

Deepfakes and Disinformation: New Threats to Democracy and Information Security in Indonesia

**Endang Fatmawati^{1*}, Abd Muhni Salam², Ratna Susanti³, Jumroni Sulaeman⁴,
Minan Faiz Fausta Rafa¹**

¹Universitas Diponegoro, Indonesia

²Universitas Brawijaya, Indonesia

³Politeknik Indonusa Surakarta, Indonesia

⁴UIN Jakarta, Indonesia

*Corresponding Author Email: endangfatmawati@lecturer.undip.ac.id

Abstract

This study examines deepfake enabled disinformation as a hybrid threat to democracy and information security in Indonesia. The study synthesizes 30 academic studies and 12 policy or institutional reports published between 2019 and 2026 using a Systematic Literature Review (SLR) guided by PRISMA 2020, combining thematic synthesis and policy document analysis. The findings identify five interrelated themes: technological affordances of synthetic media, selective belief and partisan amplification, trust erosion and democratic legitimacy, cyber enabled information security risks, and regulatory gaps in Indonesia. Deepfake harm does not stem solely from technical realism but emerges through interactions with political bias, weak verification literacy, platform algorithms, institutional distrust, and fragmented governance. In electoral settings, synthetic content can distort public opinion, damage candidates, weaken confidence in election bodies, and create the liar's dividend, in which authentic evidence may be dismissed as fabricated. Its security implications include identity impersonation, social engineering, institutional spoofing, and evidence manipulation. The study proposes the Democratic Information Security Resilience Framework, connecting threat actors, content forms, amplification pathways, vulnerability conditions, impact mechanisms, and resilience strategies. The findings highlight the need for coordinated governance involving technical detection, content provenance, platform accountability, electoral safeguards, data protection, public literacy, rapid institutional response, and democratic accountability.

Keywords: Deepfake, Disinformation, Democracy, Information Security, Indonesia.

INTRODUCTION

The rapid expansion of generative artificial intelligence has altered digital information production. Manipulated content that once depended largely on basic editing, selective

* Copyright (c) 2026 **Endang Fatmawati et.al**

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Received: July 1, 2026; Revised: August 30, 2026; Accepted: September 19, 2026

framing, or textual distortion can now be produced through synthetic audio, video, and image generation with realism that challenges ordinary public verification. Within synthetic media, deepfakes refer to digitally generated or manipulated audiovisual content that realistically depicts a real person saying or doing something that did not occur. Deepfakes are consequential because they can fabricate faces, voices, gestures, and bodily expressions in ways that imitate authentic audiovisual records. The political risk therefore arises not only from false content but from fabricated media appearing as evidence of real events or statements, creating uncertainty concerning the authenticity and provenance of audiovisual information (Łabuz & Nehring, 2024; Salvaudon, 2026; Shoaib et al., 2023).

Scholarly debates on deepfakes have moved beyond technical concerns about generation and detection. Early discussions emphasized threats to privacy, reputation, democratic deliberation, and security because malicious actors can manufacture seemingly credible evidence at scale (Chesney & Citron, 2019). Subsequent research shows that deepfakes do not necessarily deceive audiences; their effects can also involve uncertainty, skepticism, and declining confidence in digital information (Vaccari & Chadwick, 2020). Political influence is therefore shaped by technological realism and audience interpretation, including partisan alignment, prior beliefs, social networks, and confirmation bias (Langa, 2021). These consequences may persist when audiences become less confident in distinguishing authentic evidence from manipulation or when political actors invoke synthetic media to dismiss genuine evidence as fabricated.

The Indonesian context provides an important setting for examining these risks. Indonesia has a large electoral democracy and a highly digitalized communication environment in which political information circulates through social media and messaging applications. Previous research has documented hoaxes, polarizing narratives, identity based political communication, and coordinated online influence. During electoral periods, political messages can circulate rapidly across TikTok, Instagram, Facebook, X, and YouTube, as well as through closed interpersonal and group-based channels. These conditions can reduce the time available for verification and increase the visibility of emotionally salient political content. They do not establish specific democratic harms caused by deepfakes in Indonesia, but provide a basis for examining how deepfake enabled disinformation could interact with

existing patterns of political communication, partisan interpretation, and information circulation.

The 2024 Indonesian election is particularly relevant because AI generated and manipulated political content became increasingly visible during electoral communication. Reports and studies have identified the use or circulation of AI generated or manipulated material involving political figures. However, evidence concerning these incidents must be distinguished from broader international research and claims about effects on Indonesian voters. The existence of manipulated content demonstrates a potential information security and democratic integrity problem, but does not establish that such content deceived large numbers of voters or altered electoral outcomes. Deepfakes can also be situated within information disorder. Misinformation generally refers to false information circulated without an intention to cause harm, whereas disinformation involves deliberate creation or dissemination of false or misleading information for deceptive or harmful purposes; malinformation concerns genuine information used in a harmful context (Islam et al., 2024). A deepfake describes a technological form of manipulated or synthetic media, while disinformation describes communicative intention and use. A deepfake may constitute disinformation when deliberately produced or distributed to deceive, discredit, manipulate, or influence public perception, while the same material may be misinformation when circulated by users unaware that it is synthetic or manipulated.

The risk becomes more significant when deepfake enabled disinformation is examined through information security. Information security concerns protecting information and information systems, particularly their confidentiality, integrity, and availability. Deepfakes are most directly relevant to integrity and authenticity because they can falsify the apparent source, content, or context of communication. Synthetic impersonation may fabricate statements by public officials, manipulate evidence, facilitate social engineering, damage reputations, or create false impressions of institutional action. Their information security significance therefore lies not only in influencing public opinion but also in undermining confidence in the authenticity, provenance, and accountability of digital information (Pawelec, 2022; Whyte, 2020).

Indonesia's regulatory environment has begun to acknowledge broader ethical and governance concerns associated with artificial intelligence through the Minister of Communication and Informatics Circular Letter Number 9 of 2023 concerning Artificial Intelligence Ethics. The document establishes guidance for AI development and implementation, including transparency, security, humanity, credibility, and accountability. Nevertheless, general AI ethics guidance does not necessarily provide specific mechanisms for politically consequential deepfakes. Electoral deepfakes raise questions concerning synthetic content labeling, evidence provenance, removal procedures, platform responsibilities, political advertising transparency, data protection, and coordination among electoral institutions, regulators, cybersecurity authorities, media organizations, fact checkers, and civil society (Folorunsho & Boamah, 2025; Paterson & Hanley, 2020). Existing literature addresses these issues from several disciplinary perspectives. Technical studies focus on generation, detection models, datasets, and forensic performance (Michałkiewicz Kądziela, 2024; Ray, 2021), while political communication research examines disinformation, selective exposure, motivated reasoning, polarization, and trust in news (Repez & Popescu, 2020; Zalewski, 2022). Legal and policy scholarship considers platform regulation, data protection, AI ethics, and cybersecurity governance. An integrated analysis is therefore necessary to examine deepfakes as manipulated information objects, vehicles for disinformation, challenges to identity and content authentication, and potential sources of institutional distrust.

A further issue concerns the applicability of predominantly international evidence to democratic contexts in the Global South. Much of the established deepfake literature has developed from research and policy debates in North America and Europe, where political institutions, media systems, platform governance, and regulatory capacities may differ from those in Southeast Asia. Their relevance to Indonesia therefore needs to be examined in relation to the country's institutional and communication environment. Features such as platform dependent political communication, identity-based mobilization, uneven institutional capacity, and existing patterns of online political influence may affect how synthetic political content is produced, interpreted, circulated, contested, and denied. International findings should therefore be treated as comparative or theoretical evidence unless supported by Indonesian empirical or institutional sources.

Against this background, the central research problem is not simply whether deepfakes can deceive Indonesian citizens, but how deepfake enabled disinformation may interact with political communication, information integrity, institutional trust, and governance arrangements to create vulnerabilities for democratic information environments. This article therefore examines deepfake enabled disinformation as a potential dual threat to democratic integrity and information security in Indonesia. The analysis considers interactions among content fabrication, audience interpretation, platform amplification, institutional vulnerability, and regulatory responses. Four concerns guide the analysis: the relationship between deepfakes, disinformation, democracy, and information security; conditions under which synthetic political content can create risks for electoral and democratic integrity; the extent to which existing legal and policy instruments address those risks; and the conceptual components required to strengthen democratic information security resilience.

