Published online on the page : https://journal.makwafoundation.org/index.php/intellect

I n t e l l e c t :
Indonesian Journal of Learning and Technological Innovation

| ISSN (*Online*) 2962-9233 |

Reported Practices for Verifying AI-Generated Disinformation: A Qualitative Study of Indonesian Social Media Users

Saona[1], Nabilaa Faizatuz Zuhriyah[2], Nur Linda[3], Elvira Fitriyanti[4]*

[1]Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Cirebon, Indonesia
[2]Universitas Islam Bunga Bangsa Cirebon, Cirebon, Indonesia
[3]Institut Islam Kuningan, Kuningan, Indonesia
[4]Politeknik Siber Cerdika Internasional, Cirebon, Indonesia

Article Information

Article History:
Submitted: October 27, 2026
Revised: November 3, 2026
Accepted: November 4, 2026
Published: November 4, 2026

Keywords

Digital literacy, generative AI, disinformation, fact-checking, reported verification practices.

Correspondence

E-mail: saona0206@gmail.com*

A B S T R A C T

Content generated by AI can appear credible even when the claims it presents lack reliable supporting evidence. This qualitative descriptive study examines how 32 purposively selected adult social-media users from six Indonesian cities described recognizing potentially synthetic material, assessing credibility, and verifying information. Of these, 28 also participated in four focus group discussions. The study reports thematic analysis of interview, group discussion, and documentary material, although the original transcripts and coding records were unavailable for independent inspection. Participants sometimes relied on polished visuals, widespread sharing, familiar sources, and prior beliefs. They also reported source tracing, comparison with independent reports, contextual appraisal, and use of fact-checking services, although they applied these practices inconsistently. These accounts informed a provisional educational framework covering AI knowledge, critical verification, psychological awareness, digital ethics, and access to verification tools. The framework focuses on the difference between suspecting that content is synthetic and verifying the associated claim. The proposal may inform non-formal digital-literacy education for adults, but neither the code-to-theme analysis nor any learning effects have been independently validated.

Abstrak

Konten yang dihasilkan AI dapat tampak meyakinkan tanpa memberikan bukti yang dapat dipercaya atas klaim yang disampaikan. Studi deskriptif kualitatif ini mengkaji bagaimana 32 pengguna media sosial dewasa yang dipilih secara purposif dari enam kota di Indonesia menggambarkan pengenalan konten sintetis, penilaian kredibilitas, dan verifikasi informasi; 28 orang juga mengikuti empat diskusi kelompok terarah. Naskah sumber melaporkan wawancara, diskusi kelompok, dan dokumen pendukung yang diorganisasikan melalui analisis tematik, tetapi transkrip dan rekam pengkodean asli belum tersedia untuk pemeriksaan independen. Informan menggambarkan penggunaan tampilan visual, popularitas konten, sumber yang dikenal, dan keyakinan sebelumnya sebagai petunjuk kredibilitas. Mereka juga melaporkan penelusuran sumber, pembandingan laporan independen, pemeriksaan konteks, dan penggunaan layanan pemeriksa fakta, meskipun penerapannya tidak konsisten. Uraian tersebut menjadi dasar usulan awal lima dimensi pendidikan, yaitu pengetahuan AI, verifikasi kritis, kesadaran psikologis, etika digital, dan akses alat verifikasi. Fokus khususnya ialah perbedaan antara mengenali kemungkinan konten sintetis dan memeriksa kebenaran klaim. Usulan ini dapat mendukung perancangan literasi digital masyarakat, tetapi rantai bukti tematik dan dampak pembelajarannya belum divalidasi.

This is an open access article under the CC–BY-SA license

1. Introduction

Digital platforms allow information to circulate rapidly while making the origins of individual claims harder to trace [1]. In January 2024, Indonesia had approximately 185.3 million internet users

 Creative Commons Attribution-ShareAlike 4.0 International License. *Some rights reserved*

and 139.0 million active social-media user identities [2]; these figures indicate the scale of digital access, not the prevalence of misinformation. Generative AI also makes it easier to produce plausible text, images, audio, and video, increasing the importance of evidence-based verification [3].

The realism of synthetic media can complicate judgments of authenticity, but detecting manipulation and determining whether an accompanying claim is true are distinct tasks [4], [5], [6]. Controlled studies of political-speech deepfakes and synthetic media show the limits of unaided perceptual judgments, but their accuracy estimates do not apply to participants in this qualitative study [7], [8]. Deepfakes also raise questions about trust, context, and the social meaning of evidence beyond visual realism [9].

Credibility judgments occur within social relationships and information-sharing environments. Research on synthetic political videos documents changes in uncertainty and trust [10], while large-scale diffusion research describes the spread of false and accurate news without isolating generative-AI effects [11]. Reviews of misinformation psychology further distinguish cognitive, social, and affective influences on belief from actual sharing behavior [12]. These distinctions guide the interpretation of participants' reports rather than serving as measured predictors in this study.

