



OPEN ACCESS

TikTok Algorithms and the Escalation of Interfaith Conflict in Indonesia

Muhamad Jamil*

STAI Yayasan Tarbiyah Islamiyah Padang, Indonesia

Abdullah Khusairi

UIN Imam Bonjol Padang, Indonesia

Article History	Abstract
Received February 5, 2025	<i>This study examines the escalation of interreligious conflict on the TikTok platform, which is increasingly dominating social interaction in Indonesia. By analysing five live videos related to religious conflict (broadcast between January 2023 and December 2024) and 853,513 Indonesian-language comments through thematic coding and sentiment classification, it was found that TikTok's algorithm plays a significant role in reinforcing ideological polarization. Content nuanced with religious conflict generated an average of 3.5 times more comments, and videos containing negative emotions were shared four times more often than neutral or positive content. This digital conflict often spilled over into the real world, triggering police intervention. The response of law enforcement officers was generally reactive, unable to mitigate the conflict early. These findings emphasize the need for strategic collaboration between digital platforms, law enforcement officials, and civil society to promote digital literacy and ethical interfaith dialogue. This collaboration is expected to prevent ideological segregation and create a more constructive interaction space in the digital era.</i>
Revised June 17, 2025	
Accepted June 26, 2025	
Published July 30, 2025	

Keywords: Inter-religious conflict; social media; TikTok; polarization of society; algorithm

*Correspondence: Muhamad Jamil, STAI Yayasan Tarbiyah Islamiyah Padang, Indonesia
jamil@staiyastipadang.ac.id



© Author(s), 2025. Published by Balai Penelitian dan Pengembangan Agama Semarang.
This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

Social media has fundamentally transformed the ways in which individuals interact, communicate, and disseminate information, reshaping the social and political landscape on a global scale in a manner that has never occurred before. The roots of this digital revolution can be traced from the early platforms that enabled online

connections, such as Six Degrees in the late 1990's, to its rapid evolution through emergence massive global network such as Facebook and *microblogging sites* that define repeat news and discussion public such as Twitter (Boyd and Ellison 2008; VanDoorn and Eklund 2013). Each era introduces new communication dynamics

that shape how information is consumed and distributed, thereby significantly influence social structures (GD 2023).

In today's rapidly evolving environment, TikTok, which was launched in 2016 by ByteDance, has surprisingly emerged as a dominant global force. This platform captivates billions of users with its unique short video format, intuitive *interface*, and intuitive mechanisms and delivers very interesting as well as personalized content. TikTok's success is partly driven by a sophisticated algorithm designed aggressively to maximize user engagement and screen time (Koç 2023).

Disinformation on digital platforms has become a major catalyst for social conflict worldwide. Its devastating effects are evident in many cases around the globe. In India and Myanmar, for example, research by Arun 2019 underlines sharply how the uncontrolled distribution of misinformation on social media directly correlates with escalation of inter-community tension. This study shows the mechanism by which speed and range of fake information, often without adequate moderation, can aggravate prejudice that has been there, trigger violence based on identity, and deepen fault lines socially.

Furthermore, in the United States, Tucker et al. (2018) highlighted the role social media centers in eroding trust in the public, especially in landscape highly polarized politics. Research indicates how platforms of systematic algorithms create filter bubbles and echo chambers, amplifying existing beliefs and in active limit exposure for users to different perspectives. These fundamental changes in the methods individuals use to process information have led to a significant decline in trust in democratic institutions and mainstream mass media. These global cases firmly underline the crucial urgency to understand the dynamics of information distribution on social media, especially its

impact on social cohesion and the potential escalation of conflict in pluralistic societies (Deitelhoff and Schmelzle 2022).

Uniqueness and strengths that set it apart from the previous platform lie in the architecture of its algorithm which is very powerful and addictive. The algorithm in a way effectively prioritizes and encourages distribution content that generates involvement and triggers strong emotions (Silvanie et al. 2024). This is inherently leading to amplifying sensational, provocative, or offensive content, even controversial, because such content proves to be most effective in maximizing users retention and interactions within the platform.

From the theoretical perspective This kind of behavioral algorithmic own correlation is close to the phenomenon of ideological polarization. This concept is widely explained by the theory of filter "bubbles" and "echo chambers" (Pariser 2011; Lazer et al. 2018; Kreiss and McGregor 2024). These theories argue that social media algorithms with careful adaptive information and preferences that have been shown to users, in particular and on purpose, create an isolated information environment. In such environments, individuals are frequently exposed to views that align with their existing beliefs, thereby reinforcing their confirmation bias.

This phenomenon not only strengthens communal solidarity among—open minded groups, but simultaneously also reduces openness to different perspectives, which ultimately aggravate ideological split in society (Sunstein 1999). This mechanism explains how social media as an independent variable can effectively strengthen polarization as a mediating variable, which then substantially contributes to conflict as a dependent variable and turns online discussions into an ideological battleground (Khan 2023).

In Indonesia, a country that is fundamentally known with its highly pluralistic society and extraordinarily diverse religious landscape, TikTok has commonly gathered extraordinary popularity and very deep market penetration, especially in demography among youth. In 2021, Indonesia even occupied a global top ranking among countries with most TikTok users, with more from 22 million users active monthly (Bhandari and Bimo 2022). This shows that TikTok is not only an entertainment platform, but also a dominant public space.

Although this platform offers opportunities for self-expression and a wide range of entertainment, this widespread adoption is also accompanied by profound and urgent challenges, especially those concerning potential escalation of inter-religious conflict in the digital realm (Ibrahim 2020). While TikTok's algorithm is designed to boost user engagement, it also carries the inherent risk of deepening ideological polarization and intergroup tensions. This raises a significant threat to harmony for fragile social conditions in multicultural Indonesia since sensitive issues related to religion—such as interpretation doctrine, ritual differences, or even content that is considered insulting can with rapidly worsen atmosphere, trigger sensitive reaction and might devastate emotions of split community (Fabriar and Muhajarah 2024).

