

# When synthetic speech becomes evidence: linguistic authentication and evidentiary uncertainty in Indonesian deepfake cases

Rika Yani  
Universitas PGRI Sumenep, Indonesia

## Article Info

### Article history:

Received Apr 15, 2026  
Revised May 23, 2026  
Accepted Jun 29, 2026

### Keywords:

audio deepfake;  
evidentiary uncertainty;  
forensic linguistics;  
Indonesian legal discourse;  
synthetic speech.

## ABSTRACT

**Background:** Synthetic speech and audio deepfakes increasingly unsettle legal assumptions that a recorded voice can index speaker identity, communicative intention, and evidentiary presence in Indonesian digital-public discourse. **Objective:** This study examines how synthetic or disputed speech becomes evidentially meaningful by analysing the linguistic authentication, evidentiary uncertainty, and procedural positioning of deepfake-related public documents in Indonesia. **Method:** Using a qualitative corpus-assisted legal-linguistic design, this study analyses 24 publicly verifiable documents consisting of official enforcement releases, fact-checking records, news reports, legal-context documents, and scholarly sources. **Results:** The findings show that authentication is constructed through convergence among detector claims, source tracing, official-account verification, platform context, and institutional classification rather than through acoustic judgement alone. Evidentiary uncertainty is not treated as a defect to be removed, but as a structured stance expressed through probability, absence, methodological caution, and legal-contextual framing. **Implication:** Synthetic speech is procedurally positioned most often through public verification and fact-checking before it enters, or potentially enters, formal adjudicative settings. **Novelty:** This study offers a novel forensic-linguistic account of deepfake evidence by shifting analysis from technical detection toward the legal-discursive process through which synthetic speech is authenticated, contested, and made evidentially intelligible under uncertainty.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Rika Yani  
Universitas PGRI Sumenep, Indonesia  
Jl. Trunojoyo, Gedungan, Batuan, Sumenep, Jawa Timur, 69451, Indonesia  
Email: [rikayani665@gmail.com](mailto:rikayani665@gmail.com)

## 1. INTRODUCTION

Indonesia's deepfake problem has moved from speculative risk to an evidentiary concern in criminal justice. In 2024, Indonesia had 221,563,479 internet users, with internet penetration reaching 79.5% of the population, creating a vast environment for the circulation of manipulated audio-visual material. This scale matters because synthetic speech is not merely another form of online misinformation; it can imitate the evidential signs through which courts, investigators, journalists, and citizens recognise a speaker. A public example occurred in January 2025, when Indonesian police reported the arrest of a man accused of producing and circulating AI-generated deepfake videos that imitated President Prabowo Subianto, Vice President Gibran Rakabuming Raka, and Finance Minister Sri Mulyani to attract victims through fraudulent claims of government assistance. Such cases show that the problem is no longer limited to reputational harm or

political deception. It concerns whether a recorded voice can still be treated as a reliable trace of identity, intention, and communicative action.

Existing research has made substantial progress in explaining how deepfakes are generated, detected, and circulated, especially through machine learning and media forensics. Surveys have mapped detection architectures, benchmark limitations, multimodal forgery techniques, adversarial vulnerabilities, and the uneven robustness of audio deepfake detectors across languages, speakers, and recording conditions [1], [2], [3], [4]. More recent work has drawn attention to partial fake speech attacks and segmental speech features, showing that synthetic manipulation may affect only parts of an utterance rather than an entire recording [5], [6]. Criminal justice scholarship has also warned that deepfakes can undermine evidentiary trust and burden courts with new authentication problems [7]. Yet this literature remains unevenly connected to forensic linguistics. Less is known about how synthetic or disputed speech is linguistically represented, authenticated, contested, and translated into evidentiary value within Indonesian legal discourse.

This study addresses that gap by examining how synthetic speech becomes evidence in Indonesian deepfake-related cases. It asks how legal, institutional, and media documents represent disputed voice material; what linguistic, acoustic, and procedural criteria are invoked to claim or challenge authenticity; and how evidentiary uncertainty is expressed when speech may be cloned, edited, replayed, or partially fabricated. Rather than treating authentication as a purely technical act, this study approaches it as a layered legal-linguistic process. Voice evidence is examined through the interaction of speaker identity, speech style, transcription, expert knowledge, chain of custody, platform circulation, and legal classification. This orientation builds on recent calls to integrate sociolinguistics into audio deepfake discernment, because accent, register, language choice, prosody, hesitation, and speaking style may shape both human judgement and technical detection [8]. It also draws on Indonesian forensic-linguistic work on courtroom language, police interrogation, and legal evidence [9], [10], while situating synthetic speech within broader debates on calibrated forensic voice comparison in Indonesia [11].

The central argument is that synthetic speech unsettles evidentiary reasoning because it separates audibility from accountability: a voice may be heard, recognised, circulated, and believed without securely establishing who spoke, how the utterance was produced, or whether the recorded event occurred as represented. This does not mean that all audio evidence becomes unreliable. It means that authenticity must be reconstructed through converging linguistic, technical, procedural, and contextual indicators rather than assumed from recognisable voice quality alone. Deepfakes therefore create a double uncertainty: they make fabricated speech appear evidential, while also making genuine speech more vulnerable to denial. This study tests that argument by analysing how Indonesian public legal and institutional texts manage the instability between speech as communicative action and speech as digital artefact. Its implication is methodological as well as legal: forensic linguistic authentication in deepfake cases must move beyond speaker recognition toward a more explicit account of evidentiary uncertainty.

## 2. LITERATURE REVIEW

### 2.1. Synthetic speech

Synthetic speech in deepfake research is commonly understood as machine-generated or machine-altered speech that imitates the acoustic and linguistic profile of a human speaker. Yet the literature does not define it uniformly. Technical surveys often treat audio deepfakes as detection objects produced through text-to-speech, voice conversion, replay, or generative modelling [2], [12], [3], [13]. Broader media-forensic studies define deepfakes as multimodal manipulations that disturb the reliability of digital traces [14], [4], [15]. Philosophical and legal accounts, by contrast, emphasise deception, impersonation, reputational harm, and the wrong of making false presence appear credible [16], [7]. These differences matter because a speech deepfake is not only a technical artefact but also a communicative event that can be believed, circulated, and evidentially mobilised.

The main indicators of synthetic speech have been classified through production method, acoustic trace, linguistic plausibility, and forensic detectability. Detection studies typically examine spectral artefacts, prosodic irregularities, temporal discontinuities, phase information, segmental features, and model-specific traces [17], [18], [19], [6]. Recent research complicates this model by showing that fake audio may be partial rather than total, making utterance-level authentication insufficient [5]. Multilingual and language-sensitive work further shows that detection is shaped by language, speaking style, training data, and cross-linguistic representation [20], [21], [22]. The literature therefore supports a move from binary fake/real classification toward a more granular analysis of where, how, and under what linguistic conditions synthetic manipulation is claimed to occur.

