Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Polarisation and Identity Discourse on Malaysian TikTok: A Digital Humanities Approach

Nur Kareelawati Abd Karim

kareelawati@usim.edu.my

Universiti Sains Islam Malaysia

Jawahir Che Mustapha

Malaysian Institute of Information Technology, Universiti Kuala Lumpur

Juliana Jaafar

Malaysian Institute of Information Technology, Universiti Kuala Lumpur

Article

Keywords: Digital Humanities, digital polarisation, TikTok, computational discourse analysis, sentiment analysis, Malaysia

Posted Date: May 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9268897/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

TikTok has become a key platform for Malaysians to engage with identity, political, and religious discourse through algorithmically curated short-form video feeds. This study applies a Digital Humanities computational workflow to examine digital polarisation on Malaysian TikTok using a time-bounded corpus of 1,841 publicly accessible videos posted between 2023 and 2025 (1,530 accounts). A purposive hashtag-led sampling strategy was used to construct the corpus and extract captions, hashtags, and engagement metadata (views, likes, comments, shares). Text was pre-processed through tokenisation, lemmatisation, and stop-word removal to support reproducible analysis. Polarisation is operationalised as a platform-trace phenomenon triangulating affective signals (sequential sentiment modelling using a lexical baseline, a contextual transformer model, and multilingual aspect-based sentiment analysis), thematic concentration (keyword frequency and clustering), and visibility proxies (engagement metrics). Results show that neutral affect dominates caption sentiment (86.10%) and achieves the highest engagement, indicating a pattern of moderated affect in which polarising meanings circulate without relying on overtly negative language. Theme extraction reveals that discourse clusters most strongly in sociocultural talk (52.19%), followed by politics (19.40%), religious provocation (9.83%), disinformation/hoax framings (8.09%), racial/ethnic discourse (7.12%), and mobilisation calls (3.37%). Priority signals indicate that explicit antagonism concentrates most sharply in religious provocation, suggesting that religious identity functions as a high-stakes boundary that invites loyalty signalling and out-group suspicion. The study contributes a replicable workflow for multilingual TikTok research and demonstrates how “hidden” polarisation can persist through identity-linked themes, recontextualised fragments, and credibility contests, even when surface sentiment appears neutral.

Introduction

Polarisation has become a defining feature of contemporary digital public life. In platformed environments, disagreement is not only expressed but also structured by recommender systems, engagement metrics, and the affordances of short, highly shareable media (Arora et al., 2022; Bliuc et al., 2024). TikTok sharpens these dynamics through personalised discovery and compressed narrative formats that favour fast interpretation and rapid circulation (Jalli, 2025). Under these conditions, polarisation often takes the form of affective boundary-making rather than sustained ideological debate: posts signal belonging through moral tone, identity cues, and antagonistic “us/them” framings (Iyengar et al., 2012; Wodak & Meyer, 2016). This pattern aligns with the notion of affective publics, in which networked communication is organised through emotion, storytelling, and visibility rather than procedural deliberation (Papacharissi, 2014).

Digital Humanities (DH) is well positioned to address these developments because it treats platforms as cultural infrastructures that can be examined through born-digital traces. DH approaches emphasise transparent corpus construction, explicit documentation of pre-processing choices, and interpretive claims that remain accountable to computational results (Joyeux-Prunel, 2024). TikTok nevertheless poses practical constraints: multimodal content and opaque recommendation logic make it difficult to infer why specific videos become visible. A pragmatic DH response is therefore to analyse what can be collected reproducibly at scale—captions, hashtags, and engagement metadata—and to interpret these traces in relation to local political culture and discourse histories.

Malaysia offers an analytically rich case because ethnicity, religion, and partisan politics are structurally entwined in public discourse and are also legally and culturally sensitive. In digital settings, these sensitivities can be reconfigured into polarising storylines and mobilisation cues, shaped by the attention economy and by information disorder (Chinnasamy & Volkmer, 2023; Wardle & Derakhshan, 2017). During politically salient periods, Malaysian civil society and research organisations have underscored the need to monitor platform discourse and engagement

patterns, including around the 15th General Election (Hamzah et al., 2023). TikTok—widely used in Malaysia for entertainment, news discovery, and everyday commentary—has become a prominent venue where polarising topics circulate in vernacular styles such as humour, satire, moral reminders, and rapid "hot take" commentary (Jalli, 2025).

This article addresses two research questions: (1) How does polarised content circulate on TikTok among Malaysian users across political, religious, racial/ethnic, and sociocultural themes? (2) How do observable logics of visibility—approximated through engagement metrics and sentiment dynamics—shape exposure to polarised or emotionally charged content? In this study, polarisation is operationalised as the coupling of thematic boundary-making with affective intensification as it appears in captions, hashtags, and engagement traces (Iyengar et al., 2012; Wodak & Meyer, 2016).

The article contributes in three ways. First, it presents a replicable workflow for constructing and analysing a TikTok corpus in a multilingual setting, grounded in explicit sampling and cleaning rules (Joyeux-Prunel, 2024). Second, it demonstrates the value of a sequential sentiment strategy that moves from lexical scoring to contextual and aspect-based modelling, improving sensitivity to short, code-mixed captions (Hutto & Gilbert, 2014; Sanh et al., 2019; Pontiki et al., 2014). Third, it shows how polarisation can persist even when surface sentiment appears neutral, through identity-linked themes, recontextualised fragments, and mobilisation cues that organise antagonistic interpretation (Wardle & Derakhshan, 2017).

Context for the Study: Polarisation on TikTok across Cultural, Social, Political, and Ethnic–Racial Axes

TikTok has become a distinctive site of contemporary polarisation because it combines algorithmic discovery with short-form, performance-led communication. Its "For You" feed distributes content beyond follower networks, sorting users into interest-based micro-publics through viewing patterns, replays, and engagement. This architecture matters for polarisation. Content that provokes moral reaction, ridicule, or perceived threat often attracts longer watch time and higher comment activity, and is therefore more likely to be circulated within—and sometimes across—identity-aligned clusters. In this environment, disagreement is rarely expressed as a purely rational contest of evidence. Instead, polarisation becomes an affective struggle over legitimacy, belonging, and moral order, enacted through platform vernaculars such as hashtags, trending audios, stitches/duets, green-screen "receipts", and the performative intensity of comment sections.

For analytic clarity, the study positions TikTok polarisation along four interrelated axes—cultural, social, political, and ethnic/racial—while recognising that, in Malaysia, these axes often converge. Cultural controversies can quickly become ethnicised; political campaigning frequently mobilises identity cues; and social fragmentation provides the emotional infrastructure through which antagonistic interpretations travel and stabilise. The following subsections situate each axis as it is commonly expressed on TikTok, and explain why Malaysian TikTok is a particularly consequential arena for examining these dynamics from 2020 to 2025 (Wike et al., 2022; Tapsell, 2023; Saidin & Azrun, 2024; Kirk, 2025).

Cultural polarisation: values, heritage, and boundary work

Cultural polarisation refers to divisions structured around differences in values, norms, and identities tied to language, heritage, lifestyle, and moral attitudes. Digitally mediated cultural conflict often develops into "culture war" dynamics, where opposing sides circulate content that affirms in-group virtue while portraying out-groups as threats

to communal legitimacy. TikTok is especially conducive to such dynamics because cultural claims can be communicated rapidly through aesthetic cues—dress, music, tone, humour—and condensed into symbols that travel faster than explanation. Cultural polarisation is often visible in hashtags emphasising heritage and identity reinforcement (e.g., #adat, #budaya), framing content around communal values and cultural legitimacy. At the same time, the absence of a small number of dominant sociocultural hashtags may indicate decentralised, grassroots participation rather than centrally coordinated campaigning, even as the discourse performs boundary work by distinguishing “authentic” from “inauthentic” cultural positions.