Given the social media influences to intergroup conflict and ideological polarization, previous studies paid attention to academic literatures. Mostly, existing research has just focused on platforms such as Facebook and Twitter. As a result, the role of TikTok's interest-based algorithms—known for promoting viral, visual, and emotionally charged content—in influencing and potentially exacerbating inter-religious conflict in Indonesia has not been sufficiently explored (Koç 2023).

This research is of paramount importance, addressing significant gaps in understanding the dynamics of TikTok's unique algorithmic approach, its visually-driven format, its extensive presence in Indonesia, and the implications for platform users and law enforcement institutions responding to this phenomenon. Consequently, this study aims to provide a crucial perspective that not only fills existing literature gaps but also proves invaluable in developing a more effective and comprehensive strategy for managing and mitigating social conflicts in the modern digital age, particularly concerning sensitive religious issues in Indonesia.

Literature Review

Previous studies on social media and conflict can be grouped into three broad interrelated themes: (1) the impact of algorithms in shaping ideological polarization, (2) the role of social identity in reinforcing segregation and conflict, and (3) the scarcity of studies relevant to the Indonesian context, particularly the TikTok platform.

Early studies consistently highlighted the central role of social media algorithms in reinforcing ideological differences between groups. Lazer et al. (2018) and Flaxman et al. (2016), for example, used large-scale quantitative approaches to analyze data from Facebook and Twitter, convincingly showing how these platforms create filter bubbles and echo chambers that limit users' exposure to diverse views. While the strength of these studies lies in the strong empirical evidence on the technical mechanisms of algorithms, their crucial limitation is their focus on general platforms and their failure to explore in depth how the relationship between algorithms and specific religious expressions such as on TikTok works as a trigger for conflict.

Kreiss and McGregor (2024) further expanded this understanding through qualitative analysis, showing that recommendation systems not only create filter bubbles but also accelerate opinion segmentation by prioritizing emotional and provocative content, although the generalizability of these findings may be limited to the types of content and platforms they studied. Comparisons between these studies suggest a consensus on the existence and mechanisms of algorithm-driven filter bubbles, but there is a gap in understanding the specific impact on inter-religious conflict dynamics, a crucial area given TikTok's reputation for its highly adaptive algorithms to emotional preferences and viral content.

Using social identity theory Tajfel and Turner (2004) and group polarization theory Sunstein (1999), they explain that inter-group polarization and conflict are reinforced by individuals' need to maintain and affirm their in-group identity while discrediting their out-group. Tajfel & Turner theoretically outline how group identification can lead to inter-group bias, a fundamental framework that requires adaptation to the digital context as the anonymity and rapid dissemination of content on social media can alter the dynamics of intergroup bias. Empirical studies by Shu et al. 2017 through online experiments, and Naslund et al. 2016 through online conversation analysis, added that interactions that occur exclusively within a group can reinforce negative stereotypes towards outsiders and reduce empathy.

However, a limitation of most of these studies is their tendency to focus on political or ethnic conflicts in general, rather than on specific interreligious conflicts based on social media. The comparison shows that although social identity and polarization theories are highly relevant, their application and validation in the context of

religious dynamics on social media platforms still require further exploration, especially how social media algorithms can inadvertently bring opposing groups together or reinforce certain group identities, which can trigger conflict.

Although the social media landscape in Indonesia has been the subject of many studies, such as those conducted by Jamil et al. (2024) with a qualitative-descriptive approach on Facebook, and Hew (2018) on the impact of Twitter on public discussion, most of these studies predominantly focus on these traditional platforms. This collective limitation is their failure to capture the unique dynamics of newer and rapidly growing platforms such as TikTok, which has a fundamentally different algorithm and user base (Zulli and Zulli 2022). Unlike Facebook or Twitter, which are more text-based and social networking, TikTok, with its short-form video format and AI-driven For You Page (FYP), has a greater potential for rapid viral content dissemination without considering deep context, which can trigger instant emotional reactions and conflict (Dancharoenpol 2022; Čábyová and Krajčovič 2024).

Koç 2023 is one of the few studies that explicitly examines the TikTok algorithm and its impact on the echo chamber, but this study is limited to an analysis of the algorithm's technical functions without an in-depth exploration of its socio-cultural impacts in the Indonesian context especially in triggering inter-religious conflict, especially in relation to the pattern of user responses to provocative content and the form of intervention taken by law enforcement, which is often reactive (Permana and Aurellie 2025).

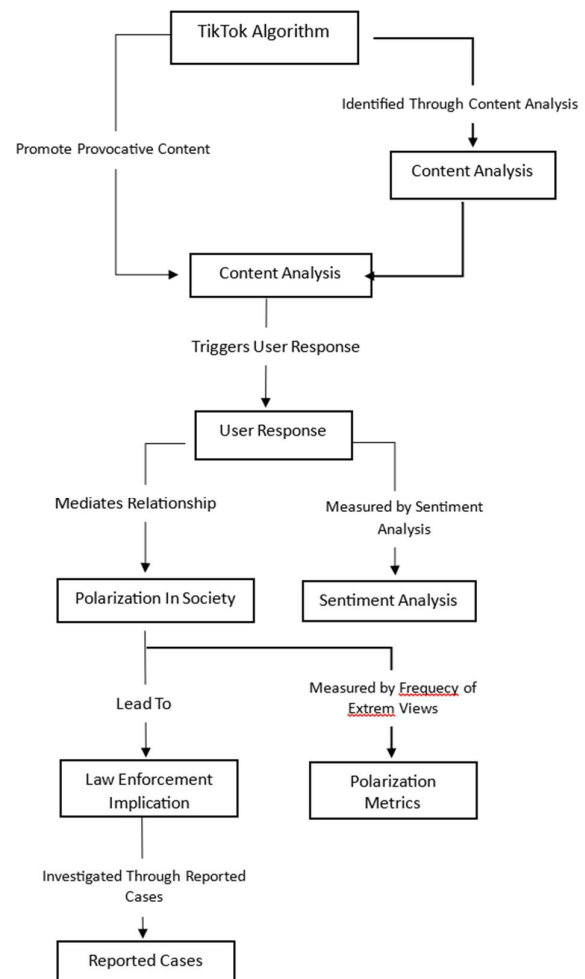
This study aims to fill this gap with a more in-depth analysis in the Indonesian TikTok context, examining how user responses contribute to the escalation of inter-religious conflict.

