Vol. 01 /No. 01 /2026

DEEPFAKE IN THE PERSPECTIVE OF HYBRID THREATS TO INDONESIA'S NATIONAL RESILIENCE

Bambang Raditya*[1], Erfanda Andi Mada Arectya[2], Kenneth Sulthon Alafi Al Hallaj [3], Eka Firmansyah[4], Uyun Khikmata Muhandis[5]

[1]Komando Distrik Militer 0830 Surabaya, Indonesia, bambangraditya@yahoo.com

[2]Universitas Muhammadiyah Surabaya, Indonesia, erfandaandi99@gmail.com

[3]Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia, elaphy07@gmail.com

[4]Universitas Muhammadiyah Palu, Indonesia, ekafirmansyah689@gmail.com

[5]Universitas Negeri Surabaya, Indonesia, uyun.22082@mhs.unesa.ac.id

*coresponding author: bambangraditya@yahoo.com

ABSTRACT: Rapid advancements in deepfake technology have enabled the creation of highly convincing synthetic media, transforming it into a disinformation tool that threatens cybersecurity, public trust, and national resilience. This study analyzes deepfake through the lens of hybrid threats and its implications for Indonesia's national resilience. Utilizing a qualitative systematic literature review with thematic content analysis, the study synthesized twenty one scholarly publications (2019-2024) from databases such as Scopus and IEEE Xplore. The results indicate that deepfake embodies a hybrid threat by combining information manipulation, technological exploitation, and psychological operations without military confrontation. Its impacts span ideological, political, social, economic, and defense security dimensions, notably through eroding public trust, intensifying social polarization, and weakening institutional legitimacy. Mitigating this threat requires an integrated strategy encompassing AI governance, regulatory adaptation, digital literacy, and multi stakeholder collaboration. Ultimately, this study contributes a conceptual framework linking synthetic media technology with hybrid threat strategic studies, which is highly relevant to the Indonesian context.

Keywords: *Deepfake, Hybrid Threats, National Resilience, Artificial Intelligence*

Received: 2026-May-11
Rev. Req: 2026-June-09
Accepted: 2026-June-22

I. INTRODUCTION

The rapid progress of artificial intelligence over the past decade has produced a class of synthetic media technologies commonly known as deepfake, capable of generating audio, image, and video content whose level of realism makes it increasingly difficult to separate from authentic recordings (Altuncu et al., 2024) [1]. What began as an experimental application of generative modeling has matured into an accessible toolkit: contemporary deepfake pipelines

require comparatively little technical expertise, computing cost has fallen, and open source generative architectures are now widely circulated, which together have accelerated both the volume and the sophistication of synthetic content in circulation. This trajectory illustrates that deepfake is no longer a narrow technical curiosity confined to computer vision laboratories, but a socio technical phenomenon with tangible consequences for the information environments that societies depend upon.

The misuse of deepfake is no longer confined to entertainment or satire; it has developed into an instrument of disinformation, identity fraud, and information manipulation with the capacity to erode public trust and to compromise the integrity of digital ecosystems (Fernández Gambín et al., 2024) [2]. Fabricated statements attributed to public officials, manipulated evidentiary material, and synthetic impersonation used for financial fraud each demonstrate how the same underlying technology can be redirected from creative or commercial use toward coercive or deceptive ends. Because these harms often surface only after synthetic content has already circulated widely, deepfake magnifies the classic asymmetry between the speed of falsehood and the slower pace of verification, a dynamic that has direct bearing on national security once it is deployed deliberately and at scale.

Security scholarship describes this broader shift as the transformation of threats to state security from the direct application of military force toward a combination of military and non military instruments, a configuration generally referred to as hybrid threats (Giannopoulos et al., 2021) [3]. A hybrid threat is deliberately designed to operate beneath the threshold of open armed conflict, blending cyberattacks, disinformation campaigns, economic pressure, and manipulation of public opinion so that an adversary's objectives can be pursued while formal attribution and conventional deterrence remain difficult to invoke (Bergaust & Sellevåg, 2023) [4]. Within such a configuration, information itself becomes a contested domain, and any technology capable of shaping what a population perceives to be true acquires strategic relevance independent of its original design intent. Deepfake, by virtue of its capacity to fabricate convincing evidence of events that never occurred, fits squarely within this category of instruments and therefore warrants sustained attention from the standpoint of state security rather than technology policy alone.

