Ideological Divergence in UK Political News Across Election Periods and Political Crises (2015–2024): A Computational Topic Modelling Study

Iola Agbalaya-Mason and Rahime Belen Saglam

School of Architecture, Computing and Engineering,
University of East London, London, United Kingdom

{iolaagbalayamason@gmail.com, r.belen-saglam@uel.ac.uk}

Abstract. This study investigates semantic patterns of ideological divergence in UK political news media between 2015 and 2024, with particular attention to major political events and election cycles. A corpus of 124,693 articles from six national newspapers was analysed using transformer-based semantic representations, BERTopic semantic clustering, and RoBERTa-based headline sentiment analysis. Rather than treating media bias solely as evaluative tone, the study models ideological bias as a semantic phenomenon observable through differences in topic salience, semantic framing structures, and temporal topic evolution across politically aligned media ecosystems. To capture both stable and event-specific dynamics, the corpus was segmented into 17 politically coherent time bins aligned with major developments including Brexit, the COVID-19 pandemic, leadership crises, and four UK general elections.

Three findings emerge. First, ideological divergence manifests primarily through semantic topic selection rather than headline sentiment, with left-leaning outlets disproportionately emphasising UK Politics and Brexit-related themes, while right-leaning outlets prioritise Scottish & Regional and International Politics. Second, the largest semantic divergences occur during periods of political crisis and instability rather than formal election campaigns, supporting conditional agenda-setting theory. Third, a dual-scale semantic modelling framework combining full-corpus and per-bin BERTopic analysis reveals transient scandal-related topics that are obscured at aggregate scale. The study contributes a longitudinal semantic analysis of UK political news framing and proposes a dual-scale semantic modelling approach for capturing both persistent and event-driven structures within large political media corpora.

Keywords: Media bias · BERTopic · Agenda-setting · UK elections · Topic modelling · Sentiment analysis · Political communication

1 Introduction

Political news media are integral to how most people understand politics. As elections approach and political divides deepen, news reporting can highlight particular

ideological viewpoints and bring certain topics to the forefront. The World Economic Forum's Global Risks Report lists misinformation, disinformation, and societal polarisation among the most significant short-term global risks, noting that advances in information technology have reduced trust and heightened both polarisation and fragmentation [1]. These risks appear on the report's medium- and long-term horizons as well, suggesting they are structurally embedded within the global risk landscape.

Within the UK, longitudinal analysis of Brexit-related discussions has highlighted increasing affective polarisation and ideological clustering, particularly around key political moments rather than through gradual accumulation [2]. News consumption habits have also shifted: the Reuters Institute's Digital News Report highlights consistent news-avoidance trends, increasing reliance on social and video platforms among young people, and general distrust of AI-generated political information [3]. These developments reduce the likelihood that citizens encounter politically diverse coverage organically, increasing the importance of systems capable of analysing how political information is semantically framed and prioritised across media ecosystems.

Although political media bias has been the subject of public discourse for many years, systematic computational analysis of UK political news coverage across multiple outlets and election cycles remains limited. Large-scale manual studies such as the Loughborough University media audits [4] analyse content in great detail but are constrained in breadth and temporal granularity. The computational literature, meanwhile, is predominantly US-focused and tends to examine either single outlets, narrow bias measures, or single timepoints [5]. Research combining longitudinal semantic analysis, computational topic modelling, and cross-outlet analysis of UK political news remains comparatively underexplored.

This study addresses these gaps by examining how semantic topic emphasis and headline sentiment differ between left- and right-leaning UK news outlets across the 2015–2024 period, encompassing four general elections and multiple major political crises. A corpus of 124,693 articles from six national newspapers is analysed using transformer-based semantic representations, BERTopic semantic clustering, and a RoBERTa-based sentiment classifier, structured around 17 event-based time bins defined by major political developments rather than arbitrary calendar divisions. The study addresses two research questions:

RQ1: How do semantic topic distributions differ between left- and right-leaning UK news outlets over time, particularly across election periods and major political events?

RQ2: What political conditions and events are associated with heightened semantic divergence between ideologically opposed news outlets?

The study makes three contributions: (1) a longitudinal computational analysis of UK political news framing spanning four election cycles, addressing a gap in sustained temporal bias analysis identified by Rodrigo-Ginés et al. [5]; (2) evidence that ideological divergence operates primarily through semantic topic selection (agenda-setting) rather than evaluative tone; and (3) a dual-scale semantic modelling approach combining full-corpus and per-bin analysis to surface both stable structural patterns and transient event-specific topics within large political media corpora.

