



Fake News Spread in Ambon's Digital Sphere: Societal Impacts and the Limits of Current Detection Approaches

Yustina Sopacua¹, Afdhal Afdhal², Rikar Sahetapi³, Mae San Agustin Lustre⁴

^{1,2}Pattimura University, Maluku, Indonesia

³State Christian Institute of Ambon, Maluku, Indonesia

⁴Central Bicol State University of Agriculture, Philippines

Abstract

The spread of fake news through online social networks poses global challenges, yet its dynamics in local media ecosystems — especially in peripheral regions like Ambon, Indonesia — remain understudied. This study explores the characteristics of fake news circulation in Ambon's digital environment, its societal impacts, and the limitations of current AI-based detection approaches. Using a mixed-method design, the research analyzes fake news content on local social media platforms, conducts stakeholder interviews, and assesses the applicability of existing detection technologies in a local context. Findings reveal that fake news in Ambon often exploits cultural, religious, and political sensitivities, leading to social fragmentation and declining trust in legitimate media. Moreover, mainstream detection systems are ineffective locally due to linguistic, contextual, and infrastructural barriers. The study highlights the need for context-sensitive, community-driven detection frameworks and contributes new perspectives on misinformation governance in localized digital ecosystems.

Keywords: Fake News, Online Social Networks, Local Media, Misinformation, Societal Impact

Paper type: Research Paper

***Corresponding author:** yustinasopacua5@gmail.com

Received: 31-05-2025; Received in Revised From 03-08-2024; Accepted: 06-10-2025; Available Online: 07-10-2025

Cite this document as: Sopacua, Yustina, Afdhal, Rikar Sahetapi, and Mae San Agustin Lustre. 2025. "Fake News Spread in Ambon's Digital Sphere: Societal Impacts and the Limits of Current Detection Approaches." *The Journal of Society and Media* 9(2):601–625. DOI: 10.26740/jsm.v9n2.p601-625.



INTRODUCTION

The spread of fake news on social media has emerged as a serious threat to social stability, community cohesion, and the quality of democracy, both globally and nationally. In Indonesia, this phenomenon has become increasingly concerning due to rising internet penetration and low levels of digital literacy among certain segments of the population. According to Rahman and Tang (2022), 67% of internet users in Indonesia reported having encountered false or misleading information via social media platforms. Meanwhile, Hui (2020) emphasized that Indonesians tend to trust information disseminated through private messaging groups such as WhatsApp, which often serve as the primary channels for the circulation of hoaxes. This issue becomes even more complex in peripheral regions such as Ambon, Maluku, which has a historical backdrop of social tensions and tightly knit community structures grounded in cultural and religious values.

Studies by Harmakaputra and Rahman (2021), Hijrah, Lestaluhu, and Jaali (2024), and Tasane and Manuputty (2024) have found that local WhatsApp groups and online discussion forums in Ambon frequently act as primary media for spreading hoaxes related to religious issues, local government policies, and public health concerns. This contributes to the risk of social fragmentation and declining trust in official institutions and credible media outlets. Unfortunately, research on fake news in Indonesia remains heavily concentrated on large urban contexts in Java and Sumatra, while the dynamics of information dissemination in Eastern Indonesia, such as Ambon, remain largely underexplored. This gap exacerbates the lack of understanding of how disinformation operates within communities characterized by distinct social, linguistic, and digital infrastructure conditions.

Globally, the phenomenon of fake news has been widely examined from sociological, political, and technological perspectives. Bastick (2021) and Chen, Chen, and Xia (2022) demonstrate how disinformation can influence public opinion and political behavior, while Beauvais (2022) and Sui, Hawkins, and Wang (2023) highlight that fake news tends to spread faster and more broadly than factual news due

to its emotional and sensational appeal. Other studies, such as those by Phippen, Bond, and Buck (2021) and Taxitari, Sitistas, and Gavriil (2025), associate the susceptibility to hoaxes with low levels of reflective thinking and a strong emotional trust in informal sources. Similar trends can be observed in several national contexts: in the United States, false information surrounding electoral campaigns has shaped partisan polarization (Allcott and Gentzkow 2017); in India, misinformation circulating through WhatsApp has been linked to communal tensions and even mob violence (Mukherjee 2020); while in the Philippines, the proliferation of fake news on social media has played a crucial role in constructing political narratives and influencing democratic processes (Delante 2024). These cases illustrate how the circulation of fake news often intersects with local political, cultural, and technological conditions, making it a global yet context-specific phenomenon.

In the technological realm, numerous approaches have been developed to detect fake news, ranging from Natural Language Processing (NLP) techniques to machine learning and deep learning methods. Li and Jia (2025) and Qasim et al. (2022), for instance, developed high-accuracy BERT-based text classification models for English-language fake news. Similarly, Aldoseri, Al-Khalifa, and Hamouda (2023), as well as Yuan et al. (2023) Emphasise the importance of integrating social context data to build more adaptive detection models. Nevertheless, most of these approaches have been tested primarily on data from the United States or other developed countries, without accounting for the linguistic and cultural challenges present in regions like Eastern Indonesia (Cury et al. 2021; Lavanya and Sasikala 2021; Vicari and Gaspari 2021).