Indonesian evidence highlights the importance of context. A national survey of rural social-media users identifies variation in critical use [13], while studies of misinformation content and user responses illustrate the range of information environments in which claims circulate [14], [15]. Research on Indonesian Generation X users examines deepfake identification as a measurable task [16]; separate surveys associate information and media literacy with judgments about hoaxes and sharing intentions [17], [18]. Their populations, instruments, and outcomes differ from the purposively selected interviewees considered here, and none can replace an analysis grounded in the original interviews.

Intervention research provides a different kind of evidence. An Indonesian experiment using brief media-literacy videos examined intentions to share misleading content [19]. Experimental work in the United States and India found that the effects of short literacy guidance varied across participant settings [20]. Meta-analyses report broadly positive but outcome- and context-dependent effects, with measured knowledge not necessarily translating into behavior [21], [22]. No educational intervention was tested in the present study; its potential contribution concerns reported needs that might inform the design of subsequent training.

Familiarity with digital media should not be equated with the ability to verify information. A comparison of literacy constructs found that information literacy was more strongly associated with recognizing false news than several other self-reported literacies [23]. Experimental work links attention to accuracy with sharing choices and reliance on emotion with susceptibility to false headlines, but the present interviews did not measure either mechanism [24], [25]. Alongside studies of synthetic-video credibility [10] and misinformation psychology [26], these findings suggest alternative explanations, not evidence of causal influences among the participants in this study.

Recent frameworks describe AI-related competencies for different audiences and purposes. Fulsher and colleagues examine generative AI in misinformation education [27], Park reviews GenAI literacy in schools [28], and Chee and colleagues discuss competency pathways [29]. UNESCO's student framework foregrounds responsible knowledge and use [30]. Their conceptual categories overlap with those proposed in this manuscript. The question, then, is not whether knowledge, ethics, or critical evaluation are novel competencies, but whether the participants' accounts suggest a defensible configuration of priorities for learning to verify information for adults outside formal education.

Education about misinformation and disinformation among adults takes various forms, while its long-term and community-level evaluation remains uneven [31]. A systematic review differentiates

effects across intervention designs [32], and an AI-literacy meta-analysis reports substantial heterogeneity across outcomes and settings [33]. A broader toolbox distinguishes verification, accuracy prompts, inoculation, and labeling by their intended mechanisms [34]. A validated misinformation-discernment measure offers a further contrast with self-reported capability [35]. These studies do not explain how the interviewees themselves negotiated visual plausibility, social pressure, and access to checking tools; understanding that connection motivates this study.

Research on online source evaluation offers ways to translate these questions into assessable learning activities. Observations of professional fact-checkers emphasize lateral reading rather than trusting the appearance of a webpage [36], and instructional research has tested how those practices can be taught [37]. Practical debunking guidance [38] and inoculation approaches that prepare learners to recognize manipulation tactics [39], [40] offer complementary mechanisms. Indonesian research on professional fact-checking procedures adds an institutional perspective [41]. These approaches were not documented in the fieldwork. Instead, this paper considers whether participants' accounts point to priorities for adapting them to non-formal adult learning.

The study addresses four questions: (1) How do participating social media users understand AI-generated disinformation and attempt to recognize it? (2) What cognitive, technical, and social factors do they associate with vulnerability? (3) Which verification practices and obstacles do they describe? and (4) What implications do these accounts suggest for a contextual digital-literacy framework?

The intended contribution is a focused account of how selected adult social-media users describe the shift from recognizing a potentially synthetic item to checking the evidence behind its claims. Their reported barriers are organized into a provisional five-dimensional set of learning priorities. This interpretation may inform contextual task design, but the study did not develop a module, observe performance, or measure learning outcomes. Evidence from Indonesian interventions may inform future testing, but it does not validate the present proposal [19].

2. Research methodology

A qualitative descriptive design was used to explore how participants described encountering questionable AI-related content, judging its credibility, and deciding whether to verify or share it [42], [43]. The four research questions guided the organization of interviews and FGD accounts into recognition, perceived vulnerability, verification practices, and potential educational implications. Each participant's account was the unit of interpretation, rather than a standardized detection score. Documents provided background on fact-checking services, but the available records do not establish whether they independently corroborated particular participant statements.

The study reports 32 purposively selected adult social-media users in Jakarta, Surabaya, Medan, Semarang, Makassar, and Denpasar. The stated eligibility criteria were an age of 18-55 years, at least two hours of social-media use per day, and prior exposure to potentially misleading material. Twenty-eight of these 32 participants also joined four FGDs of seven participants each; they were not a separate sample. Recruitment channels, group allocation, city-by-city characteristics, sample-sufficiency reasoning, and the fieldwork period are not reported in the available materials. These omissions limit assessment of sample variation and require confirmation from original study records.

Three sources of material were reported: individual interviews, four FGDs, and digital-literacy or fact-checking documents, including Mafindo and Cek Fakta resources. Interview topics concerned recognition, credibility cues, previous checking attempts, and decisions about sharing; The FGDs explored related issues in a group setting. The present record does not identify which statements came from which source type, how group disagreement was documented, or which documents were included in the analysis. Interview/FGD prompts, session details, translation and recording procedures, and a documentary inventory must be provided from authentic research records; these details have not been reconstructed.