The COVID-19 period offers a clear illustration of how quickly cultural meaning can attach to public issues online. Across countries, public health measures such as mask-wearing acquired cultural significance that signalled political and moral alignment, turning a health practice into an identity marker through hashtags, memes, and antagonistic humour. This “digital polarisation of mask-wearing” in 2020 followed partisan and ideological lines, revealing how social media bubbles can reinforce values such as trust in science, views on authority, and notions of communal responsibility versus personal liberty (Jackson et al., 2024). For TikTok, the implication is that cultural polarisation often emerges not only through topic choice, but through the moral framings and identity cues that content formats make unusually easy to reproduce and intensify.

In multicultural societies, cultural polarisation frequently centres on language and education. In Malaysia, debates over vernacular schools versus a single national system, and controversies surrounding Jawi teaching, can split sharply along ethnic lines. Disputes such as non-Muslims learning Jawi in schools or non-Muslims using the word “Allah” often erupt on social media, revealing deeper anxieties over cultural continuity and recognition (Goh, 2023; Saidin & Azrun, 2024). These controversies also create openings for political actors to amplify cultural conflict for mobilisation, reframing cultural difference as a zero-sum contest over national belonging. As a result, TikTok’s culture-coded discourse frequently carries the weight of group identity, where “protecting culture” becomes a shorthand for opposing a perceived out-group.

Social polarisation: fragmentation, civility decline, and affective distance

Social polarisation refers to a breakdown of social cohesion and civil dialogue that goes beyond ideology alone. It includes fragmentation along class, geography, and generation, as well as changes in the tone of everyday interaction—particularly the decline of civility and the growth of hostile communication norms. Cross-national surveys show that many publics perceive the internet and social media as divisive, with substantial shares reporting that online political talk has become less civil and more hostile (Wike et al., 2022). TikTok’s communicative style can intensify this form of polarisation because conflict itself is highly “performable”: creators can stage ridicule, moral condemnation, or humiliation in ways that generate engagement and invite pile-ons in comment threads.

In Malaysia, social polarisation is strongly shaped by generational and urban–rural divides. Younger Malaysians often consume media and adopt expressive styles that differ from older publics. During the 2022 election period, younger first-time voters encountered edgy TikTok videos and memes that challenged established authorities, while older voters were more exposed to WhatsApp forwards or Facebook content aligned with conventional political narratives (Tapsell, 2023). This segmentation does not merely reflect different “opinions”; it produces parallel interpretive worlds. Urban groups—often multiethnic and globally oriented—tend to promote governance reform and minority rights in digital spaces, while rural networks—more homogeneously Malay and religiously conservative—frequently emphasise community rights, affirmative action, and moral order. When these publics meet on TikTok,

interactions can quickly turn into relationship-level hostility, especially as algorithmic distribution places unfamiliar audiences in the same comment arenas.

Affective polarisation is a critical dimension of this process: groups begin to feel emotionally distant and develop negative attitudes toward one another, even when disagreement is not about concrete policy. In Malaysia, after GE15, TikTok reportedly circulated user-generated videos referencing the May 13, 1969 racial riots, warning that similar chaos could return under certain political outcomes, thereby intensifying fear and mistrust across communities (Kaur & Yusof, 2023). This illustrates how social fragmentation and affective antagonism can be activated through historical memory and threat narratives, turning "social distance" into a platform-visible dynamic of suspicion and hostility.

Political polarisation: campaigning, micro-publics, and negative framing

Political polarisation concerns intensifying divides in partisan or coalition loyalties, declining shared ground, and growing distrust of political opponents. Globally, social media has been shown to intensify polarisation by amplifying partisan hostility and extreme interpretations, even if it does not "cause" polarisation in isolation (Barrett et al., 2021). Political actors increasingly professionalise digital campaigning through micro-targeted messaging, influencer partnerships, and emotionally charged frames that are customised to different micro-publics, which can fragment electorates and erode shared political reality (Papathanassopoulos & Giannouli, 2025).

In Malaysia, political polarisation from 2020 to 2025 unfolded amid political instability, pandemic governance disputes, and a highly competitive GE15 campaign. TikTok became a central arena for campaigning and citizen interpretation, as platform developments such as short-form video and livestreaming became integral to electoral strategy (Tapsell, 2023). Political polarisation online frequently overlapped with religious and racial cleavages, but it also extended to policy-driven conflict during COVID-19, where lockdowns, emergency ordinances, and executive power were framed through sharply divergent narratives (Sani & Kasim, 2024; Lim, 2024). TikTok's format encourages political persuasion through compressed storytelling and simplified moral contrasts, which can make negative campaigning unusually efficient: opponents are framed not simply as mistaken, but as corrupt, anti-religion, extremist, or existentially dangerous.

The "Green Wave" and shifting patterns of youth political engagement further illustrate TikTok's role in political polarisation. Analysts describe PAS and its allies' effective use of TikTok to reach young and rural Malay audiences, circulating narratives of dignity, change, and communal protection (Tapsell, 2023; Shukri, 2024). In parallel, opposition-aligned publics used platform spaces to circulate counter-frames warning against authoritarianism or "Taliban-like" governance, contributing to mutual fear and delegitimation. In this way, TikTok political polarisation operates through factional publics, where different communities experience the same political moment through different platform-distributed frames, which increases the likelihood of antagonistic misrecognition across groups.

Ethnic and racial polarisation: stereotypes, threat narratives, and 3R sensitivities

Ethnic and racial polarisation involves divisions between groups defined mainly by race or ethnicity, often intensified by historical grievances and socio-economic differences. Digitally, such polarisation is amplified by the rapid spread of racially charged content, including overt hate speech, coded dog-whistles, and emotionally provocative stories that cast groups as threats. Globally, cases such as Myanmar and Ethiopia have shown how platform amplification

and moderation blind spots can escalate ethnic hostility with severe offline consequences (Oates, 2025; Schissler, 2025; Ayalew, 2025). While Malaysia's context differs, the mechanism of online escalation—stereotyping, threat framing, and mobilisation through identity cues—remains instructive for understanding TikTok dynamics.

Ethnic polarisation in Malaysia is deeply intertwined with political competition and the governance of diversity. From 2020 to 2025, online discourse frequently split into Malay/Muslim versus Chinese/Indian/non-Muslim narratives, especially during periods of political contestation. GE15 intensified these patterns, with studies documenting the strategic exploitation of racialised language and "3R" sentiments in online campaigning (Saidin & Azrun, 2024). TikTok content that insinuated communal threat, invoked historical conflict, or framed politics as a contest of ethnic dominance contributed to a toxic online environment, including episodes of intimidation narratives that circulated rapidly in youth-facing formats (Hamzah et al., 2023; Kaur & Yusof, 2023). Importantly, ethnic polarisation on TikTok is not always explicit; it often appears through coded humour, selective anecdotes, and insinuations that become legible within particular micro-publics. Comment sections then serve as escalation arenas where ambiguous prompts are interpreted through communal frames and converted into direct antagonism.

Although analytically distinct, these four axes frequently converge on TikTok because the platform rewards affective intensity and identity signalling. Cultural disputes often function as proxies for ethnic and religious anxieties; political competition mobilises these proxies; and social fragmentation supplies the interpersonal hostility that sustains antagonistic interpretations over time. The result is a platform environment where polarisation is not only expressed through ideological claims, but enacted through styles of performance, vernacular cues, and engagement-driven circulation. This matters for the study's design when examining polarisation on Malaysian TikTok requires attention to how content clusters around culture, society, politics, and ethnicity/race, and how these clusters are amplified through platform interaction patterns and algorithmic visibility (Tapsell, 2023; Saidin & Azrun, 2024; Kirk, 2025).