Research in the Indonesian context largely remains focused on major urban centres. Khalyubi and Perdana (2021) and Santoso, Setyaningsih, and Supadiyanto (2024) explored hoaxes related to political issues and elections in Jakarta and Surabaya. Anjelia, Mulyana, and Suminar (2024), along with Mahy, Winarnita, and Herriman (2022), documented the role of influencers in spreading misinformation in Yogyakarta and Bandung. In contrast, studies addressing the dynamics of hoaxes in Eastern

Indonesia, such as Ambon, are still extremely limited. Yet, as noted by Hatala and Latuconsina (2023), Kilmanun and Telussa (2025), and Solissa and Lestaluhu (2024), the complex local context — including limited digital infrastructure, low media literacy, and the uniqueness of local languages — poses unique challenges in detecting and mitigating fake news effectively.

Moreover, several studies underscore that overly technocratic detection approaches often fail to capture the socio-cultural dimensions of hoax dissemination. Asrifan (2025) and Asrifan and Alaqad (2025) argue that AI-based detection models tend to be inaccurate in recognizing local idioms, cultural references, or specific contextual meanings. As an alternative, Gong, Shang, and Wang (2024), along with Pröllochs (2022), advocate for the integration of community-driven fact-checking systems as a complement to automated mechanisms. Such approaches are considered more inclusive and adaptable to local realities, particularly in areas with strong community structures like Ambon.

While prior studies on fake news have largely concentrated on national or global levels, little attention has been paid to how disinformation functions in localized contexts with strong cultural and religious dynamics. This study is novel in positioning Ambon as a unique case where fake news intersects directly with histories of intergroup tensions, making it different from generic social media misinformation cases elsewhere. By integrating socio-cultural analysis with evaluations of AI detection systems, this research not only fills the gap in context-sensitive studies but also introduces an alternative community-based detection model that adapts to local linguistic and cultural realities.

METHODS

This study employs a mixed-methods approach, combining digital content analysis and survey-based quantitative data with qualitative in-depth interviews and focus group discussions. The quantitative component maps the frequency, themes, and patterns of fake news circulation across digital platforms in Ambon, while the qualitative component captures the perspectives of community leaders, journalists, and

local authorities regarding the socio-cultural impacts of disinformation. The integration of these methods enables both a broad statistical overview of fake news dissemination and a nuanced understanding of how it is interpreted and negotiated within local cultural and religious contexts. The decision to adopt this approach is grounded in the recognition that the spread of hoaxes is a social phenomenon that requires more than just mapping data and numerical patterns; it also necessitates a deep understanding of socio-cultural contexts, public perceptions, and the non-linear communication processes that unfold on local social media platforms. A mixed-methods strategy enables the combination of objective statistical data with narrative insights from actors directly involved, thus offering a more holistic and contextualized understanding (Gilad 2021; Hendren et al. 2023).

Ambon was selected as the research site due to its distinctive social complexity: a small city with a history of communal conflict, rich ethnic and religious diversity, and increasing digital media penetration in recent years. Although geographically situated in Eastern Indonesia — a region often marginalized in digital media studies — Ambon exhibits highly active social media dynamics. Online communities frequently engage in intense discussions about local issues on platforms such as Facebook, WhatsApp, and TikTok. The city has also become both a target and conduit for fake news dissemination, particularly around political events and major social occurrences. Thus, Ambon presents a strategic focal point for exploring how technology, culture, and local social structures intersect in the dynamics of disinformation.

Qualitative data were collected through in-depth interviews with 15 informants who were purposively selected. These informants included local journalists, digital community moderators, religious leaders, representatives from civil society organizations, academics, and local government officials involved in communication and information management. They were selected for their direct involvement in or deep understanding of information dynamics in Ambon — whether as producers, disseminators, or regulators of public information. The interviews were conducted both

in-person and online over a two-month period. Questions centred on their experiences with fake news dissemination, methods of identifying and responding to disinformation, and their perspectives on the effectiveness of current detection technologies.

Quantitative data were obtained through content analysis of 300 social media posts that were either suspected or confirmed to be fake news. These data were sourced from local Facebook groups, public TikTok discussions, and WhatsApp message chains archived by local fact-checking communities. Each post was analyzed using natural language processing (NLP) tools to trace narrative patterns, emotional tones, and the interconnectedness of accounts involved in the dissemination process. This analysis aimed to identify dominant themes, distribution patterns, and key actors within local disinformation networks. Additionally, this stage included performance testing of several AI-based fake news detection models, such as BERT and RoBERTa, which were originally trained on global datasets. The models' effectiveness was assessed in detecting hoaxes in the Indonesian language, particularly those expressed in local dialects specific to Ambon.

To ensure data validity and enhance interpretive depth, multiple triangulation strategies were employed. First, source triangulation was conducted by comparing interview results across various actor backgrounds. Second, methodological triangulation was applied by linking quantitative content analysis findings with qualitative narratives from interviews. Third, theoretical triangulation was utilized by interpreting the results through the lenses of digital communication theory, media ecology theory, and cultural sociology approaches. These steps ensured that the findings were not only descriptive but also interpretative and theoretically grounded.

RESULT AND DISCUSSION

Characteristics of Fake News Dissemination in Ambon's Digital Sphere

In a localized digital context such as Ambon, the dissemination of fake news exhibits unique patterns and characteristics that cannot be fully explained through global disinformation frameworks. Based on content analysis of 300 posts and digital

messages collected over a three-month period, the majority of hoaxes were found to spread through two primary channels: local community WhatsApp groups and neighborhood-based Facebook pages (often organized around village or neighborhood associations such as *kampung* or RT/RW units). These platforms are favored due to their ease of information sharing, high levels of engagement, and the dominance of socially intimate and trust-based community networks. Messages within these groups are frequently crafted in the form of personal narratives, often prefaced by phrases like “info from an insider” or “please forward to our family,” which lend a sense of urgency and credibility.

