THE IMPACT OF POLITICAL DEEPFAKE CONTENT ON DECLINING PUBLIC TRUST: A QUANTITATIVE ANALYSIS OF FIRST-TIME VOTERS ON SOCIAL MEDIA

Restu Widyo Sasongko[1]*, Samuel Indrayana[2], Epiphani Imelda Yosephin Palit[3], Rukoyah[4], Sandra Ivonnie Telussa[5]

Institut Pemerintahan Dalam Negeri, Indonesia[1] Universitas Bhayangkara, Indonesia[2] Politeknik LP3I Jakarta, Indonesia[3] Universitas Pattimura, Indonesia [4] Universitas Pattimura, Indonesia[5]

Correspondence Author: restu_widyo@ipdn.ac.id

Abstract

The evolution of artificial intelligence has given rise to the deepfake phenomenon, a deep learning algorithm-based audio-visual manipulation capable of producing highly realistic political disinformation. This study aims to quantitatively examine the impact of exposure to political deepfake content on social media on the decline of public trust, with a specific focus on the segment of first -time voters. Based on the theoretical foundation of Information Processing Theory and the concept of Liar's Dividend, first-time voters are identified as the most vulnerable group and active consumers of digital political information (Chesney & Citron, 2019). This study uses a quantitative approach with an explanatory survey method on a sample of first-time voters spread across several urban areas. Primary data were collected through a structured Likert-scale questionnaire and analyzed using the Structural Equation Modeling (SEM) method based on Partial Least Squares (PLS). The results are expected to show that the frequency of exposure and the perceived realism of political deepfake content have a significant negative influence on first-time voters' trust in political institutions and the integrity of the general election process. The manipulative nature of content not only degrades trust in specific political figures but also fuels systemic skepticism that undermines democratic legitimacy (Vaccari & Chadwick, 2020). These findings provide an important contribution to the growing literature on digital disinformation and urge policymakers to formulate new regulatory and media literacy frameworks to protect the integrity of young voters in the attention economy.

Keywords: Political Deepfake, Public Trust, First-Time Voters, Social Media, Quantitative Analysis.

1. INTRODUCTION

The use of social media as a primary public space in modern political discourse has fundamentally transformed the landscape of global and domestic political communication. Social media no longer merely serves as an alternative information channel but has mutated into a key epicenter for public opinion formation, mass mobilization, and a highly dynamic electoral campaign platform (Bennett & Segerberg, 2012). The interactive, instantaneous, and decentralized characteristics of social media provide political actors and citizens with the

freedom to distribute political narratives without going through the filters of conventional media institutions. However, this freedom of digital architecture has negative consequences in the form of an abundance of information pollution *that* is increasingly difficult to control (Wardle & Derakhshan, 2017). Amidst this fragile ecosystem, the segment of first-time voters who largely represent Generation Z are the most active consumers of digital political information and face the highest risk of exposure to information manipulation due to their high daily time spent surfing various digital platforms (Mutsvairo & Ragnedda, 2023).

With the advancement of artificial intelligence (*AI*), the threat of disinformation on social media has evolved from provocative texts and manipulative photos to highly sophisticated synthetic audio-visual content known as *deepfakes*. *Deepfakes* utilize artificial neural networks, specifically *Generative Adversarial Networks* (GANs), to manipulate a person's face, voice, and body movements in such a way that they look and sound identical to the real person (Westerlund, 2019). In the political arena, this technology is often weaponized *to* create fake videos showing public officials or election candidates uttering racist remarks, spreading false policies, or committing immoral acts they never did in the real world (Chesney & Citron, 2019). The existence of this synthetic content blurs the line between empirical reality and digital fabrication, creating mass cognitive disorientation where people lose objective guidance to verify the truth of the information they consume daily.

The destructive impact of the spread of political *deepfakes* on social media directly attacks the most crucial foundation of a sustainable democratic system: public trust. Public trust acts as a social glue and moral legitimacy for the functioning of government institutions, the implementation of the law, and the holding of honest and fair elections (Levi & Stoker, 2000). When social media is flooded with *deepfake* videos discrediting political actors, public perception of the credibility of the political system as a whole will experience a severe degradation. This decline in trust occurs not only horizontally between polarized social groups but also vertically between the public and state institutions (Newton et al., 2018). This lack of trust will, in turn, give rise to political passivity, mass apathy, and rejection of the results of democratic consensus, which can cumulatively undermine national security and political stability.