2 Related Work

2.1 Political Media Bias and Agenda-Setting

Rodrigo-Ginés, Carrillo-de-Albornoz and Plaza's [5] systematic review of 160 media bias studies codifies five common dimensions: selection (agenda), coverage (amount), framing (interpretation), statement or wording (tone), and visual (composition). The review identifies temporal and sustained bias analysis as understudied in the computational literature, motivating the present study's ten-year longitudinal design.

The agenda-setting framework is central to understanding selection and coverage bias. McCombs and Shaw's [6] foundational theory posits that the media affect which issues the public considers important—not through persuasion but through cues about what deserves attention. The distinction between issue salience and evaluative tone has direct implications: if outlets differ primarily in which topics they emphasise rather than in the sentiment with which they report, selection and coverage bias are more powerful mechanisms of ideological differentiation than tonal framing.

Geiß [7] demonstrates that agenda-setting effects are strongly context-dependent: they are more likely to occur and are more pronounced when issue salience spikes suddenly. This conditional model is particularly relevant to election-period analysis and directly motivates the event-based temporal segmentation used here. Berk [8] further shows that framing effects erode in broader information ecosystems, supporting the present study's focus on observable media outputs rather than inferred audience effects.

2.2 Computational Topic Modelling for News Analysis

Egger and Yu [9] assessed four topic modelling techniques and found that BERTopic and NMF produce more coherent topics than LDA and Top2Vec. Liu [10] extends this to news articles, concluding that BERTopic produces more semantically rich topics while noting LDA's advantages in compact structures. This finding is particularly useful for political news analysis, where ideologically aligned outlets may discuss similar issues using different vocabulary while sharing underlying thematic structures.

Al-Zaman and Or Rashid [11] applied BERTopic to 8,074 Rohingya-related articles from ten countries' leading newspapers, demonstrating meaningful framing differences via topic proportion disparities. Verbytska [12] applied BERTopic alongside NER and sentiment analysis to the Russia–Ukraine war, explicitly framing topic modelling as a tool for studying how media organisations construct different narratives. Cobo et al. [13] propose automatic temporal framing detection using text mining and community detection, treating frames as dynamic processes that change depending on the political environment. This directly supports the present study's use of political events as delimiters for time bins.

2.3 Sentiment Analysis and Bias Detection

Sentiment analysis is commonly used in computational studies of political media as a proxy for evaluative framing, although researchers distinguish sentiment from broader forms of ideological bias [5]. Transformer-based sentiment classifiers fine-tuned on short-form text have demonstrated strong performance on news headlines, whose compressed and attention-oriented structure resembles social media language. Beyond sentiment analysis, recent work has explored sentence-level bias detection using transformer-based models trained on annotated political corpora [14, 15, 16]. However, these approaches often require substantial domain-specific calibration and focus primarily on local linguistic markers rather than broader patterns of thematic emphasis. Accordingly, the present study uses headline sentiment as a complementary measure alongside longitudinal topic modelling rather than as a standalone indicator of ideological bias.

Collectively, existing work demonstrates the value of computational methods for analysing political media discourse, framing, and ideological bias. However, three gaps remain: longitudinal multi-outlet analysis of UK political news is limited; many studies analyse bias at either aggregate or short temporal scales; and relatively few combine transformer-based topic modelling, temporal segmentation, and comparative ideological analysis within a single framework. The present study addresses these gaps through dual-scale BERTopic modelling across politically significant time periods.

3 Data and Method

3.1 Corpus Construction

The dataset comprises 124,693 articles published by six UK national newspapers between November 2014 and August 2024. The analytical period concludes with the 2024 general election cycle. Three left-leaning outlets (The Guardian, The Independent, Daily Mirror) are contrasted with three right-leaning outlets (The Telegraph, Daily Mail, The Sun). Ideology classifications were assigned based on each outlet's established editorial positioning and historical patterns of electoral endorsement. These outlets were selected due to their national reach, consistent political positioning, and availability through accessible archival sources. Table 1 shows the outlet distribution.

Table 1. Outlet distribution by ideology in the raw corpus (before cleaning and outlier removal).

