpretation of UMAP Axes Pertaining to Tariffs

Axis	Interpretation
X	Spectrum from diffuse ideological or affective response to concrete material and geopolitical consequence framing.
Y	Spectrum from direct policy opposition to political detachment / ideological reframing.

The X-axis represents a spectrum from diffuse tariff reaction to more specific consequence-based reasoning. At one end are comments expressing general ideological, affective, or dismissive responses to tariff policy and at the other are comments anchored in concrete economic or geopolitical consequences, such as consumer prices, retaliation, supply-chain disruption, or international isolation. The Y-axis separates direct, active anti-tariff opposition from comments that are less immediately oppositional and more politically detached, ambivalent, or reframed around broader national-interest arguments. Although nationalist pro-tariff comments occupy part of this upper space, the highest-Y region is better understood as sceptical or detached observation rather than committed support.

Five cluster archetypes were identified. The largest group, Blunt Policy Opponents (41.5%), express non-specific, often dismissive opposition without engaging closely with the mechanism of harm. Skeptical Observers (35.9%) are neutral or ambivalent commenters characterised by political exhaustion rather than committed opposition or support. Trade War Realists (10.4%) oppose tariffs on the specific grounds of foreign retaliation and global isolation. Consumer Price Critics (9.4%) focus more narrowly on inflation and domestic cost-of-living impacts. The smallest cluster, Nationalist Proponents (2.8%), comprises vocal

supporters who frame tariffs as essential to economic sovereignty, representing the most ideologically coherent pro-tariff voice in the dataset despite their small share of the discourse.

Case Study 2: Stock Market Discourse

Data were collected using the query "stock market" across seven subreddits (r/wallstreetbets, r/investing, r/stocks, r/StockMarket, r/Bogleheads, r/economics, and r/personalfinance) for the period 1 March to 31 August 2025, yielding 43,774 unique top-level comments, of which 40,668 remained after cleaning; a stratified sample of 3,000 was scored using Google Gemini 3 Flash.

The rubric retained the full core framework described above and added thirteen domain-specific features covering the principal attitudinal dimensions of stock market discourse. Directional sentiment was captured through market outlook, crash or collapse fear, and speculative bubble framing (Shiller, 2015; Tetlock, 2007). Populist and conspiratorial dimensions of retail investor communities were covered by market manipulation allegations, retail-versus-institutional framing, and Wall Street hostility (Lyócsa et al., 2022). Federal Reserve and monetary policy blame was included as a distinct attribution frame, separable from partisan blame of elected actors. Wealth inequality critique drew on distributional critiques of financialised capitalism (Piketty, 2014), while market regulation support and free market or deregulation endorsement captured opposing positions in the political economy of financial markets (Helleiner, 2014). Long-term resilience framing and economic fundamentals reasoning distinguished analytically grounded discourse from sentiment-driven commentary (Barber & Odean, 2000). Market participation advocacy and alternative asset endorsement captured explicit investment advice and exit-from-equities framing common in retail investor communities.

Cluster evaluation identified five clusters, producing an opinion space whose axes reflect distinct dimensions of financial discourse. The resulting opinion space was projected on a 2D UMAP as shown in Figure 3. Table 3 shows cluster meanings and Table 4 shows the axes meanings interpreted by Gemini 3 Pro.

Figure 3
2D UMAP Projection of the Opinion Space Pertaining to Stock Market Sentiment