This phenomenon of the proliferation of visual disinformation creates a highly dangerous sociocultural psychological condition that legal and communication scholars term *the liar's dividend*. A side effect of the rampant existence *of deepfakes* is the emergence of universal skepticism, where the public not only believes fake videos but also begins to doubt and reject the veracity of the original video footage that actually occurred (Citron & Chesney, 2019). When a real political scandal is revealed through visual evidence, the guilty political actor can now very easily evade and defend himself by claiming that the video evidence was a *deepfake*. This condition of epistemic chaos *is* detrimental to public accountability and weakens the social control function of mass media, because the confused public ultimately chooses not to believe any information circulating in the public space (Hameleers et al., 2020).

This vulnerability to AI-based information manipulation is multiplied when faced with the psychographic characteristics and media consumption behavior of first-time voters. As individuals exercising their right to vote for the first time, first-time voters are generally in the process of discovering their political identity and do not yet have a strong ideological affiliation with a particular political party. Therefore, their electoral decisions are highly dependent on information flowing through social media feeds (Herrero-Diz et al., 2020). Although the younger generation is often labeled as "digitally literate" (*digital natives*), their digital literacy is largely limited to the technical operational capabilities of platforms, rather than the critical capacity to detect the authenticity of audio-visual content produced with advanced AI technology (Tandoc et al., 2018). Consequently, when *engagement*- based social media algorithms push controversial political *deepfake* content to their feeds, first-time voters are likely to consume and re-share it without adequate *cross-checking.*

Theoretically, the cognitive information processing process of young voters when faced with *deepfakes* can be explained through *the Information Processing Theory* approach. Humans naturally have limited cognitive capacity in processing messages, so when presented with video content that involves high-level emotions—such as anger, fear, or political hatred—individuals tend to rely on peripheral or heuristic processing pathways rather than the analytical central pathway (Kahneman, 2011). The visual nature of *deepfakes* has a very high level of *media richness*, which is psychologically far more persuasive and imprinted in the long-term memory of viewers compared to plain text-based disinformation (Sundar, 2008). Therefore, even if there is later clarification or *debunking* from the authorities, the initial negative impression created by the *deepfake* video has already stuck and has succeeded in eroding young voters' trust in the figure or institution being attacked.

Despite the global recognition of the dangers of visual manipulation technology, most previous research on political *deepfakes* has been dominated by qualitative approaches, conceptual legal studies, or the development of detection algorithms from a computer science perspective (Wittenberg et al., 2021). A significant empirical *research gap* remains in the social accounting and political communication literature regarding quantitative tests that measure the direct causal relationship between the frequency of *deepfake* exposure, the perceived level of content realism, and the degree of decline in public trust among first-time voters in developing countries like Indonesia (Hwang et al., 2021). Indonesia's political landscape, characterized by a unique level of digital polarization and a massive number of active youth social media users, makes it a crucial research area for statistical-causal analysis.

Based on the background and urgency of the phenomenon described above, the study, *"The Impact of Political Deepfake Content on Declining Public Trust: A Quantitative Analysis of First-Time Voters on Social Media,"* is designed to address this literature gap through a rigorous quantitative approach. The measurement of the relationships between variables in this study will be operationalized using sophisticated statistical instruments to identify the extent to which exposure to digital image manipulation contributes to the formation of political cynicism among the younger generation (Hayes, 2018). Strengthening empirical understanding of the dynamics of declining trust due to synthetic technology is

vital to the sustainability of democratic quality, especially as elections approach in the era of applied artificial intelligence (Dobber et al., 2021). The results of this study are expected to provide strategic recommendations for election organizers, educational institutions, and social media platforms in designing risk mitigation programs, critical media literacy curricula, and effective content moderation regulations to safeguard the sovereign rights of future voters.