## 2.2. Linguistic authentication

Linguistic authentication refers to the process through which speech is evaluated as attributable, credible, and evidentially meaningful. In forensic settings, authentication is not identical to automated detection. It may involve speaker comparison, transcription, pragmatic interpretation, acoustic examination, contextual corroboration, and expert explanation. Agoestyowati (2024) [11] frames forensic voice comparison in Indonesia as requiring calibrated acoustic evidence, while Mallinson et al. (2024) [8] argue that sociolinguistics has been underused in audio deepfake discernment. Indonesian legal-linguistic studies similarly show that evidence is shaped by institutional discourse, courtroom reasoning, and legal language rather than by linguistic form alone [9], [23], [10]. The key limitation in existing work is that technical authentication and legal authentication are often discussed separately, although deepfake cases require both.

The aspects of linguistic authentication can be organised into acoustic, sociolinguistic, pragmatic, and procedural dimensions. Acoustic indicators include pitch, rhythm, spectral features, segmental continuity, and signs of synthesis or editing. Sociolinguistic indicators include accent, dialect, register, code choice, speaking style, and interactional appropriateness, all of which may influence perceived authenticity [8], [22]. Pragmatic indicators concern whether speech acts, stance, politeness, hesitation, or institutional roles fit the alleged speaker and situation, an issue related to Indonesian interrogation and legal-discourse studies [9], [24]. Procedural indicators include chain of custody, metadata, expert qualification, corroborating evidence, and admissibility. This layered model avoids treating voice as a simple biometric object and instead views authentication as an interpretive practice.

## 2.3. Evidentiary uncertainty

Evidentiary uncertainty describes the condition in which the status, source, production process, or probative value of evidence cannot be settled without interpretive and procedural mediation. Deepfakes intensify this uncertainty because they create plausible speech events while also allowing genuine recordings to be dismissed as fabricated. Criminal justice research stresses that deepfakes threaten trust in evidence, burden investigative institutions, and complicate admissibility and evaluation [7]. Reliability-oriented surveys similarly show that detection systems remain vulnerable to dataset bias, adversarial manipulation, generalisation failure, and real-world recording variation [25], [1], [3]. Legal studies on Indonesian deepfake risks point to regulatory gaps in politics, national security, public image rights, and accountability (Lestari & Ibrahim, 2025), (Lianingsih & Jaizul, 2025), (Triana, 2026). Thus, uncertainty is both technological and juridical.

The indicators of evidentiary uncertainty include epistemic stance, source verification, institutional authority, corroboration, and legal classification. Epistemic stance appears in expressions of suspicion, probability, limitation, and expert dependency; these forms matter because legal actors rarely state uncertainty neutrally. Source verification focuses on whether a recording can be connected to an original event, device, platform, or speaker, an issue increasingly addressed in speech-deepfake verification research (Negroni, 2025), (Guo, 2026). Institutional authority concerns whose authentication counts: police, forensic experts, judges, platforms, or fact-checkers. Corroboration concerns whether voice evidence is supported by metadata, testimony, visual material, or other documents. This study therefore positions synthetic speech evidence as a legal-linguistic object: not merely audio to be detected, but uncertain speech to be authenticated, narrated, and contested within Indonesian evidentiary practice.

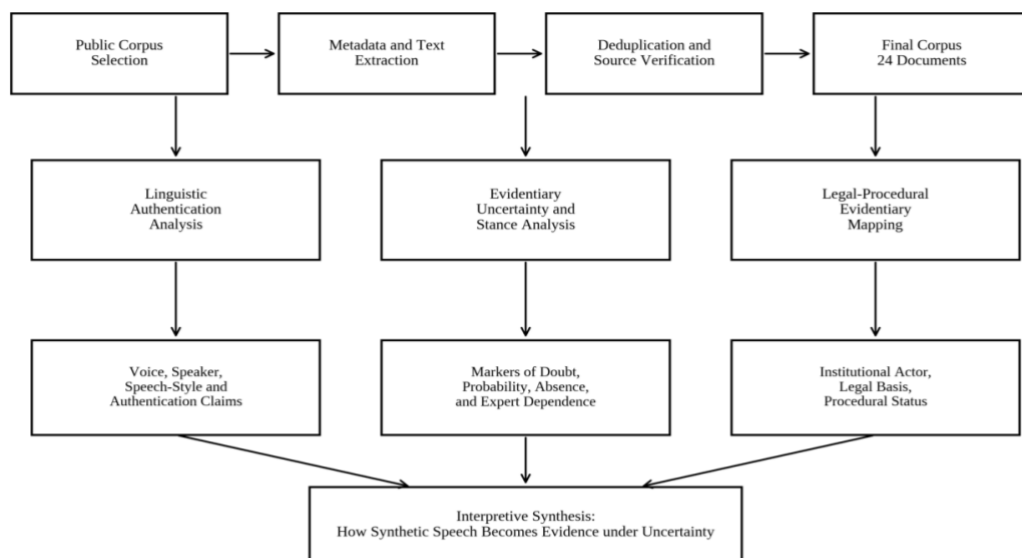
## 3. METHOD

The material object of this study is a documentary corpus of publicly accessible Indonesian materials in which synthetic speech, deepfake audio, or AI-manipulated voice is represented as a legal, institutional, or evidentiary issue. The unit of analysis is not the alleged offender, victim, or technical audio file, but the document-based evidentiary passage: a segment in which voice, authenticity, manipulation, verification, legal risk, or institutional response is explicitly formulated. The final corpus contains 24 documents collected from official enforcement releases, fact-check records, news reports, legal instruments, and scholarly references. This composition allows this study to examine how synthetic speech becomes evidentially meaningful across legal, institutional, and public verification discourse.

This study adopts a qualitative corpus-assisted legal-linguistic design. The approach is suitable because the research problem concerns how authenticity, uncertainty, and evidentiary value are constructed in discourse, not whether each audio item is technically fake. Corpus assistance is used to organise metadata, retrieve recurrent terms, compare categories, and maintain transparency across documents; interpretation remains qualitative and theoretically guided. This design follows recent work that treats deepfakes as criminal-justice and forensic problems rather than merely computational detection tasks [7], [3]. It also incorporates sociolinguistic concerns about language, accent, and speaking style in audio deepfake discernment [8].