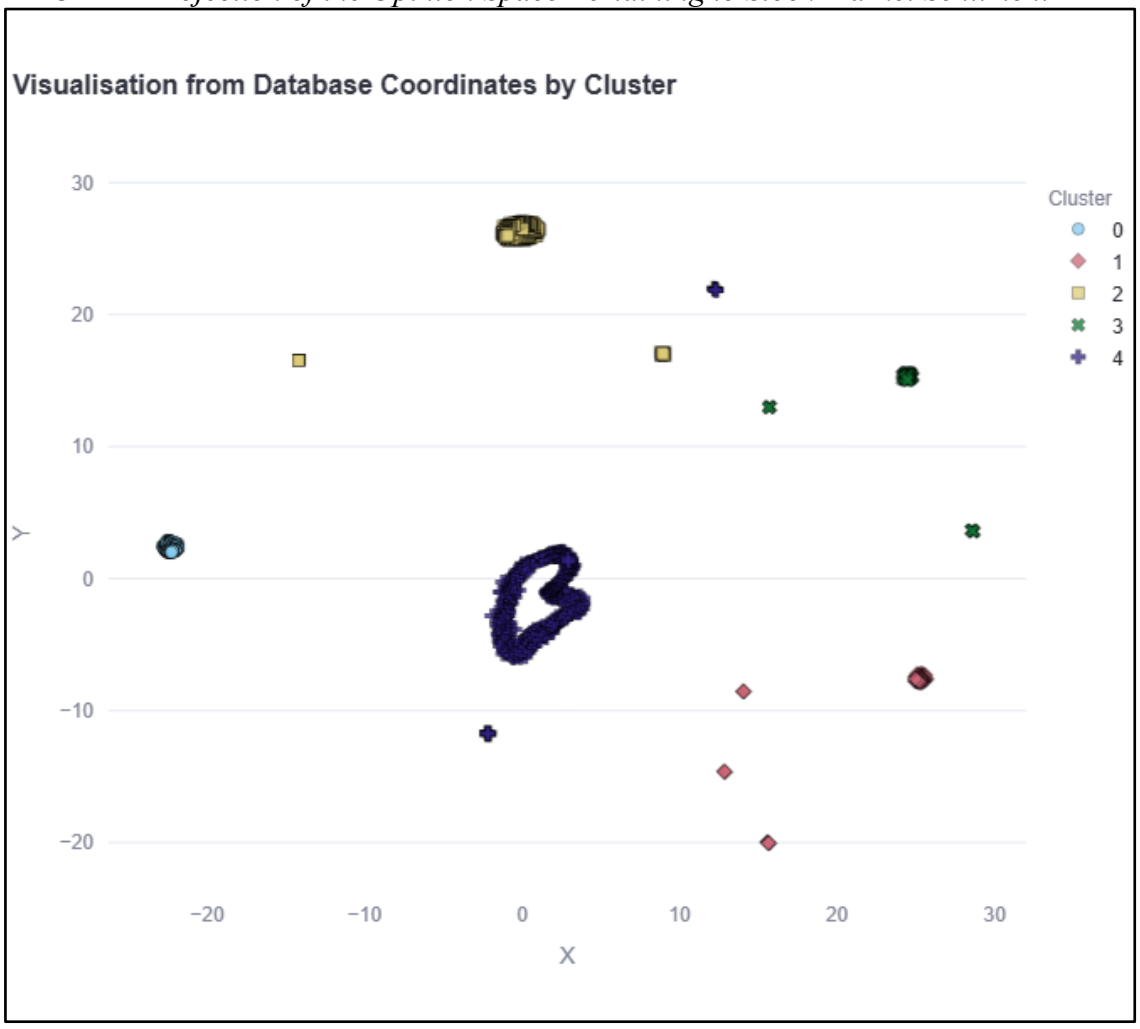


Table 3
LLM Interpretation of Clusters Pertaining to Stock Market Sentiment

Cluster	Label	%	Distinguishing Characteristics
0	Optimistic Investors	10.9%	This cluster actively encourages market participation, devoid of crash fear or political attribution.
1	Pure Doomsayers	9.1%	This cluster unanimously fears a market crash, with minimal political attribution or manipulation allegations.
2	Political Blamers (Neutral Market)	23.0%	This cluster unanimously attributes market conditions to politics, without crash fear or strong advocacy.

3	Political Doomsayers	8.0%	This cluster combines unanimous crash fear with political blame, often expressed verbosely.
4	Market Observers (Neutral)	48.9%	This largest cluster offers general market observations, free from strong opinions or political links.