The article makes three principal contributions. First, it integrates perspectives on information disorder, audience interpretation, epistemic uncertainty, and information security to explain why deepfakes extend beyond direct deception toward uncertainty, distrust, and challenges to the authenticity of digital evidence. Second, it applies these perspectives to Indonesia as a Global South democratic context while distinguishing Indonesian evidence from international findings. Third, it develops the Democratic Information Security Resilience framework by synthesizing literature and policy evidence around technical detection, content provenance, public literacy, platform accountability, electoral safeguards, data protection, institutional communication, and AI governance. The framework is presented as a conceptual and theoretically informed policy framework rather than an empirically validated governance model. Its purpose is to connect technical, political, and regulatory responses around protecting the informational conditions necessary for citizens to evaluate political claims, maintain confidence in legitimate evidence, and participate meaningfully in democratic life.

LITERATURE REVIEW

Deepfake as Synthetic Media

Deepfake refers to AI generated or AI manipulated audiovisual content that realistically imitates a person's face, voice, gestures, or expressions. Unlike conventional edited content,

deepfakes commonly rely on deep learning techniques to generate or modify media in ways that can make fabricated representations appear authentic to ordinary viewers (Emeç, 2025; Kumar et al., 2024). Deepfakes should therefore be understood as a specific form of synthetic media, particularly associated with the realistic manipulation or synthesis of human likeness and behavior. Their significance lies in their capacity to imitate the evidentiary qualities of authentic audiovisual material, creating uncertainty about whether an apparent record accurately represents a real person, statement, or event.

The literature distinguishes deepfakes from cheapfakes and the broader category of synthetic media. Cheapfakes generally involve relatively simple manipulation, such as cutting or rearranging video segments, altering captions, slowing or accelerating footage, or presenting genuine material in a misleading context. Synthetic media represents the broader category and may include AI generated or AI manipulated images, audio, video, avatars, and other forms of digitally produced content. Chesney and Citron (2019) argue that democratic harm does not necessarily depend on sophisticated technological manipulation, as relatively simple forms of falsification can also influence public perception when circulated through politically polarized networks. A highly realistic deepfake may therefore fail to persuade its audience, while a low quality manipulation may still circulate effectively when it confirms existing political beliefs or is promoted by trusted members of a social network.

Deepfake technology also has legitimate applications in education, entertainment, accessibility, and creative production. The principal concern arises when synthetic media is produced or used without appropriate consent or disclosure to deceive, defame, impersonate, intimidate, or manipulate public perception. Chesney and Citron (2019) identify potential risks to privacy, reputation, democratic discourse, and national security associated with malicious uses of deepfake technology. These risks demonstrate that the consequences of deepfakes depend on their purpose, context, and mode of dissemination rather than on synthetic generation itself.

Disinformation, Misinformation, and Malinformation

Deepfakes also need to be distinguished from the different forms of information disorder within which manipulated content may circulate. Schick (2020) distinguishes

misinformation, disinformation, and malinformation according to the nature and intention of problematic information. Misinformation generally refers to false or inaccurate information shared without an intention to deceive or cause harm, while disinformation involves the deliberate creation or dissemination of false or misleading information for deceptive or harmful purposes. Malinformation involves genuine information that is used, presented, or disclosed in a harmful or manipulative manner. These distinctions are relevant to deepfake research because the technological form of a message and the intention behind its production are analytically separate dimensions.

A deepfake can constitute disinformation when it is deliberately created or distributed to mislead voters, damage a candidate's reputation, manipulate public perception, or undermine confidence in an institution. The same deepfake may subsequently circulate as misinformation when users redistribute it without recognizing that it has been synthetically generated or manipulated. This distinction affects how responsibility and governance should be considered across different actors in the information chain, including producers, intentional amplifiers, and users who unknowingly share manipulated content.

The distinction between information disorder and the broader label of "fake news" further reinforces the need for conceptual precision. Donnelly and Wilson (2026) caution that the term fake news can encompass different forms of falsehood and manipulation, including fabrication, propaganda, satire, and misleading framing. Not every synthetic or manipulated item therefore constitutes disinformation, and not every deepfake has the same communicative purpose. A clearly identified satirical AI generated video, for example, differs substantially from a fabricated political speech presented as authentic evidence during an election. Deepfake governance consequently requires attention to the technological characteristics of synthetic content, the intention behind its use, and the context in which audiences encounter and interpret it.

Deepfake and Democratic Vulnerability

Democratic systems depend on citizens' ability to evaluate political information, assess candidates, and form informed judgments about political institutions. Deepfakes can complicate these processes by making fabricated political messages appear visually or audibly

convincing. A candidate may be portrayed as making a statement that was never made, a public official may be shown announcing a fabricated policy, or an election related institution may be represented through cloned audio or manipulated video. The democratic risk therefore extends beyond audiences simply believing fabricated content. Even when audiences recognize that content may be manipulated, convincing synthetic media can create uncertainty about the authenticity of political information and the reliability of available evidence.

Chesney and Citron (2019) argue that deepfakes can generate uncertainty and undermine confidence in information even when audiences do not fully accept fabricated content as genuine. This mechanism is closely related to the concept of the “liar’s dividend”, in which individuals facing genuine and potentially damaging evidence can dismiss that evidence by claiming that it has been artificially generated or manipulated (Sareen, 2022). The problem is therefore not limited to whether false content is believed, but also whether genuine evidence can continue to function as credible evidence after synthetic media becomes increasingly prevalent.

This distinction is important in the Indonesian context. The emergence of AI generated political content around the 2024 election provides a relevant context for examining how synthetic media may interact with political communication, but its existence should not by itself be interpreted as evidence that deepfakes substantially altered voter behavior or electoral outcomes. Muhtadi and Wihardja (2024) indicate that exposure to and belief in deepfake content can be associated with partisanship, suggesting that political predispositions may influence how audiences interpret synthetic political information. Rather than assuming that Indonesian voters uniformly accept or reject deepfakes, their effects may depend on content characteristics, political predispositions, source credibility, and the informational environment in which the content circulates.

Deepfakes consequently do not need to persuade all citizens to create democratic risks. Their potential significance may also arise when they reinforce existing political divisions, increase uncertainty surrounding competing claims, or provide political actors with an opportunity to challenge authentic evidence. In a politically polarized information environment, persistent uncertainty may make citizens more dependent on partisan cues and

trusted social groups when evaluating contested information. The resulting problem concerns not only the circulation of false information but also the ability to establish what constitutes credible evidence.

Deepfake as an Information Security Threat

Deepfake can also be conceptualized as an information security threat when information security is understood as the preservation of the authenticity, integrity, provenance, and trustworthiness of information used for decision making. Deepfakes pose a security concern because they can fabricate the apparent source and evidentiary status of a communication. A synthetic video or cloned voice may cause audiences to believe that a particular person made a statement when that person did not, compromising the reliability of information used by citizens, public institutions, organizations, and other decision makers. The security problem therefore concerns not only the content of a message, but also how its authenticity and origin are established.

Identity impersonation represents a significant risk. Deepfake video and voice cloning technologies can imitate political leaders, public officials, journalists, executives, or other identifiable individuals. In public sector communication, fabricated statements may create confusion regarding official decisions, policies, or emergency information. In electoral contexts, manipulated representations may damage candidates' reputations, create false impressions about political actors, or undermine confidence in electoral institutions. Similar technologies can facilitate fraud and social engineering in organizational and financial environments by making fraudulent communications appear to originate from trusted individuals.

Deepfakes may also affect institutional trust because the apparent authenticity of official communication is part of the information environment on which institutional action depends. A fabricated announcement attributed to an election authority, a manipulated video purporting to show electoral misconduct, or a cloned statement attributed to a public official could create competing claims about what an institution has actually said or done. However, the extent of this effect depends on audience interpretation, source credibility, speed of

dissemination, and institutional responses. The existence of a technically convincing deepfake should therefore not automatically be equated with successful public influence.

This perspective positions deepfake as more than a conventional media manipulation problem. Its security implications arise from the interaction between synthetic content, identity impersonation, information provenance, and the capacity of individuals and institutions to establish whether a communication is authentic (Neyazi et al., 2024; Yamaoka Enkerlin, 2019). The concept of information security used in this study therefore encompasses the informational conditions required for trustworthy communication and decision making, providing a conceptual basis for connecting deepfake with democratic resilience.