For Indonesia, the relevant analytical frame is national resilience (*ketahanan nasional*), understood as the state's capacity to safeguard stability across the ideological, political, economic, social-cultural, and defense security dimensions against a diverse array of threats, including those originating in cyberspace and the information domain (Ikhsanni et al., 2024) [5]. National resilience in this sense is not limited to the physical protection of territory or institutions; it also encompasses the cohesion of public trust, the legitimacy of governance processes, and the capacity of citizens to make informed judgments in an increasingly mediated public sphere. The dissemination of deepfake based falsehoods has the potential to erode public confidence in state institutions, deepen social polarization, and distort public policy deliberation whenever such content is not identified and countered in a timely manner (Deppe & Schaal, 2024) [6]. Under conditions of political contestation or national crisis, when public attention is heightened and verification capacity is often stretched thin, this vulnerability

becomes especially acute, which underscores the urgency of strengthening national resilience against artificial intelligence enabled threats.

Existing scholarship on deepfake nevertheless remains dominated by technical concerns, particularly the mechanisms of content generation and the corresponding methods for automated detection (Abbas & Taeihagh, 2024) [7]. Studies that explicitly connect deepfake with the conceptual apparatus of hybrid threats, and that trace its implications for Indonesia's national resilience specifically, remain comparatively limited, leaving a gap between the rapidly evolving technical literature and the strategic studies literature that is tasked with interpreting such developments for policy purposes. Much of the available detection oriented literature treats deepfake primarily as an engineering problem to be solved through improved classifiers and larger training datasets, an emphasis that, while valuable, tends to bracket out the political, social, and institutional consequences that unfold once synthetic content escapes the laboratory and enters a contested public sphere. Bridging this gap is not merely an academic exercise: the pace at which generative AI capability continues to improve suggests that the strategic significance of deepfake is likely to grow rather than diminish, which makes early conceptual clarity a prerequisite for coherent policy design rather than an optional refinement.

This study addresses that gap through a qualitative library research approach, and is intended to analyze deepfake from the perspective of hybrid threats to Indonesia's national resilience so as to contribute a conceptual foundation for the further development of defense studies scholarship and to inform the design of policy responses appropriate to the Indonesian context. Three questions guide the analysis that follows: first, which characteristics of deepfake justify its classification as an instrument of hybrid threats; second, across which dimensions of national resilience are the consequences of deepfake most acutely felt; and third, what combination of governance, regulatory, and societal measures is most consistent with the existing literature on countering synthetic media risk. By organizing the discussion around these three questions, this study aims to move beyond a purely descriptive account of deepfake technology toward an analytically structured contribution that strategic studies scholars and policymakers alike can draw upon.

II. METHOD

This study employed a qualitative approach using the library research method, which concentrates analytical attention on the synthesis of scholarly publications concerning deepfake, hybrid threats, and national resilience rather than on the collection of primary field data (Snyder, 2019) [8]. Library research of this kind is well suited to a topic whose empirical manifestations are widely dispersed across disciplines, since it allows the researcher to construct a coherent conceptual framework through the systematic identification, evaluation, and integration of previously published findings (Snyder, 2019) [8]. This methodological choice reflects the exploratory conceptual nature of the research problem, in which the primary task is to connect two bodies of knowledge, namely the technical literature on synthetic media and the strategic studies literature on hybrid threats and national resilience, that have so far developed largely in isolation from one another.

The search and selection of literature followed procedures adapted from the PRISMA 2020 reporting guideline, which structures a review into successive stages of identification, screening, eligibility assessment, and inclusion so that the process of literature selection remains transparent and reproducible (Page et al., 2021) [9]. Consistent with this guideline, the flow of records was documented at each stage using the logic embedded in the PRISMA2020 flow diagram framework, which was designed to make the reporting of systematic reviews more consistent and auditable (Haddaway et al., 2022) [10]. Systematic search procedures of this type were selected because they reduce the risk of selective or unrepresentative citation, a risk that is particularly salient for a rapidly evolving topic such as artificial intelligence enabled disinformation, where the literature published in any given year can differ substantially from that of the year before (Xiao & Watson, 2019) [11].