The study describes thematic analysis through familiarization, coding, theme development, review, definition, and reporting, consistent with the approach developed by Braun and Clarke [44]. The reported results are organized around four inquiry areas; they are not accompanied by a verifiable codebook. No transcripts, researcher memos, code-to-excerpt examples, decisions about contradictory cases, or documentation of whether coding was inductive or question-guided were available. The links between participant quotations, summary categories, and the proposed five dimensions therefore remain preliminary interpretations rather than an independently auditable analytical chain.

Qualitative credibility requires clear distinctions between original accounts, researchers' interpretations, and any independent records. Using several data-collection formats does not, by itself, demonstrate formal triangulation, participant checking, reflexivity, or inter-coder agreement. The available documents do not establish which safeguards were used, including procedures for informed consent, recruitment permissions, data anonymization, retention, and ethics approval or exemption. The authors must supply accurate details from contemporaneous records. Accordingly, the manuscript makes no retrospective claim that specific quality-assurance or ethical procedures took place.

3. Results and Discussion

3.1. Results

3.1.1. Reported Recognition of AI-Generated Disinformation

The interview summaries suggest varying familiarity with generative AI and uncertainty about recognizing synthetic images or videos. Some informants knew the term "deepfake" without identifying a clear way to check particular claims; others treated appearance or popularity as cues to credibility. These are descriptive interpretations of reported experiences, not measured detection rates.

This finding is reflected in the statement of a 28-year-old female informant from Jakarta who described her experience upon seeing an AI-generated video: *"I thought the video was completely real. Her face looked perfect and her movements were natural. It wasn't until I was told the video was created using AI that I noticed a slight inconsistency around the eyes. However, if you don't look closely, that inconsistency isn't noticeable." Another informant said: "I know there's AI technology that can create fake videos, but I don't know how to tell the difference between real and fake ones. Usually, I trust a video if it looks clear and has been shared by many people."*

These two excerpts illustrate how polished presentation and repeated sharing can serve as credibility cues. Their sources, participant codes, accuracy as verbatim quotations, and any translation from Indonesian must be checked against the original recordings or transcripts before submission. The quotations cannot establish how common these views were across the 32 informants. They are treated as illustrative cases rather than evidence of theme prevalence or detection accuracy.

Table 1. Findings on Users' Understanding of AI-Based Disinformation

Aspect	Reported finding	Interpretive concern
AI familiarity	General awareness of AI and deepfakes	Synthetic-media features not well understood
Visual cues	Reliance on image and video quality	Manipulation not consistently recognized
Perceived credibility	Popularity and sharing used as cues	Social approval mistaken for evidence
Verification	Authenticity checks used inconsistently	Practical verification skills need support

The most defensible finding is that participants described uncertainty and the use of surface cues. The study cannot determine how frequently these responses occurred or whether any particular social or cognitive factor caused mistaken judgments.

3.1.2. Reported Factors Associated with Vulnerability

The accounts point to five recurring concerns: limited understanding of synthetic content, trust in belief-consistent information, reliance on visual polish, dependence on social-media feeds, and social pressure to share. These are descriptive categories, not five measured or independent predictors.

Participants also described engagement levels and familiar contacts as cues to trustworthiness.

Table 2. Factors Contributing to Vulnerability to AI-Based Disinformation

Factor	Reported characteristic	Possible interpretation
Limited AI knowledge	Difficulty understanding synthetic content	Technical knowledge gaps
Belief-consistent trust	Greater trust in agreeable claims	Possible influence of prior beliefs (not a measured bias)
Visual credibility	Realistic images seen as authentic	Reliance on surface appearance
Platform reliance	Social media as a primary source	Need for independent corroboration
Social pressure	Sharing encouraged within groups	Social norms shape decisions

3.1.3. Reported Verification Practices and Barriers

The summaries describe four verification practices: checking an originating account, comparing independently published reports, examining the claim and its context, and consulting fact-checking tools. These are self-reported practices, not demonstrations of accurate verification.

Participants reported using these practices inconsistently, often citing time pressure, difficulty identifying trustworthy sources, unfamiliarity with tools, or habitual forwarding. The reasons were reported in interviews or FGDs, not experimentally manipulated.

One informant explained the strategy they use when encountering surprising information: *"If there's surprising news, I usually check it first on Google or see if major media outlets like Kompas or Detik are also reporting it. If they aren't, I start to suspect that the news is a hoax."*

Meanwhile, another informant explained that using fact-checking services is still considered quite complicated: *"I've heard of Mafindo, but I've never used it. It seems complicated to have to check everything one by one. Sometimes information comes in so quickly that I don't have time to check it before sharing it."*

The excerpts illustrate contrasting responses: seeking corroboration and reluctance to use a checking procedure perceived as cumbersome. Source identifiers, verbatim accuracy, and translation procedures require verification against original transcripts. The excerpts are retained without invented codes or frequency estimates.

