BETWEEN PROMISE AND PERIL: INTEGRATING ARTIFICIAL INTELLIGENCE INTO INDONESIA'S COUNTER-VIOLENT EXTREMISM STRATEGY

Raneeta Mutiara

e-mail: raneetamutiara001@suss.edu.sg, ORCID ID: https://orcid.org/0009-0000-2076-7944

Singapore University of Social Sciences (SUSS), Singapore

Received: March 22, 2026 | Revised: June 18, 2026 | Accepted: June 30, 2026

UDC 004.8:343.326

DOI: https://doi.org/10.33445/psssj.2026.7.2.2

Abstract

This qualitative research examines the ways in which Artificial Intelligence (AI) is being utilised to prevent and counter violent extremism (P/CVE) in Indonesia, as well as the obstacles that hinder its broader implementation. The study aims to answer two key questions: (1) How can AI concretely enhance Indonesia's P/CVE efforts? (2) What limitations and challenges emerge when incorporating AI into these efforts? For the study, a purposive sample of fifteen AI experts was recruited from the National Counter-Terrorism Agency (BNPT), the Counterterrorism Special Detachment 88 (Densus 88), academics, AI startups, civil society organisations (CSOs) focused on digital rights, and former terrorist offenders. Semi-structured interviews were transcribed and analysed through thematic coding in Atlas.ti. The findings indicate that AI currently aids Indonesian P/CVE operations in three main ways: monitoring social media activities, complementing conventional P/CVE efforts, and detecting online extremism efficiently. The challenges include algorithmic biases, a lack of public trust towards the government's use of AI, and legal barriers to integrating AI into P/CVE efforts. The study also underscores useful recommendations, including collaboration among relevant stakeholders, adherence to ethical principles, and investment in public training on AI use. This research presents timely insights for policymakers, technology developers, and civil society by exploring how AI can be ethically and effectively incorporated into P/CVE strategies in Indonesia. The findings aim to inform the development of more inclusive, rights-respecting AI frameworks that balance national security needs with public trust. Beyond local relevance, the study contributes to global debates by offering a bottom-up perspective from the Global South that challenges Western-centric models of technology governance. It demonstrates how Indonesia's experience provides a transferable framework for other nations navigating the dual promise and peril of AI in securitisation, while also interrogating the competing logics of security and human rights within multi-stakeholder governance models for emerging technologies.

Key words: Artificial Intelligence; Countering Violent Extremism; Indonesia; Technology Governance; Global South.

Introduction

Artificial Intelligence (AI) has quickly come to be seen as a transformative technology in multiple fields, including healthcare, education, finance, and many others (Akinnagbe, 2024). In the security field, AI is increasingly discussed as a potential tool for tackling one of the most serious global challenges of the twenty-first century: violent extremism (Craanen et al., 2025). Governments, technology developers, and civil society organisations (CSOs) are all exploring how AI-enabled systems might help detect and remove extremist material from the online space, reduce the dissemination of terrorist propaganda, detect individuals who are at risk, and examine their online activities (Nelu, 2024). At the same time,

critical conversations are occurring about the risks and limitations of AI, particularly focused on algorithmic biases, data privacy, and the potential for misuse by ill-intended actors (UNCRI & UNCCT, 2021). In this context, AI's potential and pitfalls have emerged as a key area of argument in global preventing and countering violent extremism (P/CVE) conversations.

In Southeast Asia, and particularly Indonesia, the urgency of this conversation is particularly pronounced. Indonesia has experienced repeated threats of terrorism and violent extremism, which have ranged from Jemaah Islamiyah (JI) in the early 2000s to more current Islamic State (IS) inspired networks (Satria, 2022). Recruitment, radicalisation, and extremist propaganda have increasingly pivoted online, facilitated by social media platforms, encrypted communication, and digital anonymity (Ismail, 2024; Zeiger & Gyte, 2021). Despite the Indonesian Government responding through various P/CVE mechanisms, including law enforcement, economic, deradicalisation, socio-cultural empowerment, and counter-narrative campaigns, the use of advanced technologies such as AI remains underexplored (Aminah et al., 2023). Current approaches are overly dependent on long-standing approaches in traditional intelligence (law enforcement) work and face-to-face interventions, without addressing the vulnerability created within the fast-evolving digital space of extreme content. Even though the online counternarrative approach has gained popularity in Indonesia, the belief that "social media is yet to become a major tool for terrorist recruitment" is still rampant (Grealy, 2018).

Worldwide, the vast majority of scholarship on AI and P/CVE has been influenced by Western experience and policy frameworks that focus on the characteristics of AI as enhancing content moderation, surveillance, and predictive policing (see Bloch-Wehba, 2021; Raji & Sholademi, 2024). This scholarship primarily sees the capabilities of AI without sufficiently accounting for the contextual realities of the Global South, such as a lack of human and capital resources, fragile digital governance structures, and deep-rooted public distrust of governments, which creates a very different landscape for adopting AI in countries such as Indonesia. In addition, discussions about security versus human rights in transitional democracies are also complicated by state overreach and abuse of surveillance technologies. This is an important gap in the academic literature, and it is particularly important to understand AI for P/CVE through an Indonesian lens, because it is critical to situate AI within Indonesia's broader socio-political, legal, and cultural context.

This study seeks to fill this gap by looking at how AI is being used to counter violent extremism in Indonesia, and what constrains its wider application. Specifically, it poses two interconnecting questions: (1) how AI can practically improve Indonesia's counter-violent extremism measures, and (2) what the limitations and challenges are in incorporating AI into these measures. The study has a purposive sample of fifteen AI experts, consisting of officials from the National Counter-Terrorism Agency (BNPT), members of the Counterterrorism Special Detachment 88 (Densus 88), academia, developers from AI startups, CSOs focusing on digital rights, and former terrorist offenders. The research captures a very broad range of perspectives. A series of semi-structured interviews was conducted and subsequently processed as thematic codes using Atlas.ti, including the prospects and constraints of using AI.

The results show that AI engagement with P/CVE in Indonesia is threefold: monitoring social media activity; complementing traditional P/CVE activities; and introducing a more efficient method for detecting various forms of online extremism. In addition to technical and policy challenges, there will be social and ethical concerns that impact the overall effectiveness of AI in its role in P/CVE. The study finds that significant challenges remain, such as algorithmic biases that may provide incorrect content or individual identification, the lack of public appreciation for how the government is using AI, and the legal and regulatory barriers that slow adoption.

The study's contribution is important for both academic debate and effective policy development. For policymakers, it provides practical recommendations for stakeholder

collaboration, ethical AI usage, and investment in training to instil public trust. For scholars, the study advances theoretical debates by breaking from the tradition of inclusion of a contemporary bottom-up Global South perspective to counter existing Western-centred models of AI governance. Indonesia's experience showcases how emerging economies can be reactive to advanced and inviting technology while meeting the needs of national security, public trust, and human rights concerns. By starting at the bottom with these perspectives, the data contribute not only to the local understanding of AI in P/CVE but also to a more global understanding of the governance of emerging technologies in securitisation.

Literature Review

1. AI and P/CVE: Global Debates

AI in P/CVE has become a dedicated theme within the global security discussion. Scholars and practitioners have been keen to outline AI's potential for enhancing surveillance, detecting extremist content on the internet, and enabling predictive analytics for risk assessment (Akilli, 2024). Similarly, sites are looking to leverage AI algorithms to quickly find and remove extremist material before it goes viral. Governments, technology companies, and CSOs are beginning to look into AI early warning systems as a means of predicting a potential radicalisation trajectory (Irfan et al., 2023). More practically, AI has also been assisting law enforcement and intelligence agencies, sifting through huge amounts of digital data and pre-flagging potential threats faster and more efficiently than a team of analysts could potentially achieve alone.

However, despite these advances, the deployment of AI to counter violent extremism continues to face difficulty. A wealth of research exists on algorithmic bias. Minorities, for instance, are more likely to be designated as security threats (Hobart, 2025). Critics claim these biases arise through an unrepresentative sample for training data, non-transparent decision-making, and perpetuate systemic inequality under the veil of technological neutrality (Min, 2023). In addition, a reliance on automated systems raises concerns about false positives, compromising trust from the public and risking the stigmatization of entire communities (Ferrara, 2024). Dilemmas arise related to AI-driven surveillance technologies, which provide considerable advantages in security and crime deterrence but pose important ethical and legal issues concerning individual privacy rights and the risk of misuse (Ünver, 2024).