Outlet	Ideology	Articles
The Guardian	Left	37,023
The Independent	Left	27,898
Daily Mirror	Left	15,033
The Telegraph	Right	26,560
Daily Mail	Right	8,125

The Sun	Right	14,229
Total		128,868

Two retrieval channels were used. Guardian content was collected via the Guardian Open Platform API using a keyword-based full-text query combined with a Politics section filter. Articles from the remaining five outlets were retrieved via Nexis Uni, a licensed academic database, using a headline-restricted keyword query targeting political terms. Both channels covered the same date range (7 November 2014 to 4 January 2025). After removing 194 articles with missing publication dates and truncating at August 2024 due to reduced coverage from the Daily Mail and The Sun, the cleaned corpus comprised 124,693 articles. The retrieval asymmetry between channels is acknowledged as a limitation.

3.2 Topic Modelling Pipeline

Article body text was embedded using the all-MiniLM-L6-v2 sentence transformer. Raw, unprocessed text was used after an early iteration demonstrated that stopword-stripped text produced approximately 50% outliers and a single dominant topic—sentence transformers require natural language input, not keyword-stripped text. UMAP reduced embeddings to five components with cosine distance, and HDBSCAN clustered the reduced embeddings with a minimum cluster size of 100. This embedding-based approach allows semantically related articles to cluster together even when they do not share identical vocabulary. A custom stopword list excluded outlet names and news boilerplate from the CountVectorizer representation.

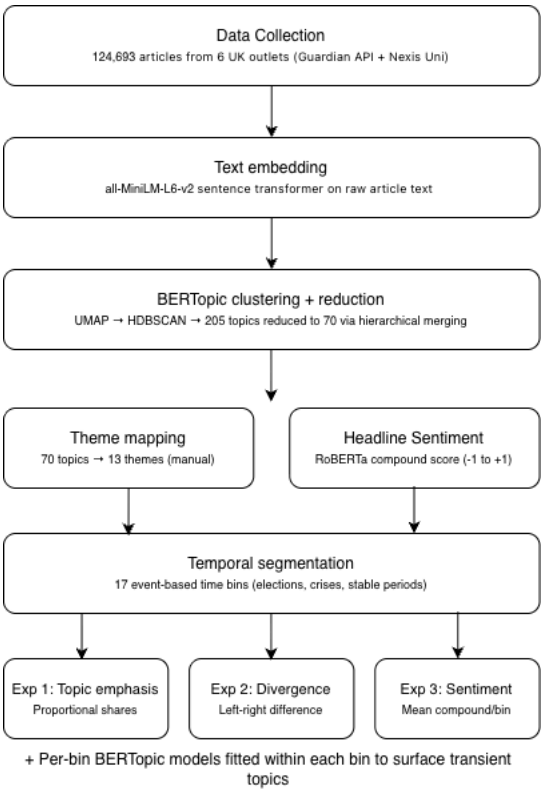

Fig. 1. Methodological pipeline: data collection, embedding, BERTopic clustering, hierarchical reduction to 70 topics, manual grouping into 13 themes, sentiment scoring, temporal binning, and experiment computation.

The initial model produced 205 topics, reduced to 70 via BERTopic's hierarchical merging. This level was selected after qualitative comparison at 50, 80, and 100 topics: at 50, distinct political actors merged; at 80–100, near-duplicate clusters emerged. The outlier rate remained constant at 36.3% across all reduction levels. Outlet-level outlier rates ranged from 31% (Telegraph) to 39% (Guardian), confirming no systematic ideological bias in outlier exclusion. The 70 topics were manually grouped into 13 interpretive themes based on inspection of top keywords and representative articles.

Complementary per-bin BERTopic models were fitted to articles within each of the 17 time bins to surface transient topics whose signal is diluted across the full corpus. The same pipeline architecture was used with adapted parameters (lower minimum cluster sizes). Per-bin models produced between 8 and 58 topics, with outlier rates ranging from 5.0% to 25.8%.

3.3 Headline Sentiment Analysis

Headline sentiment was scored using cardiffnlp/twitter-roberta-base-sentiment-latest, a RoBERTa model fine-tuned on approximately 124 million tweets [17]. A compound score was derived by subtracting the negative probability from the positive, yielding a normalised score between −1 and +1. Analysis was restricted to headlines as they represent the most visible editorial framing decision and are structurally suited to a short-text model.

3.4 Temporal Segmentation and Experiment Design

The corpus was divided into 17 time bins with boundaries defined by major political events rather than arbitrary calendar divisions (Table 2). Each bin encompasses a politically coherent period: an election cycle or a stable inter-election window characterised by a dominant political context. This approach is supported by evidence that framing effects and agenda-setting are context-dependent [7, 13].

Table 2. The 17 event-based time bins used for temporal analysis.