AI Governance and Regulatory Responses in Indonesia

Indonesia has started to address AI related risks through policy and legal instruments. The Minister of Communication and Informatics Circular Letter Number 9 of 2023 on Artificial Intelligence Ethics introduces principles such as transparency, accountability, credibility, and security. Law Number 27 of 2022 on Personal Data Protection is also relevant because deepfake technologies can involve personal images, voices, and biometric attributes without consent. These instruments provide foundations for addressing AI related and deepfake related risks concerning responsible AI use, personal data, and individual rights.

The current framework still leaves several areas requiring clarification. Indonesia does not yet have a specific legal definition that directly addresses political deepfakes as a distinct form of synthetic media. Cases involving deepfake content may therefore be addressed through electronic information law, data protection, defamation, fraud, election law, or platform governance, depending on the characteristics and consequences of each case. Although these instruments can cover different elements of harmful synthetic content, they do not necessarily operate as an integrated framework for deepfake related risks.

Another concern involves labeling, provenance, and platform accountability. AI generated political content requires disclosure mechanisms to help citizens identify whether video, image, or audio content has been synthetically produced or manipulated. However, labeling alone is insufficient because deepfake content can be reposted, compressed, edited, or redistributed through different communication channels, including closed messaging

groups. Governance therefore needs to combine detection, content provenance, public literacy, official verification channels, electoral safeguards, and platform responsibility.

Regulation must also address harmful deception without unnecessarily restricting satire, criticism, journalism, or legitimate political expression. A broad anti disinformation policy may create risks of overregulation if it lacks clear definitions, proportionality, and independent oversight. Deceptive impersonation, non consensual synthetic media, fraud, and electoral manipulation should therefore be distinguished from parody, satire, or transparent creative content (Badan Siber dan Sandi Negara Republik Indonesia, 2024). This distinction is important because efforts to strengthen information security should also preserve conditions necessary for legitimate democratic communication.

Research Framework

This article proposes a research framework that connects deepfake enabled disinformation with democratic vulnerability and information security risks in Indonesia. The framework synthesizes relationships identified in the literature between synthetic content, the circulation of misleading information, political interpretation, amplification, institutional vulnerability, and governance capacity. Deepfake is positioned as the technological entry point of synthetic manipulation, while disinformation represents one form of information disorder through which deliberately misleading synthetic content can enter political communication. Democratic vulnerability and information security risks represent the principal areas of potential impact, shaped by platform amplification, partisan interpretation, verification capacity, institutional trust, and regulatory capacity. The framework is presented in Figure 1.

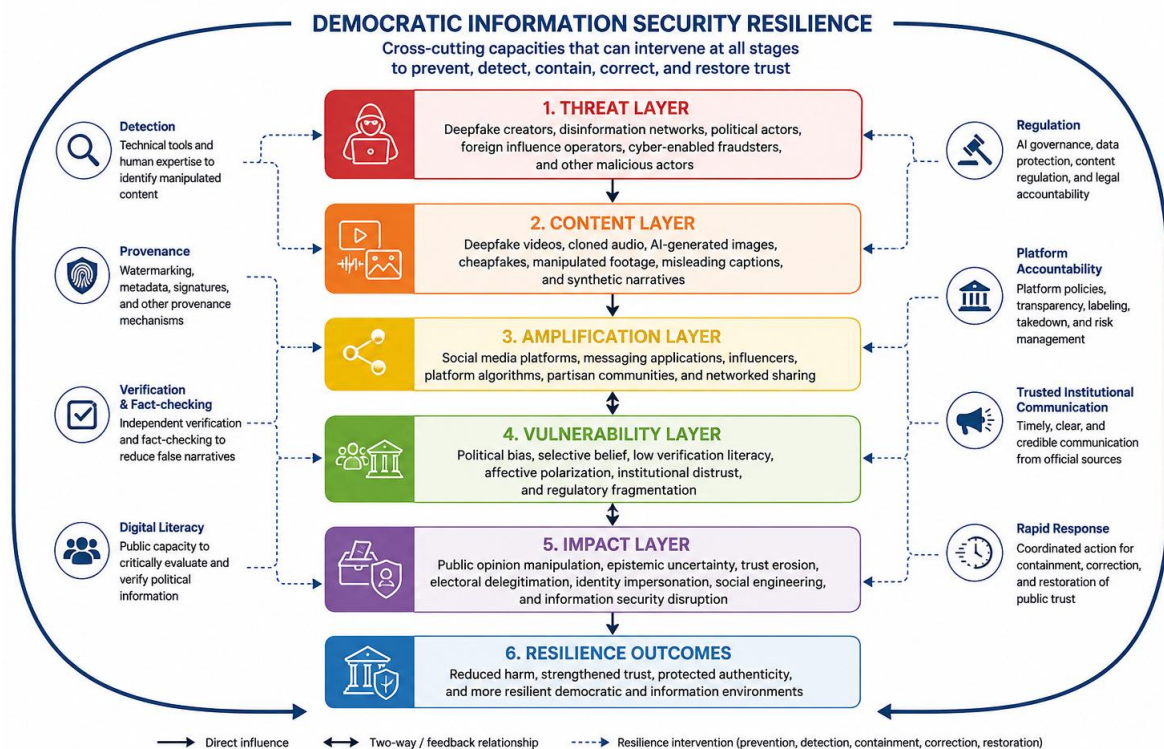


Figure 1. Research Framework

Based on this framework, deepfake is not treated as a purely technical problem but as a layered democratic and information security issue. The framework identifies how synthetic content may become politically consequential through political preferences, platform amplification, limited verification capacity, institutional trust, and regulatory conditions. The framework therefore connects the characteristics of synthetic content with the social, political, technological, and institutional conditions through which it circulates.

The framework also connects these processes to information security by emphasizing authenticity, identity verification, information provenance, institutional communication, and public trust. Fabricated representations of political actors can create uncertainty regarding whether statements or actions are authentic, making it more difficult to distinguish genuine evidence from fabricated evidence. This creates an information security concern because trustworthy decision making depends on the ability of individuals and institutions to establish the authenticity and reliability of information.

The framework supports the development of the Democratic Information Security Resilience model through several interconnected levels. These include technical detection, legal clarity, content labeling, information provenance, media literacy, platform responsibility, protection of personal data, and rapid institutional communication. Technical and provenance mechanisms support content verification; legal and regulatory measures establish responsibilities and boundaries; literacy strengthens citizens' capacity to evaluate information; platform measures address harmful amplification; and institutional communication provides authoritative clarification when contested information emerges.

Such an approach allows Indonesia to address deepfake risks without reducing the issue to criminal enforcement or technological detection alone. The framework provides an analytical basis for understanding how deepfake enabled disinformation can develop into broader democratic and information security risks and identifies points at which resilience-oriented interventions can be introduced.

RESEARCH METHODS

Research Design

This study employed a Systematic Literature Review (SLR) guided by the PRISMA 2020 reporting framework to map, evaluate, and synthesize academic debates on deepfake enabled disinformation, democratic vulnerability, information security, and AI governance. PRISMA 2020 provides a transparent procedure for source identification, screening, eligibility assessment, and final inclusion (Ahmad et al., 2024).

The review combined thematic synthesis and policy document analysis. Thematic synthesis identified recurring patterns related to synthetic media, information disorder, selective belief, electoral integrity, identity impersonation, institutional trust, and cyber enabled influence operations. Policy document analysis examined Indonesia's regulatory readiness in addressing AI generated political manipulation. This design connected international academic debates with the Indonesian policy landscape while distinguishing academic literature from policy and institutional documents.

Data Sources and Search Strategy

The academic literature was collected from major scholarly databases, including Scopus, Web of Science, ScienceDirect, SAGE Journals, SpringerLink, Taylor & Francis, IEEE Xplore, and ACM Digital Library. Google Scholar was used as a supplementary source. Additional policy and institutional materials were collected through relevant institutional websites and citation searching. Academic publications and policy or institutional documents were treated as distinct evidence categories.