The data sources consisted of national and international journal articles, conference proceedings, academic books, policy documents, and reports issued by international organizations, provided that each source bore a demonstrable relevance to deepfake, hybrid threats, or national resilience. The search was conducted across the Scopus, ScienceDirect, SpringerLink, Taylor & Francis Online, IEEE Xplore, Google Scholar, and Garuda databases using combinations of the keywords "deepfake," "synthetic media," "hybrid threat," "hybrid interference," "artificial intelligence," and "national resilience." The literature search was limited to publications from 2019 through 2024 in order to capture the period in which generative AI capability, and consequently deepfake related risk, expanded most rapidly, while still allowing inclusion of foundational conceptual sources published slightly earlier when these remained the most authoritative statement of a given concept.

Records retrieved through the initial search were screened in three stages. At the identification stage, duplicate records and sources without any demonstrable relevance to the research focus were removed. At the screening stage, titles and abstracts were reviewed against the inclusion criteria, namely thematic relevance to deepfake, hybrid threats, or national resilience; publication in a peer reviewed or otherwise credible outlet; and availability of the full text. At the eligibility stage, the remaining full texts were assessed for methodological transparency and for the substantive contribution each source could make to the conceptual synthesis, following the extraction and presentation logic recommended for reviews of this kind (Pollock et al., 2023) [12]. Twenty one sources satisfied all inclusion criteria and were retained as the analytical corpus for this study; sources without a traceable digital object identifier, or whose bibliographic details could not be independently verified, were excluded at this stage to preserve the accuracy of citation.

The retained literature was analyzed using qualitative thematic content analysis, following the coding, categorization, interpretation, and synthesis procedure through which recurring patterns of meaning are identified across a body of text (Braun & Clarke, 2006) [13]. Coding began with open, line by line labeling of statements concerning the technical characteristics of deepfake, the conceptual features of hybrid threats, and the dimensions of national resilience; related codes were then grouped into broader categories, and categories were iteratively compared across sources until three stable analytical themes emerged,

corresponding to the three subsections presented in the following section. This technique was applied to trace the conceptual linkage between deepfake technology development, the defining characteristics of hybrid threats, and the corresponding implications for Indonesia's national resilience, thereby producing a systematic conceptual construction rather than a purely descriptive summary of the literature.

To strengthen the trustworthiness of this qualitative synthesis, coding decisions were cross-checked against the original source text on at least two occasions, and any statement whose interpretation remained ambiguous after re-reading was either discarded or coded conservatively rather than forced into a predetermined category. This study is nevertheless subject to certain limitations inherent in the library research design. Because the analysis relies exclusively on secondary, previously published material, it cannot independently verify the empirical claims made within each source, and the conceptual synthesis presented here should accordingly be read as a structured interpretation of the existing literature rather than as a report of new primary evidence. In addition, restricting the search to English and Indonesian language sources indexed in the databases listed above may have excluded relevant work published in other languages or through informal channels, a limitation common to library research of this kind and one that future studies incorporating primary or comparative data collection could help to address.

III. RESULT AND DISCUSSION

The Characteristics of Deepfake as a Hybrid Threat

The development of deepfake technology shows that AI based synthetic media has evolved from simple visual manipulation into a capability that can produce audio, image, and video representations of high realism, making such content increasingly difficult to distinguish from authentic material (Altuncu et al., 2024) [1]. This progression has been driven in large part by advances in generative modeling architectures, which have simultaneously improved output fidelity and lowered the barrier to entry for content creation, so that the production of convincing synthetic media no longer requires the specialized expertise it once did (Alanazi & Asif, 2024) [14]. The convergence of these two trends, rising realism and falling technical cost, marks the point at which deepfake shifted from a laboratory curiosity into a technology with genuine strategic value in the digital security environment.

The threat character of deepfake does not rest solely on its capacity to falsify an individual's identity, it extends to its capacity to shape public perception by circulating information that appears credible within digital spaces (Sandoval et al., 2024) [15]. This dual capacity is what positions deepfake within the category of hybrid threats, because it is able to combine information manipulation, exploitation of digital technology, and psychological operation in order to influence the social and political stability of a state without resorting to direct armed confrontation (Bergaust & Sellevåg, 2023) [4]. Consequently, the primary target of deepfake enabled operations is not confined to information systems in the technical sense, it also encompasses the cognitive space of a society, that is, the collective judgments and perceptions through which citizens interpret political and social reality.