The forms of content disseminated vary widely, but there is a strong tendency toward long-form text accompanied by images or screenshots of manipulated news articles. In particular, fake news in Ambon often incorporates local idioms and cultural symbols — for example, employing the term *basudara* (a term denoting brotherhood) in provocative contexts or distorting religious quotations to incite tension. In several instances, the narratives evoke historical memories of conflict, such as by linking contemporary events with communal violence from the early 2000s, thereby triggering collective anxiety and emotional responses. For example, hoaxes that associate disaster aid distribution with particular religious groups often inflame discriminatory sentiments, which, while difficult to verify, are easily accepted by communities with entrenched social memories.

Table 1. Characteristics of Fake News Dissemination in Ambon

Characteristic	Findings
Main dissemination platforms	WhatsApp (65%), Local Facebook (30%), Local TikTok (5%)
Content formats	Long-form text (40%), Image + text (35%), Short videos (15%), Screenshots (10%)

Dominant themes	Religion (38%), Local politics (25%), Social aid (18%), Ethnicity/Culture (14%), Public health (COVID-19/stunting) (5%)
Initial sources of dissemination	Community figures (25%), Group admins (20%), Anonymous users (55%)
Language/narratives used	Mixed Indonesian-Ambon dialect, local idioms, and religious symbolism

Source: Content analysis data, 2025

These data clearly indicate that the spread of hoaxes is deeply embedded within the local socio-cultural context, where trust in informal actors — such as religious leaders, customary figures, and even Facebook group administrators — is remarkably high. This aligns with findings by Abiri and Buchheim (2022), Lecheler and Laura Egelhofer (2022), and S  e (2021), who argue that disinformation is not merely about the factual accuracy of content, but more profoundly about who disseminates the information and in what social context it circulates. In Ambon, information credibility is not defined by media authority but by social relationships and affective affiliations with the message sender. Messages originating from “insiders” within the community tend to be trusted, even when their content has not been verified.

This relational nature of fake news dissemination highlights that the digital ecosystem in regions like Ambon is not a neutral public space, but rather a social field shaped by collective emotions, historical trauma, and informal trust structures. In such an ecosystem, purely technological approaches to detecting and combating hoaxes fall short, as they fail to grasp the social nuances underpinning the spread of information. Understanding local idioms, community histories, and informal social networks is essential to comprehending why certain information — however factually implausible — circulates and gains traction. These findings reinforce the importance of mapping

disinformation not merely as a digital artifact, but as a social process embedded in localized networks of meaning and trust.

Patterns of Sensitivity: The Exploitation of Cultural, Religious, and Political Issues

In the dynamics of fake news dissemination in Ambon, it becomes evident that the most viral content is not merely that which is sensational, but rather that which touches upon deeply sensitive points embedded in the collective identity structure of the community. Based on content analysis of over 80 social media posts and 27 widely circulated WhatsApp messages from local community groups over the past six months, the most extensively disseminated fake news narratives were those portraying particular communities — defined by religion, ethnicity, or political affiliation — as either “under threat” or “being oppressed.” Themes such as inequality in the distribution of government aid, allegations against religious symbols, and suspicions of dominance by specific groups in government positions were found to be dominant.

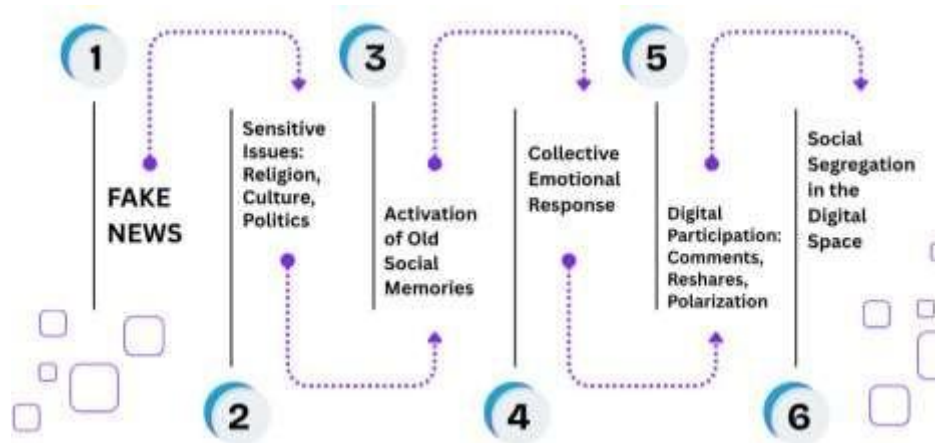
This pattern does not occur randomly. As one senior journalist from a local media outlet explained during an interview, “The news that spreads fastest usually touches on old fears or anger in the community. If an issue relates to past conflict, it immediately goes viral — especially if it’s spread by someone perceived as credible within the community.” This indicates that the spread of fake news follows historical lines of social division deeply rooted in the city’s past. In another interview, a local youth leader remarked, “Sometimes the information is clearly illogical, but when it comes from a religious figure or an elder we respect in the group, people believe it instantly without questioning it.”

These observations support Papacharissi (2020) framework of affective publics, which asserts that digital participation is often mobilized not by rational argumentation, but by collective affect or emotion grounded in shared experiences. Emotions such as anger, fear, and pride in group identity are the primary fuel behind the circulation of

fake news. In Ambon — a city still marked by the collective memory of communal conflict from 1999 to 2002 — social trauma remains embedded in public consciousness. Digital narratives that reactivate such memories possess the capacity to evoke powerful emotional responses. One example involved a fake news post alleging discrimination in disaster aid distribution to Muslim residents, accompanied by an image from the conflict era. The post spread to more than 40 WhatsApp groups in a single day and generated dozens of angry comments and accusations directed toward specific religious communities.