Table 4
LLM Interpretation of UMAP Axes Pertaining to Stock Market Sentiment

Axis	Interpretation
X	Optimistic market engagement to pessimistic crash fear
Y	Degree of political attribution for market conditions

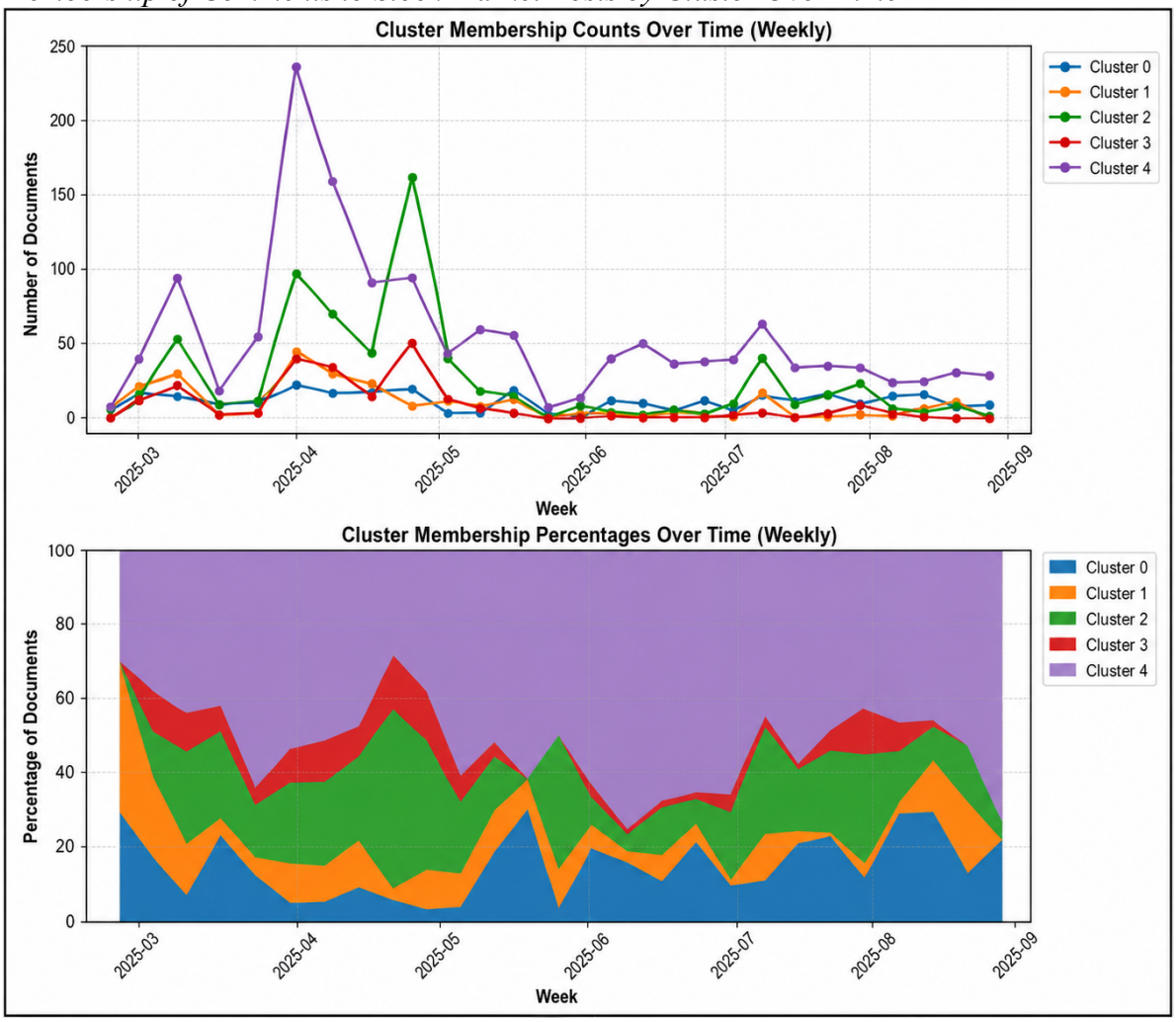
The X-axis broadly separates market-engagement discourse from crash-anxiety discourse, ranging from comments encouraging participation or expressing confidence in the market to those centred on pessimism, collapse, or financial risk. The Y-axis captures the degree of political attribution, distinguishing comments that treat market conditions as general financial events from those that explicitly connect them to political actors, policy decisions, or partisan blame.