The global discussion continues to oscillate between the enthusiasm that AI could contribute to improved security and the scepticism over the ethical and social consequences of AI. Academics are urging placing AI utilisation within transparent governance frameworks that address a delicate balance between national security measures and individual rights (Mosa et al., 2024). Unfortunately, literature provides a predominantly Western perspective, including the United States and Europe, where AI utilisation has been adopted within content moderation paradigms and predictive policing. It remains unclear to what extent these approaches can be adapted to alternative socio-political frameworks, particularly those situated around the Global South.

2. Southeast Asia and Indonesia: P/CVE Landscape

For decades, Southeast Asia has represented a significant region of concern when it comes to violent extremism and related terrorist activity (Barton, 2024). Countries such as Indonesia, the Philippines, and Malaysia have faced both transnational terrorist groups and domestic extremist cells. In Indonesia, violent extremism has evolved from the large-scale attacks organised by Jemaah Islamiyah in the early 2000s to decentralised and digitally-facilitated networks associated with the Islamic State (ISIS) or affiliated groups (Nuraniyah, 2019). New social media platforms, encrypted applications, and online forums have presented new mechanisms for recruitment, propaganda, and coordination. Moreover, women and youth are often specifically targeted using gender-specific behaviour and emotionally appealing narratives (Scheuble & Oezmen, 2022).

In Indonesia, the state response has been a diverse suite of P/CVE strategies. For example, the national programmes for deradicalisation (as implemented by BNPT) are contrasted with counter-terrorism measures (as undertaken by the Densus 88) (Hasibuan & Tijow, 2024). P/CVE also hinges on the role of civil society in terms of counter-narratives, rehabilitation, and community resilience (Sumpter, 2024). However, while the variety of measures is encouraging, the digital space remains a significant point of vulnerability. Extremist organisations in Indonesia have demonstrated that they can effectively utilise the internet and social media sites, enabled by applications such as Facebook, WhatsApp, and TikTok, to spread digital messages in a fast and subtle manner (Ismail, 2024; Riyanta, 2022).

While AI has been eyed as a useful monitoring tool for online extremist content, applications of AI into Indonesia's suite of P/CVE strategies remain negligible and non-digital, with most counter-radicalisation and deradicalisation continuing to centre on non-analysis practices (Mutiara, 2025). Most notably, non-digital measures such as rehabilitating radicalised individuals through religious re-education in prison or community settings, or engaging communities in face-to-face dialogue and awareness raising, continue to dominate the landscape (Ilyas & Athwal, 2021; Widya, 2020). The lack of quality infrastructure, low levels of available funding, and insufficient relationships between the state and new technology companies continue to hinder the application of AI monitoring technology (Mutiara, 2025). Additionally, the requirement and therefore discretion needed to facilitate security data, and the public trust it requires, complicates even the subsequent steps of sharing existing relevant data to effectively realise AI (Rulinawaty et al., 2024). Collectively, this creates a paradox, whereby officials and civil society actors can cite examples of AI, but there is little recognition or actual use of the power of AI for this threat.

3. Technology Governance and Human Rights

The utilisation of AI in security contexts raises larger questions about governance, ethics, and human rights (Jones, 2023). Academics argue that while AI can be used to address legitimate security objectives, the consideration of AI needs to be properly embedded in governance structures that support civil liberties (Arora et al., 2025). In authoritarian or weakly regulated contexts, there is a risk that AI will be utilised for political surveillance and repression under the guise of counterterrorism (Kaskina & Cvetkovska, 2024). This tension is particularly acute in Southeast Asia, where weak democratic institutions are interspersed with strong security apparatus.

In Indonesia, the situation is more complicated. Indonesia is the largest Muslim-majority democracy in the world, yet it maintains the values of diversity, human rights, and opposition to extremism. (Sumarno & Affianty, 2024). CSOs raise the alarm for overstretch by raising issues on surveillance and data privacy (ICT Watch, 2024). The introduction of AI-based tools into this context complicates the debate. With a lack of adequate safeguards, AI could actually aggravate distrust between citizens and the state, thereby undermining the legitimacy of P/CVE.

Global human rights frameworks, such as the United Nations and the Global Counterterrorism Forum, emphasise the ideas of fairness, accountability, and transparency in AI usage for security purposes (Land & Aronson, 2020). However, adopting these ideas and principles in practice is complex, particularly when resources are low and governance is still being developed. Scholars have suggested a multi-stakeholder process to involve the government, private sector actors, and CSOs to design and oversee AI systems (Criado et al., 2025). Such processes are vital in the context of Indonesia, where trust in government initiatives is tenuous, and civil society plays a supporting role in P/CVE.

4. Identified Research Gap

The body of research discusses both the potential and risks of AI in countering violent extremism, but there are some gaps to consider. First, most studies are based in the West, with very

few empirical studies conducted in the Global South, meaning that debates are primarily informed by the experiences of countries with developed technological infrastructures, and there are still some unknowns about how AI can be effectively and ethically assessed in their unique socio-political environments. Second, existing research frequently discusses the views of the government and of technology companies, with less attention on others, including CSOs or people who have directly experienced violent extremism. This limited framing excludes deliberations of the complex, multi-stakeholder dynamics that are needed to forge strategies for P/CVE that are effective, sustainable, and that respect rights.

Third, regarding Indonesia, there is a significant gap in empirical studies examining how AI is currently being used, either implicitly or explicitly, in P/CVE work, while policy documents and government statements acknowledge technology as a relevant factor. Only a few studies have focused more broadly on examining how AI is actually being operationalised, what challenges it entails as part of that work, and what ethical challenges are associated with AI integration, which is a gap that this study takes into account. In providing a bottom-up perspective from Indonesia that includes input from government agencies, academics, technology start-ups, CSOs, and former offenders to inform current understanding of AI in P/CVE, this study enriches the local significance of AI in P/CVE and contributes to global literature by emphasising non-Western narratives and providing a framework that can be transferrable in influence for other countries attempting to navigate the promise and peril of AI in securitisation.

Materials and methods

1. Research Design

This study utilised a qualitative research design to investigate the ways in which AI is used in P/CVE in Indonesia, and to understand the challenges of its broader use. A qualitative framework was chosen, as the phenomenon being studied is inherently multifaceted, contextual, and influenced by the perceptions, practices, and interactions of a number of actors (Tenny et al., 2024). Quantitative research may provide insight into the prevalence of particular practices or attitudes, but may fail to adequately address the complexity of practice and contested meanings surrounding new technologies in sensitive fields such as counter-terrorism (Sapkota, 2024).

This study focused on qualitative inquiry, in the form of semi-structured interviews, as this approach allows researchers to explore in depth the experiences, beliefs, and judgements of participants, while still retaining the flexibility to respond to emergent themes (Roller & Lavrakas, 2015). This flexibility is particularly important in this area, where rigidly structured surveys may miss nuances or details of the issue, and participants are unlikely to provide valid responses without a measure of trust and rapport with the researcher (Morell, 2023). The inductive nature of the study permitted findings to emerge from the data, rather than being tied solely to pre-established theoretical designs, while still being connected to existing conversations in the academic literature of AI, technology governance, and P/CVE (Casula et al., 2021).

2. Sampling and Participants

Purposive sampling was utilised to confirm that participants were engaged in research questions relevant to the topic, possessed knowledge about AI, and exhibited experiences with P/CVE work (Reddy & Ramasamy, 2016). The selection process involves deliberate decisions made by the researcher due to the characteristics of the participants (Tongco, 2007). When employed properly, purposive sampling outperforms probability sampling because expert participants are more knowledgeable and perceptive than random target populations (Bernard, 2018; Klar & Leeper, 2019; Tremblay, 1957). This method is also helpful for studies with limited funding and scarce resources (Campbell et al., 2020).

Participants were recruited through purposive expert sampling using three strategies: (1) direct invitations through the researcher's professional networks in Indonesia's P/CVE and AI communities, (2) organisational outreach to relevant institutions and civil society actors, and (3) snowball referrals whereby interviewed participants recommended additional experts meeting the inclusion criteria.