Table 3. Digital Literacy Strategies and Information Verification Practices

Strategy	Reported practice	Reported barrier
Source checking	Review account identity and origin	Popularity can substitute for credibility
Cross-checking	Compare reports across outlets	Reliable sources not always clear
Content review	Examine logic and context	Requires critical evaluation
Fact-checking	Use Mafindo or reverse-image search	Time and tool familiarity barriers

The findings capture how participants described their information-checking habits. No controlled activity measured whether they could complete source tracing, corroboration, or claim evaluation accurately. Future work should assess those skills directly rather than infer competence from stated intentions.

3.1.4. Provisional Digital-Literacy Framework Based on Reported Themes

The four areas of reported experience informed a provisional five-dimensional proposal for community digital-literacy education: (1) AI knowledge, (2) critical verification, (3) psychological awareness, (4) digital ethics, and (5) access to verification tools. These dimensions are proposed educational categories, not five empirically validated factors. Their separation reflects different possible learning tasks rather than proof of independent constructs.

AI knowledge concerns what synthetic-media systems can and cannot establish. Critical verification requires tracing sources and corroborating claims even when a post appears authentic. Psychological awareness involves reflecting on belief-consistent trust, emotion, and uncertainty; digital ethics concerns responsible sharing, correction, and effects on others. Tool access refers to finding and using checking services, whereas verification competence also requires interpreting their evidence. These distinctions clarify the framework without claiming that an instrument measured each dimension.

The proposed mapping follows the categories reported in the manuscript: Table 1 describes recognition cues and uncertainty; Table 2 lists perceived credibility and social influences; and Table 3 identifies stated checking strategies and barriers. Table 4 translates those summaries into possible instructional priorities. In particular, reported social pressure informs the ethics category, while difficulty using checking services informs the tool-access category. This is a transparent conceptual inference from summary material, not an independently verified excerpt-to-code-to-theme audit.

Table 4. Provisional mapping of reported themes to proposed learning dimensions (not an evaluated program)

Proposed dimension	Basis in reported summaries	Proposed learning activity (not evaluated)
AI knowledge	Limited grasp of synthetic content	Explain AI capabilities and limits
Critical verification	Difficulty judging source credibility	Practice tracing and corroboration
Psychological awareness	Prior beliefs and emotional responses	Reflect on bias and uncertainty
Digital ethics	Group pressure to share posts	Practice responsible sharing
Tool access	Limited use of checking services	Demonstrate accessible verification tools

Note: The middle column summarizes categories already reported in the manuscript; the right column contains proposed instructional activities. The table is a conceptual synthesis, not an excerpt-level coding audit or a validated instrument.

3.2. Discussion

3.2.1. Recognition, Credibility, and Evidence-Based Claim Verification

Experimental research on deepfake detection shows the limits of relying on visual realism, but does not establish participants' individual abilities [7].

Two judgments need to be distinguished: whether the media itself was synthetically produced and whether its attached statement is supported by independent evidence. Perceptual experiments find substantial difficulty distinguishing authentic from generated media [8], yet a credibility assessment can also involve checking source history, context, and corroboration. The interview accounts primarily concern perceived difficulties and strategies, not to a verified performance outcome.

Synthetic-media recognition and claim verification should therefore be treated as separate educational objectives. The first concerns the cues and limitations of media-production technology; the second requires locating and evaluating evidence. Research distinguishing information literacy from other literacy constructs supports the need for clearer operational definitions [23]. Lateral reading provides a teachable example of independent source verification, although its use was not observed in this study [36].

For future training, learners could first explain why a plausible-looking video is insufficient evidence and then test the claim through independently sourced material. In studies of online civic reasoning, explicit instruction in fact-checking strategies improved aspects of source evaluation [37]. Those findings suggest a possible teaching sequence, but provide no evidence that participants in this study improved.

This distinction also marks the conceptual boundary between technical AI knowledge and critical verification: knowing that a model can synthesize media does not establish whether a claim is accurate. Reviews of the psychological basis of misinformation caution that perceived credibility also depends on attention, familiarity, and prior beliefs [12].

Dwivedi et al. [4] discuss the broader opportunities and risks of AI across disciplinary and social settings. Their analysis supports a wider view of AI literacy that includes responsibility and evaluation, although it does not directly establish participants' verification performance.

Vaccari and Chadwick reported effects on uncertainty and perceived credibility in an experiment with synthetic political video [10]; multimodal experiments further distinguish the influence of audio and visual signals on judgments [7]. Unlike those studies, the present study relies on retrospective accounts without controlled exposure or accuracy measurement. The comparison therefore concerns the cues considered, not comparable effect sizes.

Frequent use of AI tools should not be confused with an ability to verify claims critically. Human perceptual-detection studies illustrate the difficulty of recognizing synthetic material [8], while evidence-based claim checking can require non-AI methods such as source tracing and independent corroboration.

3.2.2. Social and Psychological Interpretations of Vulnerability

In this interpretation, "confirmation bias" is a possible theoretical explanation for belief-consistent trust, not a psychological characteristic measured in the study.

Accounts of familiar sources and preferred explanations suggest that prior beliefs could shape how closely a claim is examined. Studies associating emotional processing with misinformation belief [25] and reviews of psychological drivers [12] offer plausible explanatory lenses. Neither study, however, establishes that a measurable bias or emotional mechanism operated in these particular interviews.