**Table 1.** Dataset composition

| Code     | Data Type | Source                               | Location                               | Period    | Number | Characteristics                                                                                         |
|----------|-----------|--------------------------------------|----------------------------------------|-----------|--------|---------------------------------------------------------------------------------------------------------|
| SS-POL   | Primary   | Polri official enforcement release   | National online space                  | 2025      | 1      | Deepfake-related enforcement material involving public officials                                        |
| SS-FC    | Primary   | TurnBackHoax, ANTARA/JACX, Tirto.id  | National online space                  | 2024–2026 | 11     | Fact-checks on AI voice, manipulated audio, public-figure impersonation, scams, and synthetic video     |
| SS-KOM   | Primary   | Kementerian Komunikasi dan Digital   | National online space                  | 2024–2025 | 3      | Official hoax clarifications and deepfake-related public warnings                                       |
| SS-NEWS  | Secondary | ANTARA News                          | Jakarta/national online space          | 2025      | 1      | Report on Polri-Komdigi cyber patrol against deepfake circulation                                       |
| CTX-INST | Context   | ANTARA News                          | National jurisdiction                  | 2023–2026 | 2      | Institutional risk framing of audio, video, and image deepfakes                                         |
| LEG      | Context   | JDIH BPK / Government legal database | National jurisdiction                  | 2022–2025 | 3      | Legal framework on electronic information, procedure, and personal data                                 |
| ACAD     | Context   | Scholarly literature                 | Not jurisdiction-specific              | 2024–2025 | 4      | Theoretical and methodological references on deepfake detection, criminal justice, and sociolinguistics |
| TOTAL    | Mixed     | Publicly verifiable sources          | Indonesia / relevant global literature | 2022–2026 | 24     | Corpus for legal-linguistic analysis of synthetic speech evidence                                       |

**Figure 1.** Operational analysis matrix

The information sources were selected to represent three evidentiary layers. Primary data consist of official enforcement releases, government hoax clarifications, and fact-check records that directly discuss synthetic voice, manipulated audio, or AI-generated public-figure impersonation. Secondary data consist of news reporting that documents institutional responses without functioning as the main evidentiary object. Context data consist of statutes, regulatory materials, and scholarly literature needed to interpret evidentiary procedure, authentication, and technical limitation. This layered source design prevents official law, journalistic reporting, and academic interpretation from being treated as equivalent evidence. It also allows this study to distinguish case representation, legal framework, and methodological context.

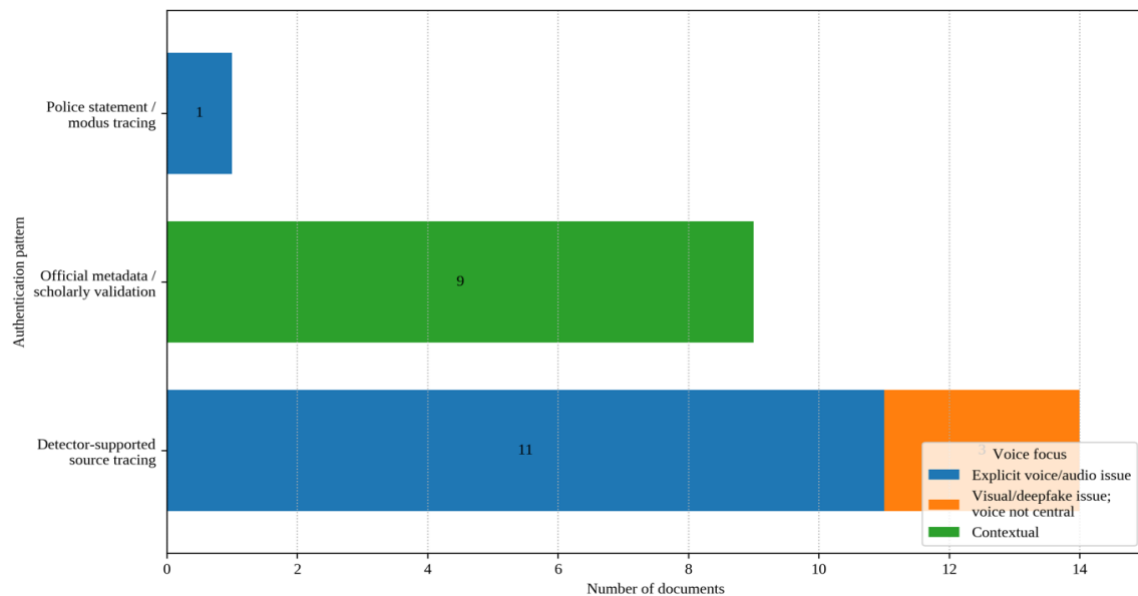
Data collection proceeded through a structured public-source protocol. Searches were conducted using Indonesian and English keywords, including “deepfake suara,” “AI voice,” “suara AI,” “video deepfake,” “hoaks suara,” “public figure deepfake,” “bantuan pemerintah palsu,” and “forensic voice comparison.” The period was set from 2022 to 2026 to capture the legal framework on personal data and electronic information, recent deepfake incidents, and current technical literature. Each item was logged with title, date, source, URL, source type, data type, case cluster, modality, legal relevance, quotation, summary, and retrieval context. Duplicate items were removed unless retained as corroborating documents within the same case cluster.

Data analysis involved three sequential procedures. First, linguistic authentication analysis identified how documents constructed authenticity through voice reference, speaker resemblance, transcription, detector claims, source tracing, original-media comparison, or institutional denial. Second, evidentiary uncertainty and stance analysis examined how documents expressed probability, suspicion, absence of source, expert dependency, and limits of verification. Third, legal-procedural mapping classified how synthetic speech was framed as fraud evidence, hoax material, public-risk communication, legal context, or technical-methodological reference. These procedures draw on forensic voice comparison, legal discourse analysis, and recent audio deepfake research [11], [6], [10]. Figure 2 presents the operational flow from corpus construction to interpretive synthesis.

## 4. RESULTS

### 4.1. From voice resemblance to linguistic authentication

Synthetic speech becomes evidentially visible in this corpus through the language of authentication rather than through direct access to the contested audio alone. The central pattern is that public documents rarely rely on a single evidentiary basis. Instead, they combine detector outputs, source tracing, official-account verification, comparison with original media, institutional denial, and legal-contextual classification. This pattern is methodologically significant because deepfake speech is not treated merely as sound to be heard, but as a digital object whose credibility depends on how its source, circulation, and technical status are reconstructed. This aligns with recent work showing that speech deepfake detection depends on datasets, detection architecture, language conditions, and real-world recording variability rather than on a simple fake/real distinction [2], [3], [13].