The sample included fifteen individuals from various stakeholders: representatives from the government (BNPT), law enforcers (Densus 88), academic experts on AI and P/CVE research, practitioners from AI start-ups, CSOs working on human rights and online safety, and former extremist offenders providing grounded perspectives on the effectiveness of technology-enabled interventions (see Figure 1). As for participants' demographics, 60% of them are male and 40% are female (see Figure 2), with more than 70% of them acquired P/CVE ground experiences and practices (see Figure 3).

Figure 1: Participants' Composition by Industries

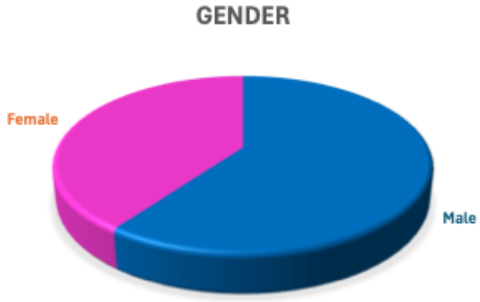

Figure 2: Participants' Composition by Gender

Figure 3: Participants' Composition by P/CVE Experience

The researcher purposefully assembled this group to understand the challenges of multi-stakeholder dynamics of AI considerations in P/CVE. By including the institutional actors and the lived experience, the researcher sought to circumvent the tight state-focused narrative and seek out interactions between policy, practice, ethics, and trust in society. This approach captures the complexities of the diverse, context-dependent, and intricate lived experiences of people involved in both AI and P/CVE infrastructures in Indonesia (Bogdan & Biklen, 1982; Lehnert et al., 2016).

Inclusion criteria required that participants either have direct professional engagement in AI development or P/CVE practice in Indonesia or have knowledge of and experience with radicalisation, deradicalisation, or the use of digital technologies to commit extremism. The researcher recruited participants from professional networks, snowball referrals, and outreach to organisations. This recruitment method was necessary due to the sensitivity of the topic and the difficulty of accessing experts in both AI and P/CVE.

Sampling adequacy was determined through the principle of information power, whereby the inclusion of highly specialised expert participants enabled the study to generate rich and conceptually relevant insights despite a modest sample size. Given the narrow research aim, the specificity of participant expertise, and the depth of semi-structured interviews conducted, the sample was considered sufficient to support meaningful qualitative analysis and theoretical interpretation.

3. Data Collection

The data collection phase was conducted over a period of three months. Semi-structured interviews were selected as the primary data collection method for two reasons. First, participants had the opportunity to express their viewpoints in their own words (Karatsareas, 2022). Second, semi-structured interviews provided individuals the freedom to personalise their responses while maintaining comparability across cases (Ahlin, 2019). The semi-structured interview script included open-ended questions categorised around three themes: (1) current applications of AI technologies within P/CVE, (2) perceived challenges and limitations of AI technologies, and (3) perspectives on future opportunities and ethics.

Interviews took place over encrypted videoconferencing platforms for security and accessibility purposes. Each interview lasted between 60 and 90 minutes in Bahasa Indonesia to facilitate comfortable and precise expression from participants. With their consent, all interviews were audio-recorded and then transcribed in Bahasa Indonesia before being translated into English for analysis. A standardised interview guide was utilised. Although the guide covered essential topics comprehensively, the interviewer had the flexibility to adapt questions based on each respondent's experiences and insights.

To mitigate potential social desirability bias, participants were given assurance that all responses were acceptable, there would be no right or wrong answers, and all contributions would be anonymous. The researcher was explicit that this study was independent, entirely academic, and not a government or institutional study. These additional measures established trust and encouraged full disclosure of opinions and experiences.

4. Data Analysis

Thematic analysis, supported by Atlas.ti software, was used in this study to analyse qualitative information obtained from the interviews. The analysis of the data was completed in accordance with Braun and Clarke's (2006) process, which includes: (i) data familiarisation; (ii) generating initial coding; (iii) searching for themes; (iv) review of themes; (v) defining and naming themes; and (vi) reporting. The transcripts were imported into the qualitative analysis software, Atlas.ti, to systematically handle the datasets, allowing researchers to efficiently organise interviews using advanced tools for coding, retrieval, and visualisation, which greatly improves the rigor and productivity of the analysis (Mastrobattista et al., 2024; Ñañez-Silva et al., 2024).

The first round of open coding resulted in a comprehensive list of codes associated with AI-related applications, challenges, and governance discussions. The open codes were then systematically refined, first through a process of axial coding, where related codes were aggregated and organised into high-level themes, such as "creates counternarratives," "budget restriction," and "amend the regulations." Although Atlas.ti offers valuable tools for coding and visualisation, the analysis remains primarily a human endeavour. Researchers immersed themselves in the data, interpreted meanings, and made conceptual decisions at each stage of the process. Therefore, Atlas.ti acts mainly as a tool for organisation and visualisation, while the researcher interprets and integrates theories.

5. Ethical Considerations

This study did not require formal institutional ethics committee approval because it constituted low-risk social research involving expert participants speaking in their professional capacity. The

research complied with the ethical principles of voluntary participation, informed consent, confidentiality, and data protection, consistent with the Singapore University of Social Sciences' research ethics guidelines. All participants were informed of the study's academic purpose, their right to withdraw at any stage, and the anonymisation procedures applied to interview data.

Due to the delicate nature of examining violent extremism and security-related technologies, an extensive set of ethical protocols was employed. Participants were asked for their explicit consent before their interviews were audio-recorded, ensuring that recordings only took place with their voluntary agreement, which aligns with the concepts of *response-ability* and *thinking with care* in research ethics (Klykken, 2021). Consent was verbally obtained. While informed consent is essential, it is not a cure-all, as simply informing subjects and requesting their acceptance of risk is insufficient (Gordon, 2020). That risk must be decreased as much as possible while still aligning with the research's scientific objectives. Furthermore, it is wise to remember the words of Henry Beecher: "The most effective safeguard is the presence of a knowledgeable, conscientious, compassionate, and responsible investigator" (Harkness et al., 2001).

Protecting the confidentiality and privacy of participants is essential, as they may share deeply personal and sensitive details (Kang & Hwang, 2023). Participants were given the choice to disclose or withhold their identities. For those who chose anonymity, it is maintained by using pseudonyms and removing any identifiable information in transcriptions or reports to safeguard their identities (Walford, 2005). In this instance, identities were replaced with generalised categories (i.e., "academic expert" or "CSO representative"). Digital data, including audio recordings and transcripts, is kept on encrypted, password-protected devices that are only accessible to the researcher to protect them from unauthorised access and possible breaches (Goyal, 2022).

The study involves vulnerable populations, such as former extremists, who may have endured trauma. There have been studies exploring what participants perceive as distressing when participating in research involving sensitive subjects, which may stem from discussing a sensitive topic for a prolonged duration (Carter-Visscher et al., 2011). To safeguard the research participants, they were explicitly assured that the research is solely for academic purposes and has no connection to law enforcement, prosecution, or surveillance activities. This reassurance is essential for fostering trust and ensuring that participants feel secure and respected when sharing their experiences.

Results

The analysis using Atlas.ti reveals the distribution of responses across different code categories (see Figure 4). It shows that participants described more challenges (red) than potential uses (green) of AI to combat violent extremism. Yet, this critical stance is counterbalanced by the broad set of recommendations (blue) that participants made. While participants were cautious about the limitations of AI, they were equally constructive in providing recommendations to support the safe, effective, and contextualised use of AI.