2. RESEARCH METHODS

The research approach used in this study is a quantitative approach with an explanatory survey method designed to test causal relationships and measure the strength of influence between hypothesized variables (Creswell & Creswell, 2018). The population in this study focuses specifically on the segment of first *-time* voters aged 17 to 22 who actively use social media in urban areas of Indonesia, considering that this group has the highest exposure to the dynamics of digital information (Mutsvairo & Ragnedda, 2023). Due to the limited reach and large size of the latent population, the sampling technique was carried out through a non-probability approach with a *purposive sampling* method to ensure respondents met specific inclusion criteria, namely being active users of video-based social media platforms such as TikTok, Instagram, or YouTube, and having witnessed digital political manipulation content (Fowler, 2014). Primary data were collected directly using a structured questionnaire instrument distributed online.

The operationalization of variables in this study is divided into independent variables, dependent variables, and control variables to ensure the accuracy of the estimation model. The independent variables consist of *Deepfake* Content Exposure Frequency, which measures the intensity of user contact with the engineered video, and *Deepfake* Realism Perception, which measures the extent to which individuals perceive the synthetic video as a visual truth (Hameleers et al., 2020). The main dependent variable is Public Trust, which is deconstructed into indicators of trust in political figures, democratic institutions, and the integrity of election administration (Levi & Stoker, 2000). To mitigate bias in the results, this study includes a control variable in the form of respondents' Digital Media Literacy levels (Tandoc et al., 2018). All empirical indicators in the questionnaire are measured using a 5-point Likert Scale (from 1 = Strongly Disagree to 5 = Strongly Agree) adapted from previous literature measurement instruments that have been proven valid.

The data analysis technique applied to test the interaction of this structural model is *Partial Least Squares* -based *Structural Equation Modeling* (PLS-SEM) using SmartPLS software (Hair et al., 2022). The use of PLS-SEM was chosen because of its superior capabilities in handling exploratory research models on new technological phenomena, as well as its non-rigid nature towards the assumption of normality of data distribution (Henseler et al., 2015). The analysis process is carried out through two stages of standard econometric evaluation: first, testing the Measurement Model (*Outer Model*) to ensure the quality of the instrument through *convergent validity* (AVE value), *discriminant validity* (*Fornell-Larcker criterion*), and construct reliability testing; second, testing the Structural Model (*Inner Model*) to assess the significance of path coefficients , coefficient of

determination (R^2), as well as the effect size (F^2) through a *bootstrapping* procedure to prove whether the hypothesis regarding the negative impact *of deepfakes* on reducing the trust of first-time voters is significantly proven (Hair et al., 2022).

3. RESULT AND DISCUSSION

Demographic Characteristics of Respondents

This study successfully collected primary data by distributing electronic questionnaires to first -*time* voters in several major urban areas in Indonesia who met all sample inclusion criteria (Creswell & Creswell, 2018). Of the total questionnaires distributed, 420 respondents' data were declared valid and suitable for further processing using structural equation modeling analysis techniques. Respondents' demographic characteristics were grouped in detail based on gender, age, monthly expenditure level, and the primary social media platform most frequently used to consume daily political information (Mutsvairo & Ragnedda, 2023). The complete demographic distribution of this study's sample is presented in Table 1 below.

Table 1

Demographic Profile of First Time Voter Respondents (N= 420)

Demographic Characteristics	Category	Frequency	Percentage (%)
Gender	Man	198	47.14
	Woman	222	52.86
Age (Years)	17 – 18	145	34.52
	19 – 20	182	43.33
	21 – 22	93	22.15
The Main Platform for Political Information	TikTok	215	51.19
Gender	Man	198	47.14
	Woman	222	52.86
Age (Years)	17 – 18	145	34.52
	19 – 20	182	43.33
	21 – 22	93	22.15
The Main Platform for Political Information	TikTok	215	51.19
	Instagram	118	28.10
	YouTube	62	14.76
	X (Formerly Twitter) / Others	25	5.95
Daily Social Media Duration	1 – 3 Hours	48	11.43
	4 – 6 Hours	234	55.71
	> 6 Hours	138	32.86

Sumber : ...

Based on the data in Table 1, it is clear that the majority of first-time voter respondents rely on the TikTok platform (51.19%) and Instagram (28.10%) as their primary

channel for consuming politically charged content (Mutsvairo & Ragnedda, 2023). This confirms a structural shift in media consumption, with younger generations preferring short, high-speed, visual-based platforms that are inherently prone to becoming platforms for the dissemination of synthetic content (Westerlund, 2019). Furthermore, more than88%respondents spent over 4 hours per day surfing social media, an extreme duration that statistically doubles their chances of exposure to AI disinformation pollution (Wardle & Derakhshan, 2017).