The use of deepfake in contemporary conflict settings demonstrates that the technology is increasingly deployed as a vehicle for disinformation, identity concealment, digital fraud, and influence operations directed at public opinion (Shoaib et al., 2023) [16]. Its integration with social media platforms accelerates the circulation of synthetic content, thereby expanding the window during which a false perception can take hold before verification by the public or by relevant authorities becomes possible. This dynamic illustrates that deepfake displays characteristics consistent with the multidimensional threat pattern that defines the hybrid threat concept, in which no single countermeasure, whether technical, legal, or societal, is sufficient on its own.

The threat character of deepfake becomes further compounded because it is able to exploit technological weaknesses, user behavior, and low digital literacy simultaneously, so that the resulting impact is not merely technical but also social, political, economic, and security related in nature (Seow et al., 2022) [17]. This compounding effect indicates that deepfake can no longer be positioned solely as a cybersecurity issue, it should instead be treated as part of a strategic threat that requires a multidisciplinary approach within the framework of national resilience (Abbas & Taeihagh, 2024) [7]. These characteristics form the analytical foundation for examining the implications of deepfake for Indonesia's national resilience in the following subsection.

Implications of Deepfake for Indonesia's National Resilience

The development of deepfake technology has broadened the spectrum of threats to information security, because synthetic media is capable of producing visual and audio representations that are difficult to distinguish from authentic information, thereby increasing the complexity of information verification in the digital space (Fernández Gambín et al., 2024) [2]. This characteristic means that the dissemination of false information not only spreads more rapidly, but is also more readily accepted by the public because it is supported by visual and audio quality that closely resembles genuine conditions. This shift demonstrates that deepfake has become a threat to the information systems that underpin national resilience, since a resilient state ultimately depends on a public sphere in which citizens can distinguish credible information from fabricated content.

The political dimension experiences a marked increase in vulnerability once deepfake is deployed to construct narratives that shape public perception of political actors, state institutions, or the democratic process itself (Deppe & Schaal, 2024) [6]. Artificial intelligence based information manipulation has the potential to diminish public trust in official information, thereby widening the space in which disinformation and social polarization can flourish, particularly during periods of political contestation or national crisis. This condition demonstrates that the information space has become one of the strategic elements in maintaining national stability, since a decline in trust toward formal information channels tends to be filled by unverified or deliberately fabricated alternatives.

The implication for the social dimension is visible in the growing difficulty citizens face in distinguishing authentic from synthetic information as the quality of deepfake increasingly

resembles genuine content (Ma et al., 2024) [18]. A diminished capacity for information verification can accelerate the spread of hoaxes, reinforce social polarization, and shape collective perceptions that diverge from factual reality, thereby affecting social cohesion as one of the essential elements of national resilience. This dynamic shows that the threat posed by deepfake is determined not only by the sophistication of the underlying technology, but also by the level of digital literacy present within society, since the same synthetic content will have markedly different effects depending on the audience's capacity for critical evaluation.

The economic dimension is affected in a comparatively less visible but no less consequential manner, since deepfake enabled fraud, including synthetic voice or video impersonation used to authorize fraudulent financial transactions or to damage the reputation of corporate actors, imposes direct material costs while also eroding the confidence that underpins digital commerce and financial technology adoption (Ma et al., 2024) [18]. Where public confidence in digital transactions declines, the broader digital economy, which Indonesia has identified as a strategic growth sector, may see its expansion constrained by a fundamentally information security problem rather than a purely economic one, illustrating how the boundaries between the economic and information dimensions of national resilience have become increasingly porous in the AI era.

The defense and security dimension confronts new challenges once deepfake is employed as part of information operations that support digital propaganda, identity manipulation, or attacks on the credibility of state institutions (Ikhsanni et al., 2024) [5]. The complexity of this threat indicates that strengthening national resilience requires the integration of synthetic-media detection capability, stronger information governance, and enhanced cybersecurity capacity in order to reduce the impact of AI based information operations, in line with the broader conceptual understanding of hybrid threats as instruments deliberately designed to remain below the threshold of armed conflict while still producing strategically significant effects (Giannopoulos et al., 2021) [3]. Taken together, these implications show that deepfake has developed into a multidimensional threat that requires a cross-sectoral response within the framework of national resilience.

Strategies for Strengthening National Resilience against Deepfake Threats

Strengthening national resilience against deepfake requires an artificial intelligence governance framework capable of integrating regulation, accountability, transparency, and risk mitigation into the management of synthetic media (Novelli et al., 2024) [19]. The development of generative technology indicates that policy effectiveness is determined not only by the capacity to identify manipulated content, but also by oversight mechanisms able to ensure that all relevant stakeholders comply with digital security standards. This dynamic shows that technology governance is an integral part of strengthening national resilience, since regulation designed without corresponding institutional capacity for enforcement is unlikely to keep pace with the speed at which generative technology continues to evolve.