Conceptualising Digital Polarisation for Digital Humanities Research

Digital polarisation refers to the widening separation of publics into more internally aligned and externally opposed camps in online environments. It is commonly associated with ideologically homogeneous communities, reduced cross-group exchange, and the consolidation of "echo chambers" in which users repeatedly encounter content that affirms existing commitments while marginalising competing viewpoints (Hartmann et al., 2025; Pratap & Pathak, 2025; Sustainability Directory, 2024). As these patterns stabilise, disagreement is less likely to operate as deliberation and more likely to take the form of identity defence and mutual delegitimation, producing a durable "us versus them" orientation that makes empathy and reciprocal recognition harder to sustain.

This study adopts a broader understanding of digital polarisation as public fragmentation into mutually suspicious camps, sustained by identity-based evaluations and by media infrastructures that reward boundary work (Bliuc et al., 2024; Iyengar et al., 2012). Rather than treating polarisation only as ideological distance, the analysis foregrounds affective and discursive polarisation. It shifts in how groups feel about one another, how legitimacy is assigned, which identities are made salient, and how antagonism is normalised through language and framing (Arora et al., 2022; Wodak & Meyer, 2016). This emphasis aligns with approaches that treat polarisation as a patterned communicative achievement—something enacted through naming, moral evaluation, and recurrent narrative templates—rather than a simple distribution of attitudes.

For Digital Humanities (DH) research, the challenge is to conceptualise polarisation in ways that are both computationally tractable and culturally interpretable. In this article, platform traces—captions, hashtags, and

engagement metrics—are treated as thin but meaningful evidence of how issues are framed and how belonging is signalled, while acknowledging that TikTok discourse is multimedia, context-dependent, and often strategically ambiguous (Papacharissi, 2015). Reproducibility is therefore approached not only as re-running code, but as documenting corpus provenance, time stamps, pre-processing decisions, and interpretive validation procedures that keep computational claims accountable to the data (Joyeux-Prunel, 2024).

Platform affordances, algorithms, and echo-chamber dynamics

Platform affordances shape what becomes visible and which communicative styles are rewarded. Engagement-optimised ranking can privilege emotionally legible and morally framed content because it prompts interaction, rapid sharing, and repeated circulation (Barrett et al., 2021; Papacharissi, 2015). On TikTok in particular, meaning is frequently compressed into recognisable cues—sound templates, trending formats, brief captions, and reusable visual conventions—so affective alignment can travel without extended argument (Arora et al., 2022). These features do not guarantee polarisation, but they create favourable conditions for it by lowering the cost of identity signalling and raising the visibility of reactive content.

This has direct implications for operationalisation. Caption sentiment alone may under-estimate polarisation on a video-first platform because antagonism can be distributed across audiovisual cues, stitch/duet commentary, and comment dynamics. Conversely, apparently neutral captions may function as strategic ambiguity, allowing users to gesture towards sensitive issues while retaining plausible deniability. DH approaches therefore benefit from linking computational outputs to interpretive reading and discourse analysis, especially in multilingual and code-mixed settings where meaning often depends on context and insinuation (Joyeux-Prunel, 2024; Wodak & Meyer, 2016).

Engagement indicators (views, likes, comments, shares) are not direct measures of persuasion, but they provide a practical proxy for visibility and reactability—what the platform has circulated and what publics have found worth responding to. In polarised contexts, engagement can reward not only outrage but also identity-affirming humour, moral reminders, and “common sense” performances, which helps explain why heavily circulated posts may still present as neutral at the caption layer (Papacharissi, 2015; Barrett et al., 2021). Empirical work on engagement-driven ranking systems similarly shows that optimisation for attention can prioritise content that provokes strong emotional responses, thereby granting disproportionate visibility to divisive material (Barrett et al., 2021; Germano et al., 2025; Stray et al., 2023).

At the same time, causal claims require care. Polarisation predates social media in many contexts and is shaped by long-standing dynamics such as partisan sorting, economic inequality, and offline identity politics (Arora et al., 2022; Barrett et al., 2021; Fortunato & Pecoraro, 2022). In this view, platforms are less the origin of division than an amplifier that accelerates its circulation and intensifies its affective charge. Experimental and comparative evidence nonetheless suggests that platform use can measurably reinforce polarised tendencies for some users and issues, supporting the claim that engagement-maximising infrastructures can contribute to feedback loops of intensification even when they do not create the underlying fractures (Barrett et al., 2021).

Information disorder as a polarisation mechanism

Information disorder frameworks distinguish misinformation, disinformation, and malinformation, emphasising that harm can arise not only from falsity but also from strategic circulation, framing, and decontextualisation (Wardle & Derakhshan, 2017). These dynamics interact with polarisation because credibility judgements become identity-laden: aligned sources are trusted as morally legitimate, while opposing voices are dismissed as immoral, foreign,

or anti-community (Iyengar et al., 2012). In Malaysia, studies of digital deliberation and election-related communication show how platform discourse can become rapidly moralised and camped, even when underlying claims remain contested (Chinnasamy & Volkmer, 2023; Hamzah et al., 2023).

Polarised communities are especially susceptible to information disorder because echo chambers supply both the motivation and the interpretive frames for accepting and sharing claims that confirm group narratives (Matheus, 2025; Budak et al., 2024). Yet recent scholarship also cautions against overgeneralising from "exposure" alone, noting that misinformation exposure is often unevenly distributed and highly context-dependent, and that strong causal claims require careful design (Budak et al., 2024). For this study, information disorder is therefore treated primarily as a discursive mechanism: labelling something a hoax, scandal, or deception can function as delegitimation, mobilisation, or boundary enforcement regardless of whether the underlying claim is verifiably false (Wardle & Derakhshan, 2017; Wodak & Meyer, 2016).

Finally, platform governance and fact-checking infrastructures shape what counts as credible during crises and political events, foregrounding the platformisation of verification work and the outsourcing of "truth maintenance" to platform partners and user practices (Cotter et al., 2022). In Southeast Asia, TikTok exemplifies this complexity as a site where civic mobilisation, misinformation, and counter-misinformation practices co-exist, complicating simple distinctions between participation and manipulation (Jalli, 2025).

Critical discourse perspectives and recontextualisation

Digital polarisation frequently extends beyond formal politics into cultural and social domains, where identity disputes are narrated as moral conflicts. In multiethnic societies, language, education, and symbolic heritage can become recurring flashpoints, and online discourse can intensify group-based grievance narratives (Saidin & Azrun, 2024). Critical Discourse Studies (CDS) provides tools for analysing these dynamics by focusing on how polarisation is enacted through nomination and predication (who is named and how), pronouns, metaphors, and argumentative topoi that construct in-groups/out-groups and naturalise moral hierarchies (Wodak & Meyer, 2016).

A key mechanism here is recontextualisation that includes fragments of speeches, headlines, or religious quotations are lifted from their original setting and redeployed as proof for antagonistic interpretations, often stripping nuance while intensifying moral urgency (Wodak & Meyer, 2016). On TikTok, recontextualisation is accelerated through *stitching* and *duets* that embed commentary into the circulation of an original clip, while captions and hashtags provide rapid cues for how audiences should interpret what they see. Persona-based credibility further shapes reception, because platform-specific self-brand labour encourages performances of authenticity and moral alignment that can operate as interpretive shortcuts in short-form contexts (Scolere et al., 2018). Accordingly, this article treats polarisation as observable in patterned linguistic cues (derogatory labels, defence frames, hoax accusations) and in the distribution of themes where identity boundaries are most salient, motivating the combined use of keyword/hashtag signals, sentiment dynamics, and thematic clustering (Joyeux-Prunel, 2024; Wodak & Meyer, 2016).