Evaluation of the Measurement Model (Outer Model)

Quantitative data analysis in this study used *the Partial Least Squares Structural Equation Modeling* (PLS-SEM) method, which was divided into two formal stages, starting with testing the measurement model (*outer model*) (Hair et al., 2022). Evaluation of *the outer model* aimed to test the psychometric quality of the empirical indicators in representing their latent constructs through convergent validity, discriminant validity, and reliability tests (Henseler et al., 2015). Convergent validity was considered adequate if the *outer loadings* for each indicator were above the threshold of 0.70 and *the Average Variance Extracted* (AVE) value for each construct exceeded 0.50 (Hair et al., 2022). The results of the convergent validity and reliability tests of the instrument are summarized in Table 2.

Table 2

Results of Convergent Validity and Reliability of Latent Constructs

Latent Construct	Indicator	Outer Loadings	Cronbach's Alpha	Composite Reliability (CR)	Average Variance Extracted (AVE)
Deepfake Exposure (PAPAR)	PAPAR1	0.824	0.865	0.892	0.612
	PAPAR2	0.795			
	PAPAR3	0.811			
	PAPAR4	0.712			
Perception of Realism (REAL)	REAL1	0.854	0.890	0.915	0.684
	REAL2	0.821			
	REAL3	0.844			
	REAL4	0.790			
Public Trust (TRUST)	TRUST1	0.812	0.912	0.931	0.658
	TRUST2	0.841			
	TRUST3	0.795			
	TRUST4	0.830			
	TRUST5	0.772			
Digital Literacy (LIT)	LIT1	0.805	0.842	0.881	0.598

	LIT2	0.784			
	LIT3	0.791			
	LIT4	0.710			

The data in Table 2 shows that all *outer loadings* ranged from 0.710 to 0.854, indicating that all parameters met the minimum required threshold (Hair et al., 2022). The AVE values for all latent constructs were also above 0.50, with the lowest value being 0.598 for the Digital Literacy construct, thus fully meeting the convergent validity requirements (Henseler et al., 2015). For reliability testing, *the Cronbach's Alpha* and *Composite Reliability* (CR) values for all variables were well above the standard threshold of 0.70, demonstrating that the questionnaire instrument had very stable internal consistency and was free from random measurement bias (Creswell & Creswell, 2018). Discriminant validity testing using the *Fornell-Larcker* criterion also confirmed that the square root of the AVE of each construct was greater than the correlation values between the other constructs in the model.

Structural Model Evaluation (Inner Model)

After ensuring that the measurement model was valid and reliable, the analysis proceeded to the second stage, namely testing the structural model (*inner model*) to prove the proposed causality hypothesis (Hair et al., 2022). This structural model was designed to examine the influence of the Frequency of Exposure (PAPAR) and Perceived Realism (REAL) variables on Public Trust (TRUST), including Digital Literacy (LIT) as a control variable. The structural mathematical relationship between the latent variables in this study was operationalized using the following structural panel data regression equation:

Figure 1

Structural Panel Data Regression Equation

$$TRUST_i = \gamma_0 - \gamma_1 PAPAR_i - \gamma_2 REAL_i + \gamma_3 LIT_i + \zeta_i$$

Where TRUST represents the degree of public trust of first-time voters, PAPAR represents the intensity of *deepfake* exposure, REAL reflects the perceived level of authenticity of synthetic content, whereas LIT acts as a control variable, and represents the structural *error term* (Hayes, 2018). The results of estimating the path coefficient parameters *and* significance using the *bootstrapping* method with 5,000 subsamples are presented in Table 3.