Strengthening regulation is a critical component given that deepfake development proceeds

more rapidly than the adaptation of legal instruments governing the misuse of artificial intelligence in the digital space (Broinowski & Martin, 2024) [20]. The absence of instruments specifically tailored to synthetic media widens the space in which such technology can be misused for disinformation, political manipulation, or attacks on the legitimacy of state institutions. This condition demonstrates that synchronizing regulatory frameworks with the pace of technological development constitutes a precondition for strengthening national resilience, particularly in a jurisdiction such as Indonesia where digital governance instruments remain under active development.

National resilience is likewise shaped by society's capacity to recognize the characteristics of digital information, which positions digital literacy as a strategic instrument for reducing the impact of deepfake dissemination (Alanazi & Asif, 2024) [14]. Improving the public's ability to verify information and to identify AI manipulated content contributes to the formation of an information space that is more resilient to AI based disinformation. This condition indicates that strengthening societal capacity is an inseparable part of any strategy for confronting hybrid threats, since technical detection tools alone cannot substitute for a public capable of exercising informed skepticism toward the content it encounters.

Strengthening national resilience requires collaboration among government, industry, academia, digital-platform providers, and the public in building mechanisms of oversight and accountability for the use of artificial intelligence (Passos et al., 2024) [21]. Such a collaborative approach enables the development of ethical standards, synthetic-media labeling mechanisms, and stronger accountability among technology developers, so that the risk of deepfake misuse can be minimized in a more systematic manner (Sandoval et al., 2024) [15]. Detection research further suggests that technical countermeasures alone, however accurate under laboratory conditions, tend to lag behind the pace at which new generative techniques emerge, which reinforces the case for treating detection technology as one component of a wider resilience strategy rather than as a stand alone solution (Passos et al., 2024) [21].

In the Indonesian context specifically, this multidimensional strategy implies concrete institutional tasks: harmonizing existing electronic information and data-protection regulations with emerging synthetic media risks, building detection and verification capacity within relevant government bodies and media organizations, embedding digital and AI literacy content within formal education pathways, and establishing coordination channels between the state, digital platforms, and civil society organizations so that emerging deepfake incidents can be identified and addressed before they escalate into broader crises of public trust. Taken together, this analysis indicates that strengthening national resilience against deepfake threats requires the integration of technology governance, regulation, digital literacy, and cross sectoral collaboration as a single, sustained strategic effort, rather than as a set of isolated interventions pursued independently of one another.

IV. CONCLUSION

The development of deepfake technology has transformed the character of digital threats

into something considerably more complex through its capacity to generate manipulated audio, image, and video content that is difficult to distinguish from authentic information. This characteristic positions deepfake within the category of hybrid threats, since it exploits the information space to shape public perception, manufacture disinformation, and disrupt stability without resorting directly to military force. The status of deepfake as a non conventional threat demonstrates that the advance of digital technology has, in turn, broadened the spectrum of threats to Indonesia's national resilience.

The implications of deepfake for national resilience are not confined to cybersecurity, they also extend to the ideological, political, social, economic, and defense-security dimensions through the dissemination of manipulative information capable of eroding public trust and reinforcing societal polarization. This condition shows that a vulnerability located within the information space can evolve into a strategic vulnerability if it is not matched by adequate capability in detection, verification, and information management. This dynamic reaffirms that national resilience in the digital era is heavily shaped by the capacity of both the state and society to confront artificial intelligence based threats.

Strengthening national resilience against deepfake threats requires a multidimensional approach that integrates artificial intelligence governance, regulation, cybersecurity, digital literacy, and collaboration among stakeholders. The integration of these components builds a national capacity that is more adaptive to the evolving character of hybrid threats while simultaneously enhancing the resilience of the national information space. This study demonstrates that the management of deepfake related risk can no longer be regarded as a purely technological issue, it must instead be understood as part of a broader strategy for sustaining Indonesia's national resilience amid the transformation of the global security environment. Future research would benefit from complementing this conceptual synthesis with empirical case studies of deepfake incidents in the Indonesian context, so as to test and refine the framework proposed here.