The search strategy used five keyword clusters: deepfake and synthetic media; disinformation and information disorder; democracy and electoral integrity; information security and cybersecurity; and Indonesia, Southeast Asia, or the Global South. The main search string was:

```
("deepfake" OR "synthetic media" OR "AI-generated content")  
AND  
("disinformation" OR "misinformation" OR "information disorder")  
AND  
("democracy" OR "election integrity" OR "political communication")  
AND  
("information security" OR "cybersecurity" OR "digital trust")  
AND  
("Indonesia" OR "Southeast Asia" OR "Global South")
```

Because the main search string required terms from all five clusters simultaneously, additional searches were conducted to capture studies addressing specific dimensions of the research problem, including democratic resilience, information security, AI governance, and selective belief. The additional search strings included:

```
("deepfake" AND "election" AND "Indonesia")  
("AI-generated disinformation" AND "democratic resilience")  
("synthetic media" AND "information security")  
("deepfake" AND "AI governance" AND "regulation")  
("disinformation" AND "selective belief" AND "Indonesia")
```

Inclusion and Exclusion Criteria

Sources were included when they were published between 2019 and 2026, discussed deepfake, synthetic media, AI generated content, disinformation, misinformation, or information disorder, and were relevant to democracy, electoral integrity, political communication, information security, cybersecurity, or AI governance. Eligible sources included peer reviewed journal articles, reputable conference proceedings, academic books, policy reports, official documents, or credible institutional reports available in English or Indonesian.

Sources were excluded when they discussed artificial intelligence without a clear link to deepfake or disinformation, focused only on technical detection without democratic or security implications, lacked relevance to election integrity or information security, appeared as opinion-based material without academic or policy grounding, duplicated records from other databases, or lacked accessible full text.

Screening Process

The screening process followed the PRISMA 2020 flow. A total of 342 records were identified from academic databases. Before screening, 88 duplicate records were removed and 29 records were removed because they fell outside the publication period of 2019-2026 or were not available in English or Indonesian. After these removals, 225 records remained for title and abstract screening.

Of the 225 screened records, 156 were excluded because they did not meet the review focus. A total of 69 reports were sought for retrieval, while 6 could not be retrieved. The remaining 63 reports were assessed for full text eligibility, resulting in 33 exclusions due to limited relevance to democratic or information security issues, general AI discussion, insufficient analytical grounding, or inadequate full text quality. This resulted in 30 academic studies included in the systematic review.

Additional sources were identified through institutional websites, policy sources, and citation searching. This route produced 56 reports sought for retrieval, of which 3 could not be retrieved. The remaining 53 reports were assessed for eligibility, resulting in 12 policy or

institutional reports after excluding sources with limited institutional credibility, limited relevance to Indonesia, Southeast Asia, or the Global South, or duplicated policy content.

The final corpus consisted of 42 items, comprising 30 academic studies and 12 policy or institutional reports. The two evidence categories were analyzed separately, with academic studies used to synthesize theoretical, empirical, and methodological debates, and policy and institutional reports used to examine Indonesia's governance response and regulatory readiness.

Quality Appraisal

Academic sources were assessed using five criteria: clarity of research purpose, appropriateness of method or analytical approach, strength of evidence or argument, relevance to deepfake enabled disinformation and information security, and contribution to theoretical or policy discussion. Each criterion was scored on a three point scale: 0 for weak or unclear, 1 for partial, and 2 for adequate. The appraisal assessed the quality and analytical usefulness of sources and prioritized stronger sources in the thematic synthesis rather than applying statistical weighting.

Policy documents and institutional reports were assessed separately based on institutional credibility, legal or policy authority, relevance to Indonesia or the broader Global South, clarity of regulatory content, and usefulness for analyzing AI governance. This distinction reflected the different functions and characteristics of policy documents and empirical journal articles.

Data Extraction

Data were extracted using a structured matrix. For academic studies, the matrix included author and year, publication type, country or regional focus, research objective, method, theoretical lens, type of deepfake or synthetic media, form of disinformation, democratic risk, information security risk, governance response, major findings, and research gaps.

For policy documents and institutional reports, the extracted data included issuing institution, year, policy scope, relevance to artificial intelligence or digital governance, treatment of synthetic media or disinformation, provisions related to personal data or identity

protection, platform accountability mechanisms, and implications for election integrity or information security.

Data Analysis

The extracted data were analyzed through thematic synthesis. Initial coding identified recurring concepts across the literature, including synthetic media realism, manipulated evidence, identity impersonation, partisan amplification, selective belief, liar's dividend, institutional trust erosion, social engineering, platform accountability, and AI governance. These codes were grouped into five themes: deepfake as synthetic media, disinformation and information disorder, democratic vulnerability, information security risk, and AI governance in Indonesia.

The analysis examined relationships among synthetic media, disinformation mechanisms, platform amplification, partisan interpretation, verification capacity, uncertainty, institutional trust, and regulatory conditions. Policy document analysis compared these findings with Indonesia's regulatory landscape, focusing on legal definition and accountability for political deepfakes, content labeling and provenance mechanisms, and coordination among relevant institutions, platforms, media, and civil society.

Methodological Trustworthiness

Several procedures were used to strengthen the trustworthiness of the review. The search strategy was documented through keyword clusters, database selection, and eligibility criteria, while the screening process followed the PRISMA 2020 flow. The quality appraisal rubric helped distinguish stronger academic sources from weaker or purely descriptive materials, and the extraction matrix provided consistency in comparing sources across disciplines and evidence types.

The integration of policy document analysis reduced the risk of treating deepfake as a purely technical or communication issue. By combining academic synthesis with policy analysis, the study developed the Democratic Information Security Resilience framework from scholarly literature and governance evidence. The framework organizes relationships across

synthetic content production, disinformation circulation, platform amplification, democratic vulnerability, information security risk, and regulatory response.

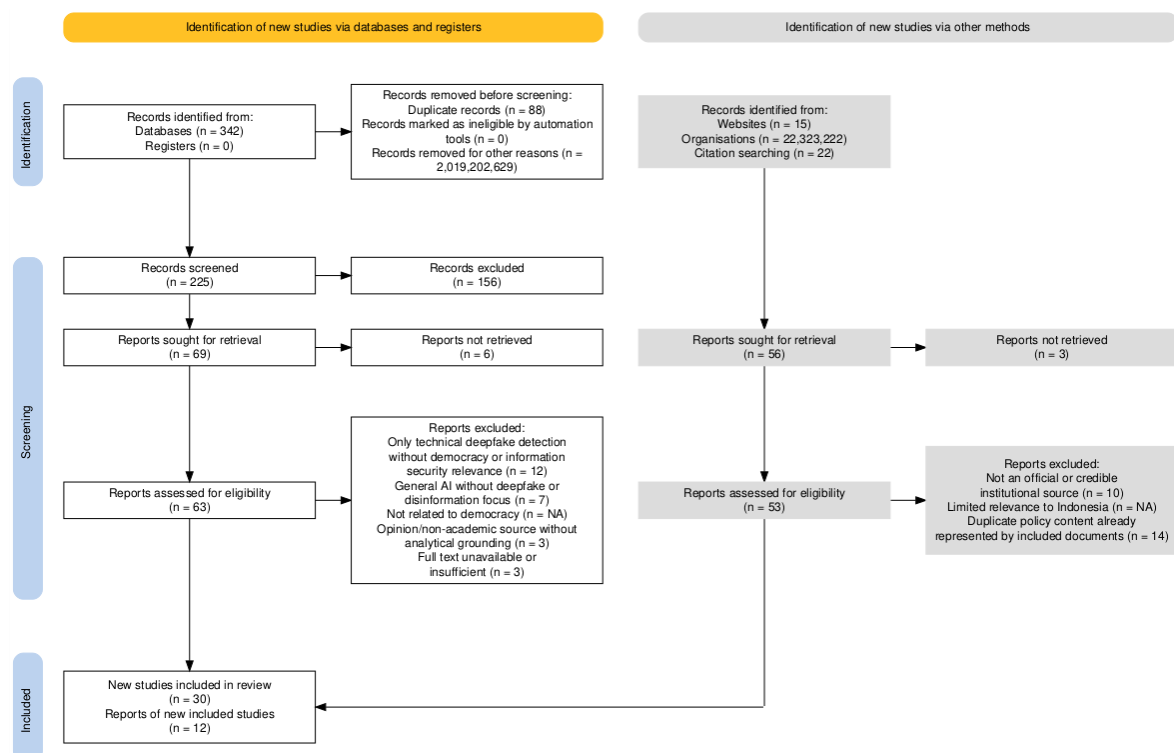


Figure 2. PRISMA Flow Diagram

Source: Authors' synthesis based on the reviewed academic and policy or institutional evidence

RESULTS AND DISCUSSION

Overview of the Review Corpus

The PRISMA based review produced a final corpus of 42 items, consisting of 30 academic studies and 12 policy or institutional reports. The academic studies were used to synthesize scholarly debates on deepfake enabled disinformation, democratic vulnerability, information security, and AI governance. The policy and institutional reports were examined separately to assess governance instruments addressing synthetic media, electoral manipulation, platform accountability, and cybersecurity related risks in Indonesia.