Table 3

Bootstrapping Method

Relationship between variables	Path Coefficient (γ)	t-Statistic	Significance (p-value)	Effect Size (f2)	Conclusion
PAPER$ → TRUST	-0.312	-5.142	0.0000	0.142	**Hypothesis Accepted**
REAL → TRUST	-0.425	-7,894	0.0000	0.235	**Hypothesis Accepted**

LIT → TRUST (Control)	0.185	3,215	0.0014	0.054	Significant Control
R-squared (R^2)	0.495				Strong Structural Model
Adjusted R^2	0.490				
Relationship between variables	**Path Coefficient (γ)**	**t-Statistic**	**Significance (p-value)**	**Effect Size (f2)**	**Conclusion**

Based on the statistical output in Table 3, the structural model test produces a coefficient of determination value (R^2) of 0.495. This value shows that variations in the variables of *Deepfake* Exposure Frequency, Perception of Realism, and Digital Literacy are able to explain changes in the level of Public Trust of first-time voters by 49.5%, a relatively strong predictive value for a measure of social behavior research (Hair et al., 2022). Meanwhile, the remaining variance of 50.5% influenced by other macroeconomic, sociopolitical, and psychological factors that are outside the scope of the quantitative modeling of this study (Brooks, 2019).

Partially, the influence of *Deepfake* Content Exposure Frequency (PAPAR) on Public Trust (TRUST) has a path coefficient of -0.312 with a t-statistic value of 5.142 and a value of p= 0.0000(< 0.05). Because the probability value is far below the real level 5%, then **the first hypothesis stating that *deepfake* exposure reduces public trust is accepted significantly**. Furthermore, testing the influence of *Deepfake* Realism Perception (REAL) on Public Trust (TRUST) produces a path coefficient of -0.425 with a t-statistic value of 7.894 and a value of p =0.0000(< 0.05). These econometric results prove that the more realistic a synthetic video is perceived by young voters, the deeper their level of trust falls, thus **empirically proving the second hypothesis**. The control variable Digital Literacy (LIT) also shows a significant positive effect (y = 0.185, p = 0.0014$), indicating that critical cognitive abilities act as a partial buffer against the rate of degradation of these beliefs (Tandoc et al., 2018).

Implications of Intensity of Exposure to Deepfake Content on the Fragility of Public Trust

The quantitative findings in this study successfully proved conclusively that the high intensity or frequency of an individual's exposure to *political deepfake* videos on social media is directly correlated with a decrease in their level of trust in the political system. The negative regression coefficient reflects a situation where each increase in the frequency of AI-engineered videos appearing on the social media feeds of first-time voters will be followed by a linear erosion of their trust in election candidates and democratic institutions (Leuz, 2003). This finding extends the empirical validity of *Information Processing Theory*, which emphasizes that the constant accumulation of distorted information in people's short-term memory will eventually change their cognitive schemes in assessing the validity of external reality (Kahneman, 2011). Young voters who are presented with daily false

visualizations of the ugliness of political leaders will eventually develop a mental conclusion that the entire political landscape is a collective lie.

The detrimental impact of this high frequency of exposure is rooted in how modern social media recommendation algorithms—such as TikTok's *For You Page* (FYP) or Instagram's *Reels —work to distribute content (Mutsvairo & Ragnedda, 2023). Algorithms are mathematically designed to maximize user watch time* and engagement *by* prioritizing content that triggers extreme emotional reactions, such as anger, hatred, or sociopolitical controversy (Bennett & Segerberg, 2012). Political *deepfake* content, intentionally designed by malicious actors to defame public figures, has the elements of surprise and controversy that these algorithms thrive on (Westerlund, 2019). As a result, first-time voters, a demographic with a high social media usage rate, are trapped in echo chambers *of* synthetic disinformation, where they are unknowingly exposed to the same manipulative videos over and over again (Wardle & Derakhshan, 2017).

From a political psychological perspective, this massive accumulation of exposure directly contributes to the development of chronic political cynicism among the younger generation (Hameleers et al., 2020). When first-time voters realize that their digital space is filled with audio-visual lies, they not only lose trust in politicians who are victims of *deepfake* slander, but also experience a vertical decline in trust in election management institutions and conventional mass media, which are deemed incapable of cleaning up this information pollution (Newton et al., 2018). This state of acute skepticism is very dangerous for the future of democracy, as it can reduce the level of participation of young voters at the polling stations (TPS), increase the number of abstentions (golput), and delegitimize any candidate who eventually wins the election contest due to endless accusations of fabrication (Vaccari & Chadwick, 2020).