Here we see that fake news functions not merely as a conduit of false information, but as a catalyst for the reactivation of latent social memory. A local academic noted, “When a hoax touches on identity issues, it’s like reopening an old wound. People stop thinking about whether the information is true or false; it becomes about how it makes them feel.” These narratives generate an emotional ecosystem that accumulates and manifests through comments, reposts, or even calls for digital exclusivism against other groups. This pattern is illustrated in the following picture 1.

Picture 1



Sensitivity Pattern of Fake News in Ambon

Source: Content Analysis Data, 2025

The schema above highlights how fake news does not remain within the realm of discourse, but translates into digital practices that reinforce existing social segregation. In the context of Ambon, the digital sphere is not merely a space for the convergence of users and information — it is an affective field where memory, identity, and power relations interact with high intensity. This pattern of sensitivity is critical, not only for mapping the spread of fake news, but also for understanding its socio-cultural impact, especially on the collective memory and conflict history of the Maluku region. Thus, fake news dissemination should not be seen solely as a product of misinformation. Rather, it is a process that engages the emotional roots of society, reshapes conflict narratives, and reinforces social boundaries in subtler yet tangible forms within the digital space.

Social Fragmentation and the Decline of Trust in Mainstream Media

Amid the intensified circulation of fake news in Ambon's digital sphere, one of the most palpable and alarming impacts is the deepening of social fragmentation and the declining public trust in official media outlets. This finding emerges not only from content analysis of disinformation prevalent in local Facebook and WhatsApp groups, but also from interviews with residents, journalists, and community activists. Viral messages often contain narratives accusing mainstream media of being “government puppets” or “political operatives,” and claim that the real truth circulates through informal, underground channels.

As one informant, a senior local journalist (MRA), noted: “Many residents now trust broadcasts in their family WhatsApp groups more than the news we publish — especially if our reports don't criticize certain parties.” This statement is corroborated by observations in several Ambon-based public Facebook groups monitored over a three-month period. In these forums, posts discrediting local media tended to receive more engagement — comments, shares, and reactions — than clarification posts from the media outlets themselves. For instance, a post in the group *Ambon Satu Hati* accusing a local media organization of “covering up inequality in post-earthquake aid

distribution” garnered over 700 interactions within just two days — significantly outperforming the media outlet’s response post.

Table 2.

Content Analysis of Fake News Targeting Mainstream Media in Ambon

Narrative Category	Number of Cases	Dominant Platforms	Primary Emotions	Source/Identity of Sender
Media seen as pro-government	17	WhatsApp Groups, Facebook	Anger, disappointment	Local group admins, community figures
Media accused of hiding post-disaster aid info	11	Facebook, TikTok	Suspicion, sadness	Anonymous/public accounts
Media accused of religious bias	7	WhatsApp, Facebook	Anger, fear	Mass broadcasts with unclear origin
Local media compared to “outside media”	6	Facebook, TikTok	Inferiority, frustration	Local influencers
Media seen as Jakarta elites' tool	4	WhatsApp, Facebook	Collective anxiety	Forwarded chain messages

Source: Content Analysis Data, 2025

The data above show that the most widespread narrative accuses local media of being aligned with government interests, particularly in contexts such as disaster coverage or aid distribution. This fuels emotional responses — mainly anger and disappointment — that rapidly spread through community-based groups. There is also a strong tendency to place greater trust in information from informal figures, such as group admins, religious leaders, or even anonymous accounts that deliver emotionally charged and confrontational messages.

Field observations conducted in several post-earthquake neighborhoods reveal that residents frequently referred to the “latest broadcast” in their WhatsApp groups when asked about policy updates or aid distribution. In one instance in the Karang Panjang area, official information delivered by village officers was rejected outright by residents who claimed it “didn’t match the info from our RT group.” This situation reflects an escalating distrust toward formal sources of authority and a growing dependence on peer-to-peer digital networks for news and guidance.

This phenomenon aligns with the arguments put forward by Fawzi et al. (2021), and Mangold, Bachl, and Prochazka (2024), who emphasize that contemporary trust in media is increasingly shaped not by journalistic standards, but by emotional affinity and social networks. In the context of Ambon — where historical trauma, information inequality, and limited digital literacy serve as structural backdrops — official media sources are often perceived as “distant” or “unsympathetic to the common people.” Fake news thus not only disrupts public communication flows but also produces and reproduces fragmented social spaces, in which communities cluster based on whom they trust and what kind of information they choose to consume. This presents a serious challenge to the development of a healthy information ecosystem, particularly in peripheral regions such as Ambon. It underscores the urgency of adopting community-based approaches to rebuild the legitimacy of mainstream media, and to foster media literacy that is culturally sensitive and emotionally resonant.

Limitations of AI-Based Detection: Contextual and Linguistic Inequality

The limitations of artificial intelligence (AI)-based detection technologies in addressing the spread of fake news in Ambon have emerged as a significant finding of this study. An analysis of locally circulating hoaxes, using transformer-based models such as BERT and RoBERTa, revealed notably low detection accuracy. These models consistently struggled to identify hoaxes composed in mixed linguistic registers — a blend of standard Indonesian, Ambonese Malay, and other local dialects rich in implicit meanings. In several instances, sarcastic or ironic expressions embedded in local vernacular were mistakenly classified as “neutral,” while factual messages employing harsh but contextually appropriate terms were misidentified as “fake news.”