Bin	Label	Start	End	Type
1	Pre-GE2015 Campaign	26 Mar 2015	6 May 2015	Election
2	GE2015 Election Window	7 May 2015	18 Jun 2015	Election
3	Post-GE2015 Settling	19 Jun 2015	31 Dec 2015	Non-election
4	EU Referendum Campaign	1 Jan 2016	23 Jun 2016	Non-election
5	Post-Referendum Fallout	24 Jun 2016	31 Dec 2016	Non-election
6	Pre-GE2017 Non-Election	1 Jan 2017	27 Apr 2017	Non-election
7	Pre-GE2017 Campaign	28 Apr 2017	7 Jun 2017	Election
8	GE2017 Election Window	8 Jun 2017	20 Jul 2017	Election
9	Brexit Negotiations	21 Jul 2017	31 Dec 2018	Non-election
10	Brexit Crisis / Leadership	1 Jan 2019	30 Oct 2019	Non-election
11	Pre-GE2019 Campaign	31 Oct 2019	11 Dec 2019	Election
12	GE2019 Election Window	12 Dec 2019	2 Feb 2020	Election
13	COVID Period	3 Feb 2020	31 Dec 2021	Non-election
14	Johnson / Truss / Sunak	1 Jan 2022	31 Dec 2022	Non-election
15	Sunak Gov. / Pre-Elec. Drift	1 Jan 2023	22 May 2024	Non-election
16	Pre-GE2024 Campaign	23 May 2024	3 Jul 2024	Election
17	GE2024 Election Window	4 Jul 2024	15 Aug 2024	Election

Three experiments were computed arithmetically on the fixed topic labels and sentiment scores already assigned by the full-corpus models. **Experiment 1 (Topic Emphasis by Ideology):** proportional topic shares per ideology per bin. **Experiment 2**

(Temporal Divergence): left–right difference scores per theme per bin. **Experiment 3 (Headline Sentiment):** mean compound sentiment per ideology group per bin.

3.5 Sensitivity Analysis

The Independent, sometimes characterised as centrist, was coded as left-leaning based on its editorial positioning. A sensitivity analysis recomputed all theme proportions with The Independent excluded to verify robustness.

4 Results

4.1 Analytical Dataset Overview

Following removal of outlier articles (topic −1) and exclusion of articles outside the defined temporal bins, the analytical sample comprised 75,401 articles distributed across 17 event-based periods. Article distribution remained broadly balanced, although reduced Daily Mail and Sun coverage increased left-leaning proportions above 60% in bin 16 and 67% in bin 17.

4.2 Overview of Extracted Themes

The 70 BERTopic clusters were manually grouped into 13 higher-level interpretive themes based on semantic similarity and inspection of representative articles (Table 3). These themes form the basis for the subsequent comparative and temporal analyses.

Table 3. The 13 interpretive themes derived from manual grouping of 70 BERTopic clusters.

Theme	Example Keywords	Description
UK Politics	labour, corbyn, starmer, tory, johnson	Domestic party politics, leadership, parliamentary affairs
Brexit & EU	eu, brexit, deal, trade, withdrawal	EU membership, referendum, negotiations
Scottish & Regional	snp, scottish, sturgeon, independence	Scottish independence, devolved governance, NI politics
International Politics	trump, biden, china, russia, war	Foreign affairs, global conflicts
Economy & Public Services	nhs, tax, health, budget, spending	Health service, taxation, fiscal policy
Political Controversy	antisemitism, inquiry, scandal	Party scandals, internal disputes
Society & Social Issues	police, crime, housing, abuse	Policing, crime, social welfare

Health & Crisis	covid, pandemic, lockdown, vaccine	Public health emergencies, pandemic response
Immigration & Security	immigration, border, asylum, terror	Immigration policy, national security
Environment & Energy	climate, energy, nuclear, green	Climate policy, energy, environmental regulation
Education & Policy	schools, education, university	Education system, curriculum, policy
Culture & Media	bbc, media, football, press	Media industry, cultural topics, sport
Institutions & Governance	court, law, commons, lords	Legal system, parliamentary procedure, constitutional

4.3 Topic Emphasis by Ideology

Clear and persistent differences in thematic emphasis emerged between left- and right-leaning outlet groups across the 17 time bins. Left-leaning outlets devoted a larger proportion of coverage to UK Politics (e.g., 25.5% vs. 16.2% in the COVID period) and Brexit-related topics (11.2% vs. 2.7%), while right-leaning outlets disproportionately emphasised Scottish & Regional (12.3% vs. 21.6%) and International Politics (8.5% vs. 20.3%). These patterns were not confined to single periods but persisted across the majority of time bins.