**Figure 2.** Linguistic authentication patterns by voice focus

**Table 2.** Linguistic authentication patterns in the corpus

| Subcategory                                  | Operational indicator                                                                                                                                                              | Frequency | Percentage | Data variation                                                                                                                                                | Evidentiary phrase                                                                                                  | Data codes                                                                                                                                    |
|----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| Explicit voice/audio issue                   | AI detector, source-video comparison, Google Lens, official-account absence, platform trace, or original-media comparison used to assess synthetic voice/audio                     | 11        | 45.8%      | AI voice, manipulated audio, synthetic public-official speech, celebrity-voice impersonation, institutional speech attribution                                | “suara Prabowo ... terdeteksi dihasilkan oleh AI generatif”; “probabilitas 99,3 persen sebagai hasil manipulasi AI” | SS-2025-003; SS-2024-001; SS-2024-002; SS-2024-003; SS-2026-001; SS-2026-002; SS-2025-004; SS-2025-005; SS-2025-006; SS-2025-007; SS-2025-002 |
| Visual/deepfake issue with voice not central | Deepfake identification is based on AI/deepfake claim, visual anomaly, scam framing, or official hoax classification, while explicit acoustic analysis is limited or absent        | 3         | 12.5%      | Investment scam, quiz/deepfake public-figure persona, fabricated assistance/school-fee content                                                                | “No direct quote retained; verify page before quotation”                                                            | SS-2025-008; SS-2025-009; SS-2024-004                                                                                                         |
| Contextual evidentiary framework             | Official legal metadata, statutory context, institutional risk framing, or scholarly source used to define the conditions under which synthetic speech may become legally relevant | 9         | 37.5%      | Electronic information law, criminal procedure, personal-data protection, criminal-justice risk, sociolinguistic detection, forensic voice-comparison context | “mulai berlaku pada tanggal 02 Januari 2026”; “melindungi Data Pribadi”; “Sekarang suara kita bisa ditiru”          | CTX-2023-001; CTX-2026-001; LEG-2024-001; LEG-2025-001; LEG-2022-001; ACAD-2024-001; ACAD-2025-001; ACAD-2024-002; ACAD-2025-002              |
| Enforcement-facing authentication            | Police statement, modus description, victim recruitment path, social-media dissemination, and WhatsApp trace used to identify deepfake content as part of a criminal scheme        | 1         | 4.2%       | Public-official impersonation, fake government assistance, social-media circulation, WhatsApp-based victim contact                                            | “Isi konten menawarkan bantuan pemerintah.”                                                                         | SS-2025-001                                                                                                                                   |
| —                                            | —                                                                                                                                                                                  | 24        | 100%       | —                                                                                                                                                             | —                                                                                                                   | —                                                                                                                                             |

The dominant pattern is detector-supported source tracing, which accounts for 14 of 24 documents, or 58.3% of the corpus. Within this group, 11 documents explicitly address voice or audio, while 3 documents concern deepfake or visual manipulation in which voice is not central. The second largest group, official metadata and scholarly validation, accounts for 9 documents, or 37.5%, and functions as the legal and methodological background for understanding electronic evidence, personal data, criminal procedure, and forensic uncertainty. Only 1 document, or 4.2%, relies primarily on police statement and modus tracing. The distribution shows that public authentication discourse is not yet centred on forensic voice comparison alone. It is distributed across fact-checking practices, institutional explanation, legal framework, and technical-methodological context.

Theoretically, this pattern supports the argument that linguistic authentication in deepfake cases is a layered evidentiary practice. Voice is not authenticated only by acoustic similarity to an alleged speaker; it is made credible or doubtful through relations among speech, source, platform, metadata, institutional authority, and legal classification. This finding is consistent with Mallinson et al.'s (2024) argument that sociolinguistic expertise should be integrated into audio deepfake discernment, because voice cannot be reduced to biometric signal alone. It also resonates with Sandoval et al. (2024) [7], who show that deepfakes challenge criminal justice by weakening trust in evidence and complicating verification. In this corpus, authentication emerges as a process of convergence: the evidentiary status of synthetic speech depends on how multiple indicators are assembled into a defensible claim. Evidentiary uncertainty in synthetic-speech claims

#### 4.2. Evidentiary uncertainty and patterns in linguistic authenticity

Uncertainty appears in the corpus as an organised evidentiary stance rather than as a simple absence of knowledge. Documents do not merely state that an audio or video is fake; they mark different degrees and sources of confidence through detector scores, comparison with original media, institutional statements, legal metadata, or methodological caution. This pattern matters because synthetic speech evidence is rarely available to the public as a complete forensic object. It is usually encountered through mediated claims: a fact-check finding, a police statement, a legal category, or a technical reference. This supports the broader concern in deepfake-forensic literature that detection outputs must be interpreted alongside source provenance, language conditions, and real-world recording constraints [3], [7], [6].