The dominant cluster, Market Observers (48.9%), offer general commentary free from strong opinions or political framing. Political Blamers (23.0%) attribute market conditions to political actors without expressing strong crash fear or investment advocacy. Optimistic Investors (10.9%) actively encourage market participation and are largely absent of both crash fear and political attribution. Pure Doomsayers (9.1%) fear a crash but without political framing, distinguishing them clearly from Political Doomsayers (8.0%), who combine crash fear with political blame and tend toward more verbose and emotionally charged expression.

Temporal Analysis

Mapping cluster membership across weekly posting volumes between February and August 2025 reveals a clear three-phase pattern in how the Reddit community responded to the tariff crisis. Figure 4 shows the population of clusters in weekly time bins.

Figure 4
Membership of Comments to Stock Market Posts by Cluster Over Time



Mapping cluster membership across weekly posting volumes between March and August 2025 suggests a three-phase pattern in how Reddit communities responded to the tariff-related market shock. In the weeks before Liberation Day on 2 April, posting volumes were already elevated and were dominated by neutral observation, indicating anticipatory attention rather than widespread panic. During the acute crash period of 2 to 14 April, markets fell sharply, but neutral observers continued to account for a substantial share of the discourse. Crash-fear and political-blame clusters became more visible, but they did not simply track market losses in a linear way.

Political-blame discourse appears to peak in the week of 21 April, after markets had begun to recover. This suggests a possible lag between the immediate affective response to market disruption and the later attribution of responsibility to political actors. Rather than panic and blame emerging simultaneously, the pattern indicates that emotional and interpretive responses to economic shocks may operate on different timescales, with political framing consolidating after the initial market reaction.

By summer, weekly posting volumes had fallen substantially, and neutral observation accounted for 73% of cluster membership. This indicates rapid disengagement once the acute phase of the crisis had passed. Overall, the episode suggests that politically salient economic

events can produce a short-lived but structured burst of discourse, which then recedes toward a more neutral baseline, while smaller ideologically distinctive clusters remain present but proportionally limited.

Downstream Evaluation, RAG Chatbot

Cluster IDs can be used as positive filters on the knowledge base of embedded post-comment pairs, enabling a retrieval-augmented generation (RAG) chatbot in which a target archetype is selected and a new comment is generated in response to a new post by retrieving post-comment pairs with the Maximal Marginal Relevance (Carbonell & Goldstein, 1998) from that cluster and adding them to an LLM prompt. This positive filtering approach is described by Drumm and Tate (2026), where categorised comments are used to simulate responses to environmental messaging. Pseudo comment generation offers a means of downstream evaluation of the scoring and clustering pipeline presented in this paper. Figures 5 and 6 show outputs from the RAG chatbot. Initial line-up tests suggest that generated comments are difficult to distinguish from real ones pertaining to a given archetype, though systematic evaluation of this claim remains ongoing for future publication.