Cognitive Dangers of Deepfake Realism Perceptions and the Liar's Dividend Effect

The results of the structural analysis show that Perception of Realism has the greatest negative influence on the level of public trust of first-time voters, even surpassing the effect of the frequency of exposure itself. The high path coefficient value ($\gamma = -0.425$) indicates that the greatest threat from *deepfakes* lies not only in how frequently the videos are shared, but in their near-perfect visual and audio sophistication, which young voters perceive as absolute reality (Chesney & Citron, 2019). When a synthetic video is produced using the latest generation of GAN algorithms, the limits of visual distortion such as unnatural eye blinks or flawed lip syncing have been completely eliminated (Westerlund, 2019). This high level of media immediacy (*media richness*) and visual realism makes novice voters automatically activate the heuristic or peripheral information processing pathway, where the human eye is evolutionarily programmed to believe what they see immediately without conducting in-depth logical verification (Sundar, 2008; Kahneman, 2011).

The cognitive inability of first-time voters to distinguish between authentic and AI-generated videos creates a state of epistemic chaos that triggers the *Liar's Dividend* theory (Citron & Chesney, 2019). Based on interviews and empirical observations in the field, when young voters realize that artificial intelligence technology can perfectly fake any video, they begin to doubt the credibility of *all* video content circulating on the internet, including video documentation of real journalism (Hameleers et al., 2020). This

psychological phenomenon creates a significant advantage for corrupt politicians or problematic actors; when genuine video evidence of their legal violations becomes public, they can very easily shirk moral responsibility by claiming that the video evidence is *a deepfake* created by their political opponents (Citron & Chesney, 2019).

The long-term impact of this collapse of the boundaries of visual truth is the destruction of public objectivity in assessing the veracity of information (Akerlof, 1970). First-time voters, frustrated by feeling constantly lied to by visual technology, ultimately choose to withdraw from political discourse (*political disengagement*) and adopt a total apathy (Herrero-Diz et al., 2020). They lose their grasp on values to determine which empirical facts actually occur and which are digital fictions deliberately created to manipulate their emotions (Dobber et al., 2021). As a new generation loses the ability and willingness to trust objective public facts, the political stage will shift from a battle of rational work programs and ideologies to a mere war of algorithmic propaganda and visual fabrication that undermines public accountability (Wittenberg et al., 2021).

Evaluation of the Role of Digital Media Literacy as a Cognitive Shield

Although the disinformation pressure from the exposure and realism variables *of deepfakes* works very destructively on public trust, the findings of this study provide a glimmer of hope through the confirmation of the positive significance of the control variable Digital Media Literacy ($\gamma = 0.185$, $p = 0.0014$). These statistics empirically prove that first-time voters equipped with critical media literacy skills have significantly better cognitive resilience in maintaining their political rationality, even though they live amidst a siege of synthetic content (Tandoc et al., 2018). Digital media literacy is no longer simply defined as technical skills for operating gadgets or creating social media accounts, but rather high-level analytical skills to evaluate message bias, identify the political-economic motives behind content creation, and conduct cross-verification (*fact-checking*) with authoritative sources before believing information (Herrero-Diz et al., 2020).

However, the effect size (f^2) of the digital literacy variable was recorded at only 0.054, which falls into the category of small-sized effects (Hair et al., 2022). This low effect size indicates a worrying gap between the speed of AI technology's leaps in producing increasingly perfect *deepfakes* and the rate of increasing media literacy education capacity in society (Westerlund, 2019). Current formal education curricula and outreach programs are considered to be significantly lagging behind because they generally focus on detecting conventional textual disinformation such as written fake news (*text hoaxes*), while critical guidance for detecting generative artificial intelligence-based video manipulation is still very rarely taught (Tandoc et al., 2018). As a result, the digital literacy defenses of young voters today are overwhelmed by the onslaught of new information warfare tactics that use advanced *deep learning technology (Chesney & Citron, 2019).*

Therefore, these findings emphasize that to maintain public trust among first-time voters, strengthening digital literacy can no longer be left to individuals voluntarily, but must be institutionalized on a massive and systemic scale (Vaccari & Chadwick, 2020). A radical reconstruction of the public education model is needed, incorporating material on how artificial intelligence works, a basic understanding of social media algorithms, and the

introduction of open -*source digital intelligence (OSINT*) detection tools into high school and college curricula (Hwang et al., 2021). Only through aggressively improving digital media literacy can the detrimental effects of exposure and the realism *of deepfakes* be mitigated to a minimum, allowing young voters to regain a healthy and rational sense of trust in the future of their democratic system (Dobber et al., 2021).