For example, in a widely circulated WhatsApp chain message stating, “Katong pung bantuan banya diatas kertas, seng pernah sampe di Batu Meja” (loosely translated: “Our aid is only on paper, never reached Batu Meja”), the RoBERTa model categorized the message as neutral and informational. Yet, in local context, this phrase carries a serious accusation of aid mismanagement. The failure here is not merely linguistic — it is a failure to comprehend the cosmology of meaning, that is, how local communities interpret and construct meaning based on their lived social realities.

This finding strongly supports the concept of Contextual AI Failure as proposed by Jones-Jang and Park (2022), which highlights the inability of AI systems to function effectively when applied outside the socio-cultural frameworks in which they were trained. While technically robust, models like BERT and RoBERTa were developed and tested using corpora vastly different from the communicative ecology of peripheral societies such as Ambon. As a result, these detection systems are context-blind and consistently fail to recognize disinformation when it is embedded in local symbols, cultural irony, or socially charged humor, all of which often carry political meanings at the community level. To reinforce this argument, the table below summarizes the accuracy testing results of AI detection models on local hoax content:

Table 3.

Accuracy of AI Models in Detecting Local-Language Hoaxes in Ambon

Detection Model	Language Type of Content	Sample Size	Detection Accuracy	Dominant Error Type
BERT (pre-trained)	Formal Indonesian	30	83%	Minimal
BERT (pre-trained)	Mixed Indo–Ambonese	40	42%	Misclassification of sarcasm & local idioms
RoBERTa (Indo)	Formal Indonesian	30	85%	Minimal
RoBERTa (Indo)	Mixed Indo–Ambonese + Local Dialects	40	37%	Failure to detect culturally embedded emotional narratives

Source: Internal testing on 140 local hoax samples, Ambon, 2024

The underperformance of these models is further corroborated by interviews with local fact-checking community members. One informant noted, “The detection systems work fine for Jakarta news, but once the word *katong* appears, the model gets confused.” This statement reveals a clear epistemic gap between the operational logic of AI models and the communicative practices of local communities. Additional observations show that hoaxes in Ambon are frequently adorned with cultural symbols unrecognized by AI systems. For example, images of specific geographic locations or references to historical events — particularly the communal conflict of the early 2000s — are often used to amplify the emotional impact of fake news. These elements, however, remain entirely outside the scope of text-based AI detection systems. This failure underscores the pressing need to recalibrate global AI models when applied in culturally distinct regions. The development of detection systems that are sensitive to

local linguistic nuances and socio-political history is not merely a technical necessity but also an ethical imperative. Only through such contextual sensitivity can AI technologies equitably serve the information needs of diverse communities and contribute meaningfully to democratic communication in peripheral regions.

Toward Context-Aware Detection: The Role of Community Verification and Socio-Cultural Intelligence

Efforts to develop context-sensitive fake news detection systems in Ambon cannot rely solely on algorithmic capabilities. As previously discussed, the inability of AI-based systems to recognize localized narratives opens up space for alternative approaches that are more adaptive to the social dynamics on the ground. A consistent theme emerging from interviews with informants is the critical importance of community engagement in the verification process. As YSN, a community media activist in Nusaniwe District, noted, “Sometimes it’s not that the system is unsophisticated, but that we’re not invited to help check. We on the ground know the context better.” This perspective highlights the gap between technological sophistication and sensitivity to local knowledge.

Field observations across various digital interaction hubs — such as Facebook forums for Batu Merah residents and youth church communities on TikTok — revealed that verification often occurs informally. Users regularly comment, question, or provide clarifications based on personal experience or community knowledge. Though unstructured, these processes function as informal “social filters” that help limit the spread of false content. However, a significant limitation is the absence of institutional support to formalize and sustain such grassroots verification within the digital ecosystem.

In this context, the concept of socio-technical hybridity, as proposed by Duckert (2023), becomes especially relevant. Rather than reinforcing the binary between human and machine, this framework emphasizes that the most effective information systems are those that integrate computational power with the social relationships and knowledge embedded within communities. In the Ambonese setting, this implies

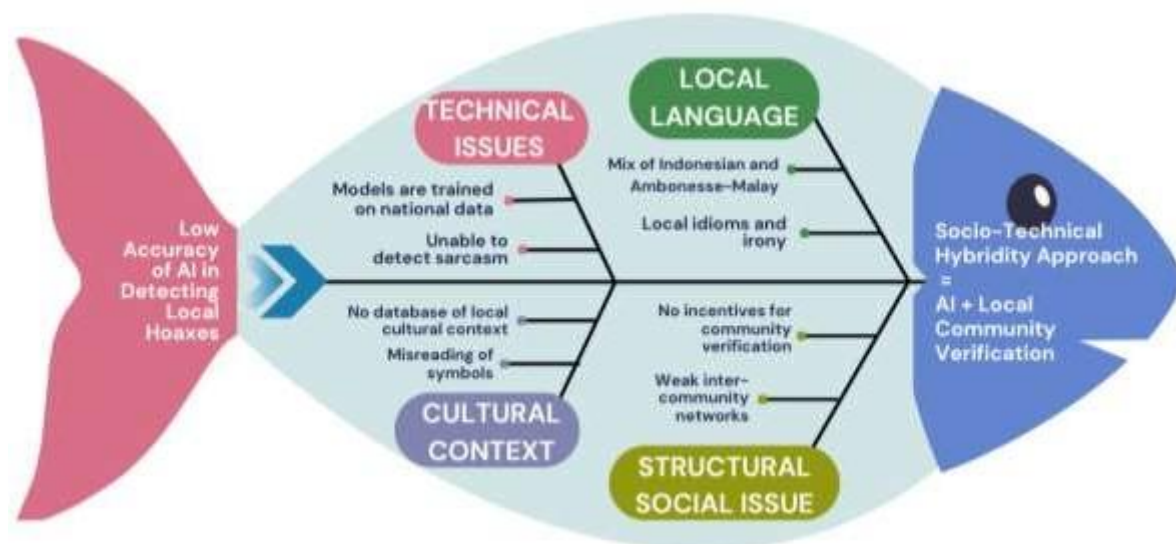
combining AI-based early warning systems with community figures — such as religious leaders, youth representatives, or digital group administrators — as on-the-ground verification agents endowed with social legitimacy.