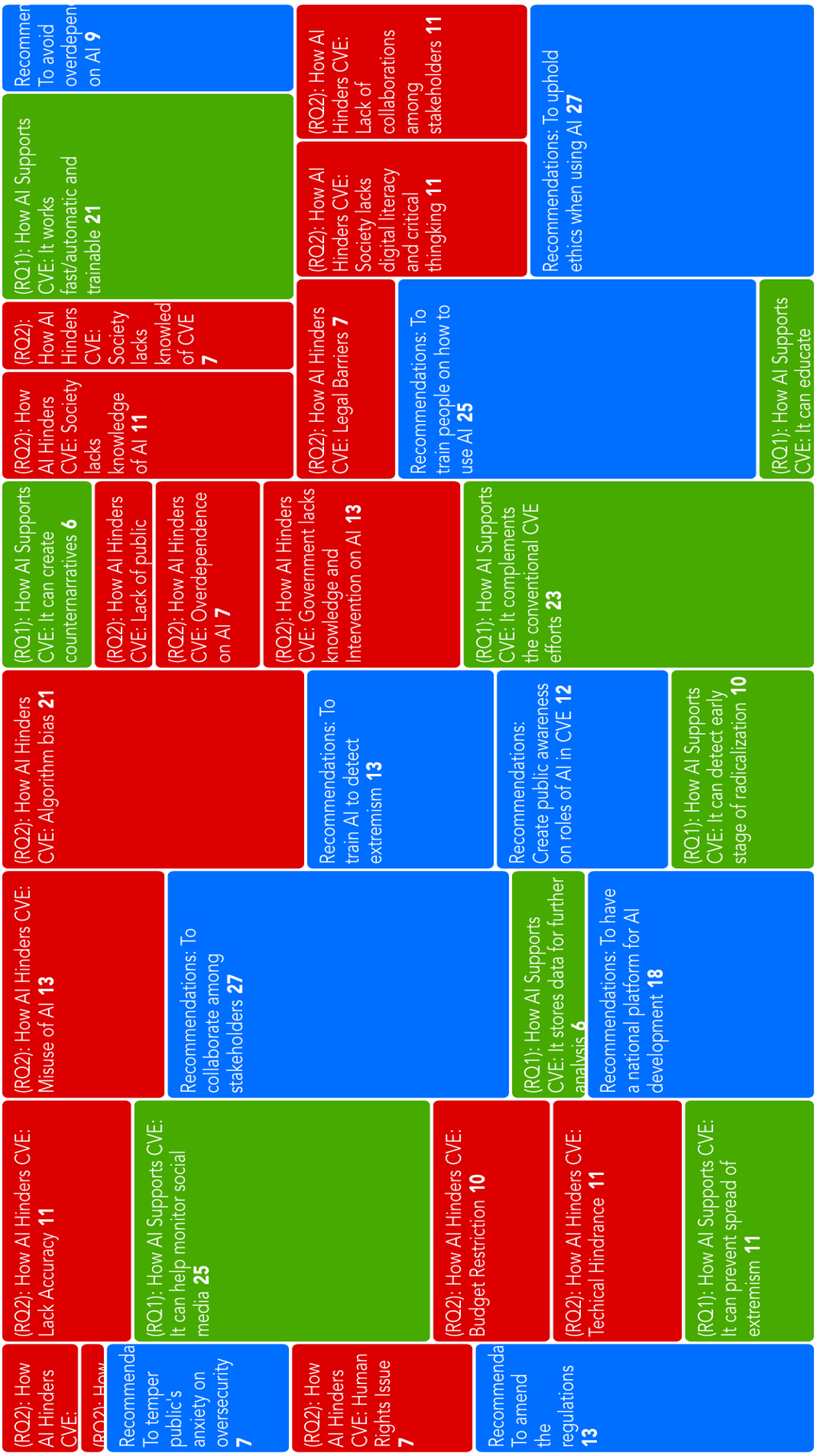

Figure 4: Distribution of Responses Across Different Code Categories

1. The Potential Use of AI in P/CVE in Indonesia

Participants showed enthusiasm regarding the utilisation of AI to bolster Indonesia's P/CVE efforts (see Figure 5). The most prominent themes that emerged from the data analysis include AI as a tool to create counternarratives and early detect online radicalisation. AI is also perceived as a means to educate the public on violent extremism, assist in monitoring social media, prevent the spread of extremism, complementing conventional P/CVE efforts, and is fast and trainable. Although most participants agreed that this technology is still nascent in application, their reflections suggested three primary possibilities for application: enhancement of monitoring, augmentation of conventional P/CVE activities, and improved detection or response efficiency.

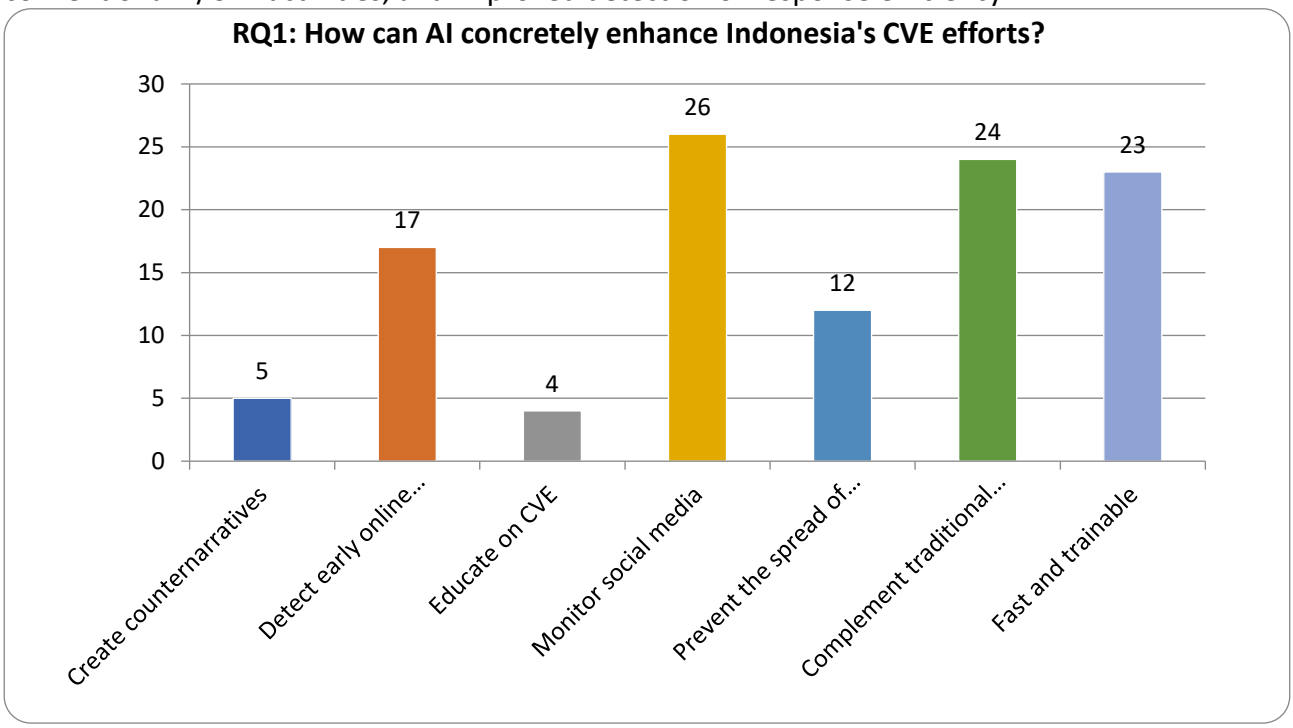

Figure 5: Prominent Themes on How AI Enhances Indonesia's P/CVE Efforts

A number of respondents emphasised the pertinence of AI as an effective means of monitoring cyberspace, where the narratives of Jihadist and far-right movements can develop. They spoke about watching online platforms, social media, forums, and video content with radical language and symbols. As one respondent from an AI startup remarked, "*The algorithm is quite sophisticated now, even AI can pick out words off parts of video*," as an example of its ability to identify content with Jihadist or far-right extremist language in multimedia.

Fajar Pahlawan Husain, the Founder of an AI-powered edutech platform in Indonesia, *Tugasisten*, expressed his vision of harnessing AI-powered keyword and sentiment analysis to identify emerging themes and trends in real-time: "*If there was a violent extremism issue, we could see how often it's mentioned in negative content or keywords across social media by someone*." He felt that if spikes in words such as *jihad* or *ISIS* were seen, it might be possible to identify "hot spots" for intervention earlier on. He also observed that preventative measures could be initiated before online narratives grew: "*Preventive measures can be put in place quickly, should the issue be spreading rapidly on social media*".

Notably, participants underscored that AI could also be used to track behavioural trends. This could potentially allow AI to identify individuals who are simply "*leaning to certain content or ideas*" by seeing how they interact, and allow for an opportunity to intervene before they become radicalised. Some interviewees repeatedly noted that AI is not a substitute for the existing P/CVE programmes, but rather a valuable complement. As one interviewee said, "*AI is a complement and*

an aid". They envisioned AI being used to supplement BNPT's counternarrative campaigns by identifying trends in misinformation, which could then be countered through additional targeted positive narratives.