The final corpus provides an interdisciplinary basis for examining deepfake as a technological, political, informational, and regulatory problem. The distinction between

academic and policy evidence was maintained throughout the analysis, with academic studies informing the thematic synthesis and policy or institutional reports informing the assessment of governance responses and regulatory readiness in Indonesia.

Metadata of Included Sources

The 30 academic studies covered deepfake, synthetic media, political disinformation, selective belief, democratic legitimacy, information security, and AI governance. The 12 policy and institutional documents addressed the regulatory and governance landscape related to artificial intelligence, disinformation, electoral integrity, personal data protection, platform accountability, and digital security.

Table 1. Metadata of Key Included Sources

ID	Author(s) / Institution	Year	Source Type	Method / Approach	Main Focus	Relevance to This Study
S1	Chesney & Citron	2019	Journal article	Legal-conceptual	Deepfake, privacy, democracy, national security	Liar's dividend and evidentiary distrust
S2	Paris & Donovan	2019	Research report	Conceptual and media analysis	Deepfakes, cheapfakes, visual manipulation	Distinguishes AI based and low cost manipulation
S3	Westerlund	2019	Review article	Literature review	Deepfake technology and social risk	Early technological and societal risks
S4	Vaccari & Chadwick	2020	Journal article	Experimental study	Deepfake, uncertainty, trust in news	Uncertainty without full belief in fabricated content
S5	Tolosana et al.	2020	Review article	Technical survey	Face manipulation and deepfake detection	Technical affordances and detection challenges
S6	Mirsky & Lee	2021	Review article	Technical survey	Deepfake creation and detection	Generation methods, detection, and limitations
S7	Groh et al.	2022	Journal article	Experimental study	Human and machine detection	Limits of visual inspection and verification
S8	Nguyen et al.	2022	Review article	Technical survey	Deep learning, deepfake creation, detection	Realism, detection, and adversarial development
S9	Pawelec	2022	Journal article	Normative-theoretical analysis	Deepfake and democracy	Democratic vulnerability and political communication ethics

ID	Author(s) / Institution	Year	Source Type	Method / Approach	Main Focus	Relevance to This Study
S10	Masood et al.	2023	Review article	Systematic technical review	Deepfake generation, detection, countermeasures	Technical production and detection challenges
S11	Ferrara	2024	Analytical study	Conceptual and case-based analysis	Generative AI and election interference	AI enabled influence operations
S12	Muhtadi & Wihardja	2024	Policy research / empirical analysis	Survey-based analysis	Deepfake, partisanship, selective belief	Indonesia specific exposure and susceptibility
P1	Kementerian Komunikasi dan Informatika Republik Indonesia	2023	AI ethics circular letter	Policy document analysis	AI ethics, transparency, accountability, security	AI governance principles and regulatory gaps
P2	Badan Pengawas Pemilihan Umum Republik Indonesia	2022	Law	Legal document analysis	Personal data protection	Synthetic media and biometric identity misuse
P3	Badan Siber dan Sandi Negara Republik Indonesia	2024	Election governance document	Policy document analysis	Campaign integrity and election supervision	AI generated political content and election security
P4	Law Number 27 of 2022 on Personal Data Protection	2024	Institutional report	Cybersecurity policy analysis	Cybersecurity and digital resilience	Identity impersonation and cyber enabled influence
P5	Freedom House	2024	Institutional report	Policy and rights analysis	Internet freedom and disinformation	Digital rights and online disinformation in Indonesia
P6	The Organization for Economic Co-operation and Development	2024	Policy research report	Empirical and policy analysis	Deepfake, Indonesian election, selective belief	Partisan exposure and electoral vulnerability
P7	Verdoliva	2019	Research report	Conceptual analysis	Deepfakes, cheapfakes, media manipulation	Contextual nature of visual disinformation
P8	World Economic Forum	2024	Global risk report	Risk analysis	AI-generated misinformation and disinformation	Broader global risk context
P9	OpenAI	2024	Institutional threat report	Threat analysis	Deceptive AI use and influence operations	AI misuse in coordinated

ID	Author(s) / Institution	Year	Source Type	Method / Approach	Main Focus	Relevance to This Study
						information operations
P10	UNESCO	2023-2024	Governance guideline	Policy framework analysis	AI responsibility, platform governance, digital integrity	Governance principles for Indonesia

Source: Authors' synthesis based on the reviewed academic and policy or institutional evidence.

Table 1 presents selected key sources, while the complete metadata is provided in the appendix. The academic literature shows an interdisciplinary distribution, with technical studies focusing primarily on the production and detection of synthetic media, while political communication studies examine belief, trust, and content circulation. Policy and governance documents identify regulatory principles and institutional responsibilities. This distribution indicates a need to connect technological, political, information security, and governance perspectives in analyzing deepfake enabled disinformation in Indonesia.

Thematic Synthesis

The thematic synthesis generated five interconnected findings: technological affordances of deepfake, selective belief and partisan amplification, trust erosion and democratic legitimacy, cyber enabled information security risk, and regulatory and governance gaps in Indonesia. These themes indicate that deepfake enabled disinformation can move from synthetic content to political consequences through circulation, audience interpretation, institutional conditions, and governance capacity. This process is not automatic or uniformly causal, as its effects depend on social, political, technological, and institutional conditions. Table 2 summarizes the thematic findings, their analytical interpretation, and their implications for Indonesia.

Table 2. Thematic Synthesis and Analytical Implications

Theme	Key Finding	Analytical Interpretation	Implication for Indonesia
Technological affordances of deepfake	Deepfake combines visual realism, voice imitation, low-cost production, and cross-platform circulation.	Risk arises from the interaction of synthetic realism, accessibility, circulation, and social vulnerability, not	Platform based political communication may facilitate manipulated audiovisual content

Theme	Key Finding	Analytical Interpretation	Implication for Indonesia
		technical sophistication alone.	when verification is limited.
Selective belief and partisan amplification	Audiences may be more receptive to synthetic content that confirms political identity, prior distrust, or group-based expectations.	Deepfake enabled disinformation operates through political interpretation and social identity as well as apparent realism.	Partisan communities, volunteer networks, and closed messaging groups may strengthen the visibility and perceived credibility of synthetic narratives.
Trust erosion and democratic legitimacy	Deepfake can weaken trust in media, candidates, election bodies, and authentic evidence.	Harm may emerge through uncertainty, reduced confidence in evidence, and the liar's dividend rather than direct belief alone.	Manipulated or disputed content may reduce confidence in electoral integrity, especially when institutional responses are delayed or contested.
Cyber-enabled information security risk	Deepfake can falsify identity, manipulate apparent evidence, support social engineering, and impersonate officials or institutions.	Synthetic media threatens authenticity, integrity, identity verification, provenance, and institutional communication.	Electoral bodies, cybersecurity agencies, and digital regulators need rapid verification and trusted public communication.
Regulatory and governance gap	Existing legal and policy instruments address AI ethics, personal data, electronic information, and platform governance, but political deepfake remains weakly defined.	Fragmented governance can create uncertainty over responsibility and delay verification, response, accountability, and correction.	Indonesia needs clearer rules concerning AI-generated political content, provenance, platform responsibility, and inter-agency coordination.

Source: Authors' synthesis based on the reviewed literature and policy or institutional evidence.

The five themes provide the analytical basis for the Democratic Information Security Resilience Framework. Technological affordances explain the characteristics of synthetic content, while selective belief and partisan amplification explain how political interpretation and circulation can shape its reception. Trust erosion identifies democratic consequences, while information security risk captures threats to authenticity, identity verification, and institutional communication. The regulatory and governance theme identifies institutional capacities and gaps that shape prevention, verification, and response.

Technological Affordances of Deepfake

The reviewed literature indicates that deepfake becomes politically consequential because it alters the evidentiary character of digital content. Unlike edited footage, misleading captions, or decontextualized images, deepfake can simulate a person's face, voice, expression, and gesture, allowing fabricated content to appear as audiovisual evidence (Ahmad et al., 2024; Venema, 2023). However, apparent realism alone does not determine political influence, as effects also depend on how content is encountered, interpreted, and circulated.