Conceptual Framework

This conceptual framework aims to map and analyze the complex connections between the TikTok algorithm, the provocative spread of content, user responses, societal polarization, and the implications for law enforcement within the context of interfaith conflict in Indonesia. Unlike previous studies that often focus on traditional social media platforms, this framework uniquely highlights the dynamics of TikTok's algorithm, which drives interactions based on user interests and has the potential to accelerate ideological polarization and conflict escalation.

The emphasis on TikTok and its offline implications in Indonesia represents a significant departure in the approach to this study. In general, this specific framework identifies the TikTok algorithm as the primary influencing factor in the distribution of provocative content. Subsequently, this provocative content elicits user responses, which serve as a mediating focus, thereby reinforcing or hastening the occurrence of societal polarization, which is the main impact of concern. Ultimately, the escalation of societal polarization and the conflicts that arise online can lead to the enforcement of real-world laws, which we also consider as an external focus of this digital conflict. This framework aims to describe several key components and their interrelated interactions., such as depicted in Figure 1. Framework Conceptual Research below this:

Figure 1. Conceptual framework



Source: Authors' conceptualization

Initially, the primary focus of the analysis will be on the influence of the TikTok algorithm and provocative content. The TikTok algorithm is characterized as a recommendation system that personalizes content to engage users based on their historical interactions (e.g., likes, shares, and watch duration). This study is predicated on the assumption that the algorithm inherently tends to favor provocative content, specifically videos or comments that include expressions of hatred, misinformation, or disinformation regarding other religious groups, or narratives that incite hostility. The material will be identified and evaluated through thematic content analysis of viral videos and

comments, and assessed based on engagement metrics (number of views, comments, and shares).

Subsequently, the focus will be on mediation, its outcomes, user responses, and public polarization. The term 'response user' pertains to the audience's engagement and their emotional reactions to provocative contexts, which will be evaluated through sentiment analysis (positive, negative, and neutral) of comments and shared patterns. This research does not assume that response mediates the relationship between the TikTok algorithm and societal polarization. Polarization is characterized by an increase in ideological segregation, a reinforcement of identity, and heightened hostility towards external groups (Amit and Venzhik 2024). The symptoms of this phenomenon will be assessed by the frequency of extreme expressions, the stigmatization of other groups, and the observed decline in tolerance within online comments and discussions.

Then, as the focus shifts towards the conclusion, the implications of enforcement law will be examined. The term 'implications of enforcement law' pertains to the official actions (such as police reports, investigations, and arrests) undertaken by authorized parties in response to the escalation of conflict on TikTok. Our research will investigate how this online conflict escalates into offline consequences that necessitate legal intervention, as measured by reported cases to law enforcement and documented responses in enforcement law data.

To understand a comprehensive dynamics explained in framework this, data analysis will be based on two main mutually complementary theories. First, the Polarization Theory by Sunstein (Sunstein 1999). This theory states that when individuals with the same view interact exclusively in a group, they tend to become more extreme than before interaction. This

is aggravated in an online environment where algorithms can create 'space echo' and 'filter bubble', limiting exposure of individuals to diverse perspectives (Mahmoudi et al. 2024; Rodillosso 2024; Coady 2024). In the context of TikTok, the algorithm is designed to maximize engagement and relevance content in a way No direct push confirmation bias reinforcement. Users who are consistently exposed to content that confirms their religious views will tend to become more dogmatic and lacking tolerance to other beliefs.

Therefore, this theory will be applied for interpreting data about how content provocatively recommended by the algorithm strengthens existing beliefs and encourages group users to more position extremes, as seen from analysis sentiment and narrative in comment.

This also refers to the Social Identity Theory by Tajfel and Turner (2004) as quoted by Begby (2022). This theory states that the individual to obtain a part drafts himself from membership in group social, and they tend to support and differentiate his group alone (*in-group*) from other groups (*out-group*) (Bail et al. 2018). The identification process often involves comparison social that can trigger sentiment negative to *out-group*.

In the TikTok environment, the algorithm can facilitate formation of strong religious groups with recommended content that strengthens in-group solidarity. When the content provocative challenging belief or symbol *out group* appear, things That trigger mechanism identity social this, improve cohesion *in-group* but in a way simultaneously to worsen tension and hostility. With Thus, the theory This will be used for analyzing how language used in comment reflect identification *in-group* and *out-group*, as well as how bias and stereotypes group reinforced by recommended TikTok content.

RESEARCH METHOD

This study uses a qualitative approach to examine interfaith conflicts that occur in user interactions on the TikTok platform. This research was conducted from January 2023 to December 2024, focusing on live video content uploaded during this specific period. Data were obtained through direct observation and discourse analysis of the narratives of the debaters in live videos republished on YouTube. The main focus of the study was open debates that consistently triggered disputes, both in the form of online discussions and impacts that led to reports to the authorities.

Data collection began with the identification and selection of TikTok live videos that were relevant to interfaith debates. Only debates that triggered significant conflict were considered, especially if they led to increased tensions between religious groups or reports to the police. The criteria for selecting videos were based on the number of simultaneous viewers between 1,000 and 20,000, not on the total number of views. The live videos analyzed specifically were videos that were originally broadcast on TikTok and then re-uploaded to YouTube between January 2023 and December 2024. Live videos that met these criteria were analyzed in depth to understand communication patterns, debate strategies, and impacts on audience interactions. In addition, 5 live broadcast recordings that were reuploaded to YouTube by the admin or debate participants were used as additional references in the analysis.