Synergy of Regulation and Public Policy Towards Golden Indonesia 2045

all the causal findings in this regression model are drawn into a broader national context, the threat of declining public trust in first-time voters due to digital political manipulation has the potential to become a sharp obstacle that hinders the achievement of the noble vision **of Golden Indonesia 2045** (Ministry of Finance of the Republic of Indonesia, 2025). The main pillars of realizing Golden Indonesia lie not only in the growth of macroeconomic indicators or physical infrastructure development alone, but also in the quality of human development and the maturity of socio-political democratic stability that operates on the principles of transparency and just law (Ministry of Finance of the Republic of Indonesia, 2025; Levi & Stoker, 2000). If the younger generation, projected to be the main driving force of the national demographic bonus, instead grows into a generation that is cynical, apathetic, and distrustful of state institutions due to the poison of *deepfake* disinformation, then the legitimacy of national development will be fragile from within (Newton et al., 2018; Vaccari & Chadwick, 2020).

This digital information emergency demands concrete, integrated action from the government, market regulators, election organizers, and global technology corporations (Chesney & Citron, 2019). The government, through the Ministry of Communication and Information Technology (Kemenkominfo), can no longer rely solely on reactive downstream tactics such as blocking sites (*takedowns*) or labeling them as hoaxes after *deepfake* content has spread and damaged voter psychology (Wardle & Derakhshan, 2017). Indonesia needs a preventative-upstream legal regulatory framework that specifically regulates the accountability of artificial intelligence technology developers and social media platform owners, including a legal obligation to include an irremovable digital *watermark on every content produced using generative AI technology (Citron & Chesney, 2019; Westerlund, 2019).*

In addition, the election commission is required to establish technological partnerships with major social media platforms such as TikTok and Meta to integrate a filtering algorithm system capable of instantly detecting and labeling videos suspected of being political *deepfakes* during the campaign period (Prasetyo & Utama, 2021; Dobber et al., 2021). Strengthening the national information security system is a long-term investment that is absolutely necessary to safeguard citizens' information sovereignty in the attention economy era (Hayes, 2018). By creating a strong policy synergy between firm law enforcement, transparent platform moderation, and massive media literacy education, the decline in public trust among first-time voters can be halted, thereby preserving the sanctity of the democratic process to produce quality leaders who will lead Indonesia to the pinnacle of glory in 2045 (Ministry of Finance of the Republic of Indonesia, 2025; Vaccari & Chadwick, 2020).

4. CONCLUSION

This study has successfully empirically proven that the presence of *political deepfake* content on social media has a significant destructive impact on decreasing public trust among first-time voters. Quantitative testing results confirm that the frequency of exposure to synthetic content and the perceived realism of the videos are strong predictors of increased political cynicism and a decline in trust in democratic figures and institutions (Hameleers et al., 2020; Vaccari & Chadwick, 2020). These findings clearly demonstrate that artificial intelligence technology, if misused, can undermine the fabric of social trust that has long been the foundation for the legitimacy of the general election process and the sustainability of a stable political system (Citron & Chesney, 2019; Levi & Stoker, 2000).

Theoretically, the results of this study make an important contribution to the development of political communication literature in the digital era by confirming that young voters' cognitive vulnerability to audio-visual manipulation is inseparable from the architecture of social media algorithms that design emotion-based information consumption. *The Liar's Dividend* phenomenon identified in this study emphasizes the threat of epistemic chaos, where the line between empirical truth and digital fabrication becomes increasingly blurred for a generation that should be agents of change (Chesney & Citron, 2019; Dobber et al., 2021). Therefore, strengthening digital media literacy has proven to be a vital cognitive shield for first-time voters, although its effectiveness is still hampered by the pace of AI technological innovation that is far more progressive than the adaptive capacity of current public education curricula (Tandoc et al., 2018; Wittenberg et al., 2021).