Ideological Coverage Gap: All Themes × All Periods

Colour shows direction and magnitude of left–right difference in topic share

	Time Bin																
Theme	Pre-GE2015 ..	GE2015 Election ..	Post-GE2015 ..	EU Referendum Ca..	Post-Referendum ..	Pre-GE2017 ..	Pre-GE2017 ..	GE2017 Election ..	Brexit Negotiations	Brexit Crisis / L..	Pre-GE2019 ..	GE2019 Election ..	COVID Period	Johnson Crisis / T..	Sunak Government ..	Pre-GE2024 ..	GE2024 Election ..
UK Politics	-0.11	-0.11	-0.08	-0.05	-0.12	-0.05	-0.10	-0.04	-0.06	-0.05	-0.03	-0.10	-0.09	-0.17	-0.13	-0.10	-0.13
Brexit & EU	0.04	-0.02	-0.01	-0.07	-0.06	-0.11	-0.06	-0.07	-0.16	-0.08	-0.02	-0.06	-0.08	-0.03	-0.03	0.00	-0.02
Political Controversy	-0.03	-0.04	-0.02	-0.06	-0.04	-0.07	-0.03	-0.02	-0.03	-0.03	0.00	-0.01	-0.01	0.00	-0.02	-0.03	-0.06
Society & Social Issues	-0.02	-0.04	-0.01	-0.02	-0.01	0.00	-0.01	0.00	-0.02	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0.01	0.01
Health & Crisis	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	-0.01	-0.02	-0.01	-0.06	-0.01	0.00	0.00	0.00
Immigration & Security	-0.01	0.02	0.00	-0.03	0.00	0.00	-0.02	0.03	0.01	0.01	0.01	-0.02	0.00	-0.01	-0.02	-0.01	-0.02
Education & Policy	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.02	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01	0.01	0.01
Environment & Energy	-0.01	0.02	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	-0.02	0.00	0.00	0.00	0.01	-0.01	0.00
Economy & Public Services	-0.01	0.01	-0.01	0.00	0.00	0.02	0.01	0.02	0.00	0.00	-0.02	0.02	0.00	0.02	0.01	-0.01	0.00
Institutions & Governance	0.01	0.01	0.00	-0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	-0.01	-0.01	0.00
Culture & Media	0.02	0.05	0.02	0.03	0.03	0.02	0.01	0.01	0.03	0.02	0.01	0.04	0.04	0.03	0.00	0.00	-0.02
International Politics	0.01	0.04	0.01	0.06	0.12	0.10	0.06	0.07	0.14	0.07	0.01	0.05	0.12	0.12	0.08	0.09	0.14
Scottish & Regional	0.13	0.07	0.11	0.16	0.09	0.09	0.14	-0.01	0.08	0.08	0.11	0.11	0.09	0.04	0.12	0.06	0.09

Fig. 2. Coverage gap heatmap across all 13 themes and 17 time bins. Negative values (blue) indicate left-leaning emphasis; positive values (red) indicate right-leaning emphasis. Darker colours represent larger gaps.

The heatmap (Fig. 2) reveals that UK Politics has the largest left-leaning skew across all 17 bins, while Scottish & Regional and International Politics are right-leaning throughout. The largest inter-bin divergence occurs in Brexit & EU during bins 9–10 (Brexit Negotiations and Brexit Crisis), where the left–right gap reaches −0.16. Outlet-level comparisons further illustrate these differences: during the COVID period, The Guardian published 1,084 UK Politics articles while The Telegraph published only 351.

4.4 Temporal Dynamics of Ideological Divergence

Temporal analysis revealed that ideological divergence intensified during periods of political instability rather than during formal election campaigns. The left–right coverage gap for UK Politics remained consistently negative (left-heavy) across all 17 bins, ranging between −0.04 and −0.22 (Fig. 3). The largest divergences occurred at bin 14 (Johnson Crisis / Truss / Sunak, −0.17) and bin 15 (Sunak Government, −0.13).