The next process was the collection and analysis of comments and debater narratives in the selected videos. Not only comments in the live chat, but also direct quotes from the debaters were examined to identify argumentation patterns that had the potential to trigger conflict. Transcription was performed on key statements in the debate, including the use

of provocative rhetoric, comparisons of religious teachings, and audience responses.

To improve the accuracy of data selection, comments analytically were filtered based on Indonesian language and relevant hashtags that are often used in religious discussions. Sentiment analysis was performed on comments and audience responses using a semi-automatic approach with the help of Python-based NLTK software. Manual validation was performed on comments containing nuances of regional languages or special terms to ensure the analysis runs with higher accuracy.

A discourse analysis approach was applied to understand how debaters construct their arguments and the extent to which their communication styles contribute to the escalation of conflict. Interaction patterns between speakers and audiences were evaluated to see if there were patterns of provocation, emotional pressure, or narratives that reinforced polarization between religious communities. Data were analyzed using an inductive strategy, where key themes were determined based on communication patterns and the characteristics of the debate that developed.

In comparing Muslim and non-Muslim communities, an inferential approach was used to identify interaction patterns based on the comments they left. Identification of user religious affiliation was carried out through usernames, use of certain hashtags, and the content of comments that reflected religious identity. This approach allows for analysis of how polarization forms in digital discussions and the extent to which religious group differences influence their engagement patterns on TikTok.

In conducting this research, ethical aspects were taken as the main consideration. User anonymity was

maintained by disguising personal identities in the analysis and reporting process. Given that the data used came from public content available on social media, individual consent was not required in accordance with the ethical guidelines for research based on public data. However, data interpretation was carried out by considering the sensitivity of the issue so that the analysis remains objective and does not trigger negative impacts on the communities involved.

Although this research was conducted with a systematic approach, there are several limitations that need to be considered. Sentiment interpretation can be biased due to the use of informal language, slang, or regional dialects that are difficult to classify automatically. In addition, the dynamics of the TikTok algorithm that continues to change can affect interaction patterns and conflicts that occur, so the results of this study have limitations in terms of generalization to long-term trends. Identification of user religious affiliation also has limited accuracy, considering that determining groups based on usernames and hashtags is not always as accurate as direct demographic data.

Taking all these aspects into account, this study aims to provide a deeper understanding of inter-religious conflict in the TikTok digital space, as well as how the

debaters' narratives contribute to shaping the interactions that occur within it.

RESULT AND DISCUSSION

This study analyzed 5 live video recordings related to inter-religious conflict on TikTok, which were re-uploaded to YouTube and triggered significant conflict that resulted in a police report. These live videos were uploaded between January 2023 and December 2024, with selection criteria based on the number of simultaneous viewers between 1000 and 20,000. The comments analyzed were Indonesian language comments that were relevant to religious debates and conflicts. Identification was carried out through keyword filtering and the use of specific religious hashtags that often appeared in related discussions. From the 853,518 comments collected, an analysis was carried out to understand user interaction patterns, sentiment distribution, and the impact of debaters' narratives on conflict escalation.

Characteristics of Live Video Collection Data and User Engagement

Table 1 presents a summary of the data collected as well as the average user engagement based on the 5 videos analyzed.

Table 1. Summary of live video collection data and average user engagement

No	Category Content	Number of Videos	Number of Comments	Number of simultaneous viewers	Average likes	Average share
1	Trinity vs Tawhid	2	65.423	310.010	6.300	6.000
2	Use of the word "Allah"	2	769.232	7.713	100.600	1.000
3	Authenticity of the holy book	1	18.863	1.745	1.458	190
Total Categories		5	853.518	319.468	108.358	7.190

The analysis results show that content that is provocative or questions religious concepts tends to get higher engagement than content that is educational or neutral. In general, the number of comments increases significantly when the narrative in the debate starts to enter a more confrontational topic.

Distribution of Comment Sentiment

Sentiment analysis of 853,518 comments shows the distribution of user responses in three main categories: positive, neutral, and negative. Negative sentiment is most dominant in discussions related to religious terminology and the authenticity of holy books, while neutral comments are more common in discussions that are argumentative without direct provocation.

Table 2. Distribution of comment sentiment on live videos of inter-religious conflict on TikTok

Comment sentiment	Number of comments	Percentage (%)	Comment characteristics
Negative	582.463	68.2	Contains sharp criticism, condemnation of the debate opponent, or expressions of emotion such as anger and disagreement.
Neutral	210.679	24.7	Contains arguments that are informative or non-partisan, often in the form of academic questions and answers.
Positive	60.376	7.1	Supporting tolerance, appreciating other points of view, or encouraging non-

			confrontational discussion.
Total	853.518	100	-

The analysis results show that negative sentiment has the largest proportion (68%), especially in videos that show debates with confrontational narratives from debaters. In contrast, positive comments appear more often in discussion sessions that have a more diplomatic delivery style, where speakers try to build dialogue rather than simply sharpen doctrinal differences.

Thematic findings from comment analysis

Of the 853,518 comments, three main themes were identified based on content patterns and user interactions in TikTok live videos related to religious conflict.

Table 3. Thematic distribution of comments and characteristics of interaction patterns

No	Identified themes	Number of comments	Percentage of comments (%)	Characteristics of the observed patterns
1	Strengthening of belief	319.230	37,4	Active users reinforce religious doctrine, promote dogma, or respond to criticism of their beliefs.
2	Provocation by debater narrative	276.626	32,4	The narratives from the debaters directly triggered an increase in conflict-based comments in the live chat.
3	Community fragmentation	228.020	26,7	Users interact more with groups of the same religious affiliation, strengthening social

				boundaries between groups.
4	Law enforce- ment actions	29.642	3.5	Certain live videos cause escalation of conflict to the point of being reported to the authorities.
Total Themes		815,518	100	

From the data, it can be seen that the narratives of the debaters in the live video directly contribute to the intensity of audience interaction, especially when they present sharper arguments or corner certain perspectives.