Figure 5
RAG Chatbot Emulating the Nationalist Proponents Cluster

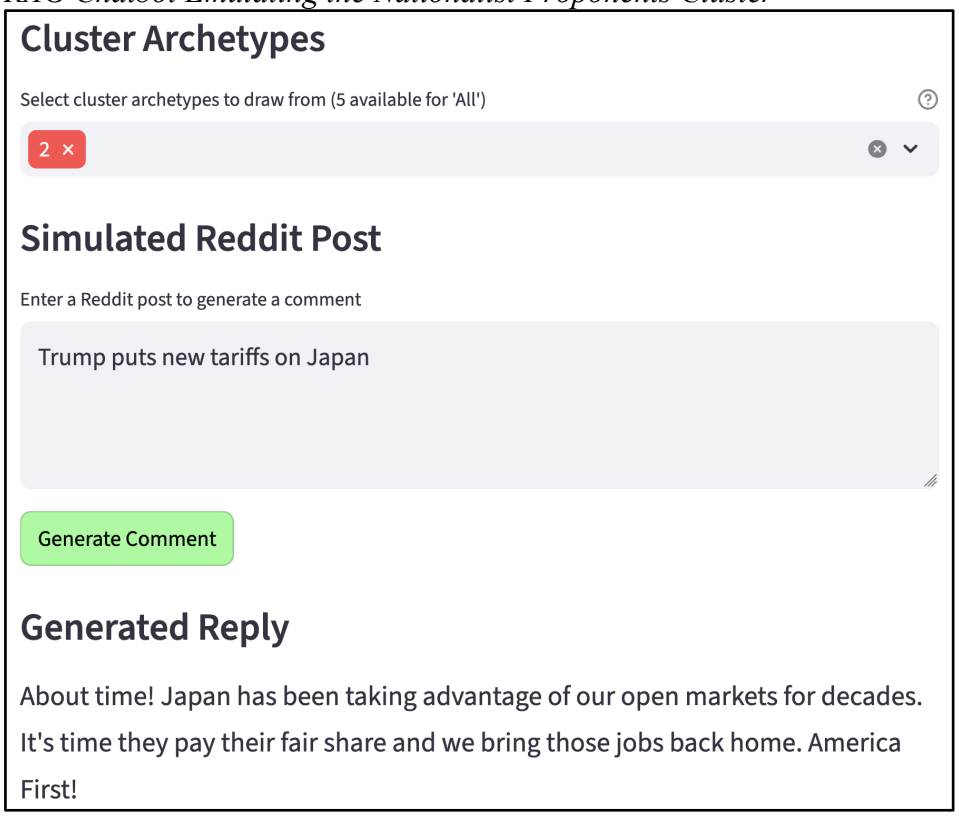
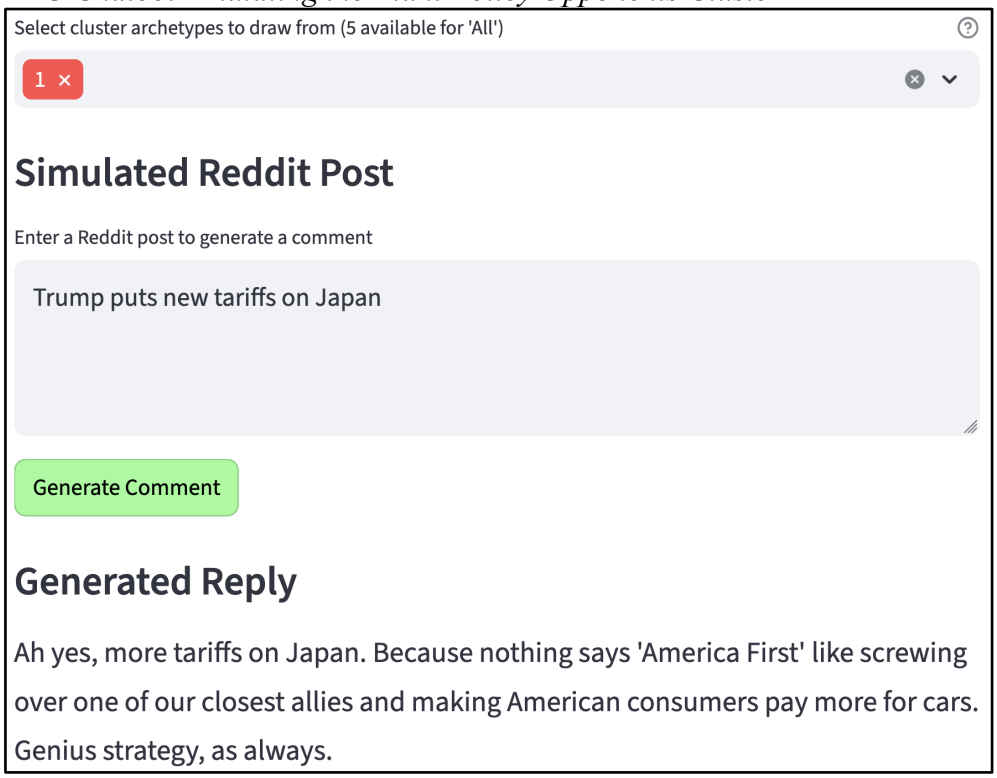