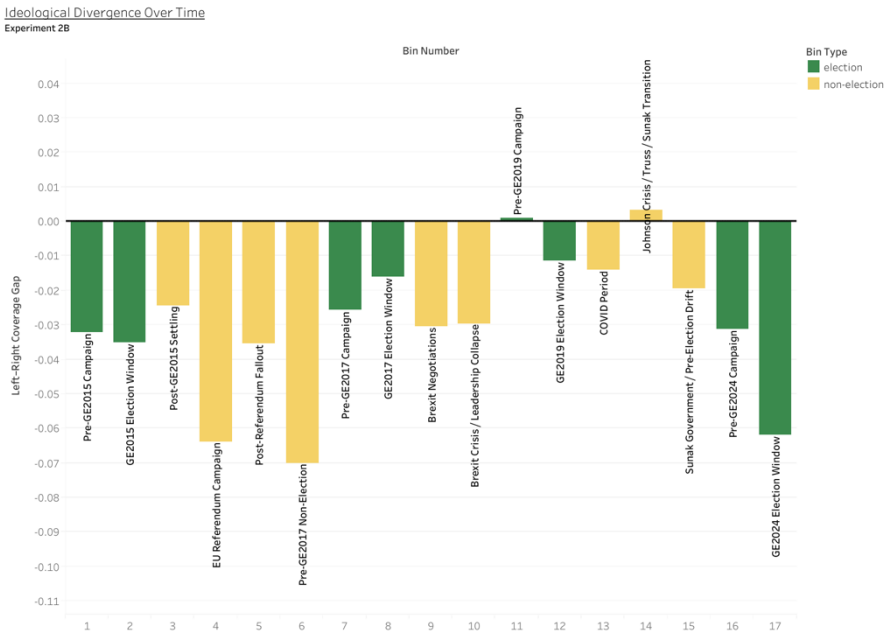

Fig. 3. Left–right coverage gap for UK Politics across 17 time bins. Green bars = election periods; yellow = non-election. The gap is consistently left-heavy but does not systematically widen during elections.

Contrary to expectations, election periods were not consistently associated with the largest divergences. UK Politics divergence during the GE2017 Election Window (−0.04) was actually lower than the preceding non-election bin (−0.10). Brexit & EU showed nearly double the divergence during non-election periods (0.07) compared with election periods (0.04), driven by high-divergence post-referendum and negotiation bins. This pattern suggests that editorial agenda-setting in UK political news is more responsive to political crises than to election campaigns.

4.5 Headline Sentiment

No substantive difference in headline sentiment was found between ideological groups. Mean compound scores were −0.261 (left) and −0.245 (right), yielding an aggregate difference of just 0.017. In virtually every bin, both groups tracked near-

identically (Fig. 4). This supports the conclusion that ideological bias operates primarily through topic selection rather than evaluative tone.

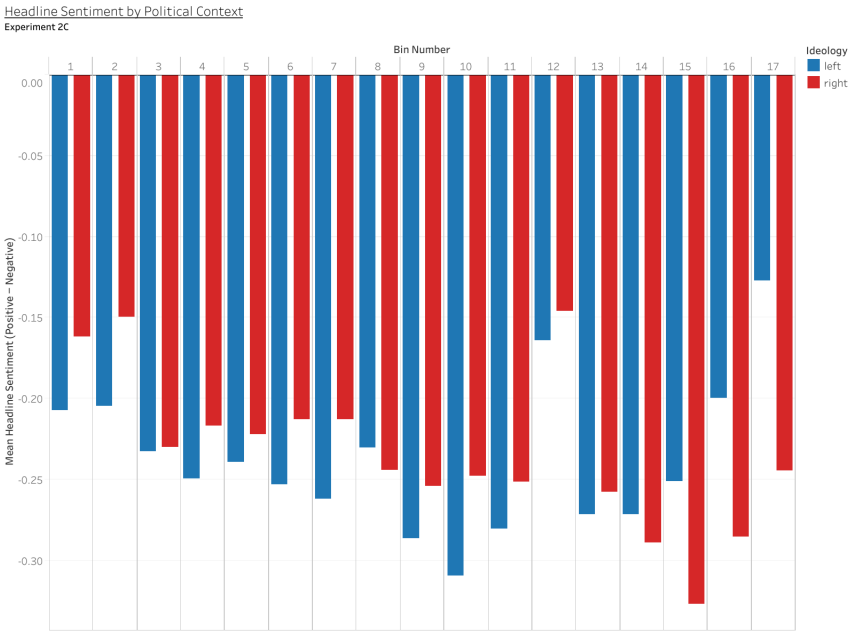

Fig. 4. Mean headline sentiment by ideology across all 17 time bins. Both groups track near-identically, with a negligible aggregate difference of 0.017.

4.6 Per-Bin Transient Topics

Complementary per-bin modelling revealed six transient topics not captured by the full-corpus model. These were classified as transient when semantically coherent topics emerged within individual time-bin models but did not correspond to any stable theme in the aggregate structure. All related to politically concentrated controversies involving the Conservative government, and all skew toward left-leaning outlets (64–79%), as summarised in Table 4.