Interactions in the live chat show an interesting pattern: when the debate takes place in a more diplomatic tone, comments remain argumentative but do not lead to direct confrontation. Conversely, when one party uses more provocative rhetoric, the number of negative comments increases significantly, indicating that users are more reactive to the style of the debate than to the content of the argument itself.

In the case of the video that resulted in a police report, comments in the live chat not only show support or criticism of the debaters, but also reflect the social tensions that are developing outside the digital platform. Many users discuss the legal implications or social consequences of the debate, indicating that the impact of this live video extends to a larger social space.

The role of the TikTok algorithm in strengthening ideological polarization

This study clearly shows that algorithm TikTok recommendations play a basic role and are active in strengthening ideological polarization between religious groups in Indonesia. Analysis in-depth video content and commentary identify domination themes key such as "strengthening "faith" and "strengthen

identity" (as summarized in Table 2 in section observation result This in a way consistent identify that algorithm in a way intelligent serve content that is precision in harmony with view beginning and preferences that have been shown users, regardless from whether they support majority religion doctrine or minority.

For example, videos that are explicit emphasize draft Trinity, Christian symbols, or practice religious certain tend more often appear on the timeline Christian users who have interaction with similar content, strengthen Affiliation identity there. On the contrary, video is empowering draft monotheism, Islamic values, or Muslim rituals will be watched by Muslim users.

This pattern in a way systematic reinforced by metrics engagement is prioritized by TikTok's algorithm, where content that triggers response emotional and more relationships strong, even If content the nature polarized or break split, often get improvement long visibility and reach bigger. This creates an environment where identity religion is not only confirmed but also strengthened through consumption of homogeneous content.

Mechanism recommends advanced personalization This in a way effectively creates room echo chambers and filter bubbles are increasingly dense and homogeneous on the TikTok platform. The "For You Page" (FYP) algorithm, which is at the heart of experience TikTok users, designed with great precision for predict preference content based on history interaction user previously, including likes, comments, shares, and times watch. As a result, users in a way gradually only will be exposed to information and perspectives that confirm belief ideological those who have there is, while view alternative, different nuance, or even the different one's opinion will in a way be systematically filtered and not visible. Environment isolated information not only limits

diversity thinking but also reduces the chance for individuals to interact with challenging ideas, so that hinder formation understanding cross-group and encourage isolation cognitive.

This strengthening process polarization is exacerbated by inherent confirmation bias in psychology human beings, where individuals in a way selective searching, interpreting, and remembering supporting information believe those who exist (Sunstein 1999). TikTok's algorithm, with its accuracy in serving this, not only responding to inherent bias but also in an active strengthening of it, creates a circle bait come back positive that drives groups to position increasingly ideological extremes. Findings this consistent with study previously about segregation ideologically on social media, as done by Cinelli et al. 2021 who highlighted formation groups ideologically homogeneous on other platforms through interaction based on interest.

However, research this gives a contribution unique to show how homophily (tendency) for interacting with individuals who have similarities) and formation room echo operate in a way specific in context conflict interfaith on TikTok, a domain that is often loaded with strong emotions and identity.

Unlike web -based platforms, text or images, dynamic short video formats, rapid virality, and provocative visual potential from TikTok significantly accelerate the polarization process. This content "strengthens identity visual group," as identified in analysis thematically, utilizing strong visual elements like religious symbols, group rituals, or representation identity certain. Visualization not only strengthens bonds and solidarity in a group but also, in a way implicit or even explicit, involves rejection or *othering* to identity groups outside.

This is reflected in pattern interactions in which users in a way actively tend to consume, interact with, and share fully content in line with their religious affiliation, which in turn in a way progressive deepen segregation ideologically and drastically reduces opportunity for constructive dialogue and understanding together between different religious groups. Implications from this is the creation of more digital boundaries sharp between communities, inhibiting cohesion socially.

This analysis firmly shows that the TikTok algorithm is not just a neutral tool or passive in the digital ecosystem. On the contrary, he functions as agent active and strong driving force in to form a digital environment where polarization ideological religion can not only grow but also growing fertile. The impact beyond preference content merely; he in a way directly influenced formation identity collective, strengthening in-group/out-group boundaries, and changing perception about the "other" in pluralistic society. These highlights need an urge to understand how technology, in particular algorithms, can in a way not on purpose to worsen split social, especially in issues sensitive like religion, demanding more ethical and regulatory consideration.

Emotional manifestations and impacts in online religious conflicts

Sentiment analysis of 582.463 TikTok user comments clearly shows that 68,2% of the comments have negative sentiments (as presented in Figure 2 in the results section). The high proportion of negative sentiments is significantly concentrated in discussions involving fundamental medical debates, such as the comparison of the Trinity versus monotheism, the use of sensitive religious terminology such as the use of the word "Allah" by non-Muslims, and the authenticity of each religion's holy book.

This emotional polarization is not simply an expression of intellectual disagreement; rather, it is a direct manifestation of a perceived threat to group identity, a defensive response to what is perceived as an aggression against fundamental beliefs.

The mechanism behind this emotional intensity is complex. Conflicting content that is often confrontational or demeaning to the opponent's beliefs directly triggers a strong emotional reaction from users. The use of specific provocative words, phrases, and visuals are key triggers. For example, comments that explicitly use derogatory terms such as “heretic,” “infidel,” or “idolatrous” that are often found in our data serve as rhetorical “bullets” that injure identity and provoke anger. More than just emotional terminals, these words are tools for classifying tension points. Similarly, videos that feature aggressive gesture caricatures or narratives that denigrate other religious doctrines can amplify this effect, triggering emotional responses of anger, frustration, anxiety, and even verbal aggression. These emotions then trigger a conflict spiral in which users respond with greater or greater emotional intensity.