Figure 6
RAG Chatbot Emulating the Blunt Policy Opponents Cluster



Discussion

The two case studies demonstrate that OSM can extract structured, interpretable attitudinal landscapes from large volumes of unstructured social media discourse. This section reflects on the approach's strengths, limitations, ethical considerations, and future directions.

The most significant strength of the OSM pipeline is its ability to operate at a scale and analytical depth that neither manual coding nor conventional computational methods can easily achieve simultaneously. Manual annotation of 3,000 comments against 40+ features is not practically feasible within typical research timelines; keyword and sentiment approaches can achieve scale but sacrifice the theoretically grounded feature set that OSM applies. The combination of structured rubric design with LLM scoring occupies a productive middle ground, inheriting the interpretability of codebook-driven content analysis while gaining the scalability of large language models.

The rubric-as-instrument approach also offers methodological transparency that unconstrained LLM use does not. Because every feature is explicitly defined and every score accompanied by a natural language justification, the framework is auditable in a way that black-box classification is not, enabling targeted rubric refinement and principled error analysis.

The validity results are encouraging. With LLM-versus-human disagreement below 9.7% against a human-versus-human baseline of 10.3%, LLM scoring is operating within normal inter-annotator variance, suggesting the structured rubric is doing meaningful work in constraining outputs.

Gower distance K-Medoids clustering handles the mixed feature types in the rubric natively, and medoid selection ensures every archetype is grounded in a real document rather than an abstract centroid. The temporal analysis adds a further dimension, detecting not only what attitudinal groups exist but when they are active relative to external events. The RAG generation component extends the pipeline from analysis into simulation, anchoring synthetic outputs to real discourse from specific regions of the opinion space rather than to demographic proxies.

LLM scoring, however carefully structured, remains a form of model-generated annotation rather than direct measurement. The LLM exhibits systematic failure modes, including attributing features to the broader topic of a post rather than to the comment itself, relying on surface-level syntactic cues, and inferring values or attitudes without sufficient textual evidence. Hence the approach presented in this paper should be understood as a method for large-scale statistical and interpretive analysis rather than as a source of absolute or individually definitive classifications.

Reddit's demographic skew means the opinion spaces reflect a particular online community rather than the general public and the use of publicly available social media data raises ethical considerations that warrant acknowledgement. Although Reddit comments are publicly accessible, users may not anticipate that their posts will be subject to systematic ideological analysis. No personally identifiable information was retained in the pipeline, and all analysis was conducted at the aggregate level of clusters rather than individual users. The generation of synthetic comments conditioned on real discourse raises a distinct set of concerns: outputs that are indistinguishable from authentic comments could in principle be misused for astroturfing, influence operations, or the artificial amplification of particular viewpoints. The present work is oriented toward analysis and evaluation rather than deployment. But the dual-use potential of archetype-conditioned generation is a genuine risk.

Conclusion

The Opinion Space approach offers a reproducible, scalable framework for discourse analysis that takes ideological, moral, and rhetorical complexity seriously as analytical targets. The rubric-as-instrument model bridges the interpretive depth of qualitative coding and the scale requirements of large dataset analysis, and the methodology is domain-agnostic, i.e. the same pipeline that maps tariff opinion can be redeployed to immigration, climate, or any politically contested domain. The RAG generation component offers a principled approach to synthetic social data and provides some downstream validation.