The political risk therefore cannot be judged only by technical quality. Paris and Donovan (2019) show that manipulated content can produce public harm when it reaches a receptive audience. A synthetic video may generate political consequences because it appears during a politically sensitive moment, targets a distrusted figure, or confirms existing partisan expectations. The relevant mechanism is therefore not only whether audiences can detect manipulation, but whether the content fits existing expectations and social interpretations.

Circulation can further intensify this risk. Synthetic content can be produced, reposted, compressed, re captioned, and moved across platforms before verification is completed, creating an asymmetry between rapid manipulation and slower correction. During elections, this timing gap can allow content to influence political perception or generate uncertainty before corrections reach affected audiences (Balogun et al., 2025; Chesney & Citron, 2019).

Selective Belief and Partisan Amplification

Deepfake does not operate through deception alone. Citizens interpret political information through prior beliefs, identity, trust, and attachment to political groups. Synthetic media may become more credible when it confirms what an audience already suspects or expects, meaning that political effects depend on the interaction between synthetic content and audience interpretation.

Gregory (2022) found that deepfake can increase uncertainty and reduce trust in news even when viewers do not fully accept fabricated content. This suggests that its consequences do not depend exclusively on whether audiences believe a synthetic video. Uncertainty can

also weaken the corrective capacity of journalism, fact checking, and official clarification when competing claims remain plausible to different audiences.

In Indonesia, Zoidze and Abuselidze (2025) show that exposure and susceptibility to deepfake content during the 2024 election were shaped by partisanship. Synthetic content became more persuasive when it aligned with political preferences or pre existing distrust. This does not demonstrate that deepfakes uniformly changed voter behavior or electoral outcomes, but indicates that partisan orientation can condition how synthetic political content is received and evaluated.

Platform architecture can further shape this process. Public platforms may reward attention and emotional engagement, while closed messaging groups rely more on interpersonal and community based trust. Content circulated through trusted social relationships may therefore receive greater initial credibility than institutional corrections. Technical detection alone cannot address these conditions, making credible verification, trusted communication, and mechanisms to limit misleading amplification necessary.

Trust Erosion and Democratic Legitimacy

Trust erosion emerged as a central finding across the reviewed studies. Deepfake can create uncertainty around audiovisual evidence, and this doubt may remain in public circulation even after fabricated content is debunked. Democratic consequences therefore do not depend only on whether citizens believe synthetic content, but also on whether they remain uncertain about the credibility of political evidence.

The concept of the liar's dividend illustrates this risk. Makhniashvili (2025) argues that the existence of deepfake technology allows political actors to deny authentic evidence by claiming that it is fabricated. Deepfake can therefore make false content appear credible while making genuine content appear questionable, reducing citizens' confidence in distinguishing truth from manipulation.

The risk is particularly relevant to electoral institutions. Fabricated statements from election officials or manipulated footage of ballot handling can damage public trust before verification is issued, while distrust may remain among groups already inclined to question electoral fairness. In Indonesia, this risk intersects with identity politics, personality centered

campaigns, polarized digital networks, and distrust toward political institutions. Fact checking remains necessary, but its effectiveness depends on timing, institutional credibility, and audience willingness to accept correction (Chaturvedi et al., 2025; Tahraoui, Krätzer, Dittmann, et al., 2023).

Deepfake as Cyber-Enabled Information Security Risk

The review supports treating deepfake as an information security risk because synthetic media threatens authenticity, integrity, identity verification, and institutional communication. Deepfake can falsify the identity of a speaker, alter the perceived origin of a message, and manipulate the evidentiary status of digital content.

Identity impersonation represents a clear security threat. Voice cloning and synthetic video can imitate public officials, candidates, election administrators, journalists, executives, or civil society figures. In electoral politics, fabricated statements can damage reputations, mislead voters, or undermine trust in election management bodies. In public administration and organizational settings, impersonation can support social engineering, fraud, unauthorized instructions, and reputational attacks.

Evidence manipulation further weakens the assumption that audiovisual materials can serve as evidence when properly verified. Deepfake can produce fabricated evidence while creating doubt toward genuine evidence, requiring stronger authentication mechanisms, official verification channels, and rapid response procedures.

Synthetic media can also form part of broader influence operations involving fake accounts, coordinated posting, targeted advertising, partisan influencers, and repeated narratives across platforms. Its effects depend on timing, audience segmentation, emotional framing, and network credibility. The security response therefore requires coordination among election bodies, cybersecurity agencies, digital regulators, media institutions, civil society, and platforms, including early detection, trusted authentication, public communication, and institutional coordination (Nasiri & Hashemzadeh, 2025; Verdoliva, 2020).

Regulatory and Governance Gap in Indonesia

The policy document analysis indicates that Indonesia has relevant instruments addressing parts of the deepfake problem. AI ethics guidance introduces principles of

transparency, credibility, accountability, security, and human centered use. Personal data protection law is relevant because deepfake may use faces, voices, images, and biometric traces without consent, while electronic information regulation and election governance documents address harmful digital content, fraud, manipulation, campaign conduct, and public trust in electoral processes.

The main difficulty lies in fragmentation. Political deepfake does not yet have a clear regulatory category, while a single incident may involve personal data misuse, impersonation, electoral manipulation, platform amplification, reputational harm, cybersecurity risk, and public disorder. Unclear coordination can create uncertainty over who verifies content, orders takedown, preserves evidence, informs the public, and holds perpetrators accountable.

Labeling and provenance also remain underdeveloped. Watermarking, metadata standards, cryptographic signatures, platform labels, and verified official channels can help identify synthetic or altered content, although their effectiveness may decline when content is downloaded, re uploaded, compressed, edited, or circulated through closed messaging groups (Masood et al., 2023; Pawelec, 2022).

Regulatory design must also protect democratic rights. Rules should address deceptive impersonation, non consensual synthetic media, fraud, electoral manipulation, and malicious identity fabrication while distinguishing harmful deception from satire, parody, journalism, artistic expression, and legitimate political criticism. Indonesia's governance response therefore requires precision, transparency, independent oversight, and clear distinctions between harmful deception and protected speech.

Toward Democratic Information Security Resilience

The findings support the Democratic Information Security Resilience Framework as a conceptual framework for understanding how deepfake enabled synthetic media can generate interconnected democratic and information security risks. Deepfake represents the technological entry point, while disinformation mechanisms, platform amplification, partisan interpretation, institutional vulnerability, and governance capacity shape whether synthetic content becomes politically consequential. These relationships should not be understood as

a fixed causal sequence because their effects depend on technological, social, political, and institutional conditions.

Democratic vulnerability may emerge through polarization, weakened trust in media, challenges to electoral integrity, and the liar's dividend. Information security risk emerges through identity impersonation, social engineering, and manipulation of the authenticity and provenance of evidence. Governance responses therefore require coordinated intervention through AI regulation, content provenance, platform accountability, public literacy, electoral safeguards, cybersecurity coordination, and trusted institutional communication.

The findings show that deepfake enabled disinformation represents a hybrid threat that is technological in production, political in target, psychological in reception, security related in its implications for authenticity and identity verification, and regulatory in its demand for coordinated response. The framework therefore connects technical detection with verification, institutional communication, platform accountability, provenance mechanisms, and safeguards for democratic expression. As the framework is derived from existing literature and policy evidence rather than empirical validation, its propositions require future testing through electoral cases, audience interpretation, institutional responses, and governance interventions.

Proposed Conceptual Framework: Democratic Information Security Resilience

The synthesis of 30 academic studies and 12 policy or institutional reports supports the Democratic Information Security Resilience Framework. The framework conceptualizes deepfake enabled disinformation as an interconnected risk involving threat actors, synthetic or manipulated content, amplification, vulnerability, democratic and information security impacts, and resilience responses. Rather than assuming a fixed causal sequence, it explains how these dimensions interact to determine whether synthetic media becomes politically consequential in Indonesia.

The framework treats deepfake as an entry point into a broader system. Its effects depend on how manipulated content is produced or circulated, reaches audiences through communication networks, interacts with partisan interpretation or distrust, and encounters

verification or institutional response capacity. Resilience therefore operates across these conditions through prevention, verification, containment, correction, and restoration of trust.