Affective polarisation and identity signalling in Malaysia's platformed publics

Affective polarisation foregrounds how group-based feelings and moral evaluations intensify beyond policy disagreement, producing distrust and hostility that can raise the stakes of everyday discourse (Iyengar et al., 2012). In digital environments, such polarisation can be sustained by repeated cues of threat, disrespect, and out-group blame, generating feedback loops of identity reinforcement and social distance (Arora et al., 2022; Bliuc et al.,

2024). Comparative examples underline the potential severity of these processes when disinformation and extremist framing consolidate around identity conflict (Jackson et al., 2024; Barrett et al., 2021). In Malaysia, identity discourse is often anchored in ethnoreligious categories and claims of moral authority, which can be activated through seemingly ordinary sociocultural debates; research on election-related online communication shows how participation can become emotionally charged as publics interpret events through competing identity frames (Chinnasamy & Volkmer, 2023; Hamzah et al., 2023).

For TikTok specifically, polarisation may appear less as continuous explicit antagonism and more as periodic surges around identity flashpoints, with alignment signalled through humour, moral reminders, and strategic hashtagging. Such signalling can co-exist with neutral caption sentiment, reinforcing the need for theme-linked interpretation rather than sentiment scores alone (Papacharissi, 2015; Barrett et al., 2021). This also speaks to the "polarisation paradox": networked media can widen access to information and cross-group contact, yet simultaneously deepen interpretive silos and intensify conflict, depending on platform design, user demographics, and the surrounding political culture (Costantino, 2024; Ferreira & Ferreira, 2025). While findings vary across contexts and platforms, there is growing agreement that contemporary incentive structures and recommender systems shape the conditions under which polarisation becomes visible, durable, and emotionally charged (Gomila Olvera, 2025). Taken together, this conceptual grounding frames digital polarisation as an interplay of platform affordances, information disorder, discursive recontextualisation, and affective identity signalling. With this foundation in place, the article turns next to the methodological approach used to construct and analyse the Malaysian TikTok corpus.

Methods

This study implements a Digital Humanities (DH) computational workflow that draws on communication studies and computer sciences to examine digital polarisation through reproducible platform traces. The pipeline integrates hashtag-led corpus construction, staged text and metadata pre-processing, sequential sentiment modelling, computational theme extraction, and engagement-based validation, consistent with post-computational reproducibility principles (Joyeux-Prunel, 2024). Polarisation is operationalised as a multi-layered phenomenon that is observable (but not exhaustively measurable) in TikTok traces via the triangulation of: (1) affective signals (caption sentiment and aspect-linked sentiment), (2) thematic concentration (computationally derived discourse themes and their distribution), and (3) visibility signals (views, likes, comments, shares) that proxy circulation in TikTok's attention economy. This design enables cautious, descriptive inference about where polarising dynamics cluster and how they circulate, without claiming causal access to TikTok's opaque recommendation system.

Because platform environments are volatile and data access is time-sensitive, replicability is strengthened through documentation at each stage (hashtag discovery pathways, scraping parameters, filtering rules, and modelling choices) and by anchoring analysis to a time-bounded corpus (2023–2025). In line with DH arguments, reproducibility is treated as transparent provenance and time-stamping—alongside documented analytic decisions—rather than code re-runs alone (Joyeux-Prunel, 2024).

Alongside the platform-trace computational analysis, the study includes a qualitative, desk-based integrative evidence synthesis to contextualise digital polarisation between 2020 and 2025. This component does not involve participants or primary qualitative data collection; instead, it synthesises peer-reviewed literature, policy reports, platform documentation, and credible grey literature to support comparative interpretation of global trends and Malaysia's local dynamics (Snyder, 2019; Whittemore & Knafl, 2005). The integrative synthesis strengthens interpretive validity by clarifying how secondary evidence informs the reading of computational outputs, particularly for politically sensitive claims.

The overall design is descriptive and explanatory in a limited sense. Computational summaries (sentiment distributions, thematic clusters, keyword/hashtag “priority signals”, engagement patterns) are interpreted through discourse-informed reading, with attention to multilingual and code-mixed contexts. The analytic window (2023–2025) captures recent Malaysian TikTok discourse in a period when short-form video is central to civic talk and identity signalling. The unit of analysis is the TikTok post, represented by its caption (description), extracted hashtags, and engagement metadata. Engagement metrics are treated strictly as visibility proxies—signals of circulation and reactability—rather than direct measures of persuasion, agreement, or offline impact.

Desk-based integrative evidence synthesis (2020–2025)

The evidence synthesis proceeded along two streams: (1) global digital polarisation research and policy evidence and (2) Malaysia-focused materials capturing localised polarisation dynamics. The global stream reviewed peer-reviewed publications (2020–2025) in media studies, political communication, sociology, and digital culture, identified through Scopus, Web of Science, and Google Scholar searches. Policy and institutional evidence was drawn from official repositories of reputable organisations (e.g., Pew Research Center, Freedom House, the Brookings Institution, and the Reuters Institute), alongside platform documentation such as transparency reports, content policies, and archived statements from Meta, TikTok, and Twitter/X.

The Malaysia stream included scholarly publications in Malaysian and Southeast Asian journals; civil society and fact-checking outputs (e.g., Centre for Independent Journalism, MyCheck Malaysia, and IDEAS); official regulations and statements from the Malaysian Communications and Multimedia Commission (MCMC) and the Ministry of Communications and Digital; and verified reporting from major Malaysian news outlets (e.g., The Star, Astro Awani, Malaysiakini, and Sinar Harian). Malaysia is treated as the primary interpretive case because digital polarisation often concentrates around race, religion, and politics, with heightened salience during and after the 2022 General Election (GE15).

Analysis used thematic coding with a combined deductive–inductive logic. A deductive frame was derived from key strands in the literature and used to code materials into: cultural, religious, racial/ethnic, and political polarisation; digital mobilisation and boycotts; disinformation and hoax framings; and platform dynamics (with specific attention to TikTok, Facebook, and WhatsApp). Subthemes were added inductively where repeated patterns emerged. A cross-source matrix recorded how themes appeared across evidence types and enabled consistency checks across coding rounds (Snyder, 2019; Whittemore & Knafl, 2005).

The synthesis used a two-direction comparative logic. Horizontally, it contrasted global patterns and Malaysian cases within each theme. Vertically, it examined interactions between structural forces (platform incentives, regulatory environments, offline cleavages) and user-facing practices (identity signalling, mobilisation, boundary work). Given differences in scale and method between global survey-based evidence and Malaysia’s more event-centred materials, the synthesis prioritises interpretive comparison over statistical generalisation. Triangulation across evidence types was used to strengthen validity, and politically sensitive claims were cross-checked against independent reports, multiple news sources, and official statements where available. No human subjects participated in the study, and all materials were publicly accessible at the time of analysis.

Corpus construction and sampling strategy (2023–2025)

Corpus construction followed a purposive, hashtag-led strategy aligned to five polarisation-relevant categories: neutral, politics, religion, racial/ethnic, and sociocultural. For each category, eight hashtags were identified through iterative manual content discovery using a link-tracing technique: starting from a seed post, related hashtags and

keywords were followed across connected content to expand the list. This approach approximates how users encounter polarising discourse via networked linkage rather than isolated search, while keeping discovery logic auditable.

Hashtag-led sampling can over-represent highly “taggable” conversations and under-represent untagged talk; the corpus is therefore treated as an analytically useful slice of Malaysian TikTok discourse rather than an exhaustive representation of the platform. This limitation is addressed through transparent reporting of discovery logic and inclusion criteria, consistent with DH expectations for corpus provenance (Joyeux-Prunel, 2024).