Table 4. Transient topics surfaced by per-bin BERTopic models, with left-leaning coverage share.

Topic	Bin	N	Left %	Top Keywords
Truss/Kwarteng mini-budget	John-son/Truss/Sunak	975	67.8%	truss, tax, liz, kwarteng
Cummings/Barnard Castle	COVID Period	282	79.1%	cummings, durham, dominic
Paterson sleaze scandal	COVID Period	194	73.2%	paterson, standards, sleaze

Downing St flat/Geidt	COVID Period	169	64.5%	flat, refurbishment, geidt
Southport riots/disorder	GE2024 Elec. Window	131	72.5%	police, disorder, violence
Election betting scandal	Pre-GE2024 Campaign	131	68.7%	gambling, commission, bet

Although all transient topics showed stronger left-leaning representation, this should be interpreted cautiously. The Guardian contributes approximately 49% of all left-coded articles, meaning any transient topic is statistically more likely to contain Guardian content.

4.7 Sensitivity Analysis

To assess robustness, all proportional theme calculations were recomputed with The Independent excluded from the left-leaning category. All key findings remained stable. No result reversed direction or materially changed interpretation.

5 Discussion

Three findings merit discussion. First, the dominance of topic-level divergence over sentiment-level divergence provides computational support for agenda-setting theory [6] in the UK newspaper context. The outlets differ primarily in what they cover rather than how they cover it. This is consistent with Berk's [8] finding that framing effects erode in complex information environments, suggesting that selection and coverage bias are more powerful mechanisms of ideological differentiation than tonal framing. The practical implication is that media literacy interventions focused on differences in topic selection may be more effective than those focused on evaluative language.

Second, divergence intensifying around political crises rather than election campaigns is consistent with Geiß's [7] conditional agenda-setting model, which predicts stronger effects when issue salience spikes suddenly. Brexit coverage divergence peaked during negotiations, not during election windows. This has methodological implications: studies using election periods as natural boundaries for bias analysis may miss the most substantial divergences. One possible explanation is that rapidly evolving political crises create greater opportunities for selective issue prioritisation and editorial differentiation across ideologically opposed outlets.

Third, the per-bin modelling demonstrates the value of complementary temporal scales. The six transient topics—all associated with Conservative government controversies—were invisible at aggregate scale but emerged clearly within politically coherent time windows. This demonstrates that ideological divergence operates across multiple temporal scales: structurally persistent thematic differences coexist with short-lived but highly concentrated event-driven divergence. These transient discourse formations are particularly important for understanding political polarisation, as

rapidly evolving controversies appear to intensify selective issue prioritisation across ideologically opposed outlets. This dual-scale strategy extends Cobo et al.'s [13] approach by combining stable aggregate topics with event-specific clusters within a single analytical framework.

5.1 Limitations

Several limitations should be considered. First, the binary left–right ideology classification simplifies a more complex editorial landscape, although the sensitivity analysis demonstrated robustness. Second, article distribution was uneven, particularly in GE2024 bins where reduced Daily Mail and Sun coverage affected ideological balance. The Guardian's substantial contribution may have amplified certain transient topics. Third, the study relies on descriptive proportional comparisons without formal statistical significance testing. Finally, the corpus used two different retrieval mechanisms, introducing methodological asymmetry, and the sentiment model's training domain (tweets) may limit sensitivity to British political headline register.

6 Conclusion

This study examined how topic emphasis and headline sentiment differed between left- and right-leaning UK news outlets between 2015 and 2024, spanning four general elections and multiple periods of political instability. Using BERTopic-based topic modelling, transformer-based sentiment analysis, and event-driven temporal segmentation, 124,693 articles from six national newspapers were analysed across 17 politically coherent time periods.

Three main findings emerged. First, ideological divergence manifested primarily through differences in thematic emphasis rather than headline sentiment, supporting agenda-setting perspectives on political media bias. Second, the largest divergences occurred during periods of political crisis and institutional instability rather than during formal election campaigns, suggesting that media polarisation intensifies most strongly during moments of heightened uncertainty and elite conflict. Third, complementary per-bin modelling revealed transient event-specific discourse formations that became diluted within aggregate corpus-level analysis, demonstrating the value of combining long-term and temporally localised computational approaches.