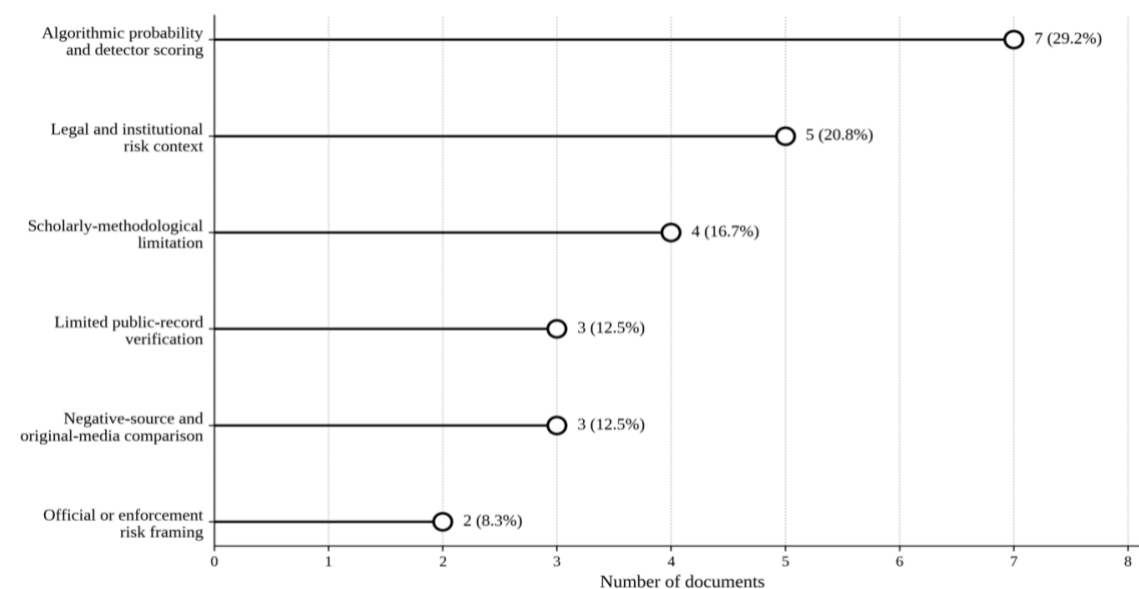


Figure 3. Evidentiary uncertainty patterns in the corpus

**Table 3.** Evidentiary uncertainty and stance patterns in the corpus

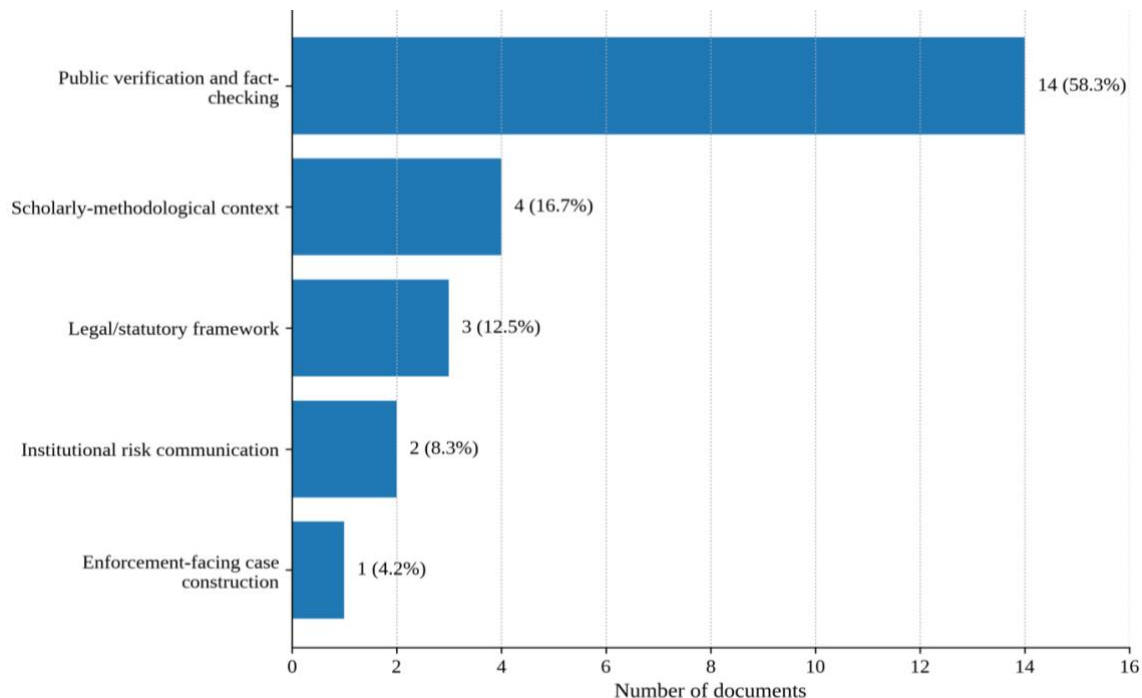
| Subcategory                                                 | Operational indicator                                                                                                                                                    | Frequency | Percentage | Data variation                                                                                                             | Example / evidentiary phrase                                                                                  | Data codes                                                                                                  |
|-------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|------------|----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Detector-mediated uncertainty                               | Use of AI detector, probability score, percentage, “terdeteksi,” “terindikasi,” or tool-based classification to support a claim of synthetic speech or manipulated media | 7         | 29.2%      | AI voice, deepfake public-official speech, celebrity voice, health testimonial video, synthetic hearing video, policy hoax | “99,9 persen merupakan hasil buatan AI”; “probabilitas 99,3 persen”; “76 persen suara ... dihasilkan oleh AI” | SS-2025-003;<br>SS-2024-002;<br>SS-2024-003;<br>SS-2026-001;<br>SS-2026-002;<br>SS-2025-004;<br>SS-2025-005 |
| Absence-based verification                                  | Use of missing official source, original-video comparison, altered audio identification, synthetic-content label, rebuttal, or claim-source mismatch                     | 3         | 12.5%      | Manipulated product promotion, public-assistance hoax, corporate/political accusation                                      | “tidak ditemukan”; “bukan berasal dari rekaman asli”; “konten ... dihasilkan menggunakan teknologi AI”        | SS-2024-001;<br>SS-2025-006;<br>SS-2025-007                                                                 |
| Institutional assertion of risk                             | Police report, cyber-patrol statement, displayed evidence, suspect initials, alleged modus, victim recruitment path, or prevention-oriented language                     | 2         | 8.3%       | Public-official impersonation, fake government assistance, WhatsApp recruitment, cyber patrol                              | “Isi konten menawarkan bantuan pemerintah”; “membuat video, gambar, atau audio palsu”                         | SS-2025-001;<br>SS-2025-002                                                                                 |
| Public hoax classification with restricted extractable text | Official or fact-check classification exists, but the available public text contains limited directly quotable authentication language                                   | 3         | 12.5%      | Investment-platform scam, quiz deepfake, aid-and-school-fee hoax                                                           | “No direct quote retained; verify page before quotation”                                                      | SS-2025-008;<br>SS-2025-009;<br>SS-2024-004                                                                 |
| Procedural and regulatory uncertainty                       | Legal database metadata, statutory framework, electronic-information context, personal-data protection, criminal-procedure relevance, or institutional warning           | 5         | 20.8%      | Deepfake risk discourse, electronic information, criminal procedure, biometric/personal data, AI-related public warning    | “Sekarang suara kita bisa ditiru”; “mulai berlaku pada tanggal 02 Januari 2026”; “melindungi Data Pribadi”    | CTX-2023-001;<br>CTX-2026-001;<br>LEG-2024-001;<br>LEG-2025-001;<br>LEG-2022-001                            |
| Technical and forensic caution                              | Reference to language sensitivity, generalisability, criminal-justice threat, forensic voice comparison, segmental features, or sociolinguistic discernment              | 4         | 16.7%      | Detection limitation, forensic audio method, criminal-justice risk, sociolinguistic contribution                           | “limitations”; “generalisability”; “risk”; “segmental features”                                               | ACAD-2024-001; ACAD-2025-001;<br>ACAD-2024-002; ACAD-2025-002                                               |
| —                                                           | —                                                                                                                                                                        | 24        | 100%       | —                                                                                                                          | —                                                                                                             | —                                                                                                           |

The strongest pattern is algorithmic probability and detector scoring, appearing in 7 documents, or 29.2% of the corpus. Legal and institutional risk context follows with 5 documents, or 20.8%, while scholarly-methodological limitation accounts for 4 documents, or 16.7%. Negative-source comparison and limited public-record verification each appear in 3 documents, or 12.5%, and official or enforcement risk framing appears in 2 documents, or 8.3%. The distribution shows that evidentiary uncertainty is not handled through one dominant authority. Fact-checkers tend to rely on detector language and source comparison; enforcement-facing sources rely on modus and institutional assertion; legal-context sources define procedural relevance; scholarly sources explain methodological limits. The corpus therefore shows distributed rather than centralised authentication.