Representatives from government entities envisioned blending AI outputs with population data to provide deeper insights. As an illustration, they pointed to examples that could cross-reference flagged accounts with demographic data so that authorities could provide tailored responses to new risk threats. The representatives also indicated that AI could enhance early detection by conducting deeper and faster analyses for more data-driven decision-making for security agencies than human reviews.

A further proposed benefit of AI is its capability for automation and scale. A number of participants articulated that it was something "*automatic and [does not] need any human involvement*", which can analyse big data across multiple datasets in order to identify meaningful patterns within the data quickly. By decreasing the time and processing needed to work through data and analyse it, AI could create opportunities for the analyst to focus more on actions: "*If we can cut the process to accelerate the analysis and data collection, we can focus on producing products that can help combat terrorism*". Additionally, participants noted AI's capacity to train, meaning that AI could learn from or be based upon its exposure to established datasets and previous cases - enhancing predictability over time. The capacity of AI to train was viewed as key to predicting the ever-evolving strategies of extremists.

2. Challenges in Integrating AI into Indonesia's P/CVE Strategies

Participants were honest about the obstacles to using AI within Indonesia's P/CVE framework, despite some optimism (see Figure 6). The most frequently cited challenges encompass algorithm bias, budget constraints, the government's limited understanding and intervention in AI, legal barriers, insufficient collaboration among stakeholders, public distrust towards the government, low levels of critical thinking and digital literacy among the public, and society's overall lack of knowledge on the subject. Interviewees articulated three major concerns: issues associated with algorithmic bias, insufficient public trust, and regulatory and institutional challenges.

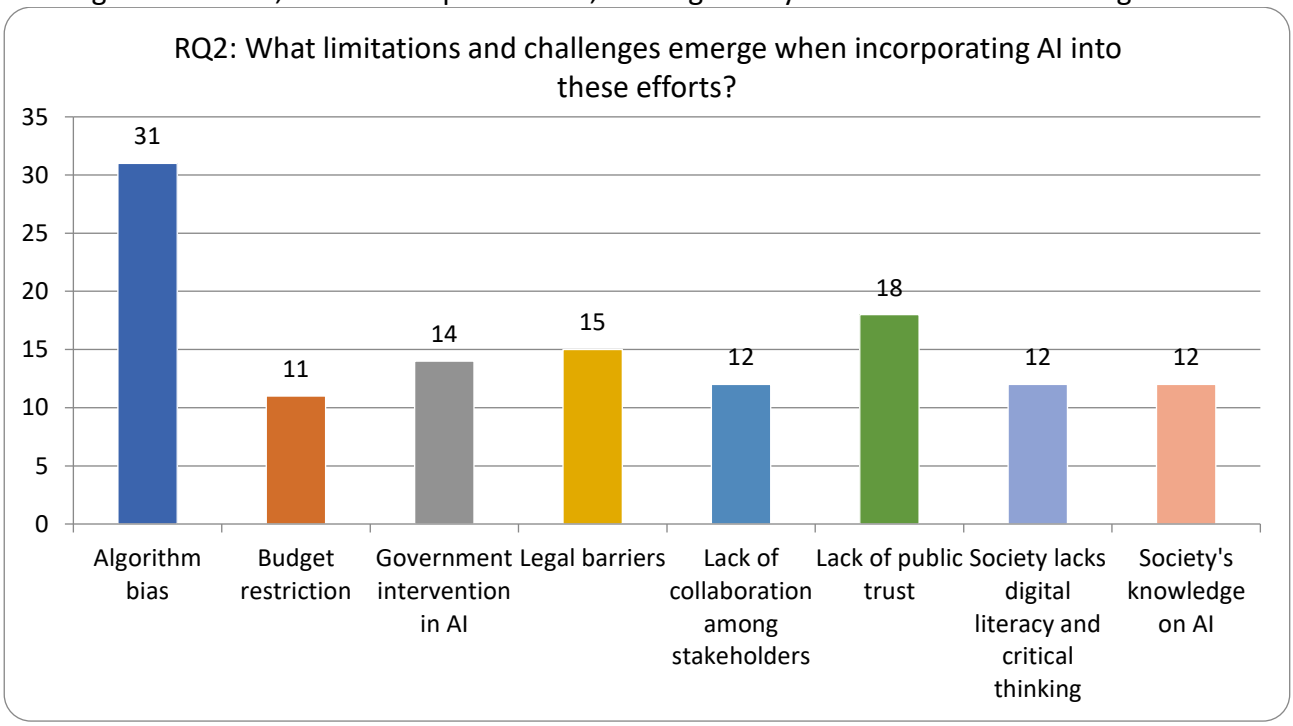

Figure 6: Prominent Themes on the Challenges in Incorporating AI into Indonesia's P/CVE Efforts

A significant obstacle identified was AI's incapacity to navigate Indonesia's linguistic complexities. A researcher at Padjadjaran University in West Java, Indonesia, highlighted that most early AI displays were trained using English datasets, which diminished trust in their local application: "*I have used analysis in Indonesian several times, but it's limited. That's because the datasheets or the databases in their AI are mostly based in English*."

With hundreds of dialects in their regions, participants stated that AI consistently misinterprets or misses the subtleties and expressions specific to their context: "*They haven't learned the regional languages yet*"; "*AI just has to learn more before it can understand*." With these gaps in AI technology comes a greater chance for false positives, which undermines confidence in any automated monitoring of their work. Assistant Professor Roosalina Wulandari from the Department of Psychology, Bina Nusantara University (BINUS) in Jakarta framed the underlying dilemma this way: "*Algorithms always work as a double-edged sword*."

Some of the participants said people are worried about the surveillance aspect. Many Indonesians expressed trepidation about monitoring AI-enabled systems: "*They do not want to be monitored by the system*." A representative from BNPT commented on people worrying about data privacy: "*They are very concerned about data privacy*." Significant distrust creates risks to legitimacy and effectiveness. Without public support, AI-based P/CVE approaches may be seen as governmental overreach, undermining resilience against, rather than building resilience against, violent extremism.

Lastly, participants highlighted the lack of clarity around legal frameworks. In Indonesia, the country passed the Personal Data Protection (PDP) Law in No. 27 in 2022, but its interpretation remains ambiguous. Another representative from BNPT stated, "*There should be a balance between civil liberty, freedom of expression, and privacy rights, which can be challenging when it comes to security issues*". He added, "We are discussing the conflict between security issues versus privacy issues; it can sometimes be a zero-sum game." Vague regulatory boundaries stifle innovation for technology developers and promote caution around government agencies. Participants were concerned that without clear regulation, AI use in P/CVE could be punitive, too prohibitive, or too permissive.

3. Cross-Cutting Insights

In addition to its potential and challenges, the participants also provided additional contextual information on how AI integration into P/CVE should be approached in the case of Indonesia (see Figure 7). The main themes identified in this analysis include the importance of raising public awareness about AI and P/CVE, the need to amend regulations when necessary, and encouraging stakeholders to strengthen their collaboration. For Indonesia to develop a national AI centre, it should address public anxieties about security, incorporate local languages and dialects into AI, train the public on AI usage, and uphold ethical standards in AI deployment. Among these, three themes are evident: the importance of collaboration, the priority of ethics, and the importance of education and training.