More broadly, the findings highlight the importance of temporally sensitive computational approaches for understanding ideological differentiation within contemporary media ecosystems. Rather than functioning as static and consistently polarised entities, news outlets appear to shift their editorial divergence dynamically in response to crises, political salience, and rapidly evolving public controversies. By combining longitudinal topic modelling with event-based temporal analysis, the study demonstrates how computational methods can capture both persistent and transient structures within political discourse, contributing to broader computational social science research on media polarisation and evolving information ecosystems.

References

1. Elsner, P., Atkinson, R., Zahidi, S.: Global Risks Report 2025. World Economic Forum (2025). https://www.weforum.org/publications/global-risks-report-2025/
2. Wellings, T., Hänggli Fricker, R., Pournaras, E., Fu, W.: Brexit on Twitter: unraveling the dynamics of polarization over time. In: Proc. 25th Annual Int. Conf. on Digital Government Research, pp. 855–866 (2024). https://doi.org/10.1145/3657054.3657269
3. Newman, N., Ross Arguedas, A., Robertson, C.T., Nielsen, R.K., Fletcher, R.: Reuters Institute Digital News Report 2025. Reuters Institute for the Study of Journalism, University of Oxford (2025). https://doi.org/10.60625/risj-8qqf-jt36
4. Loughborough University: Report 5: 30 May – 3 July 2024 — General Election 2024. Centre for Research in Communication and Culture (2024). https://www.lboro.ac.uk/news-events/general-election/report-5-2024/
5. Rodrigo-Ginés, F.-J., Carrillo-de-Albornoz, J., Plaza, L.: A systematic review on media bias detection: what is media bias, how it is expressed, and how to detect it. Expert Syst. Appl. 237, 121641 (2024). https://doi.org/10.1016/j.eswa.2023.121641
6. McCombs, M.E., Shaw, D.L.: The agenda-setting function of mass media. Public Opin. Q. 36(2), 176–187 (1972). https://doi.org/10.1086/267990
7. Geiß, S.: The media's conditional agenda-setting power: how baselines and spikes of issue salience affect likelihood and strength of agenda-setting. Commun. Res. 49(2), 296–323 (2022). https://doi.org/10.1177/0093650219874968
8. Berk, N.: The impact of media framing in complex information environments. Polit. Commun. 1–17 (2025). https://doi.org/10.1080/10584609.2025.2456519
9. Egger, R., Yu, J.: A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Front. Sociol. 7 (2022). https://doi.org/10.3389/fsoc.2022.886498
10. Liu, Y.: Comparison of LDA and BERTopic in news topic modeling: a case study of the New York Times' reports on China. Pac. Int. J. 7(3), 47–51 (2024). https://rclss.com/index.php/pij/article/view/616
11. Al-Zaman, M.S., Or Rashid, M.H.: The humanitarian crisis in the media: framing analysis of Rohingya-related international news using BERTopic. Journalism Studies 26(1), 1–23 (2025). https://doi.org/10.1080/1461670X.2024.2412205
12. Verbytska, A.: Topic modelling as a method for framing analysis of news coverage of the Russia–Ukraine war in 2022–2023. Lang. Commun. 99, 174–193 (2024). https://doi.org/10.1016/j.langcom.2024.10.004
13. Cobo, M.J., De Mascellis, A.M., Misuraca, M., Spano, M.: Automatic temporal framing detection for news analysis. Qual. Quant. (2025). https://doi.org/10.1007/s11135-025-02492-1
14. Spinde, T., Plank, M., Krieger, J.-D., Ruas, T., Gipp, B., Aizawa, A.: Neural media bias detection using distant supervision with BABE. In: Findings of EMNLP 2021, pp. 1166–1177 (2021). https://doi.org/10.18653/v1/2021.findings-emnlp.101
15. Lin, L., Wang, L., Zhao, X., Li, J., Wong, K.-F.: IndiVec: leveraging LLMs for media bias detection with fine-grained indicators. In: Findings of EACL 2024, pp. 1038–1050 (2024). https://doi.org/10.18653/v1/2024.findings-eacl.70
16. Krieger, J.-D., Spinde, T., Ruas, T., Kulshrestha, J., Gipp, B.: A domain-adaptive pre-training approach for language bias detection in news. arXiv:2205.10773 (2022). https://doi.org/10.1145/3529372.3530932

17. Loureiro, D., Barbieri, F., Neves, L., Anke, L.E., Camacho-Collados, J.: TimeLMs: Diachronic language models from Twitter. In: Proc. 60th ACL: System Demonstrations, pp. 251–260 (2022). https://doi.org/10.18653/v1/2022.acl-demo.25