Theoretically, this pattern indicates that deepfake speech evidence operates through calibrated uncertainty. A detector percentage does not automatically become legal proof, and an institutional statement does not by itself resolve acoustic authenticity. Instead, each document positions uncertainty differently: probabilistic, source-based, procedural, institutional, or methodological. This is consistent with forensic and criminal-justice scholarship arguing that deepfakes do not only produce false media but also destabilise evidentiary confidence in genuine media [7], [25]. It also supports Mallinson et al.'s (2024) argument that linguistic and sociolinguistic expertise should be integrated into audio deepfake discernment. In this study, evidentiary uncertainty is therefore not treated as a methodological weakness; it is the object of analysis.

#### 4.3. Institutional and procedural treatment of deepfake speech evidence

Synthetic speech is procedurally positioned through a public verification ecology before it reaches the stricter evidentiary environment of formal adjudication. The largest share of the corpus, 14 of 24 documents, places deepfake voice or manipulated audio within fact-checking and media verification. This indicates that the first institutional response to synthetic speech is often corrective and classificatory: identifying a claim as false, tracing its source, or warning users against scams. Only one document in the corpus is enforcement-facing, while legal instruments and scholarly sources function as contextual scaffolding. This distribution is consistent with criminal-justice research showing that deepfakes create evidentiary pressure before a court formally determines authenticity, admissibility, or probative value [7], [4].



**Figure 4.** Procedural positioning of synthetic speech evidence

**Table 4.** Legal-procedural positioning of synthetic speech evidence

| Institutional actor                                  | Operational indicator                                                                                                                                                     | Frequency | Percentage | Data variation                                                                                                                                                        | Platform trace | Legal-procedural relevance                                                                         | Data codes                                                                                                                                                                                                                  |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fact-checker / media                                 | Public-facing verification of synthetic voice, manipulated audio, deepfake video, source mismatch, detector result, or official-account absence                           | 14        | 58.3%      | AI voice, manipulated audio, public-official impersonation, celebrity deepfake, public assistance hoax, commercial scam, health misinformation, policy misinformation | Present        | Pre-legal authenticity screening; public correction before or outside formal adjudication          | SS-2025-002;<br>SS-2025-003;<br>SS-2024-001;<br>SS-2024-002;<br>SS-2024-003;<br>SS-2026-001;<br>SS-2026-002;<br>SS-2025-004;<br>SS-2025-005;<br>SS-2025-006;<br>SS-2025-007;<br>SS-2025-008;<br>SS-2025-009;<br>SS-2024-004 |
| Law enforcement                                      | Police report, suspect initials, alleged modus, social-media dissemination, WhatsApp contact route, victim recruitment path, or seizure/evidence description              | 1         | 4.2%       | Public-official impersonation, fake government assistance, social-media circulation, WhatsApp-based contact                                                           | Present        | Synthetic media is positioned as part of a reported criminal scheme, not merely as misinformation  | SS-2025-001                                                                                                                                                                                                                 |
| Legal database / state legal source                  | Statutory material concerning electronic information, criminal procedure, personal data, or evidentiary framework                                                         | 3         | 12.5%      | UU ITE 2024, KUHAP 2025, PDP 2022                                                                                                                                     | Not applicable | Provides legal context for electronic evidence, procedural treatment, and identity/data protection | LEG-2024-001;<br>LEG-2025-001;<br>LEG-2022-001                                                                                                                                                                              |
| Government / public institution / state news context | Public warning, institutional framing, prevention discourse, or national-risk explanation concerning deepfake audio, video, or image                                      | 2         | 8.3%       | Public warning about imitated voice, regulatory and institutional risk framing                                                                                        | Not applicable | Frames synthetic speech as a governance and public-risk problem                                    | CTX-2023-001;<br>CTX-2026-001                                                                                                                                                                                               |
| Academic literature                                  | Methodological discussion of criminal-justice risk, forensic voice comparison, sociolinguistic discernment, segmental features, detection limits, or language sensitivity | 4         | 16.7%      | Criminal justice risk, forensic audio method, sociolinguistic contribution, technical detection context                                                               | Not applicable | Provides interpretive tools for evaluating the limits of public-source authentication              | ACAD-2024-001; ACAD-2025-001;<br>ACAD-2024-002; ACAD-2025-002                                                                                                                                                               |
| —                                                    | —                                                                                                                                                                         | 24        | 100%       | —                                                                                                                                                                     | —              | —                                                                                                  | —                                                                                                                                                                                                                           |

The dominant pattern is public verification and fact-checking, representing 58.3% of the corpus. These documents commonly include platform traces, such as TikTok, Facebook, Instagram, YouTube, or WhatsApp, and they position synthetic speech as a circulating object that must be corrected before it produces further harm. Legal and scholarly context together account for 29.2%, showing that the corpus also contains a strong interpretive layer: statutes define procedural relevance, while academic studies explain technical and sociolinguistic limitations. Institutional risk communication accounts for 8.3%, and law-enforcement case construction accounts for 4.2%. The pattern is therefore asymmetrical: most documents classify and contextualise synthetic speech publicly, while only a small part presents it as an enforcement-facing evidentiary object.

Theoretically, this pattern shows that synthetic speech evidence does not move directly from digital circulation to legal proof. It passes through intermediary regimes of verification, warning, classification, and institutional framing. Fact-checkers and media actors operate as early evidentiary translators, converting platform content into a documented claim about manipulation. Law enforcement translates synthetic speech into alleged modus, harm, and victimisation. Legal instruments translate it into questions of electronic information, procedure, and identity protection. Scholarly sources translate it into methodological uncertainty. This supports the argument that authentication is not merely acoustic or computational, but institutional and procedural. In this study, synthetic speech becomes evidence through a chain of discourse practices that assign it source, status, risk, and legal relevance.