Experimental research shows that the relationship between accuracy judgments and sharing decisions can be altered by shifting attention toward accuracy [24]. Sharing behavior was not directly recorded in this study, so any comparison must remain at the level of reported decision-making rather than observed influence.

Group discussion can reveal disagreement, but participants may also defer to perceived group consensus. The research record does not identify dissenting views within FGDs or document facilitator effects. Research on the social diffusion of misinformation underscores the importance of interpersonal relationships although the dynamics in these particular groups remain unverified [45].

The pattern is consistent with a broader distinction between credibility judgments and social actions: recognizing uncertainty does not guarantee that a person will refrain from forwarding a claim. Prior experimental findings on emotion [25] and accuracy attention [24] concern measured outcomes in other samples; the categories in this study remain interpretive descriptions.

The concern participants expressed for relatives or acquaintances offers one possible reason for forwarding apparently useful information without independent checks. Research on responses to false news suggests that personal relevance can affect decisions to correct content [45]. That observation offers a comparator, not a verified explanation of the behavior of this sample.

Literacy training focused only on familiarity with software may miss the reflective and interpersonal challenges described by informants. A relevant educational aim would be to practice questioning source authority, checking evidence, and discussing corrections respectfully. A framework of separate information-literacy competencies helps avoid equating technical confidence with verified competence [23].

The term "belief-consistent trust" describes participants' accounts of accepting familiar or agreeable messages. Confirmation bias offers a possible theoretical explanation, but it was neither operationalized nor measured. Research relating emotional reliance to misinformation belief suggests another mechanism worth investigating [25]; it cannot be assumed to have occurred in this sample.

Pennycook and Rand review the role of reasoning in evaluating fake news [26]. A separate experimental study finds that attention to accuracy can alter sharing choices [24]. Together, they suggest alternative explanations to test, without validating a causal account of the interview themes.

The accounts are best understood as descriptions of how shortcuts to trust may intersect with social expectations. Without authentic coded extracts and contrasting cases, the analysis cannot rank the relative influence of different factors.

The importance participants placed on close relationships suggests a practical design problem for future training: how to request evidence or communicate uncertainty without escalating conflict. Such an exercise should be assessed in a later intervention rather than treated as a measured benefit of this research.

3.2.3. Verification Practices and Digital-Literacy Learning

A study of Indonesian fact-checking principles provides a useful institutional comparison, but does not establish which procedures were used by these interviewees [41].

Interventions that direct attention to accuracy suggest a possible training approach, although it would require separate testing in this setting [24].

Source verification could be taught as a sequence: locate the original claim, examine its source, find independent coverage, and assess the supporting evidence. Professional fact-checkers have been observed reading laterally [36], and related online-reasoning strategies have been evaluated in educational interventions [37]. In this study, the sequence is a possible educational application, not an action observed in a standardized task.

Easy access to a tool should not be mistaken for the ability to verify information. A user may know of a tool but find its procedures cumbersome or fail to interpret its results. Validated assessments of misinformation discernment provide one potential way to test future competence claims [35]; no comparable assessment was administered to the present informants.

A critical account of verification therefore needs both epistemic judgment and a workable routine. The finding that information literacy may differ from general or self-rated media literacy [23] supports specifying observable future tasks rather than equating familiarity with ability.

Dwivedi et al. [4] emphasize that the social implications of AI require perspectives beyond technical performance. For digital-literacy education, this suggests combining tool familiarity with discussion of risks, uncertainty, and the quality of evidence.

Experimental evidence on accuracy prompts [24] and curriculum-based online reasoning [37] points to different strategies for addressing misinformation. Whether either can address the particular obstacles reported in these interviews is a question for future research.

3.2.4. Interpretation, Scholarly Positioning, and Educational Implications

The literature supplies alternative explanations and established instructional approaches. Reviews of misinformation psychology address emotion and belief revision [12]; studies of accuracy prompts examine sharing judgments [24]; and intervention reviews differentiate skill-building, inoculation, and labeling [34]. None verifies the coding chain of these interviews. The fieldwork material must be examined before the proposed mapping can be claimed as empirically grounded at the level of individual participants.

Comparison with established frameworks shows substantial conceptual overlap. Park's school-oriented GenAI-literacy review [28], the competency pathways described by Chee and colleagues [29], and UNESCO's student guidance [30] already address AI knowledge, evaluation, and ethics. The specific contribution proposed here is narrower: contextualizing familiar competencies around reported Indonesian adult experiences of visual plausibility, belief-consistent trust, social pressure, and barriers to using verification tools. This contextual mapping should not be described as a new general theory.

Research on AI risks highlights technical, organizational, and ethical consequences [4]. In this study, that perspective supports integrating tool use with decisions about evidence and sharing rather than presenting digital ethics as an empirically new competency.

The proposed combination also overlaps with established intervention approaches, including lateral reading, source evaluation, and accuracy-focused prompts [34]. Its contribution should therefore be assessed for its contextual interpretive value rather than for an unsupported claim to novel pedagogical mechanisms.