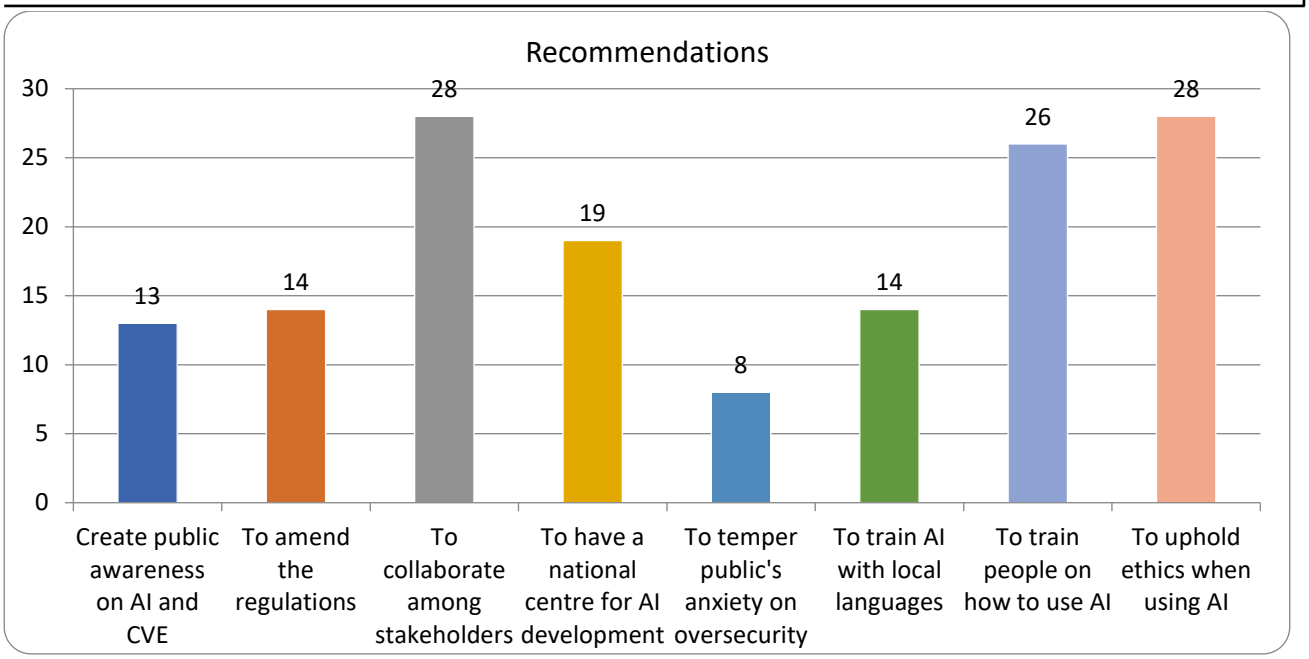

Figure 7: Recommendations from Participants

The interviewees consistently pointed to the importance of shared responsibility in AI for P/CVE, with no actor being able to manage all relevant aspects effectively alone. Such responsibility could fall on government, industry, academia, civil society, and communities. As summarised by Yudhistira from an AI startup, Billix Creative: "*Both sides need to learn from each other. The government needs to learn from developers about the importance of AI technology, and the developers need to learn about laws and regulations"*. Another research participant emphasised the need for the government to be both genuinely collaborative: "*The government now needs to be brave enough to listen to academia and also brave enough to listen to industry. Collaboration should not just be on paper*". Research participants also suggested creating collaborative forums where various stakeholders across sectors and contexts could share knowledge, share resources, and request both experiences and the use of many experiments completed in public.

Similarly, ethical issues were brought up. A few participants talked about the dangers of AI in political surveillance and privacy violations. They advocated for human rights and local knowledge to be integrated into AI applications: *"Ideally, where tools with AI capabilities… than terms and conditions would be aligned with any regulations within respective contexts*". Stanislaus Riyanta, a researcher, lecturer, and expert in intelligence, security, and terrorism from the University of Indonesia, emphasised the significant role of strong ethical reviews, community engagement, and transparency: "*In-depth ethical reviews and community engagement are also necessary to ensure fair and responsible use of AI*".

Lastly, it was indicated by participants that training and awareness-raising will be important. AI was conceptualised as not simply a technical tool, but a system that requires someone skilled to supervise it. As one participant said, "*We will really need personnel … specifically in the prevention of terrorism. They need training*". An activist from ruangngobrol.id, Ani Ema Susanti, added, "*We educate them about misinformation, disinformation, and how AI works … to make young people aware of the danger of manipulative content*". Participants suggested that addressing the intrusion of AI into everyday lives should include creating digital literacy campaigns, formal training programmes, and new subject-matter integration to develop resilience at the societal level. Many

participants indicated that a lack of attention to training or awareness-raising will result in AI being misapplied, misunderstood, or mistrusted.

To expand upon the description of these dynamics, we conducted a code co-occurrence analysis (see Figure 8). The Sankey diagram illustrates the participant's pattern of categorizing challenges alongside recommendations, for instance, instances raising the concern with distrust of the public relationship with AI, and sections noting the need for training at a local level, raising public awareness, and developing ethical guidelines. In a similar fashion, lack of collaboration was categorised alongside the suggestions for multi-stakeholder forums and regulation to tighten this measure. This demonstrates the participant's ability to not only identify barriers but also suggest pathways to move forward in a constructive manner.

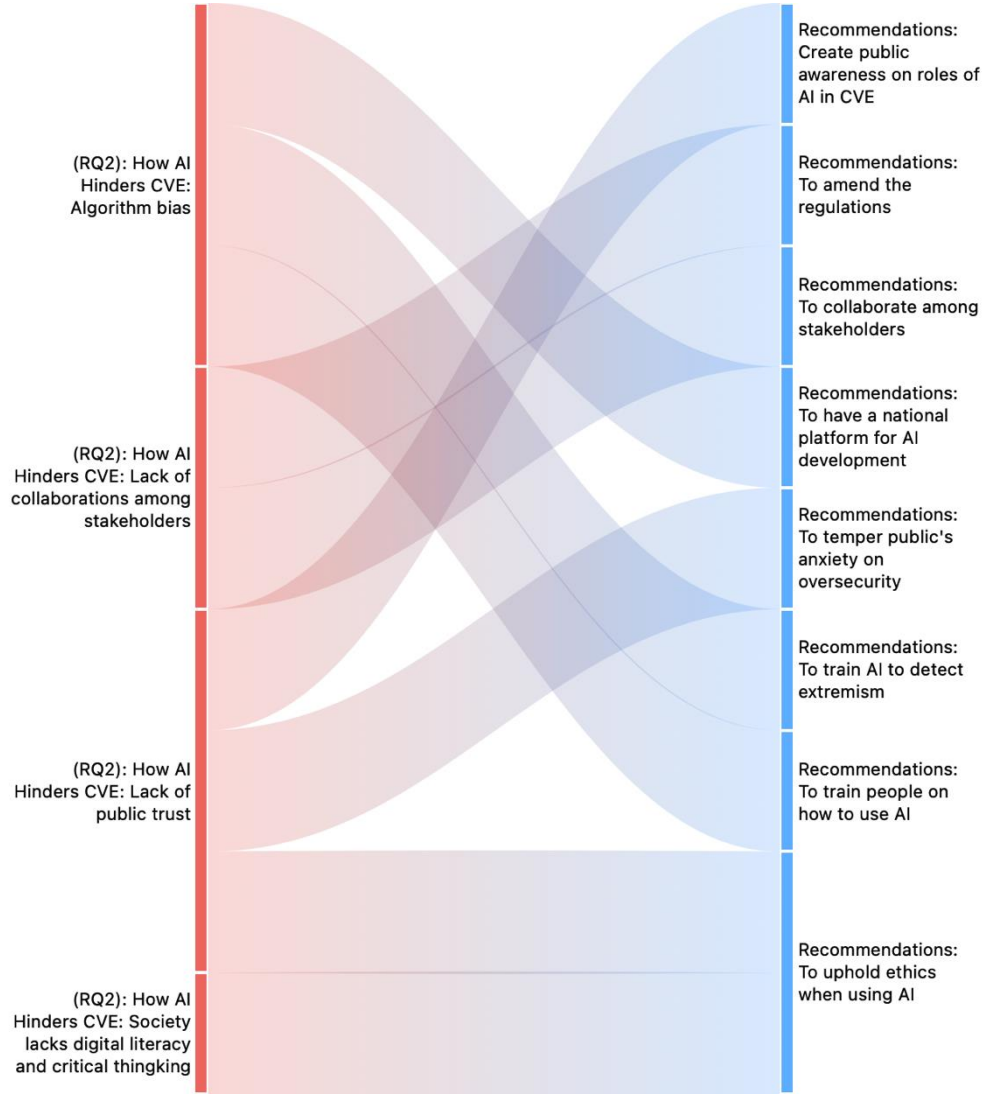

Figure 8: Code-Occurrence Analysis Showing the Strong Correlations Among Variables (Identified Challenges and Recommendations)

Discussion

1. The Promise of AI in P/CVE

The objective of this research was to explore ways in which AI can be incorporated into Indonesia's P/CVE responses. Globally, AI is employed in P/CVE for content moderation and predictive analytics. Indonesian participants reflected these opportunities, noting AI could be employed as a device for extremist keyword detection and behavioural pattern tracking, and highlighting potential online hotspots. However, they asserted that AI should not replace community engagement, religious re-

education, and rehabilitation. This view is different from some Western discourses that propose AI as a solution measured against traditional techniques, indicating a much healthier, balanced view of intervention pitfalls as well as context sensitivity in Indonesia.