## 5. DISCUSSION

The first finding shows that linguistic authentication in Indonesian deepfake-related documents is built less through isolated acoustic judgement than through convergence among detector outputs, source tracing, official-account verification, and comparison with original media. Its implication is significant: synthetic speech becomes evidentially intelligible only when voice is connected to a broader chain of digital provenance. This complicates conventional assumptions in forensic voice work, where speaker identity is often approached through acoustic similarity, voice quality, and expert comparison. In this corpus, however, voice is rarely allowed to stand alone. It is authenticated or challenged through surrounding traces: platform circulation, missing source material, public rebuttal, and institutional classification. This finding supports Li et al. (2024) [3], Pham et al. (2024) [13], and Khanjani et al. (2023) [2], who emphasise the technical difficulty of speech deepfake detection, but it extends their argument by showing that public evidentiary reasoning depends on more than detection architecture. The function of linguistic authentication is therefore reconstructive, not merely classificatory.

This pattern appears to arise from the unstable evidentiary status of synthetic speech in public legal discourse. Deepfake voice is persuasive precisely because it imitates recognisable features of a speaker, but public documents cannot usually demonstrate authenticity through full forensic examination. As a result, authentication shifts toward a practical structure of corroboration: a detector score is strengthened by source comparison; a claim becomes doubtful when no official trace exists; an alleged voice is treated as suspect when circulation context conflicts with institutional records. This does not prove that detector-supported verification is legally sufficient. Rather, it shows why public authentication often operates as layered plausibility. The finding partly contrasts with forensic audio studies that foreground segmental, spectral, and model-specific features [6], [17], because such features are not always visible in public documents. Yet it is compatible with Mallinson et al. (2024) [8], who argue that speech authenticity requires sociolinguistic and contextual interpretation alongside technical analysis.

The second finding concerns evidentiary uncertainty. The corpus shows that uncertainty is not a residual weakness to be removed after verification; it is actively structured through probability, absence, institutional risk language, and methodological limitation. This matters because deepfake evidence operates in a zone between technical suspicion and legal relevance. Detector scores may increase confidence, but they also preserve uncertainty through probabilistic language. Absence-based verification may undermine a claim, but it does not provide direct proof of how the media was produced. Institutional warnings may stabilise public risk, but they do not substitute for courtroom proof. This supports Sandoval et al. (2024) [7], who argue that deepfakes threaten criminal justice by challenging the reliability of evidentiary materials, and Wang et al. (2022) [25], who stress reliability as a central issue in detection systems. The implication is that uncertainty should not be treated as methodological noise. In this study, it is the central condition through which synthetic speech becomes discussable as evidence.

The underlying structure of this uncertainty is the separation between audibility and accountability. A voice may be heard, recognised, and circulated, while the conditions of its production remain unsettled. This produces a double evidentiary problem: fabricated speech may appear authentic, while authentic speech may

become deniable. The corpus reflects this structure through several forms of stance: probabilistic detector claims, negative-source comparisons, limited public-record verification, legal-contextual framing, and scholarly caution. These patterns are consistent with Alali and Theodorakopoulos (2025) [5], who show that partial fake speech attacks complicate utterance-level certainty, and with Nguyen et al. (2025) [22], who foreground linguistic sensitivity in detection. They also complicate more solution-oriented detection literature by showing that uncertainty persists even when a document offers a verification outcome. The reason is not simply technological insufficiency. It is also procedural: public evidence is mediated by platforms, fact-checkers, law-enforcement statements, legal categories, and access limits.

The third finding shows that synthetic speech is procedurally positioned most often through public verification and fact-checking, while fewer documents present it as an enforcement-facing or courtroom-ready evidentiary object. The implication is that the earliest evidentiary work occurs outside formal adjudication. Fact-checkers, media organisations, and public institutions act as preliminary translators of deepfake speech: they convert circulating content into documented claims about manipulation, absence of source, platform trace, or public risk. This does not mean that fact-checking replaces legal proof. Rather, it indicates that deepfake speech first acquires public evidentiary status through correction and classification before it enters criminal procedure, if it enters at all. Qureshi et al. (2024) [4] and Mubarak et al. (2023) [14] similarly describe deepfakes as multimodal forensic and social-media problems, but this corpus highlights the procedural asymmetry between public verification and legal adjudication. Synthetic speech therefore circulates through a pre-legal evidentiary ecology.

This procedural distribution appears to reflect the institutional gap between technological circulation and formal evidence. Deepfake audio spreads quickly through platforms, while legal evaluation depends on slower requirements: admissibility, chain of custody, expert qualification, statutory relevance, and evidentiary weight. Indonesian legal-linguistic scholarship helps explain why this matters. Setiawan (2026) [10] shows that forensic linguistic evidence in criminal verdicts is shaped by courtroom language and legal reasoning, while Ansori (2026) [9] demonstrates that investigative transcripts are structured by institutional speech acts. In the deepfake context, however, many documents remain before that courtroom threshold. They are warnings, verifications, legal contexts, or methodological references rather than judicial determinations. This partly diverges from technical studies that treat detection as the main endpoint, because the corpus shows that detection is only one stage in a longer evidentiary pathway. The broader implication is that synthetic speech becomes evidence through institutional translation: from platform artefact, to public claim, to legal question, and only potentially to adjudicated proof.

## 6. CONCLUSION

This study shows that synthetic speech becomes evidence not through voice recognition alone, but through a layered process of linguistic authentication, evidentiary stance, and procedural positioning. The most important lesson is that a recorded voice can no longer be treated as a transparent sign of speaker identity, intention, or presence. In the Indonesian corpus, authenticity is constructed through convergence among detector claims, source tracing, platform context, institutional statements, and legal categorisation. This study contributes to forensic linguistics by shifting the question from whether deepfake speech can be detected to how disputed speech is made evidentially intelligible under uncertainty.

This study is limited by its reliance on publicly accessible documents rather than full case files, courtroom transcripts, original audio materials, or forensic laboratory reports. Consequently, this study analyses public evidentiary discourse, not the technical truth of individual recordings or judicial determinations of admissibility. Further research should examine court decisions, expert testimony, police dossiers where accessible, and controlled forensic comparison of Indonesian synthetic speech across languages, accents, platforms, and recording conditions. Future studies should also compare Indonesian cases with other multilingual jurisdictions to clarify how legal systems translate synthetic voice from digital artefact into admissible evidence.

## ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

## FUNDING INFORMATION

Author states no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

**Rika Yani:** conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing.

## CONFLICT OF INTEREST STATEMENT

Author states no conflict of interest.

## AI USE DECLARATION

The author used artificial intelligence tools to assist with the generation of illustrative figures and with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

## INFORMED CONSENT

Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible enforcement releases, fact-checking records, and news reports.

## ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible enforcement releases, fact-checking records, and news reports, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

## DATA AVAILABILITY

The corpus of documents analyzed in this study are publicly accessible and available from the corresponding author upon reasonable request.