V. REFERENCES

[1] Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: Definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, 7, Article 1400024, pp. 1-23. https://doi.org/10.3389/fdata.2024.1400024

[2] Fernández Gambín, Á., Yazidi, A., Vasilakos, A. V., Frías-Martínez, V., & Haugerud, H. (2024). Deepfakes: Current and future trends. Artificial Intelligence Review, 57(3), *Article 64*, pp. 1-30. https://doi.org/10.1007/s10462-023-10679-x

[3] Giannopoulos, G., Smith, H., & Theocharidou, M. (2021). The landscape of hybrid threats: A conceptual model (EUR 30585 EN). *Publications Office of the European Union*. https://doi.org/10.2760/44985

[4] Bergaust, J. C., & Sellevåg, S. R. (2023). Improved conceptualising of hybrid interference below the threshold of armed conflict. *Cambridge Review of International Affairs*, pp. 169-195. https://doi.org/10.1080/09662839.2023.2267478

[5] Ikhsanni, A., Prasetyo, R., & Wibowo, S. (2024). Cybersecurity dan tata kelola intelijen. *Jurnal Kajian Stratejik Ketahanan Nasional*, 7(1), pp. 1-14. https://doi.org/10.7454/jkskn.v7i1.10086

[6] Deppe, K., & Schaal, G. S. (2024). Deepfakes, cognitive warfare, and democratic resilience. *Frontiers in Political Science*, 6, pp. 1-13. https://doi.org/10.3389/fpos.2024.1392658

[7] Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. *Expert Systems with Applications*, 252, 124260, pp. 1-22. https://doi.org/10.1016/j.eswa.2024.124260

[8] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, pp. 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

[9] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71, pp. 1-9. https://doi.org/10.1136/bmj.n71

[10] Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams. *Campbell Systematic Reviews*, 18(2), e1230, pp. 1-6. https://doi.org/10.1002/cl2.1230

[11] Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1), pp. 93-112. https://doi.org/10.1177/0739456X17723971

[12] Pollock, D., Peters, M. D. J., Khalil, H., McInerney, P., Alexander, L., Tricco, A. C., Evans, C., de Moraes, É. B., & Godfrey, C. M. (2023). Recommendations for the extraction, analysis, and presentation of results in scoping reviews. *JBI Evidence Synthesis*, 21(3), pp. 520-532. https://doi.org/10.11124/JBIES-22-00123

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

[14] Alanazi, S., & Asif, S. (2024). Exploring deepfake technology: Creation, consequences and countermeasures. *Human-Intelligent Systems Integration*, 6, pp. 49-60. https://doi.org/10.1007/s42454-024-00054-8

[15] Sandoval, M. P., de Almeida Vau, M., Solaas, J., et al. (2024). Threat of deepfakes to the criminal justice system: A systematic review. Crime Science, 13, *Article 41*, pp. 1-19. https://doi.org/10.1186/s40163-024-00239-1

[16] Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. *arXiv*, pp. 1-15. https://doi.org/10.48550/arXiv.2311.17394

[17] Seow, J. W., Lim, M. K., Phan, R. C. W., & Liu, J. K. (2022). A comprehensive overview of deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing*, 513, pp. 351-371. https://doi.org/10.1016/j.neucom.2022.09.135

[18] Ma, Y., Zhang, L., Chen, H., & Wang, X. (2024). Public perception of AI-generated misinformation and deepfake content in social media. *Telematics and Informatics Reports*, 14, 100138, pp. 1-11. https://doi.org/10.1016/j.teler.2024.100138

[19] Novelli, C., Casolari, F., Taddeo, M., & Floridi, L. (2024). AI governance in the era of generative artificial intelligence: Regulatory perspectives and emerging challenges. Philosophy & Technology, 37, *Article 107*, pp. 1-25. https://doi.org/10.1007/s13347-024-00727-9

[20] Broinowski, A., & Martin, A. (2024). Governing generative artificial intelligence and synthetic media: Policy challenges and regulatory responses. *AI and Ethics*, pp. 1-14. https://doi.org/10.1007/s43681-024-00542-2

[21] Passos, L. A., Jodas, D. S., Costa, K. A. P., Eskelinen, P., Papa, J. P., & Ferrer, M. A. (2024). A review of deep learning-based approaches for deepfake content detection. *Expert Systems*, 41(8), e13570, pp. 1-25. https://doi.org/10.1111/exsy.13570