This phenomenon can be comprehensively explained through the lens of social identity theory Tajfel and Turner 2004. When individuals feel that their social identity in this context, their religious identity is being attacked, denigrated, or questioned by an out-group, they tend to adopt responses that maintain the integrity of their group. This manifests itself in two main forms: stronger in-group favoritism, where users aggressively defend and glorify their own beliefs, and increased out-group derogation, where they verbally attack, discredit, or denigrate the beliefs of other religious groups. For example, videos that explicitly “challenge the authenticity of the Bible” or “question the concept of Tawhid” (as seen in the interaction pattern

in table 2) are flooded with comments showing deep anger from believers who feel their faith is being tarnished, followed by fierce counter-responses from the other side. This creates an emotionally reinforced “us vs. them” dichotomy.

Furthermore, the nature of social media platforms such as TikTok contributes to the amplification of negative emotions. Kowert's 2020 research highlights how emotions, especially negative ones such as anger and hatred, have a high viral power in the digital environment. Users tend to be quicker to respond, interact, and share content that triggers these strong emotions, thus expanding the reach of polarizing content and amplifying negative sentiments within their networks. The presence of anonymity or angle also facilitates effective inhibition where users feel freer to express verbal aggression or extreme views that they might not express in physical spaces. As a result, religious discussions that should be a place for dialogue and understanding turn into an arena for emotional confrontation dominated by provocative rhetoric, where the goal is to understand or dialogue, replaced by the urge to “win the debate” or “subdue the opponent”. This emotional intensity not only hinders dialogue, but also has the potential to translate online tensions into real-world social conflict.

Fragmentation of Digital Communities and Lack of Interfaith Dialogue

The research findings explicitly address the formation of “isolated digital communities on TikTok, as identified in the theme of “community fragmentation” (see Table 3 in the results section). User engagement patterns show a strong tendency to interact only within their own religious ideological group, with little or no substantial dialogue with individuals or groups who hold different views. This is not

simply a personal preference but a manifestation of online homophily, where individuals naturally seek out and interact with others who share similar beliefs, values, and backgrounds. On TikTok, the algorithm reinforces this tendency, effectively limiting users' exposure to perspectives outside their ideological "circle."

The implications of this pattern of interaction are serious for the potential for constructive dialogue and interfaith peace. The lack of cross-group interaction fundamentally erodes opportunities for mutual understanding, empathy, and common ground. When discussions occur only within Echo Chambers, group views become increasingly homogenous and extreme (Sunstein 1999). Group polarization theory suggests that in environments where like-minded individuals interact exclusively, their initial views tend to become more radical and extreme. In the context of religion on TikTok this means that beliefs in a group are reinforced without any challenge or nuance from the perspective of the out-group, which is often reduced to creativity or objects of derogation.

This phenomenon is further exacerbated by the dynamics of social identity theory Tajfel & Turner, which explains that individuals are highly motivated to maintain a positive image of their own group. When interactions are limited to the in-group, group identity becomes increasingly consolidated and distinct from the out-group. This not only encourages strong in-group favoritism, but can also trigger discrimination and contempt for the out-group. The absence of cross-border dialogue means that there is no "bridge" that can challenge negative perceptions or prejudices against other groups. Users feel safer and more validated in their homogeneous community, reducing the motivation to engage in conversations

that may be uncomfortable challenging their beliefs.

Unfortunately, from the data analyzed, evidence of positive dialogue or constructive interfaith interactions is very minimal or even non-existent in the videos or comments that are the focus of this study. The discussions observed were dominated by confrontational debates, assertions of group identity, and verbal attacks on opponents, rather than attempts to seek common understanding or build bridges between beliefs. This confirms that in the context of religious conflict, the algorithm Tik Tok and its user preferences have created an environment that effectively hinders the development of healthy dialogue, encouraging segregation rather than integration. Thus, facilitating cross-cultural and religious understanding has become a platform that accelerates social fragmentation and deepens ideological divides.

The role of external intervention (law enforcement) in managing online conflicts

This study identified five live video recordings on TikTok that sparked significant religious conflict to the point of police reports, as mentioned in the results section of the data collection. The presence of law enforcement intervention in these cases highlights the threshold at which online conflict moves from a disagreement of opinion to a violation of the law that requires an external response. The role of law enforcement in this context is to respond to content that is explicitly against the law, such as blasphemy or hate speech driven by ideological polarization. Our data shows that these interventions often occur when the conflict has reached a critical point, identifying that TikTok, despite having internal moderation mechanisms, sometimes requires third-party intervention to control escalation.

The effectiveness of law enforcement interventions requires more in-depth critical evaluation. Although the law will not be able to successfully remove content that violates or prosecute certain individuals, its impact on the dynamics of polarization and conflict in the digital space is still questionable. In some cases, this intervention can provide a temporary deterrent effect, but does not fundamentally address the root cause of ideological polarization or negative sentiment. In fact, legal actions that are considered biased or unfair can trigger a boomerang effect, where users who feel injustice switch to other platforms or create new accounts to continue disseminating their views.

Observations in the context of content moderation in general show that efforts to remove content often lead to the "whack-a-mole" phenomenon where content reappears in a different form or in another place as in controversial cases on social media. This indicates that although paid content can be prosecuted, the escalation of tensions between groups is not necessarily submerged; instead, it can migrate or adapt in other areas.

Integration of literature on content moderation and social media regulation, such as research by Gorwa, Binns, and Katzenbach 2020, underlines the importance of collaboration between social media platforms and law enforcement. Korowa and colleagues further argue that creating a safer and more inclusive environment requires joint efforts to identify and enforce unlawful content. The examples of Chanel Zuma and Uniriva (as mentioned in the literature) show that such coordination can help to effectively manage and defuse inter-religious conflict.

However, the effectiveness of this collaboration is highly dependent on transparency and consistency in its implementation. When users understand how the moderation and enforcement

processes are carried out, trust can be built, which in turn can encourage better compliance. Without transparency, interventions can be perceived as acts of censorship or suppression, which can actually increase polarization and create distrust of digital authorities.