2. Confronting Challenges: Algorithm Bias, Public Mistrust, and Legal Gaps

Participants expressed that AI did not perform well in local languages and dialects, raising the risk of false positives and missing important signals. This reflects global anxieties over unrepresentative datasets, but the potential costs are magnified given the linguistic diversity across the Indonesian landscape. Many participants also expressed hesitancy about government surveillance. Concerns about privacy and past abuses anchor suspicion that AI has the potential for misuse, and if so, would somehow delegitimise P/CVE efforts. Without community support, technical changes put communities at risk of becoming counterproductive. Indonesia's PDP Law of 2022 provides a legal framework, but participants discussed the ambiguities of implementation. The absence of clear directives and rules on accountability and proportionality reduces the potential to innovate and raises the spectre of misuse. These obstacles show that technical progress is not enough; of equal importance are legitimacy, trust, and regulatory clarity.

3. Theoretical and Practical Implications

This paper makes contributions to the academic literature in several ways. It expands conversations on AI and P/CVE through a Global South perspective that pushes against assumptions of universal transferability from Western-centric realities. Second, this paper redefines the role of AI as supportive and complementary, therefore providing a view of socio-technical hybridity rather than technological determinism. Third, this paper connects security studies to the governance of technology in seeing algorithmic bias, distrust, and legal ambiguity as governance issues rather than merely technical failures. Finally, this paper highlights the “securitisation paradox” of the technology, enhancing safety but eroding rights and trust when these technologies are deployed without adequate safeguards.

These findings also facilitate lessons that could be put into practice. Policymakers and developers should consider prioritising the development of localised datasets to address linguistic bias, and while requiring more precise legal and ethical frameworks, there needs to be frameworks that provide proportionality, transparency, and accountability. Multi-stakeholder partnerships should be deliberately organised in well-resourced forums that reflect policymakers, academia, industry, and civil society. These forums could facilitate the development of public trust through public campaigns and a digital citizenship curriculum in schools so that citizens of all ages, particularly youth, can start to engage with AI and extremist narratives more critically. Finally, capacity-building initiatives should be launched to enable practitioners to develop a responsible AI context that adequately operates and is supervised.

4. Indonesia’s Contribution to Global Debates

Indonesia's experience expands the global discourse on AI, given that it has surfaced how resource limitations, ambiguity of regulations, and trust issues from society impact the adoption of AI in the Global South. In particular, rather than drawing on Western conceptions implemented from the top down, the stakeholders in Indonesia have drawn attention to engaging and trust-building processes. This sets out a transferable governance model that balances rights with security and illustrates the significance of adopting a bottom-up and context-appropriate approach to integrating emerging technology in P/CVE.

Conclusions

This investigation has explored the potential applications and lessons learned about using AI in Indonesia's P/CVE strategies. The results show that while regarded as a potentially game-changing tool for supporting social media monitoring, AI is understood as complementing the existing CVE

approach while improving the efficiency of detection and response to extremist narratives. In the same breath as these findings, participants also reflected on substantial, environment-specific barriers that would impact these potential uses - algorithmic discrimination or bias, public distrust, and ambiguity in legal definitions. After analysing the findings, each of these obstacles was followed by actionable recommendations for improvement, emphasising a call to action for multi-stakeholder collaboration, ethical checks, and support for education or training investments.

As is the case in all research, this research has limitations. Data saturation was not formally assessed, which means that future perspectives on these findings may develop other perspectives in Indonesia's P/CVE ecosystem. Nevertheless, this study has timely implications for policymakers, practitioners, and researchers. It has also contributed to the global debate by placing a perspective from the Global South and showing that in the P/CVE sector, AI must find common ground between security and rights, trust, and ethics. Future research may build on this work by using larger samples, comparative studies, and longitudinal research endeavours.

Funding

This study received no specific financial support.

Competing interests

The authors declare that they have no competing interests.

Use of Artificial Intelligence

Artificial intelligence was used for language editing, translation, and technical verification of the manuscript.

References

Ahlin, E. M. (2019). *Semi-Structured Interviews with Expert Practitioners: Their Validity and Significant Contribution to Translational Research*. SAGE Publications Ltd. DOI: https://doi.org/10.4135/9781526466037

Akilli, E. (2024). Artificial Intelligence in Counterterrorism: Navigating the Intersection of Security, Ethics, and Privacy. *Perspektive*, *73*, 1–5.

Akinnagbe, O. B. (2024). Human-AI Collaboration: Enhancing Productivity and Decision-Making. *International Journal of Education, Management, and Technology*, *2*(2), 387–417. URL: https://ejournal.yasin-alsys.org/IJEMT/article/view/4209

Aminah, S., Kartika, R. S., Susilo, S. R. T., Sipahutar, H., Asrori, Wibowo, C., Sabtohadi, J., U.R., F., Supratikta, H., Maemunah, S., & Rahayiningsih, Y. (2023). Combating Terrorism in Indonesia through Collaborative Strategy. *Migration Letters*, *21*(1), 561–571. https://doi.org/10.59670/ml.v21i1.5315

Arora, M. K., Lal, S., Singh, B., & Raghav, A. (2025). Balancing Security with Civil Liberties and Risks in AI-Driven Surveillance. In P. K. Dutta, B. Singh, C. Kaunert, & A. L. Sciacovelli (Eds.), *Security Intelligence in the Age of AI*. Emerald Publishing Limited.

Barton, G. (2024). Southeast Asia's Threat Environment in 2024. *Counter Terrorist Trends and Analyses*, *16*(4), 1–9.

Bernard, H. R. (2018). *Research Methods in Anthropology: Qualitative and Quantitative Approaches* (6th ed.). Rowman & Littlefield.

Bloch-Wehba, H. (2021). Content Moderation as Surveillance. *Berkeley Technology Law Journal*, *36*(3), 1297–1340. https://scholarship.law.tamu.edu/facscholar/1662/

Bogdan, R. C., & Biklen, S. K. (1982). *Qualitative Research for Education: An Introduction to Theory and Methods*. Allyn and Bacon.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing*, *25*(8), 652–661. https://doi.org/10.1177/1744987120927206

Carter-Visscher, R., Bell, K. M., & Blankenship, A. S. (2011). Ethical Issues in Research on Sensitive Topics: Participants' Experiences of Distress and Benefit. *Journal of Empirical Research on Human Research Ethics*, 55–64. https://doi.org/10.1525/jer.2011.6.3.55

Casula, M., Rangarajan, N., & Shields, P. (2021). The potential of working hypotheses for deductive exploratory research. *Quality & Quantity*, *55*, 1703–1725. https://doi.org/10.1007/s11135-020-01072-9

Craanen, A., Allen, E., & Atamuradova, F. (2025). *Artificial Intelligence for Counter Extremism: Exploring Threats, Challenges, Opportunities and Needs for Leveraging Generative AI in Countering Extremism.* Hedayah.

Criado, J. I., Sandoval-Almazán, R., & Gil-Garcia, J. R. (2025). Artificial intelligence and public administration: Understanding actors, governance, and policy from micro, meso, and macro perspectives. *Public Policy and Administration*, *40*(2). https://doi.org/10.1177/09520767241272921

Gordon, B. G. (2020). Vulnerability in Research: Basic Ethical Concepts and General Approach to Review. *Ochsner Journal*, *20*, 34–38.

Goyal, P. (2022). The Importance of Data Encryption in Data Security. *Journal of Nonlinear Analysis and Optimization*, *13*(1), 1–11.

Grealy, K. (2018). *Indonesia: Countering A Message of Hate*. Lowy Institute. https://www.lowyinstitute.org/the-interpreter/indonesia-countering-message-hate

Harkness, J., Lederer, S. E., & Wikler, D. (2001). Laying ethical foundations for clinical research. *Bulletin of the World Health Organization*, *79*(4), 365–372.

Hasibuan, H., & Tijow, L. M. (2024). The Government's Role in Enforcing the Law Against Perpetrators of Terrorism. *Journal of Law and Sustainable Development*, *12*(1), 1–16.