The conceptual novelty of this proposal is limited. Existing models already discuss AI knowledge, evaluation, ethics, and learner competencies; the proposed contribution lies in the account of how selected Indonesian adults describe overlapping obstacles and how those obstacles might inform an educational agenda that can be evaluated. Claims about the generalizability or effectiveness of such an agenda require independent evidence.

The participants and setting also differ from those emphasized in much of the school-based GenAI-literacy literature [28], [29]. The present analysis does not demonstrate superiority over those frameworks, but suggests a narrower problem definition for community trainers: credible claim checking in everyday social-media exchanges.

A central insight is the reported gap between suspecting synthetic production and checking the accuracy of the accompanying claim. Whether a training sequence built around that distinction produces transferable skills is an open empirical question that can be assessed using appropriate verification tasks [35].

The intervention literature also clarifies what remains unknown. Boler and colleagues describe varied adult-learning approaches [31], while Droog and colleagues distinguish the effectiveness and limits of media-literacy interventions [32]. Meta-analyses emphasize variation across outcomes, program designs, and cultural contexts [33], [21], [22]. These studies offer benchmarks for future testing, but none validates the five dimensions in the current sample.

Future assessments should distinguish judgments about media authenticity, decisions about the accuracy of claims, and sharing behavior. MIST is an example of a validated discernment measure developed in other populations [35]; it was neither adapted nor administered here. Differences between such measures and the present qualitative accounts illustrate why proposed learning priorities cannot be reported as measured competencies.

A future learning module could ask adults to compare a plausible synthetic item with independently sourced reporting, discuss their reasons for trusting it, check the claim, and decide whether forwarding it would be responsible. Lateral-reading instruction [37], [36] and an Indonesia-

specific media-literacy intervention [19] provide related tested approaches. The present study proposes these activities; it did not develop or evaluate them.

A practical design hypothesis is that adult learners may benefit from explaining why they consider particular sources credible, locating corroborating evidence, and explaining the social consequences of forwarding uncertain claims. Such behaviors could be assessed in authentic tasks, as research on online-reasoning instruction has begun to demonstrate [37].

A subsequent study could evaluate a short claim-checking learning activity against existing community training, using source tracing, evidence appraisal, and sharing decisions as separate outcomes. Evidence from an Indonesian intervention shows that such designs can be studied empirically, while their effects should not be presumed for the present framework [19].

Community organizations could use the reported obstacles to prioritize accessible exercises in source checking and respectful correction. Indonesian studies of information literacy and false-information sharing offer useful comparisons [17], [18], but do not verify the accounts of these 32 participants.

For tool designers and fact-checking services, the interviews raise questions about whether users can find, understand, and act on information from checking services. Indonesian research on institutional fact-checking procedures offers context [41]; service usability and user success remain to be measured.

The next step is to verify the original fieldwork record, co-design tasks with the intended learner group, and test usability and verification performance. Indonesian surveys point to contextual variation in digital literacy [13], [18]; any later intervention should therefore report its participants and delivery conditions precisely.

3.2.5. Study Limitations and Future Validation

First, the sampling strategy and setting limit transferability. Thirty-two purposively selected users from six named cities cannot represent Indonesian social-media users nationally, and the absence of recruitment details and participant profiles further restricts assessment of transferability. Rural-literacy survey findings illustrate why the setting and population should be specified rather than generalized [13].

Second, the findings rely on self-reported accounts and group discussions. A statement about checking does not demonstrate successful verification, and FGD consensus may differ from individual views. No documented facilitator notes, dissenting-case procedure, or reflexive record was supplied to evaluate these possibilities.

Third, generative-AI capabilities and available verification tools are changing rapidly. Findings about visual cues and checking routines should be situated in the still-unconfirmed fieldwork period rather than assumed to describe stable competence.

Fourth, the proposed five-dimensional framework was not validated by experts, administered as an instrument, or tested in a learning intervention. Its instructional implications remain hypotheses for design and evaluation, not study outcomes.

Finally, transcripts and coding records, source attribution for quoted passages, translation procedures, applicable participant consent, and the fieldwork year were unavailable for independent verification. These are material evidence and transparency requirements, not errors that can be resolved through prose editing alone.

4. Conclusions

Accounts from 32 purposively selected Indonesian social-media users suggest that recognizing AI-related terminology or suspecting a synthetic image is not the same as verifying an associated

claim. Participants described using visual polish and popularity as credibility cues and reported uneven use of source checks, independent corroboration, contextual appraisal, and fact-checking tools. These findings describe reported practices rather than measured detection performance or national prevalence.

The study provisionally organizes these accounts around AI knowledge, critical verification, psychological awareness, digital ethics, and tool access. Its specific educational contribution is a set of contextual design questions for non-formal adult learning; the dimensions are not newly established competencies or an evaluated intervention. Future studies should compare this mapping with existing frameworks and measure authentic claim-checking and responsible-sharing outcomes.

Before editorial reassessment, the authors need to verify source-to-theme evidence, quote provenance and translation, the fieldwork period, recruitment and data-protection records, and applicable consent or ethics documentation. These limitations constrain the present theoretical and methodological claims even though the proposed educational questions may be useful for subsequent research.