Thus, although law enforcement interventions serve as a last resort for content that violates the law, their impact on reducing polarization in the long term is still complex. These interventions need to be accompanied by comprehensive strategies that address the problem of polarization, including increasing digital literacy, more responsible algorithm design, and promoting constructive dialogue so as not only to treat the symptoms but also to prevent the causes.

CONCLUSION

This study comprehensively underscores the crucial role of TikTok's recommendation algorithm in exacerbating inter-religious conflict and ideological polarization in Indonesia. Our findings consistently demonstrate that the algorithm actively facilitates the formation of Echo Chambers and Filter Bubbles, where users from both Christian and Muslim communities are exposed to content that exclusively confirms their ideological views. This mechanism actively serves and reinforces user confirmation bias, driving significant fragmentation within digital communities. This phenomenon empirically validates the relevance of group polarization theory in the contemporary digital context, where homogeneous interactions drive more extreme and less tolerant religious views.

The study's primary theoretical contribution lies in demonstrating how TikTok's highly personalized, engagement-driven short-form video format can generate ideological amplification and

emotional polarization at a speed and intensity not fully explored in previous literature. This study also deepens the application of social identity theory to the context of religious conflict in Indonesia, empirically demonstrating how threats to in-group religious identity, triggered by provocative online content are exacerbated by algorithms. This fuels extreme in-group favoritism and aggressive out-group denigrating, which in turn deepens the fragmentation of digital communities by exposing socio-psychological mechanisms accelerated by TikTok's unique characteristics. These algorithmic dynamics are exacerbated by the emotional content of religious conflict, where the high proportion of negative sentiment and provocative rhetoric in comments reflects reactions to threats to group identity.

With the rapid spread of viral content and TikTok's interactive features, the platform becomes an arena where negative emotions can spread and rapidly escalate conflict, potentially translating online tensions into real-world implications requiring law enforcement intervention. However, the effectiveness of law enforcement intervention requires a more in-depth critical evaluation; while crucial for explicit legal violations, reactive measures often fail to address the root causes of polarization. Limitations such as account reappearance or content migration underscore the complexity of the problem and the need for more productive and comprehensive strategies.

The findings of this study call for concrete responses from various stakeholders to create a healthier digital space. This includes reform and transparency of TikTok's algorithm to reduce the amplification of polarizing content, with the potential development of algorithms that promote exposure to diverse views. Context-based digital literacy programs are also needed, focusing on

developing critical thinking regarding provocative religious content, recognizing cognitive biases, and training in intercultural/religious empathy. Nuanced and collaborative content moderation is needed, capable of identifying "grey area" content that triggers polarization, involving religious and cultural experts. Finally, a multi-stakeholder regulatory framework should be developed by the government and regulators, encouraging platform accountability, innovation, and protecting user rights through ongoing dialogue.

While this study provides significant insights, limitations include the limited data on TikTok (thus lacking generalizability to other platforms), the focus on viral content (not capturing all user interactions), the absence of detailed demographic data, and the limitations of quantitative sentiment analysis in capturing nuances. The limited observation period also limits tracking long-term patterns of polarization. Based on these limitations, future research could explore cross-platform comparative studies, intervention research to reduce polarization, in-depth qualitative studies of user motivations, investigations of the offline impacts of online polarization, analyses of user agency in navigating filter bubbles, and longitudinal analyses to monitor the evolution of polarization.

Ultimately, while the complex social and political realities underlying religious conflict in Indonesia cannot be reduced solely to algorithmic dynamics, this study confirms that digital platforms bear a significant responsibility in shaping public discourse. Addressing religious polarization in the digital realm requires a collective and sustained effort from all stakeholders to foster healthier and more inclusive interactions.

REFERENCES

- Amit, Adi, and Eitan Venzhik. 2024. "Morality, Social Inclusion, and In-Group Superiority: The Differential Role of Individualizing and Binding Foundations in Perceptions of the Social Identity of In-Group and Out-Group Members." *Personality and Social Psychology Bulletin*. doi: 10.1177/01461672241262367.
- Arun, Chinmayi. 2019. "On Whatsapp, Rumours, and Lynchings." *Economic and Political Weekly* 54(6):30–35. https://www.researchgate.net/publication/331768800_On_whatsapp_rumours_and_lynchings
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences of the United States of America* 115(37):9216–21. doi: 10.1073/pnas.1804840115.
- Begby, Endre. 2022. "From Belief Polarization to Echo Chambers: A Rationalizing Account." *Episteme* 1–21. doi: 10.1017/epi.2022.14.
- Bhandari, Aparajita, and Sara Bimo. 2022. "Why's Everyone on TikTok Now? The Algorithmized Self and the Future of Self-Making on Social Media." *Social Media and Society* 8(1):1–11. doi: 10.1177/20563051221086241.
- Boyd, Danah M., and Nicole B. Ellison. 2008. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication* 13(1):210–30. doi: 10.1111/j.1083-6101.2007.00393.x.
- Čábyová, Ľudmila, and Peter Krajčovič. 2024. "Perspective Chapter: The TikTok Phenomenon – Harnessing Opportunities, Assessing Risks and Marketing Insights." Pp. 1–18 in *Social Media and Modern Society - How Social Media Are Changing the Way We Interact with the World Around*. <https://www.intechopen.com/chapters/1193517>. doi: 10.5772/intechopen.1006038
- Cinelli, Matteo, Gianmarco de Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The Echo Chamber Effect on Social Media." *Proceedings of the National Academy of Sciences of the United States of America* 118(9). doi: 10.1073/pnas.2023301118.
- Coady, David. 2024. "Stop Talking about Echo Chambers and Filter Bubbles." *Educational Theory* 74(1):92–107. doi: 10.1111/edth.12620.
- Dancharoenpol, Phakkhaporn. 2022. "The Influence of Perceived Value, User Engagement, and Emotions on Usage Intention of Thai Tiktok Users." Chulalongkorn University Theses and Dissertations (Chula ETD). <https://digital.car.chula.ac.th/chulaetd/8139>.
- Deitelhoff, Nicole, and Cord Schmelzle. 2022. "Social Integration Through Conflict: Mechanisms and Challenges in Pluralist Democracies." *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie* 75:69–93. doi: 10.1007/s11577-023-00886-3.
- Fabriar, Silvia Riskha, and Kurnia Muhajarah. 2024. "Tren Dakwah Nawaning Di Era New Media." *Jurnal SMART (Studi Masyarakat, Religi, Dan Tradisi)* 10(1):112–26. doi: 10.18784/smart.v10i1.2224.
- Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo

- Chambers, and Online News Consumption.” *Public Opinion Quarterly* 80(Specialissue1):298–320. doi: 10.1093/poq/nfw006.
- GD, Stasberger. 2023. “Digital Media: Shaping Communication, Culture, and Society in the Digital Age.” *Global Media Journal* 21(64):1–3. doi: 10.36648/1550-7521.21.64.388.
- Gorwa, Robert, Reuben Binns, and Christian Katzenbach. 2020. “Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance.” *Big Data and Society* 7(1):1–15. doi: 10.1177/2053951719897945.
- Hew, Wai Weng. 2018. “The Art of Dakwah: Social Media, Visual Persuasion and the Islamist Propagation of Felix Siauw.” *Indonesia and the Malay World* 46(134):61–79. doi: 10.1080/13639811.2018.1416757.
- Ibrahim, Adamkolo Mohammed. 2020. “Media Narrative Creation of Inter-Religious Conflict and Its Repercussion on National Development in Nigeria: A Dialectic Discourse on the Media’s Role as an Instrument for Conflict Construction and Management.” *New Media and Mass Communication* 91:33–46. doi: 10.7176/nmmc/91-04.
- Jamil, Nurul Akma, Lee Khuan, Cheong Ai Theng, and Siti Mariam Muda. 2024. “Facebook in Qualitative Research: Answering ‘What’ and ‘How’ Questions.” *Jurnal Ilmiah Teknologi S* 5(3):100–105. doi: 10.62527/jitsi.5.3.24.
- Khan, A. 2023. “The Impact of Social Media on Political Polarization in the United States.” *Research Journal for Social Affairs* 1(1):1–7. <https://rjsaonline.com/journals/index.php/rjsa/article/view/5>.
- Koç, Birkan. 2023. “The Role of User Interactions in Social Media on Recommendation Algorithms: Evaluation of Tiktok’S Personalization.” *Doctoral Dissertation* (July). doi: 10.13140/RG.2.2.34692.71040.
- Kowert, Rachel. 2020. “Dark Participation in Games.” *Frontiers in Psychology* 11(March). doi: 10.3389/fpsyg.2020.598947.
- Kreiss, Daniel, and Shannon C. McGregor. 2024. “A Review and Provocation: On Polarization and Platforms.” *New Media and Society* 26(1):556–79. doi: 10.1177/14614448231161880.
- Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. “The Science of Fake News: Addressing Fake News Requires a Multidisciplinary Effort.” *Science* 359(6380):1094–96. doi: 10.1126/science.aao2998.
- Mahmoudi, Amin, Dariusz Jemielniak, and Leon Ciechanowski. 2024. “Echo Chambers in Online Social Networks: A Systematic Literature Review.” *IEEE Access* 12(January):9594–9620. doi: 10.1109/ACCESS.2024.3353054.
- Naslund, J. A., K. A. Aschbrenner, L. A. Marsch, and S. J. Bartels. 2016. “The Future of Mental Health Care: Peer-to-Peer Support and Social Media.” *Epidemiology and Psychiatric Sciences* 25:113–22. doi:10.1017/S2045796015001067
- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Hiding from You*. New

- York: The Penguin Press.
- Permana, Aditya, Audrey Aurellie 2025. "Digital Anarchy and Security Dilemmas: Unraveling the TikTok Controversy in the Landscape of Cyber Realism and Hegemony Struggles between the United States and China." *Social and Political Science Faculty* 13(2):p.361-385.doi: 10.24269/ars.v13i2.10373
- Rodilosso, Ermelinda. 2024. "Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization." *Philosophy and Technology* 37(2):1–21. doi: 10.1007/s13347-024-00758-4.
- Shu, K., A. Sliva, S. Wang, J. Tang, and H. Liu. 2017. "Fake News Detection on Social Media: A Data Mining Perspective." *ACM SIGKDD Explorations Newsletter* 16(1): 22 - 36 doi: 10.1145/3137597.3137600.
- Silvanie, Astried, Rino Subekti, Dwi Sidik Permana, and Ari Kurniawan. 2024. "Tinjauan Komprehensif tentang Dampak Algoritma Media Sosial: *Jurnal Ilmiah Multidisiplin* 2(8):189–95. <https://doi.org/10.5281/zenodo.13253688>
- Sunstein, Cass. 1999. "The Law of Group Polarization." *University of Chicago Law School Chicago Unbound* (91):80–101. doi: 10.1002/9780470690734.ch4.
- Tajfel, Henri, and John Turner. 2004. *Organizational Identity A Reader*. edited by M. J. Hatch and M. Schultz. New York: Oxford University Press.
- Tucker, Joshua A., Andrew Guess, Pablo Barberá, Christian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brenden Nyhan. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." *Hewlett Foundation* (March):1–95. <http://dx.doi.org/10.2139/ssrn.3144139>
- VanDoorn, George, and Antoinette Eklund. 2013. "Face to Facebook: Social Media and the Learning and Teaching Potential of Symmetrical, Synchronous Communication." *Journal of University Teaching and Learning Practice* 10(1). doi: 10.53761/1.10.1.6.
- Zulli, Diana, and David James Zulli. 2022. "Extending the Internet Meme: Conceptualizing Technological Mimesis and Imitation Publics on the TikTok Platform." *New Media and Society* 24(8):1872–90. doi: 10.1177/1461444820983603.