Data collection via web scraping

Data were collected via web scraping using Apify’s TikTok Scraper, which extracts publicly accessible post metadata and captions from TikTok’s web interface (Apify, 2025). Ten scraping sessions were conducted over one month, with sessions run twice for each polarisation category. Each session targeted data from over 300 videos using the eight hashtags identified per category, producing over 3,000 raw entries across sessions. Scraping parameters were configured to focus on publicly available content posted between 2023 and 2025 and to operate under a capped maximum cost per session (USD 1.00), providing a documented and repeatable configuration. Scraped outputs were compiled into a structured dataset, for example Tiktok_Big_Data_with_Sentiment.xlsx containing caption text (description), account handle (raw export), post identifiers, extracted hashtags, and engagement statistics (views, likes, comments, shares), alongside sentiment scores and derived sentiment categories.

Filtering excluded non-Malaysian, commercial, or irrelevant posts to ensure the corpus reflected Malaysian discourse within the selected topical frame. After pre-processing, 1,862 clean entries were retained; subsequent deduplication and authenticity checks yielded 1,841 unique videos posted by 1,530 accounts for analysis. Across the cleaned corpus, 9,860 hashtags were extracted from video descriptions, underscoring hashtagging as a dissemination and discoverability practice within the sampled frame.

Text pre-processing followed a staged cleaning procedure designed to improve comparability and model stability while preserving culturally meaningful tokens. Punctuation and uppercase letters were removed to standardise casing. Emojis, hashtags, URLs, and non-text symbols were removed from the description field prior to tokenisation and lemmatisation. Stop-word filtering was applied using common English stop-words and high-frequency Malay discourse particles (e.g., lah, kan) that can inflate frequency counts without carrying stable thematic content. This approach reduces surface variation while retaining lexical material needed for theme extraction and discourse interpretation.

Sentiment analysis framework (VADER, DistilBERT, and ABSA)

Polarisation was assessed via sequential sentiment modelling to balance interpretability, speed, and contextual sensitivity in short, code-mixed captions. First, VADER provided a transparent lexical baseline designed for social-media text, producing polarity scores between − 1 (most negative) and + 1 (most positive) (Hutto & Gilbert, 2014). Second, DistilBERT was applied for contextual sentiment modelling to better capture meaning in short captions, including Malay–English code-mixing and cases where lexical cues are misleading (Sanh et al., 2019). Third, aspect-based sentiment analysis (ABSA) using multilingual BERT (mBERT) associated sentiment with topical aspects (e.g., government, economy, identity, lifestyle), enabling theme-level affective comparison even where overall caption sentiment appeared neutral (Pontiki et al., 2014).

Model outputs were applied sequentially and cross-checked for consistency before being stored as sentiment scores and categorical sentiment labels in the dataset. This triangulation reduces dependence on any single model

and improves robustness for multilingual, platform vernacular text.

Theme extraction, **“priority signals”, and engagement-based validation**

Theme extraction combined keyword frequency analysis with computational clustering to group captions into thematic categories, followed by interpretive labelling consistent with thematic analysis principles. The final themes reported were: politics, religious provocation, racial/ethnic, sociocultural, mobilisation/call to action, and disinformation/hoax. To connect theme prevalence to confrontation intensity at the caption layer, the analysis also used “priority signals”: keywords and hashtags were classified into high-, medium-, and low-priority categories to identify where antagonistic boundary work becomes linguistically explicit. This step supports interpretation of results that show high levels of neutral sentiment alongside identity-salient themes, by distinguishing between (a) routine topical participation and (b) caption-level markers of antagonism and delegitimation.

Engagement metrics (views, likes, shares, comments) were analysed across sentiment categories and themes to provide visibility-linked validation. The study follows the report’s four-stage analytic sequence—(1) sentiment distribution, (2) engagement comparison by sentiment, (3) aspect/theme distribution, and (4) comparative evaluation linking affective tone to theme visibility—supporting a layered DH inference strategy in which computational outputs are interpreted alongside circulation signals (Joyeux-Prunel, 2024). Data handling, modelling, and clustering were conducted in Python, while SPSS supported descriptive statistics and visualisation, enabling transparent reproduction of summary statistics and figures.

Ethics, data management, and limitations

The study uses only publicly accessible TikTok content and reports aggregate patterns without identifying users. Scraping was restricted to posts and accounts set for public viewing and accessible without login, and did not include private or restricted accounts. Ethical handling followed the report’s stated compliance with Malaysian online research expectations (including reference to MCMC ethical guidelines), including removal of personally identifiable information prior to analysis and no attempt to contact, profile, or intervene with users.

In line with post-computational reproducibility, sampling logic, timestamps, and cleaning/model decisions are documented as part of the research object (Joyeux-Prunel, 2024). Limitations include (1) caption-centred inference on a multimodal platform where audiovisual and comment-layer meaning may carry affect not visible in text; (2) sampling bias introduced by hashtag-led collection; and (3) residual modelling error in code-mixed or sarcastic captions despite triangulation. Accordingly, findings are presented as probabilistic patterns in a time-bounded corpus rather than exhaustive representations of Malaysian TikTok.

Results

This section reports findings in response to two research questions: (RQ1) How does polarised content circulate on TikTok among Malaysian users across political, religious, racial/ethnic, and sociocultural themes? and (RQ2) How do observable logics of visibility—approximated through engagement metrics and sentiment dynamics—shape exposure to polarised or emotionally charged content? The analysed corpus comprises 1,841 unique TikTok videos posted between 2023 and 2025 by 1,530 accounts; 9,860 hashtags were extracted from captions, underscoring hashtagging as a dissemination device and a mechanism for aligning posts with algorithmically legible topic clusters. In line with the DH workflow, results are presented as (1) affective distributions, (2) thematic concentrations with keyword/hashtag “priority signals”, and (3) engagement-linked validation, while maintaining interpretive caution about caption-only inference on a multimodal platform (Joyeux-Prunel, 2024).

Affective profile and engagement: moderated affect

Caption sentiment is dominated by neutral affect (86.10%), with smaller shares of provocative (7.01%), positive (3.53%), and negative (3.31%) content; neutral posts also record the highest engagement across metrics. This distribution does not imply the absence of polarisation. On TikTok, captions are often minimal, informational, or strategically ambiguous, while polarising meaning may travel through identity cues, recontextualisation, and comment dynamics that are only partially visible in the text layer (Wodak & Meyer, 2016). The dominance of neutral sentiment is therefore interpreted as a platform-specific pattern of moderated affect: polarising work can circulate in forms compatible with wide visibility and everyday participation, rather than relying primarily on overtly negative or toxic caption language (Joyeux-Prunel, 2024).

The sequential sentiment design matters here. Lexical and contextual models can register neutrality even when captions act as "hooks" that point to polarising meanings embedded elsewhere (in video rhetoric, duet/stitch commentary, or comment escalation). This supports the study's operationalisation of polarisation as multi-layered: affective signals are one window into boundary-making, but not a complete proxy for it (Iyengar et al., 2012; Wodak & Meyer, 2016).

Thematic concentration: where identity discourse clusters

Theme extraction indicates that identity-related discourse clusters most strongly in sociocultural talk (52.19%), followed by politics (19.40%), religious provocation (9.83%), disinformation/hoax framings (8.09%), racial/ethnic discourse (7.12%), and mobilisation calls (3.37%). The distribution suggests that Malaysian TikTok polarisation is not confined to explicit political debate; it is more frequently embedded in everyday cultural talk where moral evaluation, respectability, and belonging are negotiated. In multi-identity contexts, "culture" can function as a proxy arena for religious and ethnoreligious tension rather than a separate domain, making sociocultural discourse a key substrate for boundary work (Wodak & Meyer, 2016).