Table 3. Democratic Information Security Resilience Framework

Layer	Main Components	Analytical Function	Indonesian Relevance
Threat Layer	Deepfake creators, disinformation networks, political actors, foreign influence operators, and cyber enabled fraudsters	Identifies actors and capacities involved in producing or circulating manipulation	Digital political networks, campaign volunteers, influencers, and other connected actors can facilitate information circulation
Content Layer	Deepfake video, cloned audio, AI generated images, cheapfakes, manipulated footage, misleading captions, and synthetic narratives	Identifies forms through which information can be fabricated, altered, or deceptively presented	Synthetic and manipulated political content can enter public platforms and private digital spaces
Amplification Layer	Social media platforms, messaging applications, influencers, platform recommendation systems, partisan communities, and networked sharing	Explains how manipulated content gains reach, repetition, visibility, and perceived credibility	Digital platforms provide different environments for political content to circulate and be repeatedly encountered
Vulnerability Layer	Political bias, selective belief, limited verification literacy, affective polarization, institutional distrust, and regulatory fragmentation	Identifies conditions that increase exposure or reduce the capacity to challenge manipulation	Partisan interpretation, uneven verification capacity, polarization, and fragmented responsibilities may complicate responses
Impact Layer	Public opinion manipulation, epistemic uncertainty, trust erosion, electoral delegitimation, identity impersonation, social engineering, and information security disruption	Explains democratic and information security consequences	Synthetic content may undermine trust in candidates, electoral institutions, official communication, and authentic political evidence
Resilience Layer	Technical detection, content provenance, fact-checking, digital literacy, AI regulation, platform accountability, trusted institutional communication, and rapid response	Identifies capacities for prevention, verification, containment, correction, accountability, and trust restoration	Resilience requires coordination among electoral institutions, regulators, cybersecurity institutions, platforms, media, fact checkers, and civil society

Source: Authors' synthesis based on the reviewed academic and policy or institutional evidence.

The six layers represent different analytical levels of the same risk system. The Threat Layer identifies actors involved in producing or circulating manipulation, while the Content Layer identifies its technological and communicative form. The Amplification Layer explains how content gains reach and perceived credibility, and the Vulnerability Layer identifies conditions that shape audience and institutional responses. The Impact Layer captures democratic and information security consequences, while the Resilience Layer identifies capacities that can intervene across the system.

The Threat Layer concerns actors involved in creating, coordinating, or circulating deepfake enabled disinformation, while distinguishing deliberate production from unknowing recirculation. The Content Layer covers deepfake videos, cloned voices, AI generated images, cheapfakes, manipulated footage, and synthetic narratives. These forms differ in production methods and evidentiary characteristics but can function as vehicles for deceptive political narratives.

The Amplification Layer explains how manipulated content gains visibility, repetition, and perceived credibility through recommendation systems, influencers, partisan communities, messaging applications, and interpersonal sharing. The Vulnerability Layer captures political bias, selective belief, affective polarization, limited verification literacy, institutional distrust, and regulatory fragmentation that may influence whether content is accepted, rejected, questioned, or redistributed.

The Impact Layer captures consequences including distorted political perceptions, uncertainty about authentic evidence, declining trust in media or electoral institutions, electoral delegitimation, identity impersonation, and information security risks. These consequences do not require audiences to fully believe a deepfake, as uncertainty itself may weaken confidence in political evidence and institutional communication. This mechanism is particularly relevant to the liar's dividend.

The Resilience Layer represents capacities to prevent, detect, verify, contain, correct, and recover from deepfake related information risks. Technical detection, provenance mechanisms, fact checking, public verification literacy, trusted institutional communication, legal and regulatory mechanisms, and platform accountability can operate before, during, or after an incident (Islam et al., 2024; Łabuz & Nehring, 2024; Shoaib et al., 2023).

The relationships among the layers are interactive rather than linear. Threat actors provide the capacity for manipulation, content determines its form, amplification shapes exposure and repetition, and vulnerability influences interpretation and institutional response. These interactions can produce democratic and information security impacts, while resilience can intervene through detection, verification, provenance, institutional response, and limits on harmful amplification.

The framework integrates democratic vulnerability and information security into a single analytical perspective. Deepfake is treated not only as a technical detection problem or form of false information, but as a mechanism that can affect political interpretation and the ability to establish the authenticity of information. For Indonesia, the framework provides a policy oriented synthesis rather than an empirically validated causal model, emphasizing coordination among technical, legal, institutional, platform based, and civic capacities.

The framework is a conceptual and theoretically informed proposition derived from existing academic and policy evidence. Future empirical research should test its relationships through electoral cases, audience studies, platform and messaging network analysis, and assessments of institutional verification and response mechanisms.

Limitations of the Study

This study has several limitations. First, the final corpus consisted of 30 academic studies and 12 policy or institutional reports, while limited Indonesia specific empirical studies constrain the generalizability of findings. Second, the findings are influenced by database coverage, search terms, publication period, and language criteria, while differences in research design and evidence type limit direct comparison between academic and policy sources. Third, some implications for Indonesia are derived from international literature and interpreted within its political, institutional, and digital environment; identified risks should therefore not be understood as evidence of their extensive occurrence in Indonesia. Finally, the Democratic Information Security Resilience Framework is conceptual rather than empirically validated and does not establish causal relationships among its six layers. Future research should test the framework through documented deepfake cases, audience studies,

platform and closed network circulation, and evaluations of institutional verification, regulation, provenance, and communication mechanisms.

CONCLUSION

This study concludes that deepfake enabled disinformation represents a hybrid threat involving technological, political, social, institutional, and governance dimensions. Based on the synthesis of 30 academic studies and 12 policy or institutional reports, the findings show that its risks are shaped not only by synthetic media but also by partisan interpretation, platform amplification, verification capacity, institutional trust, and governance conditions. These interactions may contribute to uncertainty about political information, weakened confidence in electoral institutions, identity impersonation, and reduced trust in the authenticity of audiovisual evidence. The Democratic Information Security Resilience Framework provides a conceptual contribution by integrating these dimensions into a framework that connects democratic vulnerability with information security. Rather than treating deepfake as a linear cause of harm, the framework highlights the interaction between content, circulation, vulnerability, impact, and resilience capacities. For Indonesia, the findings indicate the importance of coordination among electoral institutions, cybersecurity agencies, digital regulators, platforms, media organizations, fact checkers, and civil society. Policy responses should strengthen detection, provenance, verification, institutional communication, and platform accountability while maintaining safeguards for legitimate political expression. The framework is not an empirically validated causal model. Future research should test its proposed relationships through documented deepfake incidents, audience interpretation, platform and closed network circulation, institutional responses, and the effectiveness of regulatory, provenance, labeling, and communication mechanisms during electoral periods.

REFERENCES

- Ahmad, J., Salman, W., Amin, M., Ali, Z., & Shokat, S. (2024). A survey on enhanced approaches for cybersecurity challenges based on deep fake technology in computing networks. *Spectrum of Engineering Sciences*, 133-149.
- Badan Pengawas Pemilihan Umum Republik Indonesia (2024). *Indeks Kerawanan Pemilu dan Pemilihan Serentak* 2024.