Several informants proposed initiatives such as forming “local fact-checking teams” composed of residents from diverse religious and ethnic backgrounds, and organizing short trainings on using online-based misinformation detection tools. As Marlina, a local journalist, explained, “If only the machine works, it cannot feel whether this information hurts the feelings of certain groups or not.” This underscores that the affective dimension of Ambonese society cannot be disregarded in the information filtering process.

Participant observation in the Komunitas Literasi Digital Muda Ambon (Young Ambon Digital Literacy Community) further revealed emerging local initiatives. These include the use of simple Google Forms to report hoaxes and sharing clarifications via Instagram Stories. However, these practices are hampered by a lack of structural support and formal recognition. These grassroots strategies have yet to be optimally incubated within existing technological architectures.

To clarify the root causes and potential solutions for the failure of hoax detection systems in Ambon, the following fishbone diagram illustrates the problem structure and its contextually relevant remedies:

Picture 2



Fishbone Diagram – Root Causes of Hoax Detection Failures in Ambon and Contextual Solutions

Source: Research data, 2025

As shown above, the most effective solution is not to replace existing technologies but to redesign systems that are more responsive to local contexts. An ideal detection model for regions like Ambon should remain open to community input, foster trust across social groups, and integrate socio-cultural intelligence alongside artificial intelligence.

CONCLUSION

This study concludes that fake news dissemination in Ambon is deeply influenced by socio-cultural dynamics rather than random information flow. Collective emotions, memories of past conflicts, and reliance on informal trust networks drive the spread and acceptance of disinformation—especially when religion, culture, and political aid are framed through local idioms and symbols. While focused on Ambon, these dynamics reflect broader contexts where strong cultural and religious identities intersect with histories of tension, showing that disinformation thrives when rooted in

local meanings and collective memories. The study also reveals the limits of universal AI-based detection systems, which often overlook local linguistic and cultural nuances. To bridge this gap, it proposes hybrid detection frameworks that combine computational tools with community-based socio-cultural knowledge—offering a more adaptive, context-sensitive model for combating disinformation and enriching digital governance debates.

Funding Acknowledgement

There is no funding for this research and there is no conflict of interest. We thank our informants who provided insight and experience that greatly assisted the research. We would also like to show our gratitude to anonymous reviewers for their comments on an earlier version of the manuscript. However, the responsibility for this research is taken by the authors.

About the Author(s)

Yustina Sopacua, S.Sos., M.Si., is a lecturer in the Communication Studies Program, FISIP, Pattimura University, where she has taught since 2002. She holds a Bachelor's degree in Social Sciences from Pattimura University and a Master's degree in Science from Hasanuddin University. Her expertise includes Cultural Communication and Media Studies. Her research can be found on Google Scholar and SINTA under her name. Contact: yustinasopacua25@gmail.com.

Afdhal, M.Si., is a lecturer in Sociology at Universitas Pattimura and National Secretary of the Perhimpunan Pendidikan dan Guru (P2G). Born in Tanah Datar, West Sumatra, on July 17, 1994, he earned a Bachelor's in Sociology Education from Universitas Negeri Jakarta (2016) and a Master's in Sociology from Universitas Indonesia (2019). He actively engages in teaching, writing, and social research. Contact: afdhal@lecturer.unpatti.ac.id.

Rikar Sahetapy, S.Sos., M.A., is a lecturer in the Communication Studies Program, Faculty of Religious Social Sciences (FISK), IAKN Ambon, since 2023. He

holds a Bachelor's degree in Social Sciences from Pattimura University and a Master of Arts from Gadjah Mada University. His expertise includes Media and Cultural Studies. His research can be found on Google Scholar under his name. Contact: rikarsahetapy@gmail.com.

Mae San Agustin Lustre is a graduate student in the Department of Landscape and Environmental Management at Central Bicol State University of Agriculture, Philippines. Her research focuses on environmental and development issues, particularly the intersections of feminist and queer movements in environmental studies. Contact: mae.lustre91@gmail.com.

REFERENCES

- Abiri, Gilad, and Johannes Buchheim. 2022. "Beyond True and False: Fake News and the Digital Epistemic Divide." *Michigan Technology Law Review* (29.1):59. doi: 10.36645/mtlr.29.1.beyond.
- Aldoseri, Abdulaziz, Khalifa N. Al-Khalifa, and Abdel Magid Hamouda. 2023. "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges." *Applied Sciences* 13(12):7082. doi: 10.3390/app13127082.
- Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31(2):211–36. doi: 10.1257/jep.31.2.211.
- Anjelia, Lydia Okva, Deddy Mulyana, and Jenny Ratna Suminar. 2024. "The Role of Indonesian Doctors as Content Creators in Fighting Health Hoaxes, Myths, and Stigma on Social Media." *Jurnal Komunikasi Indonesia* 13(2). doi: 10.7454/jkmi.v13i2.1253.
- Asrifan, Andi. 2025. "Navigating Ethical Dilemmas in AI-Powered Translation." Pp. 327–56 in *Role of AI in Translation and Interpretation*.
- Asrifan, Andi, and Mohammed H. Alaqad. 2025. "Evolution of AI in Interpretation."