To avoid treating themes as mere topic labels, the analysis incorporates priority signals—keywords and hashtags classified into high-, medium-, and low-priority categories—to identify where confrontation intensity becomes lexically explicit in the caption layer. Across themes, most keywords fall into low-priority categories, indicating that polarisation often circulates through routine descriptors and topical tags rather than through slurs alone. However, high-priority concentrations differ sharply by theme, revealing where identity boundaries harden into overt antagonistic sorting.

Cultural and social polarisation: boundary work through heritage and moral judgement

Cultural polarisation is visible through hashtags and keywords that emphasise heritage and identity reinforcement (e.g., #adat, #budaya), framing content around communal values and cultural legitimacy. The absence of a dominant set of high-priority sociocultural hashtags suggests decentralised, "grassroots" participation rather than centrally coordinated campaigning, even as discourse performs boundary work by differentiating legitimate versus deviant cultural positions.

Social polarisation becomes more explicit in high-priority sociocultural keywords dominated by public moral judgement and gender conflict—*fitnah sosial* (social slander), *hina lelaki* (insulting men), *hina perempuan* (insulting women). These cues are consistent with performative condemnation and public shaming dynamics: the content positions audiences as moral evaluators and invites alignment through judgement rather than deliberation. This

pattern helps explain why sociocultural discourse is the largest theme cluster despite relatively neutral sentiment: polarising work is enacted through routine moral policing and identity-coded respectability talk, not only through overt negative affect (Wodak & Meyer, 2016).

Religious polarisation is the most explicit antagonism at the keyword layer. Religious provocation exhibits the highest share of high-priority keywords (32%), including antagonistic and exclusionary terms such as anti-Islam, *benci Islam* (hate Islam), *hina agama* (insult religion), *kafir* (blasphemy), and *murtad* (apostasy). These markers indicate that religious identity can operate as an existential boundary that invites loyalty signalling and out-group suspicion. This aligns with affective polarisation accounts where group evaluation and distrust intensify beyond policy disagreement (Iyengar et al., 2012). In this corpus, religious polarisation is therefore the most linguistically explicit form at the caption layer, concentrating confrontation intensity in a way that differs from the more insinuation-driven style of political discourse.

Political polarisation illustrates vernacular critique and delegitimation cues. Political discourse (19.40%) is shaped more by vernacular critique, satire, and distrust cues than by the most explicit antagonistic lexicon found in religious provocation. This suggests a mechanism where polarisation is enacted through framing and insinuation—who is treated as credible, who is mocked, and which interpretations are made "common sense"—rather than primarily through direct insult. The pattern is consistent with discourse-based accounts of polarisation where legitimacy is negotiated through rhetorical positioning and narrative templates that render opponents immoral, incompetent, or threatening without necessarily using extreme lexical markers (Wodak & Meyer, 2016).

Racial/ethnic polarisation revolves around issues of belonging, entitlement, and recontextualised proof fragments. Racial/ethnic discourse (7.12%) appears as boundary work around belonging, entitlement, and historical grievance. In this thematic space, recontextualised fragments can function as "proof texts" that enable rapid judgement and moral ranking, condensing complex histories into circulating cues (Wodak & Meyer, 2016). Although the racial/ethnic theme is smaller than sociocultural and political clusters, its significance lies in how quickly identity cues can activate affective sorting and escalate interpretive conflict, particularly when linked to sensitive historical references.

Disinformation/hoax framings and mobilisation calls involve credibility contests and coordination cues. Disinformation/hoax framings (8.09%) and mobilisation calls (3.37%) operate primarily as credibility contests and coordination cues rather than as verifiable truth assessments. This is consistent with information disorder frameworks that treat "hoax" and "fake" labels as discursive resources: accusations of deception can delegitimise opponents, assert epistemic authority, and mobilise attention within polarised environments regardless of the underlying claim's veracity (Wardle & Derakhshan, 2017; Budak et al., 2024). These themes therefore matter analytically because they reveal how polarisation is sustained through disputes over credibility and collective alignment, not only through ideological disagreement.

Synthesis and answers to the research questions

Overall, the results support a pattern of "hidden" polarisation. Affective boundary-making persists even when caption sentiment appears neutral because visibility is not limited to toxic language and can be sustained through moderated identity signalling and moral framing (Iyengar et al., 2012). DH triangulation—sentiment distributions, thematic clustering, priority signals, and engagement proxies—shows that explicit antagonism concentrates in specific identity domains (especially religious provocation), while the most prevalent polarising work occurs through sociocultural discourse via everyday moral evaluation and shaming cues.

RQ1 (circulation) inform how polarised content circulates primarily through hashtag-mediated clustering and theme-based participation. Most posts engage sociocultural and political discourse as the main substrate, while episodic spikes of explicit antagonism concentrate in religious provocation and, to a lesser extent, racial/ethnic discourse. RQ2 (visibility/exposure) illustrates how exposure appears shaped by engagement-oriented visibility combined with creator practices that increase algorithmic legibility (notably extensive hashtagging and short-form performative genres). The dominance of neutral captions alongside identity-salient themes suggests that platform-mediated attention may consistently reward indirect or "safe" modes of boundary-making—humour, sarcasm, moral critique—more than overt hostility, allowing polarising meanings to circulate widely while remaining compatible with platform norms and everyday participation (Joyeux-Prunel, 2024).

Discussion

Across the corpus, neutral captions dominate and receive the highest engagement. Rather than reading this as evidence of a low-polarisation environment, the DH interpretation suggests a TikTok-specific style in which polarising work is often performed indirectly. Short captions, strategic ambiguity, and participation in trends allow users to signal alignment while avoiding overtly hostile lexical cues. In CDS terms, this is compatible with polarisation enacted through evaluative framing and moral hierarchy construction rather than direct insult: "common sense" claims and identity cues can normalise in-group legitimacy while rendering out-groups deviant or threatening (Wodak & Meyer, 2016).

The findings therefore challenge approaches that equate polarisation primarily with toxicity or negative sentiment. On TikTok, the caption layer can understate affect carried by audiovisual rhetoric and comment interaction. The sequential modelling and thematic clustering should be understood less as a sentiment "thermometer" than as a DH instrument for mapping where identity meanings concentrate and how they are packaged for circulation (Joyeux-Prunel, 2024). The thematic distribution—sociocultural dominance followed by politics and religious provocation—supports a boundary-making interpretation: polarisation is enacted through moral evaluation, respectability talk, and "who belongs" narratives rather than solely through explicit partisan conflict.

Priority signals sharpen this account by separating prevalence from confrontation intensity. Heritage hashtags #adat (custom) and #budaya (culture) indicate identity reinforcement without centralised coordination, while high-priority sociocultural keywords show how public shaming and gender conflict become repeated polarising cues. This points to an ambient form of social polarisation reproduced through routine moral policing rather than only campaign-driven antagonism.

Religious provocation stands out as the most lexically explicit domain, showing the highest concentration of high-priority antagonistic and exclusionary keywords. This suggests that religion functions as a high-stakes identity boundary in Malaysian TikTok discourse, rapidly triggering affective sorting and inviting loyalty signalling and suspicion of out-groups. This aligns with affective polarisation theory, which emphasises intensified group evaluation and distrust beyond policy disagreement (Iyengar et al., 2012). At the same time, the presence of tolerance-oriented and contesting vocabulary within the broader religious theme indicates that religious polarisation is not a single-direction trajectory. It is better understood as a contested discursive field in which antagonistic and coexistence frames compete for attention and circulation.

With respect to RQ2, the study does not claim causal identification of TikTok's recommendation system. Instead, it treats engagement as a visibility proxy and examines how creators structure posts for algorithmic legibility, particularly through extensive hashtagging. Within these constraints, the results support a cautious inference.