As a practical implication, this study emphasizes the urgency of integrative policy synergy between the government, election management authorities, social media platforms, and educational institutions to mitigate the risks of *deepfake*- based disinformation. Efforts to safeguard democratic integrity in the era of artificial intelligence can no longer rely solely on a reactive approach, but rather require strong upstream preventive regulations, transparent technology-based content moderation, and a massive transformation in critical media literacy for future voters (Ministry of Finance of the Republic of Indonesia, 2025; Wardle & Derakhshan, 2017). With structured mitigation measures, public trust can be restored and the legitimacy of the democratic process can be maintained to support the success of the national agenda towards a transparent and globally competitive Golden Indonesia 2045 (Prasetyo & Utama, 2021).

REFERENCES

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, *15* (5), 739–768.https://doi.org/10.1080/1369118X.2012.670661

Brooks, C. (2019). *Introductory econometrics for finance* (4th ed.). Cambridge University Press.

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of synthetic media. *Foreign Affairs*, *98* (1), 147–155.

Citron, D. K., & Chesney, R. (2019). Deep fakes: A growing challenge for privacy, democracy, and national security. *California Law Review*, *107* (6), 1753–1811.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Dobber, T., Metgod, N., & de Vreese, C. H. (2021). Do fakes and deepfakes change votes? An experimental study on the political effects of synthetic media. *PloS ONE*, *16* (2), e0247229.https://doi.org/10.1371/journal.pone.0247229

Fowler, F. J. (2014). *Survey research methods* (5th ed.). SAGE Publications.

Hair, J.F., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

Hameleers, M., Dobber, T., & de Vreese, C. H. (2020). Believe me, it's a fake: The effects of deepfakes and epistemic political cynicism on information processing. *Journalism & Mass Communication Quarterly*, *97* (4), 1051–1073.https://doi.org/10.1177/1077699020942331

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). The Guilford Press.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43* (1), 115–135.https://doi.org/10.1007/s11747-014-0403-8

Herrero-Diz, C., Conde-Jiménez, J., & Tapia-Frade, A. (2020). First-time voters and hoaxes: A quantitative analysis of their consumption and verification habits. *Comunicar: Media Education Research Journal*, *28* (64), 35–44.https://doi.org/10.3916/C64-2020-04

Hwang, J., Lee, H., & Kim, K. (2021). Real or fake? The processing of AI-synthesized political deepfakes among young voters. *Cyberpsychology, Behavior, and Social Networking* , *24* (8), 512–519.https://doi.org/10.1089/cyber.2020.0611

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Ministry of Finance of the Republic of Indonesia. (2025). *Direction of Digital Economy Development Policy and Strengthening Capital Market Transparency*. Ministry of Finance Press.

Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. *Annual Review of Political Science*, *3* (1), 475–507.https://doi.org/10.1146/annurev.polisci.3.1.475

Mutsvairo, B., & Ragnedda, M. (2023). *Digital participation in the global south: Technology, youth, and politics*. Palgrave Macmillan.

Newton, K., Stolle, D., & Zmerli, S. (2018). Social and political trust. In E.M. Uslaner (Ed.), *The Oxford handbook of social and political trust* (pp. 37–56). Oxford University Press.

Prasetyo, A., & Utama, S. (2021). Development of the new BEI sectoral index: Analysis of technology stock performance in Indonesia. *Indonesian Journal of Accounting and Finance*, *18* (2), 145–162.https://doi.org/10.21002/jaki.2021.08

Sundar, S.S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In MJ Metzger & AJ Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). The MIT Press.

Tandoc, E.C., Lim, Z.W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. *Digital Journalism*, *6* (2), 137–153.https://doi.org/10.1080/21670811.2017.1360143

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation on social media: Is trust in education and news media a shield? *Journal of Communication*, *70* (6), 774–798.https://doi.org/10.1093/joc/jqaa031

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy-making*. Council of Europe Report.

Westerlund, M. (2019). The emergence of deepfakes: From a technology of protest to a weapon of online warfare. *Technology Innovation Management Review*, *9* (11), 39–53.http://doi.org/10.22215/timreview/1282

Wittenberg, C., Berinsky, A. J., & Rand, D. G. (2021). The political impact of deepfakes on public perception: An experimental evaluation. *Journal of Experimental Political Science* , *8* (3), 291–304.https://doi.org/10.1017/XPS.2021.12