Pp. 63–94 in *Role of AI in Translation and Interpretation*.

Bastick, Zach. 2021. “Would You Notice If Fake News Changed Your Behavior? An Experiment on the Unconscious Effects of Disinformation.” *Computers in Human Behavior* 116(12):106633. doi: 10.1016/j.chb.2020.106633.

Beauvais, Catherine. 2022. “Fake News: Why Do We Believe It?” *Joint Bone Spine* 89(4):105371. doi: 10.1016/j.jbspin.2022.105371.

Chen, Long, Jianguo Chen, and Chunhe Xia. 2022. “Social Network Behavior and Public Opinion Manipulation.” *Journal of Information Security and Applications* 64(1):103060. doi: 10.1016/j.jisa.2021.103060.

Cury, Ricardo C., Istvan Megyeri, Tony Lindsey, Robson Macedo, Juan Batlle, Shwan Kim, Brian Baker, Robert Harris, and Reese H. Clark. 2021. “Natural Language Processing and Machine Learning for Detection of Respiratory Illness by Chest CT Imaging and Tracking of COVID-19 Pandemic in the United States.” *Radiology: Cardiothoracic Imaging* 3(1):e200596. doi: 10.1148/ryct.2021200596.

Delante, Nimrod L. 2024. “Trolling and Fake News Diminish Critical Analysis and Morality: Insights from Philippine Newspapers and Trolling Narratives.” Pp. 141–63 in *Handbook of Social Media Use Online Relationships, Security, Privacy, and Society, Volume 2*. Elsevier.

Duckert, Melanie. 2023. “The Futures of Hybrid Work: A Socio-Technical Framework for Designing Hybrid Environments.” Pp. 441–43 in *Computer Supported Cooperative Work and Social Computing*. New York, NY, USA: ACM.

Fawzi, Nayla, Steindl Nina, Obermaier Magdalena, Prochazka Fabian, Arlt Dorothee, Blöbaum Bernd, Dohle Marco, Engelke Katherine M., Hanitzsch Thomas, Jakob Nikolaus, Jakobs Ilka, Klawier Tilman, Post Senja, Reinemann Carsten, Schweiger Wolfgang, and Marc and Ziegele. 2021. “Concepts, Causes and

- Consequences of Trust in News Media – a Literature Review and Framework.” *Annals of the International Communication Association* 45(2):154–74. doi: 10.1080/23808985.2021.1960181.
- Gilad, Sharon. 2021. “Mixing Qualitative and Quantitative Methods in Pursuit of Richer Answers to Real-World Questions.” *Public Performance & Management Review* 44(5):1075–99. doi: 10.1080/15309576.2019.1694546.
- Gong, Y., L. Shang, and D. Wang. 2024. “Integrating Social Explanations Into Explainable Artificial Intelligence (XAI) for Combating Misinformation: Vision and Challenges.” *IEEE Transactions on Computational Social Systems* 11(5):6705–26. doi: 10.1109/TCSS.2024.3404236.
- Harmakaputra, Hans A., and Luthfi Rahman. 2021. “From Toleration to Solidarity: Muslim-Christian Relations in Indonesia during the COVID-19 Pandemic.” *The Muslim World* 111(4):616–38. doi: 10.1111/muwo.12416.
- Hatala, Rifaldi, and Nurainy Latuconsina. 2023. “Transformasi Digital Di Pedesaan: Evaluasi Implementasi Sistem Informasi Desa Di Kabupaten Seram Bagian Barat.” *Populis: Jurnal Ilmu Sosial Dan Ilmu Politik* 18(1):104–25. doi: 10.30598/populis.18.1.104-125.
- Hendren, Kathryn, Kathryn Newcomer, Sanjay K. Pandey, Margaret Smith, and Nicole Sumner. 2023. “How Qualitative Research Methods Can Be Leveraged to Strengthen Mixed Methods Research in Public Policy and Public Administration?” *Public Administration Review* 83(3):468–85. doi: 10.1111/puar.13528.
- Hijrah, Hijrah, Said Lestaluhu, and La Jaali. 2024. “Etnisitas, Demokrasi Lokal, Dan Komunikasi Politik: Praktik Mobilisasi Pada Pilkada Maluku Tengah.” *Populis: Jurnal Ilmu Sosial Dan Ilmu Politik* 18(2):206–20. doi: 10.30598/populis.18.2.206-220.
- Hui, Jennifer Yang. 2020. “Social Media and the 2019 Indonesian Elections: Hoax