References

[1] J. van Dijck, T. Poell, and M. de Waal, The Platform Society: Public Values in a Connective World. Oxford, U.K.: Oxford University Press, 2018.

[2] S. Kemp, "Digital 2024: Indonesia," DataReportal, 2024. [Online]. Available: https://datareportal.com/reports/digital-2024-indonesia.

[3] J. A. Goldstein et al., "Generative language models and automated influence operations: Emerging threats and potential mitigations," arXiv preprint arXiv:2301.04246, 2023, doi: 10.48550/arXiv.2301.04246.

[4] Y. K. Dwivedi et al., "So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," International Journal of Information Management, vol. 71, art. 102642, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[5] M. Westerlund, "The emergence of deepfake technology: A review," Technology Innovation Management Review, vol. 9, no. 11, pp. 39-52, 2019, doi: 10.22215/timreview/1282.

[6] J. T. Hancock and J. N. Bailenson, "The social impact of deepfakes," Cyberpsychology, Behavior, and Social Networking, vol. 24, no. 3, pp. 149-152, 2021, doi: 10.1089/cyber.2021.29208.jth.

[7] M. Groh et al., "Human detection of political speech deepfakes across transcripts, audio, and video," Nature Communications, vol. 15, art. 7629, 2024, doi: 10.1038/s41467-024-51998-z.

[8] D. Cooke, A. Edwards, S. Barkoff, and K. Kelly, "As good as a coin toss: Human detection of AI-generated content," Communications of the ACM, vol. 68, no. 10, pp. 100-109, 2025, doi: 10.1145/3729417.

[9] B. N. Jacobsen and J. Simpson, "The tensions of deepfakes," Information, Communication & Society, vol. 27, no. 6, pp. 1095-1109, 2024, doi: 10.1080/1369118X.2023.2234980.

[10] C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news," Social Media + Society, vol. 6, no. 1, pp. 1-13, 2020, doi: 10.1177/2056305120903408.

[11] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146-1151, 2018, doi: 10.1126/science.aap9559.

[12] U. K. H. Ecker et al., "The psychological drivers of misinformation belief and its resistance to correction," Nature Reviews Psychology, vol. 1, pp. 13-29, 2022, doi: 10.1038/s44159-021-00006-y.

[13] M. Akbar and K. Kusnandar, "Social media literacy among rural people in Indonesia: Empirical findings from a nationwide survey," Information Development, vol. 42, no. 4, 2026, doi: 10.1177/02666669241302053.

[14] A. A. Rosyidah, F. Fajriyah, E. A. Galuh, and D. S. Ulfa, "Exploring misinformation and disinformation towards 2024 election: Patterns and policy recommendations," Profetik: Jurnal Komunikasi, vol. 17, no. 2, pp. 269-290, 2024, doi: 10.14421/pjk.v17i2.2973.

[15] N. R. Anggraini et al., "Hoaxes in the digital era: An analysis of social media users' perceptions and attitudes," Jurnal Lemhannas RI, vol. 12, no. 4, pp. 567-580, 2024, doi: 10.55960/jlri.v12i4.944.

[16] D. F. Dhahir et al., "The relationship of digital literacy, exposure to AI-generated deepfake videos, and the ability to identify deepfakes in Generation X," Jurnal Pekommas, vol. 9, no. 2, 2024, doi: 10.56873/jpkm.v9i2.5873.

[17] E. Kurniasih, N. A. Damayani, and F. Zubair, "The role of information literacy in mitigating the spread of hoaxes on social media," Khizanah al-Hikmah, vol. 12, no. 1, pp. 1-13, 2024, doi: 10.24252/kah.v12i1a1.

[18] E. Kurniasih, N. A. Damayani, and F. Zubair, "Media literacy, information literacy, and novice voters combating online political misinformation," Jurnal Kajian Informasi & Perpustakaan, vol. 12, no. 1, pp. 67-84, 2024, doi: 10.24198/jkip.v12i1.53845.

[19] T. Ford, M. Yankoski, M. Facciani, and T. Weninger, "Online media literacy intervention in Indonesia reduces misinformation sharing intention," Journal of Media Literacy Education, vol. 15, no. 2, pp. 99-123, 2023, doi: 10.23860/JMLE-2023-15-2-8.

[20] A. M. Guess et al., "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India," Proceedings of the National Academy of Sciences, vol. 117, no. 27, pp. 15536-15545, 2020, doi: 10.1073/pnas.1920498117.

[21] G. Huang, W. Jia, and W. Yu, "Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors," Communication Research, vol. 53, no. 7, pp. 899-926, 2026, doi: 10.1177/00936502241288103.

[22] H. Cho, C. J. Carpenter, and W. Li, "Media literacy interventions: Meta-analytic review of 40 years of research," Human Communication Research, vol. 51, no. 2, pp. 57-79, 2025, doi: 10.1093/hcr/hqaf004.