Visibility on Malaysian TikTok is compatible with polarising meanings packaged in moderated affective forms. This resonates with scholarship arguing that engagement-oriented systems can privilege emotionally resonant and identity-confirming content even when it avoids overt negativity (Wardle & Derakhshan, 2017; Budak et al., 2024).

Methodologically, this is where the DH workflow contributes. By triangulating sentiment distributions, thematic clustering, priority signals, and engagement patterns, the analysis treats algorithmic exposure as an interpretive problem rather than a black-box claim. This approach aligns with post-computational reproducibility principles: the scholarly contribution includes the documented pipeline that enables scrutiny of how corpus construction, preprocessing, and modelling choices shape interpretation (Joyeux-Prunel, 2024). Disinformation/hoax framings and mobilisation calls are particularly illustrative. The study does not adjudicate truth; instead, it shows how "fake" and "hoax" labels operate as boundary work—delegitimising opponents, asserting epistemic authority, and mobilising attention—within fragmented publics (Wardle & Derakhshan, 2017). This interpretation also fits recent cautions that misinformation harms should not be reduced to a simple deficit model; rather, they should be understood in relation to identity, trust, and the political economy of attention (Budak et al., 2024).

Three limitations qualify these conclusions. First, captions represent only one layer of TikTok meaning; audiovisual and comment dynamics may carry affect and antagonism not visible in text traces. Second, hashtag-led sampling captures a structured slice rather than an exhaustive map of Malaysian TikTok, potentially over-representing highly "taggable" conversations. Third, despite triangulation, sentiment and ABSA remain vulnerable to code-mixing, slang, and sarcasm. Future research should extend this DH workflow to multimodal signals (audio reuse, visual motifs, stitching networks) and event-centred corpora aligned with specific civic moments (e.g., GE15) to strengthen claims about exposure pathways while maintaining reproducibility commitments (Joyeux-Prunel, 2024).

Conclusion

This study developed a Digital Humanities computational workflow to analyse polarisation and identity discourse on Malaysian TikTok using a corpus of 1,841 publicly accessible videos posted between 2023 and 2025. By integrating hashtag-led corpus construction, sequential sentiment modelling (VADER, DistilBERT, multilingual ABSA), computational theme extraction, priority-signal interpretation, and engagement-based validation, the study demonstrates how polarisation can be examined as a platform-trace phenomenon while retaining interpretive accountability (Joyeux-Prunel, 2024).

Empirically, the corpus is characterised by moderated affect by which neutral captions dominate and achieve the highest engagement, yet polarising meanings persist through identity signalling and boundary work. Sociocultural discourse constitutes the largest thematic cluster, indicating that polarisation is often embedded in everyday moral evaluation and respectability talk rather than confined to formal political debate. Explicit antagonism concentrates most strongly within religious provocation, where high-priority keywords indicate that religious identity can function as a high-stakes boundary inviting loyalty signalling and out-group suspicion (Iyengar et al., 2012). Disinformation/hoax framings and mobilisation calls operate primarily as credibility contests and coordination cues, reinforcing the view that information disorder forms part of polarisation's discursive infrastructure rather than merely reflecting a "truth deficit" (Wardle & Derakhshan, 2017; Budak et al., 2024).

Addressing the research questions, the findings suggest that polarised content circulates primarily through hashtag-mediated clustering and theme-based participation (RQ1), with everyday sociocultural discourse providing the main substrate for periodic spikes of identity conflict. Exposure (RQ2) appears shaped by the interaction of engagement-oriented visibility and creator tactics that optimise algorithmic legibility—supporting a pattern of "hidden"

polarisation where indirect boundary-making circulates widely without relying on overtly hostile caption language. As a DH contribution, the article advances a reproducible approach to TikTok polarisation research that treats workflows, provenance, and time-stamping as part of the scholarly output, a necessary stance in platform research where data access is volatile and interpretive claims depend on transparent documentation (Joyeux-Prunel, 2024). Future work should extend the workflow to multimodal analysis and controlled audit designs to strengthen claims about exposure pathways, while retaining interpretive sensitivity to multilingual and culturally specific identity discourse in Malaysia.

Declarations

Author Contribution

N.K. led the study conceptualisation and research design, developed the Digital Humanities framing of polarisation and identity discourse, and wrote the main manuscript text (including the theoretical framing, interpretation of results, and discussion/conclusion). J.C. and J.J. contributed equally to the computational workflow: they conducted TikTok data collection/scraping and corpus construction, performed data cleaning and text pre-processing, implemented the sequential sentiment modelling (lexical baseline, transformer-based modelling, and aspect-based sentiment analysis), carried out thematic extraction/clustering and engagement-metric analysis, and produced the tables/figures supporting the results. N.K., J.C., and J.J. jointly interpreted findings, revised the manuscript for intellectual content, and approved the final version for submission.

Acknowledgement

This article is developed through collaborative efforts and a broader research project on digital polarisation supported by the Social Research Grant/Dana Kajian Sosial (KPT.BDP(S).600-2/4/3 JLD.2(39). The author would like to thank the Ministry of Higher Education for the grant and the anonymous reviewers for their constructive feedback and insightful comments, which helped improve the quality of this manuscript.

Data Availability

The study did not involve engagement with human participants or interaction. The research uses platform-available digital materials and computational corpus processing. Thus, no participants were involved, and informed consent was unnecessary.Supporting materials are prepared for editorial and peer review. Data, scripts, and documentation are available on the submission platform. The data package includes the retained final structured dataset with sentiment fields, a deidentified reviewer copy of that dataset, the retained SPSS base analysis file, a retained raw-style TikTok scrape export and a deidentified reviewer copy of that export, together with the retained hashtag and keyword support files and explanatory documentation.For transparency, the attached materials reflect the files currently retained by the author team rather than a complete reconstruction of every workflow stage described in the manuscript. In particular, the retained package does not presently include the full series of original raw scrape exports, intermediate processing files, Python scripts, or SPSS syntax/output files. A note within the package explains the status and apparent role of each file.