<https://www.bawaslu.go.id/sites/default/files/publikasi/BUKU%20IKP%20PEMILU%20DAN%20PEMILIHAN%20SERENTAK%202024-2.pdf>

- Badan Siber dan Sandi Negara Republik Indonesia (2024). *Lanskap Keamanan Siber Indonesia 2024*. <https://alika.pesisirbaratkab.go.id/files/272/E-BOOK/262/Lanskap-Keamanaan-Siber-Indonesia-2024--BSSN.pdf>
- Balogun, A. Y., Ismaila Alao, A., & Olaniyi, O. O. (2025). Disinformation in the digital era: The role of deepfakes, artificial intelligence, and open-source intelligence in shaping public trust and policy responses. *Oluwaseun Oladeji, Disinformation in the Digital Era: The Role of Deepfakes, Artificial Intelligence, and Open-Source Intelligence in Shaping Public Trust and Policy Responses (March 04, 2025)*.
- Botha, J., & Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. *ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and Publishing Limited*, 57.
- Chaturvedi, A., Raghani, K. N., Kumar, S., Choudhary, M., & Yakub, A. A. M. (2025). *Deepfakes in Society: Risks and Realities*.
- Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, 107, 1753.
- Donnelly, N., & Wilson, R. (2026). Malinformation, Deepfakes, and Cyber Warfare: Ethical and Anticipated Ethical Issues. *International Conference on Cyber Warfare and Security*, 75-82.
- Emeç, M. (2025). AI-Driven Disinformation as A Global Cybersecurity threat to Democratic Systems. *Cyberpolitik Journal*, 10(20), 153-167.
- Ferrara, E. (2024). Charting the Landscape of Nefarious Uses of Generative Artificial Intelligence for Online Election Interference. In *arXiv*. <https://doi.org/10.48550/arXiv.2406.01862>
- Folorunsho, F., & Boamah, B. F. (2025). Deepfake technology and its impact: ethical considerations, societal disruptions, and security threats in ai-generated media. *International Journal of Information Technology and Management Information Systems*, 16(1), 1060-1080.
- Freedom House (2024). *Freedom on the Net 2024: Indonesia*. Freedom House. <https://freedomhouse.org/country/indonesia/freedom-net/2024>
- Gregory, S. (2022). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *Journalism*, 23(3), 708-729.
- Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. (2024). Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video. *Nature Communications*, 15(1), 1-16. <https://doi.org/10.1038/s41467-024-51998-z>
- Irfan, M. M. (2025). Regulation of Deepfakes and Public Expression: Balancing Freedom of Expression and Protection from Disinformation. *Opinia de Journal*, 5(2), 47-57.
- Islam, M. B. E., Haseeb, M., Batool, H., Ahtasham, N., & Muhammad, Z. (2024). AI threats to politics, elections, and democracy: A blockchain-based deepfake authenticity verification framework. *Blockchains*, 2(4), 458-481.
- Kementerian Komunikasi dan Informatika Republik Indonesia. (2023). *Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial*. https://jdih.komdigi.go.id/produk_hukum/view/id/883/t/surat%2Bedaran%2Bmenteri%2Bkomunikasi%2Bdan%2Binformatika%2Bnomor%2B9%2Btahun%2B2023

- Kumar, R., Khan, S. A., Alharbe, N., & Khan, R. A. (2024). Code of silence: cybersecurity strategies for combating deepfake disinformation. *Computer Fraud & Security*, 2024(4).
- Łabuz, M., & Nehring, C. (2024). On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*, 23(4), 454-473.
- Langa, J. (2021). Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes. *BUL Rev.*, 101, 761.
- Law Number 27 of 2022 on Personal Data Protection (2022). <https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022>
- Makhniashvili, M. (2025). Deepfake Threats And EU Law: Navigating Disinformation, Cyber Violence, and the Risks of Digital Manipulation. *L'Europe Unie*, (23), 17-32.
- Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward. *Applied Intelligence*, 53(4), 3974-4026. <https://doi.org/10.1007/s10489-022-03766-z>
- Michałkiewicz-Kądziała, E. (2024). The impact of deepfakes on elections and methods of combating disinformation in the virtual world. *Teka Komisji Prawniczej PAN Oddział w Lublinie*, 17(1), 151-161.
- Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, 54(1), 1-41. <https://doi.org/10.1145/3425780>
- Muhtadi, B., & Wihardja, M. M. (2024). Deepfakes and Selective Belief: How Partisanship Affects Voters' Exposure and Susceptibility to Deepfake Content. In *ISEAS Perspective* (Number 2024/58). ISEAS-Yusof Ishak Institute. <https://www.iseas.edu.sg/articles-commentaries/iseas-perspective/2024-58-deepfakes-and-selective-belief-how-partisanship-affects-voters-exposure-and-susceptibility-to-deepfake-content-by-burhanuddin-muhtadi-and-maria-monica-wihardja/>
- Nasiri, S., & Hashemzadeh, A. (2025). The evolution of disinformation from fake news propaganda to AI-driven narratives as deepfake. *Journal of Cyberspace Studies*, 9(1), 229-250.
- Neyazi, T. A., Nadaf, A. H., Tan, K. E., & Schroeder, R. (2024). Does trust in government moderate the perception towards deepfakes? Comparative perspectives from Asia on the risks of AI and misinformation for democracy. *Government Information Quarterly*, 41(4), 101980.
- Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep Learning for Deepfakes Creation and Detection: A Survey. *Computer Vision and Image Understanding*, 223, 103525. <https://doi.org/10.1016/j.cviu.2022.103525>
- OpenAI. (2024). *AI and Covert Influence Operations: Latest Trends*. OpenAI. https://downloads.ctfassets.net/kftzwdyauwt9/5IMxzTmUclSOAcWUXbkVrk/3cfab518e6b10789ab8843bcca18b633/Threat_Intel_Report.pdf
- Paris, B., & Donovan, J. (2019). *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*. Data & Society Research Institute. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>
- Paterson, T., & Hanley, L. (2020). Political warfare in the digital age: cyber subversion, information operations and 'deep fakes.' *Australian Journal of International Affairs*, 74(4), 439-454.

- Pawelec, M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audiovisual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, 1(2), 19. <https://doi.org/10.1007/s44206-022-00010-6>
- Prasojo, P., Ratnayake, I., & Wijerathna, K. (2026). Beyond Fake News: The Rise of Deepfakes and the Crisis of Epistemic Trust in Digital Journalism. *Journal of Social Science Utilizing Technology*, 4(2), 173-188.
- Ray, A. (2021). Disinformation, deepfakes and democracies: The need for legislative reform. *The University Of New South Wales Law Journal*, 44(3), 983-1013.
- Repez, C. P. F., & Popescu, M.-M. (2020). Social media and the threats against human security deepfake and fake news. *Romanian Military Thinking*, (4).
- Salvaudon, C. (2026). Cybersecurity and Information Warfare: A New Era of Modern Conflict. In *The New Geopolitics of Sustainability: How Big Data, AI, Green Transition and Sustainable Finance Transform the World* (pp. 9-13). Springer.
- Sareen, M. (2022). Threats and challenges by DeepFake technology. In *DeepFakes* (pp. 99-113). CRC Press.
- Schick, N. (2020). *Deep fakes and the infocalypse: What you urgently need to know*. Hachette uk.
- Shoaib, M. R., Wang, Z., Ahvanooy, M. T., & Zhao, J. (2023). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. 2023 *International Conference on Computer and Applications (ICCA)*, 1-7.
- Sophia, L. I. (2025). The social harms of ai-generated fake news: Addressing deepfake and ai political manipulation. *Digital Society & Virtual Governance*, 1(1), 72-88.
- Tahraoui, M., Krätzer, C., Dittmann, J., & Aden, H. (2023). Defending Informational Sovereignty by Detecting Deepfakes? Opportunities and Risks of an AI-Based Detector for Deepfakes-Based Disinformation and Illegal Activities. *Weizenbaum Journal of the Digital Society*, 3(2).
- The Organization for Economic Co-operation and Development (2024). *Revised Recommendation of the Council on Artificial Intelligence*. OECD. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>
- UNESCO. (2023). *Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multistakeholder Approach*. UNESCO. <https://www.unesco.org/en/internet-trust/guidelines>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- Venema, A. E. (2023). Deepfake disinformation: how digital deception and synthetic media threaten national security. In *Routledge Handbook of Disinformation and National Security* (pp. 175-191). Routledge.
- Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5), 910-932. <https://doi.org/10.1109/JSTSP.2020.3002101>
- Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 40-53. <https://doi.org/10.22215/timreview/1282>

- Whyte, C. (2020). Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy*, 5(2), 199-217.
- World Economic Forum (2024). *The Global Risks Report 2024*. World Economic Forum. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf
- Yamaoka-Enkerlin, A. (2019). Disrupting disinformation: deepfakes and the law. *NYUJ Legis. & Pub. Pol'y*, 22, 725.
- Zalewski, P. (2022). New technologies, information, disinformation, fake news, deepfake, hate. Mechanisms of manipulation in the digital space as a threat to security in the 21st Century. *Przegląd Policyjny*, 146, 293-317.
- Zoidze, G., & Abuselidze, G. (2025). Deepfake News Technologies: Contemporary Challenges for Society's Sustainable Economic Development and Security. In *Mastering Deepfake Technology: Strategies for Ethical Management and Security* (pp. 31-48). River Publishers.
- Старовойтенко, О. (2025). Ai in the electoral process: New dimensions of cyberthreats and cybersecurity. *Проблеми Політичної Психології*, 17, 230-246.