## REFERENCES

- [1] F. Abbas and A. Taeihagh, "Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence," *Expert Systems with Applications*, vol. 252, p. 124260, 2024, doi: 10.1016/j.eswa.2024.124260.
- [2] Z. Khanjani, G. Watson, and V. P. Janeja, "Audio deepfakes: A survey," *Frontiers in Big Data*, vol. 5, 2023, doi: 10.3389/fdata.2022.1001063.
- [3] M. Li, Y. Ahmadiadli, and X.-P. Zhang, "A survey on speech deepfake detection," *ACM Computing Surveys*, vol. 57, pp. 1–38, 2024, doi: 10.1145/3714458.
- [4] S. Qureshi, A. Saeed, S. H. Almotiri, F. Ahmad, and M. A. Ghamdi, "Deepfake forensics: A survey of digital forensic methods for multimodal deepfake identification on social media," *PeerJ Computer Science*, vol. 10, 2024, doi: 10.7717/peerj-cs.2037.
- [5] A. Alali and G. Theodorakopoulos, "Partial fake speech attacks in the real world using deepfake audio," *Journal of Cybersecurity and Privacy*, vol. 5, p. 6, 2025, doi: 10.3390/jcp5010006.
- [6] T. Yang, C. Sun, S. Lyu, and P. Rose, "Forensic deepfake audio detection using segmental speech features," *Forensic Science International*, vol. 379, p. 112768, 2025, doi: 10.1016/j.forsciint.2025.112768.
- [7] M. Sandoval, M. De Almeida Vau, J. Solaas, and L. Rodrigues, "Threat of deepfakes to the criminal justice system: A systematic review," *Crime Science*, vol. 13, 2024, doi: 10.1186/s40163-024-00239-1.
- [8] C. Mallinson, V. Janeja, C. Evered, Z. Khanjani, L. Davis, N. Bhalli, and K. Nwosu, "A place for (socio)linguistics in audio deepfake detection and discernment: Opportunities for convergence and interdisciplinary collaboration," *Language and Linguistics Compass*, vol. 18, 2024, doi: 10.1111/lnc3.12527.
- [9] M. Ansori, "Speech act strategies in police interrogation transcripts: Evidence from bahasa Indonesia investigative records," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 51–60, 2026.
- [10] W. B. M. Setiawan, "Forensic linguistic evidence in Indonesian courtrooms: A study of language use in criminal verdicts," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 1–12, 2026.
- [11] R. Agoestyowati, "Calibrated acoustic evidence: Legal and methodological advances in forensic voice comparison for Indonesia," *Lingua*, vol. 2024, doi: 10.61978/lingua.v2i4.987.
- [12] O. A. Shaaban, R. Yıldırım, and A. A. Alguttar, "Audio deepfake approaches," *IEEE Access*, vol. 11, pp. 132652–132682, 2023, doi: 10.1109/access.2023.3333866.
- [13] L. Pham, P. Lam, T. Nguyen, H. Tang, H. Nguyen, A. Schindler, and C. Vu, "A comprehensive survey with critical analysis for deepfake speech detection," *Computer Science Review*, vol. 57, p. 100757, 2024, doi: 10.48550/arxiv.2409.15180.
- [14] R. Mubarak, T. A. A. Alsbou, O. Alshaikh, I. Inuwa-Dutse, S. Khan, and S. Parkinson, "A survey on the detection and impacts of deepfakes in visual, audio, and textual formats," *IEEE Access*, vol. 11, pp. 144497–144529, 2023, doi: 10.1109/access.2023.3344653.
- [15] I. Amerini, M. Barni, S. Battiato, P. Bestagini, G. Boato, V. Bruni, et al., "Deepfake media forensics: Status and future challenges," *Journal of Imaging*, vol. 11, 2025, doi: 10.3390/jimaging11030073.
- [16] A. De Ruiter, "The distinct wrong of deepfakes," *Philosophy & Technology*, vol. 34, pp. 1311–1332, 2021, doi: 10.1007/s13347-021-00459-2.
- [17] G. Hua, A. Teoh, and H. Zhang, "Towards end-to-end synthetic speech detection," *IEEE Signal Processing Letters*, vol. 28, pp. 1265–1269, 2021, doi: 10.1109/lsp.2021.3089437.
- [18] A. Dixit, N. Kaur, and S. Kingra, "Review of audio deepfake detection techniques: Issues and prospects," *Expert Systems*, vol. 40, 2023, doi: 10.1111/exsy.13322.
- [19] A. Heidari, N. Navimipour, H. Dağ, and M. Unal, "Deepfake detection using deep learning methods: A systematic and comprehensive review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 14, 2023, doi: 10.1002/widm.1520.

- [20] Z. Almutairi and H. Elgibreen, "Detecting fake audio of Arabic speakers using self-supervised deep learning," *IEEE Access*, vol. 11, pp. 72134–72147, 2023, doi: 10.1109/access.2023.3286864.
- [21] N. M. Muller, P. Kawa, W. Choong, E. Casanova, E. Golge, T. Muller, et al., "MLAAD: The multi-language audio anti-spoofing dataset. In 2024 International Joint Conference on Neural Networks (IJCNN) (pp. 1–7)," *IEEE*, 2024, doi: 10.1109/ijcnn60899.2024.10650962.
- [22] B. Nguyen, S. Shi, R. Ofman, and T. Le, "What you read isn't what you hear: Linguistic sensitivity in deepfake speech detection," *arXiv*, 2025, doi: 10.48550/arxiv.2505.17513.
- [23] Sadani, "Linguistic construction of justice: A pragmatic analysis of judges' reasoning in Indonesian constitutional court decisions," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 31–40, 2026.
- [24] D. Setiaji, "Legal language and power: A critical discourse analysis of anti-corruption laws in Indonesia," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 13–21, 2026.
- [25] T. Wang, X. Liao, K.-P. Chow, X. Lin, and Y. Wang, "Deepfake detection: A comprehensive survey from the reliability perspective," *ACM Computing Surveys*, vol. 57, pp. 1–35, 2022, doi: 10.1145/3699710.

## BIOGRAPHY OF AUTHOR



**Rika Yani**    is affiliated with Universitas PGRI Sumenep, Indonesia. Her academic concerns are related to literature, language, and related humanities fields, and she is engaged in research on popular literature, political rhetoric, discourse analysis, and cultural representation. Her work examines how literary texts construct ideas of society, peoplehood, and identity in contemporary Indonesian popular culture. She can be contacted at email: [rikayani665@gmail.com](mailto:rikayani665@gmail.com)