References

1. Apify (2025) *TikTok Scraper (clockworks/tiktok-scraper)* [Software]. Retrieved February 27, 2025, from https://apify.com/clockworks/tiktok-scraper
2. Arora SD, Singh GP, Chakraborty A, Maity M (2022) Polarization and social media: A systematic review and research agenda. Technol Forecast Soc Chang 183:121942. https://doi.org/10.1016/j.techfore.2022.121942
3. Ayalew YE (2025) Linguistic blind spots in platform content moderation in Ethiopia: The case of Facebook. Mod Lang Open 2025(1).),Article468.https://doi.org/10.3828/mlo.v0i0.468
4. Barrett P, Hendrix J, Sims G (2021) September27).*How tech platforms fuel U.S. political polarization and what government can do about it*.BrookingsInstitution.https://www.brookings.edu/articles/how-tech-platforms-fuel-u-s-political-polarization-and-what-government-can-do-about-it/
5. Bliuc A-M, Betts JM, Vergani M, Bouguettaya A, Cristea M (2024) A theoretical framework for polarization as the gradual fragmentation of a divided society. Commun Psychol 2:75. https://doi.org/10.1038/s44271-024-00125-1
6. Budak C, Nyhan B, Rothschild DM, Thorson E, Watts DJ (2024) Misunderstanding the harms of online misinformation. Nature 630(8015):45–53. https://doi.org/10.1038/s41586-024-07417-w
7. Chinnasamy S, Volkmer I (2023) Public actors in new spaces: A case study of digital Malaysia in transnational public deliberation. Pac Journalism Rev 29(1–2):136–156. https://doi.org/10.24135/pjr.v29i1and2.1231
8. Costantino A (2024) December5).*The digital divide: Social media's role in polarization and democratic backsliding*.DemocraticErosion.https://democratic-erosion.org/2024/12/05/the-digital-divide-social-medias-role-in-polarization-and-democratic-backsliding/
9. Cotter K, DeCook JR, Kanthawala S (2022) Fact-checking the crisis: COVID-19, infodemics, and the platformization of truth. Social Media + Soc 8(1). https://doi.org/10.1177/20563051211069
10. Ferreira GB, Ferreira LS (2025) The polarization paradox: Social media, young voters, and the challenges to the open society. Preprints. https://doi.org/10.20944/preprints202502.1941.v1
11. Fortunato P, Pecoraro M (2022) Social media, education, and the rise of populist Euroscepticism. Humanit Social Sci Commun 9 Article 301. https://doi.org/10.1057/s41599-022-01317-y
12. Germano F, Gómez V, Sobbrio F (2025) Ranking for engagement: How social media algorithms fuel misinformation and polarization. SSRN Electron J. https://doi.org/10.2139/ssrn.5316506
13. Goh DPS (2023) Racial culture wars in Indonesia, Malaysia, and Singapore. InOxford Research Encyclopedia of Communication. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.1336
14. Gomila Olvera A (2025) *Understanding digital communities' polarization: A study of community detection algorithms, polarization metrics and applications*(Bachelor'sthesis).UniversitatPolitècnicadeCatalunya
15. Hamzah M, Naidu W, Lee SF, Naidu D, Balachandar D (2023) *Social media monitoring of Malaysia's 15th general election*.CentreforIndependentJournalismMalaysia.https://cijmalaysia.net/wp-content/uploads/2023/05/GEM-Report-ENG.pdf
16. Hartmann D, Wang SM, Pohlmann L, Berendt B (2025) A systematic review of echo chamber research: Comparative analysis of conceptualizations, operationalizations, and varying outcomes. J Comput Social Sci 8:52. https://doi.org/10.1007/s42001-025-00381-z
17. Hutto CJ, Gilbert E (2014) VADER:Aparsimoniousrule-basedmodelforsentimentanalysisofsocialmediatext.In*Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*(pp.216–225).AAAIPress.https://doi.org/10.1609/icwsm.v8i1.14550
18. Iyengar S, Sood G, Lelkes Y (2012) Affect, not ideology: A social identity perspective on polarization. Pub Opin Q 76(3):405–431. https://doi.org/10.1093/poq/nfs038

19. Jackson D, Hendrix J, Bernard T (2024) October29).*The science of social media's role in January 6*.TechPolicyPress.https://techpolicy.press/the-science-of-social-medias-role-in-january-6
20. Jalli N (2025) Viral justice: TikTok activism, misinformation, and the fight for social change in Southeast Asia. Social Media + Soc 11(1). https://doi.org/10.1177/20563051251318122
21. Joyeux-Prunel B (2024) Digital humanities in the era of digital reproducibility: Towards a fairest and post-computational framework. Int J Digit Humanit 6:23–43. https://doi.org/10.1007/s42803-023-00079-6
22. Kaur M, Yusof IM (2023) October12).*Malaysia: TikTok not abiding by laws or curbing fake news*.BenarNews.https://www.benarnews.org/english/news/malaysian/tiktok-incompliant-with-malaysian-law-minister-says-10122023111759.html
23. Kirk G (2025) July30).*Digital extremism and the rise of the green wave in Malaysia*.CenterfortheStudyofOrganizedHate.https://www.csohate.org/2025/07/30/digital-extremism-malaysia/
24. Lim M (2024) Social media and politics in Southeast Asia. Cambridge University Press
25. Matheus V (2025) January27).*Digital polarization and disinformation: How to rebuild online public debate*.HiveMind.https://en.hive-mind.community/blog/1049,digital-polarization-and-disinformation-how-to-rebuild-online-public-debate
26. Mohd Sani MAM, Kasim A (2024) MalaypoliticsandsocialmediainMalaysia.InJ.Chin&P.P.Y.Leong(Eds.),*Social media and political communities in Malaysia*(pp.54–90).ISEAS–YusofIshakInstitute.https://doi.org/10.1355/9789815203165-004
27. Oates H (2025) Digital colonialism: An analysis of Facebook's role in the Rohingya genocide. ANU Undergrad Res J 14(1):36–46. https://harrisonoates.com/papers/digitalcolonialism/
28. Papacharissi Z (2014) Affective publics: Sentiment, technology, and politics. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199999736.001.0001
29. Papacharissi Z (2015) Affectivepublicsandstructuresofstorytelling:Sentiment,eventsandmediality.*Information, Communication & Society*.Advanceonlinepublication.https://doi.org/10.1080/1369118X.2015.1109697
30. Papathanassopoulos S, Giannouli I (2025) Political communication in the age of platforms. Encyclopedia 5(2).),Article77.https://doi.org/10.3390/encyclopedia5020077
31. Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S (2014) SemEval-2014task4:Aspectbasedsentimentanalysis.In*Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*(pp.27–35).AssociationforComputationalLinguistics.https://doi.org/10.3115/v1/S14-2004
32. Pratap A, Pathak A (2025) From public square to echo chamber: The fragmentation of online discourse. arXiv. https://doi.org/10.48550/arXiv.2501.18441
33. Saidin MIS, Azrun N (2024) Digital media and religious sentiments in Malaysia: Critical discourse analysis of Pan-Malaysian Islamic Party and Democratic Action Party cyberspace campaigns in the 15th general election. Religions 15(8):920. https://doi.org/10.3390/rel15080920
34. Sanh V, Debut L, Chaumond J, Wolf T (2019) DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv. https://doi.org/10.48550/arXiv.1910.01108
35. Schissler M (2025) Beyond hate speech and misinformation: Facebook and the Rohingya genocide in Myanmar. J Genocide Res. Advance online publication https://doi.org/10.1080/14623528.2024.2375122
36. Scolere L, Pruchniewska U, Duffy BE (2018) Constructing the platform-specific self-brand: The labor of social media promotion. Social Media + Soc 4(3):1–11. https://doi.org/10.1177/2056305118784768

37. Shukri S (2024) *Why young Malay voters in Malaysia are turning green*(TrendsinSoutheastAsiaTRS12/24).ISEAS–YusofIshakInstitute.https://www.iseas.edu.sg/wp-content/uploads/2024/04/TRS12_24.pdf
38. Snyder H (2019) Literature review as a research methodology: An overview and guidelines. J Bus Res 104:333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
39. Stray J, Iyer R, PuigLarrauri H (2023) The algorithmic management of polarization and violence on social media. SSRN Electron J. https://doi.org/10.2139/ssrn.4429558
40. Sustainability Directory (2024) March31).*Digital polarization*.https://climate.sustainability-directory.com/term/digital-polarization/
41. Tapsell R (2023) Social media and Malaysia's 2022 election: The growth and impact of video campaigning. Pac Affairs 96(2):303–322. https://doi.org/10.5509/2023962303
42. Wardle C, Derakhshan H (2017) *Information disorder: Toward an interdisciplinary framework for research and policy making*.CouncilofEurope.RetrievedMay27,2025,fromhttps://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html
43. Whittemore R, Knafl K (2005) The integrative review: Updated methodology. J Adv Nurs 52(5):546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x
44. Wike R, Silver L, Fetterolf J, Huang C, Austin S, Clancy L, Gubbala S (2022) December6).*Social media seen as mostly good for democracy across many nations, but U.S. is a major outlier*.PewResearchCenter.RetrievedMay27,2025,fromhttps://www.pewresearch.org/global/2022/12/06/social-media-seen-as-mostly-good-for-democracy-across-many-nations-but-u-s-is-a-major-outlier/
45. Wodak R, Meyer M (eds) (2016) Methods of critical discourse studies, 3rded. edn. SAGE