The case studies demonstrate practical viability. Five empirically grounded archetypes were recovered from the tariff corpus revealing a discourse landscape dominated by diffuse opposition and political exhaustion with a small but coherent nationalist minority. The temporal analysis of stock market discourse showed that political blame and emotional panic operate on distinct timescales, a finding difficult to recover from sentiment trends alone.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

Claude Sonnet 4.6 was used for proof reading, tidying and reference sorting.

References

Abdurahman, S., Salkhordeh Ziabari, A., Moore, A. K., Bartels, D. M., & Dehghani, M. (2025). A primer for evaluating large language models in social-science research. Advances in Methods and Practices in Psychological Science, 8. https://doi.org/10.1177/25152459251325174

Altemeyer, B. (1981). *Right-wing authoritarianism*. University of Manitoba Press.

Altemeyer, B. (1996). *The authoritarian specter*. Harvard University Press.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, *55*(2), 773–806. https://doi.org/10.1111/0022-1082.00226

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 335–336. https://doi.org/10.1145/290941.291025

Drumm, I., & Tate, A. (2026). Simulating human responses to environmental messaging. Journal of Computational Social Science, 9, Article 23. https://doi.org/10.1007/s42001-025-00453-0

Dunwoody, P. T., & Funke, F. (2016). The Aggression-Submission-Conventionalism Scale: Testing a new three factor measure of authoritarianism. Journal of Social and Political Psychology, 4(2). https://doi.org/10.5964/jspp.v4i2.168

Helleiner, E. (2014). *The status quo crisis: Global financial governance after the 2008 meltdown*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199973637.001.0001

Hiscox, M. J. (2006). Through a glass and darkly: Attitudes toward international trade and the curious effects of issue framing. *International Organization*, *60*(3), 755–780. https://doi.org/10.1017/S0020818306060255

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. https://doi.org/10.1037/pspi0000033

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual Review of Psychology*, *60*, 307–337. https://doi.org/10.1146/annurev.psych.60.110707.163600

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474. https://doi.org/10.5555/3495724.3496517

Lyócsa, S., Baumöhl, E., & Výrost, T. (2022). YOLO trading: Riding with the herd during the GameStop episode. *Finance Research Letters*, *46*, 102359. https://doi.org/10.1016/j.frl.2021.102359

Mudde, C., & Rovira Kaltwasser, C. (2017). *Populism: A very short introduction*. Oxford University Press. https://doi.org/10.1093/actrade/9780190234874.001.0001

Mutz, D. C. (2018). Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences*, *115*(19), E4330–E4339. https://doi.org/10.1073/pnas.1718155115

Piketty, T. (2014). *Capital in the twenty-first century* (A. Goldhammer, Trans.). Belknap Press of Harvard University Press. https://doi.org/10.2307/j.ctt6wpqbc

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. https://doi.org/10.1037/a0029393

Shiller, R. J. (2015). *Irrational exuberance* (Rev. and expanded 3rd ed.). Princeton University Press. https://doi.org/10.2307/j.ctt1287kz5

Tajfel, H., & Turner, J. C. (2000). *An integrative theory of intergroup conflict*. In M. A. Hogg & D. Abrams (Eds.), Intergroup relations: Essential readings (pp. 56–65). Psychology Press / Oxford Academic. https://doi.org/10.1093/oso/9780199269464.003.0005

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thapa, S., Shiwakoti, S., Shah, S. B., Adhikari, S., Veeramani, H., Nasim, M., & Naseem, U. (2025). Large language models (LLM) in computational social science: Prospects, current state, and challenges. Social Network Analysis and Mining, 15(1), Article 4. https://doi.org/10.1007/s13278-025-01428-9

Törnberg, P. (2024). Best Practices for Text Annotation with Large Language Models. Sociologica, 18(2), 67–85. https://doi.org/10.6092/issn.1971-8853/19461

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems, 36*, 74952–74965. https://doi.org/10.5555/3666122.3669397

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science? *Computational Linguistics*, *50*(1), 237–291. https://doi.org/10.1162/coli_a_00502

Contact email: i.drumm@salford.ac.uk