Hobart, L. N. (2025). AI, Bias, and National Security Profiling. *Berkeley Technology Law Journal*, *40*(1), 1–165.

ICT Watch. (2024). *9 Rekomendasi ICT Watch untuk Masa Depan AI yang Inklusif dan Bertanggung Jawab [9 ICT Watch Recommendations for an Inclusive and Responsible AI Future]*. https://internetsehat.id/2024/03/31/9-rekomendasi-ict-watch-untuk-masa-depan-ai-yang-inklusif-dan-bertanggung-jawab/

Ilyas, M., & Athwal, R. (2021). De-Radicalisation and Humanitarianism in Indonesia. *Social Sciences*, *10*(87), 1–17. https://doi.org/10.3390/socsci10030087.

Irfan, M., Almeshal, Z. A., & Anwar, M. (2023). Unleashing Transformative Potential of Artificial Intelligence (AI) in Countering Terrorism, Online Radicalisation, Extremism, and Possible Recruitment. *Global Strategic & Security Studies Review*, *8*(4), 1–15. https://doi.org/10.31703/gsssr.2023(VIII-IV).01.

Ismail, N. H. (2024). Countering Online Radicalisation in Southeast Asia Through the 5M Framework. *RSIS Commentary*, *138*. https://www.rsis.edu.sg/wp-content/uploads/2024/09/CO24138.pdf

Ismail, N. H. (2024). Social Media's Dark Side in Online Radicalisation. *RSIS Commentary*, *167*. https://www.rsis.edu.sg/wp-content/uploads/2024/11/CO24167.pdf

Jones, K. (2023). *AI governance and human rights: Resetting the relationship*. Chatham House. https://www.chathamhouse.org/

Kang, E., & Hwang, H.-J. (2023). The Importance of Anonymity and Confidentiality for Conducting Survey Research. *Journal of Research and Publication Ethics*, *4*(1), 1–7. https://doi.org/10.15722/jrpe.4.1.1

Karatsareas, P. (2022). Semi-Structured Interviews. In *Research Methods in Language Attitudes*. Cambridge University Press. https://doi.org/10.1017/9781108864961

Kaskina, R., & Cvetkovska, A. (2024). *Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights*. European Union. https://www.europarl.europa.eu/

Klar, S., & Leeper, T. J. (2019). Identities and Intersectionality: A Case for Purposive Sampling in Survey-Experimental Research. In P. J. Lavrakas, M. W. Traugott, C. Kennedy, A. L. Holbrook, E. D. de Leeuw, & B. T. West (Eds.), *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment* (1st ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119083771.ch10

Klykken, F. H. (2021). Implementing continuous consent in qualitative research. *Qualitative Research*, *22*(5), 795–810. https://doi.org/10.1177/14687941211036399

Land, M. K., & Aronson, J. D. (2020). Human Rights and Technology: New Challenges for Justice and Accountability. *Annual Review of Law and Social Science*, *16*, 223–240.

Lehnert, K., Craft, J., Singh, N., & Park, Y.-H. (2016). The human experience of ethics: A review of a decade of qualitative ethical decision- making research. *Business Ethics: A European Review*, *25*(4), 498–537. https://doi.org/10.1111/beer.12117

Mastrobattista, L., Muñoz-Rico, M., & Cordón-García, J. A. (2024). Optimizing Textual Analysis in Higher Education Studies Through Computer-Assisted Qualitative Data Analysis (CAQDAS) with Atlas.ti. *Journal of Technology and Science Education*, *12*(2), 622–632. https://doi.org/10.3926/jotse.2825

Min, A. (2023). Artificial Intelligence and Bias: Challenges, Implications, and Remedies. *Journal of Social Research*, *2*(11), 3808–3817. https://ejournal.rifainstitute.com/index.php/jsr

Morell, S. (2023). Balancing Standardization and Flexibility: How to Get the Most Out of Your Interviews. *Qualitative and Multi-Method Research*, *21*(2), 39–41. https://doi.org/10.1017/qmm.2023.9

Mosa, M. J., Barhoom, A. M., Alhabbash, M. I., Harara, F. E., Abu-Nasser, B. S., & Abu-Nasser, S. S. (2024). AI and Ethics in Surveillance: Balancing Security and Privacy in a Digital World. *International Journal of Academic Engineering Research*, *8*(10), 8–15.

Mutiara, R. (2025). The Use of Artificial Intelligence in Countering Online Radicalisation in Indonesia. *RSIS Commentary*, *157*.

Ñañez-Silva, M. V., Quispe-Calderón, J., Huallpa-Quispe, P. M., & Larico-Quispe, B. N. (2024). Analysis of academic research data with the use of ATLAS.ti. Experiences of use in the area of Tourism and Hospitality Administration. *Data and Metadata*, *3*(306), 1–13. https://doi.org/10.56294/dm2024306

Nelu, C. (2024). *Exploitation of Generative AI by Terrorist Groups*. International Centre for Counter-Terrorism. https://icct.nl/publication/exploitation-generative-ai-terrorist-groups#:~:text=Regarding%20counter%2Dpropaganda%20and%20de,and%20analysing%20their%20online%20behaviour.

Nuraniyah, N. (2019). *The Evolution of Online Violent Extremism in Indonesia and the Philippines*. Royal United Services Institute for Defence and Security Studies.

Raji, I., & Sholademi, D. B. (2024). Predictive Policing: The Role of AI in Crime Prevention. *International Journal of Computer Applications Technology and Research*, *13*(10), 66–78.

Reddy, L. S., & Ramasamy, K. (2016). Justifying The Judgmental Sampling Matrix Organization in Outsourcing Industry. *Vidushi*, *July-December*, 17–25.

Riyanta, S. (2022). Shortcut to Terrorism: Self-Radicalization and Lone-Wolf Terror Acts: A Case Study of Indonesia. *Journal of Terrorism Study*, *4*(1), 1–20.

Roller, M. R., & Lavrakas, P. J. (2015). *Applied Qualitative Research Design: A Total Quality Framework Approach*. The Guilford Press.

Rulinawaty, R., Samboteng, L., Aripin, S., Basit, M., & Kasmad, M. R. (2024). Impact of artificial intelligence capability and public trust on service performance in Indonesia. *Premiere International Seminar on Engineering, Chemical and Biological*. AIP Conference Proceeding.

Sapkota, M. (2024). Implications and Critiques of Quantitative Research: A Systematic Review. *Journal of Learning Theory and Methodology*, *5*(3), 153–159.

Satria, A. (2022). Two Decades of Counterterrorism in Indonesia. *Counter Terrorist Trends and Analyses*, *14*(5), 7–16.

Scheuble, S., & Oezmen. (2022). *Extremists' Targeting of Young Women on Social Media and Lessons for P/CVE*. Office of the European Union.

Sumarno, & Affianty, D. (2024). The Relationship between Islam and Democracy in Indonesia: Building Harmony in Diversity. *International Journal of Political Sciences*, *10*(1), 21–28.

Sumpter, C. (2024). Decentralising and Coordinating P/CVE through the Indonesia Knowledge Hub (I- KHub). *Counter Terrorist Trends and Analyses*, *16*(4), 10–16.

Tenny, S., Brannan, J. M., & Brannan, G. D. (2024). Qualitative Study. In *StatPearls [Internet]*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/

Tongco, Ma. D. C. (2007). Purposive Sampling as a Tool for Informant Selection. *Ethnobotany Research & Applications*, *5*, 147–158.

Tremblay, M.-A. (1957). The Key Informant Technique: A Nonethnographic Application. *American Anthropologist*, *59*(4), 688–701. https://doi.org/10.1525/aa.1957.59.4.02a00100

UNCRI, & UNCCT. (2021). *Countering Terrorism Online with Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia*. UNICRI and UNCCT.

Ünver, H. A. (2024). *Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights*. European Union.

Walford, G. (2005). Research ethical guidelines and anonymity. *International Journal of Research & Method in Education*, *28*(1), 83–93. https://doi.org/10.1080/1743727052000327826

Widya, B. (2020). Deradicalization in Indonesia: Implementation and Challenge. *Journal of Terrorism Studies*, *2*(1), 32–50.

Zeiger, S., & Gyte, J. (2021). Prevention of Radicalization on Social Media and the Internet. In A. P. Schmid (Ed.), *Handbook of Terrorism Prevention and Preparedness*. International Centre for Counter-Terrorism.