- Takes the Centre Stage.” *Southeast Asian Affairs* SEAA20(1):155–74. doi: 10.1355/aa20-1i.
- Jones-Jang, S. Mo, and Yong Jin Park. 2022. “How Do People React to AI Failure? Automation Bias, Algorithmic Aversion, and Perceived Controllability” edited by M. Yao. *Journal of Computer-Mediated Communication* 28(1). doi: 10.1093/jcmc/zmac029.
- Khalyubi, Wildhan, and Aditya Perdana. 2021. “Electoral Manipulation Informationally on Hoax Production in 2019 Presidential and Vice Presidential Election in Indonesia.” *Journal of Government and Political Issues* 1Khalyubi,(2 SE-Articles):87–99. doi: 10.53341/jgpi.v1i2.17.
- Kilmanun, Febyyola Welmina, and Sandra Ivonnice Telussa. 2025. “Transformasi Informasi Digital Untuk Pelayanan Publik Dan Pembangunan Daerah Tertinggal: Studi Strategis Atas Peran Diskominfo Kabupaten Maluku Tenggara.” *Populis: Jurnal Ilmu Sosial Dan Ilmu Politik* 19(2):185–99. doi: 10.30598/populis.19.2.185-199.
- Lavanya, PM., and E. Sasikala. 2021. “Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey.” Pp. 603–9 in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*. IEEE.
- Lecheler, Sophie, and Jana Laura Egelhofer. 2022. “Disinformation, Misinformation, and Fake News.” Pp. 69–87 in *Knowledge Resistance in High-Choice Information Environments*. London: Routledge.
- Li, Xi, and Lili Jia. 2025. “English Text Topic Classification Using BERT-Based Model.” *Journal of Computational Methods in Sciences and Engineering* 25(1):669–84. doi: 10.1177/14727978251321982.
- Mahy, Petra, Monika Winarnita, and Nicholas Herriman. 2022. “Influencing the

- Influencers: Regulating the Morality of Online Conduct in Indonesia.” *Policy & Internet* 14(3):574–96. doi: 10.1002/poi3.321.
- Mangold, Frank, Marko Bachl, and Fabian Prochazka. 2024. “How News Audiences Allocate Trust in the Digital Age: A Figuration Perspective.” *Journalism & Mass Communication Quarterly* 101(4):955–80. doi: 10.1177/10776990221100515.
- Mukherjee, Rahul. 2020. “Mobile Witnessing on WhatsApp: Vigilante Virality and the Anatomy of Mob Lynching.” *South Asian Popular Culture* 18(1):79–101. doi: 10.1080/14746689.2020.1736810.
- Papacharissi, Zizi. 2020. “Affective Publics.” Pp. 61–75 in *The Oxford Handbook of Digital Media Sociology*. Oxford University Press.
- Phippen, Andy, Emma Bond, and Ellen Buck. 2021. “Effective Strategies for Information Literacy Education: Combatting ‘Fake News’ and Empowering Critical Thinking.” Pp. 39–53 in *Future Directions in Digital Information*, edited by D. Baker and L. B. T.-F. D. in D. I. Ellis. Elsevier.
- Pröllochs, Nicolas. 2022. “Community-Based Fact-Checking on Twitter’s Birdwatch Platform.” *Proceedings of the International AAAI Conference on Web and Social Media* 16(12):794–805. doi: 10.1609/icwsm.v16i1.19335.
- Qasim, Rukhma, Waqas Haider Bangyal, Mohammed A. Alqarni, and Abdulwahab Ali Almazroi. 2022. “A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification” edited by R. Taiar. *Journal of Healthcare Engineering* 2022:1–17. doi: 10.1155/2022/3498123.
- Rahman, Rofi Aulia, and Shu-Mei Tang. 2022. “Fake News and Internet Shutdowns in Indonesia: Symptoms of Failure to Uphold Democracy.” *Constitutional Review* 8(1):151–65. doi: 10.31078/consrev816.
- Santoso, Didik Haryadi, Rila Setyaningsih, and Supadiyanto Supadiyanto. 2024. “Combating The Disinformation: Verifying The Fact of Political Hoaxes in Election-2024 in Indonesia.” *Profetik: Jurnal Komunikasi* 17(2):369–84. doi:

10.14421/pjk.v17i2.3003.

Søe, Silje Obelitz. 2021. "A Unified Account of Information, Misinformation, and Disinformation." *Synthese* 198(6):5929–49. doi: 10.1007/s11229-019-02444-x.

Solissa, Friskila, and Said Lestaluhu. 2024. "Pendidikan Politik Digital Dan Tradisional: Dinamika Peran Pemuda Dalam Demokrasi Lokal Di Indonesia." *Populis: Jurnal Ilmu Sosial Dan Ilmu Politik* 19(1):16–30. doi: 10.30598/populis.19.1.16-30.

Sui, Mingxiao, Ian Hawkins, and Rui Wang. 2023. "When Falsehood Wins? Varied Effects of Sensational Elements on Users' Engagement with Real and Fake Posts." *Computers in Human Behavior* 142(4):107654. doi: 10.1016/j.chb.2023.107654.

Tasane, Rican, and Feky Manuputty. 2024. "Partisipasi Perempuan Dalam Politik: Antara Representasi Dan Kebijakan Di Kota Ambon." *Populis: Jurnal Ilmu Sosial Dan Ilmu Politik* 18(2):185–205. doi: 10.30598/populis.18.2.185-205.

Taxitari, Loukia, Thanos Sitistas, and Eleni Gavriil. 2025. "'Disinformation Aims to Mislead; Misinformation Thrives in Ignorance': Insights from Experts and Non-Experts in Greek-Speaking Cyprus." *Social Sciences* 14(3):133. doi: 10.3390/socsci14030133.

Vicari, Mattia, and Mauro Gaspari. 2021. "Analysis of News Sentiments Using Natural Language Processing and Deep Learning." *AI & SOCIETY* 36(3):931–37. doi: 10.1007/s00146-020-01111-x.

Yuan, Jingling, Nana Wang, Siqi Cai, Mincheng Chen, and Xinping Li. 2023. "AMLFN-AD: Adaptive Multi-Level Integrated Fusion Attack Detection Framework for Intelligent Building Systems." *Computer Networks* 227(1):109700. doi: 10.1016/j.comnet.2023.109700.