[23] S. M. Jones-Jang, T. Mortensen, and J. Liu, "Does media literacy help identification of fake news? Information literacy helps, but other literacies don't," American Behavioral Scientist, vol. 65, no. 2, pp. 371-388, 2021, doi: 10.1177/0002764219869406.

[24] G. Pennycook et al., "Shifting attention to accuracy can reduce misinformation online," Nature, vol. 592, pp. 590-595, 2021, doi: 10.1038/s41586-021-03344-2.

[25] C. Martel, G. Pennycook, and D. G. Rand, "Reliance on emotion promotes belief in fake news," Cognitive Research: Principles and Implications, vol. 5, art. 47, 2020, doi: 10.1186/s41235-020-00252-3.

[26] G. Pennycook and D. G. Rand, "The psychology of fake news," Trends in Cognitive Sciences, vol. 25, no. 5, pp. 388-402, 2021, doi: 10.1016/j.tics.2021.02.007.

[27] A. Fulsher, M. Pagkratidou, and P. Kendeou, "GenAI and misinformation in education: A systematic scoping review of opportunities and challenges," AI & Society, vol. 41, pp. 1373-1385, 2026, doi: 10.1007/s00146-025-02536-y.

[28] J. Park, "A systematic literature review of generative artificial intelligence (GenAI) literacy in schools," Computers and Education: Artificial Intelligence, vol. 9, art. 100487, 2025, doi: 10.1016/j.caeai.2025.100487.

[29] H. Chee, S. Ahn, and J. Lee, "A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway," British Journal of Educational Technology, vol. 56, no. 5, pp. 2146-2182, 2025, doi: 10.1111/bjet.13556.

[30] F. Miao, K. Shiohira, and N. Lao, AI Competency Framework for Students. Paris, France: UNESCO, 2024, doi: 10.54675/JKJB9835.

[31] M. Boler, H. Gharib, Y.-J. Kweon, A. Trigiani, and B. Perry, "Promoting mis/disinformation literacy among adults: A scoping review of interventions and recommendations," Communication Research, vol. 53, no. 7, 2026, doi: 10.1177/00936502251318630.

[32] E. Droog et al., "Combatting the misinformation crisis: A systematic review of the literature on characteristics and effectiveness of media literacy interventions," Communication Research, vol. 53, no. 7, 2026, doi: 10.1177/00936502251363705.

[33] S. Liu, Z. Lin, O. L. Ng, Y. Yang, and Y. Dai, "AI literacy interventions in education: A meta-analysis of effects and moderators," Educational Research Review, vol. 52, art. 100813, 2026, doi: 10.1016/j.edurev.2026.100813.

[34] A. Kozyreva et al., "Toolbox of individual-level interventions against online misinformation," Nature Human Behaviour, vol. 8, pp. 1044-1052, 2024, doi: 10.1038/s41562-024-01881-0.

[35] R. Maertens et al., "The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment," Behavior Research Methods, vol. 56, pp. 1863-1899, 2024, doi: 10.3758/s13428-023-02124-2.

[36] S. Wineburg and S. McGrew, "Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information," Teachers College Record, vol. 121, no. 11, pp. 1-40, 2019, doi: 10.1177/016146811912101102.

[37] S. McGrew, "Learning to evaluate: An intervention in civic online reasoning," Computers & Education, vol. 145, art. 103711, 2020, doi: 10.1016/j.compedu.2019.103711.

[38] S. Lewandowsky et al., The Debunking Handbook 2020. 2020, doi: 10.17910/b7.1182.

[39] J. Roozenbeek, S. van der Linden, B. Goldberg, S. Rathje, and S. Lewandowsky, "Psychological inoculation improves resilience against misinformation on social media," Science Advances, vol. 8, no. 34, art. eabo6254, 2022, doi: 10.1126/sciadv.abo6254.

[40] J. Roozenbeek and S. van der Linden, "Fake news game confers psychological resistance against online misinformation," Palgrave Communications, vol. 5, art. 65, 2019, doi: 10.1057/s41599-019-0279-9.

[41] H. Purnama, "Comparison of fact checking principles of misinformation and disinformation in social media in Indonesia," CoverAge: Journal of Strategic Communication, vol. 14, no. 2, pp. 102-114, 2024, doi: 10.35814/coverage.v14i2.6139.

[42] L. Doyle, C. McCabe, B. Keogh, A. Brady, and M. McCann, "An overview of the qualitative descriptive design within nursing research," Journal of Research in Nursing, vol. 25, no. 5, pp. 443-455, 2020, doi: 10.1177/1744987119880234.

[43] S. Hall and L. Liebenberg, "Qualitative description as an introductory method to qualitative research for master's-level students and research trainees," International Journal of Qualitative Methods, vol. 23, 2024, doi: 10.1177/16094069241242264.

[44] V. Braun and V. Clarke, Thematic Analysis: A Practical Guide. London, U.K.: SAGE Publications, 2022.

[45] E. C. Tandoc Jr., D. Lim, and R. Ling, "Diffusion of disinformation: How social media users respond to fake news and why," Journalism, vol. 21, no. 3, pp. 381-398, 2020, doi: 10.1177/1464884919868325.