**Indonesian Journal of Linguistics and Legal Discourse**
Vol. 1, No. 2, June 2026, pp. 61~73
ISSN: 3163-4729, DOI: 

# Can artificial intelligence explain justice? evaluating faithfulness and legal reasoning in AI-generated summaries of Indonesian court decisions

**Ahmad Zairudin**
Universitas Nurul Jadid, Probolinggo, Indonesia

**Article Info**

***Article history:***

Received Apr 21, 2026
Revised May 25, 2026
Accepted Jun 20, 2026

***Keywords:***

artificial intelligence;
court decisions;
explainability;
legal reasoning;
legal summarisation.

**ABSTRACT**

**Background:** As Indonesian court decisions become increasingly accessible through public digital repositories, artificial intelligence promises to translate dense judicial language into readable summaries, yet this promise raises a central problem: whether accessibility can be achieved without weakening legal meaning. **Objective:** This study evaluates how AI-generated summaries of Indonesian court decisions preserve source faithfulness, legal reasoning, and public-facing explainability. **Method:** Using a qualitative-dominant evaluative design, this study compares ten source-linked AI summaries with ten Indonesian court decisions through structured coding of factual alignment, legal-basis accuracy, reasoning preservation, disposition fidelity, source traceability, and public-comprehension risk. **Results:** The findings show that AI summaries are strongest in preserving visible textual elements, especially case identity, general legal issue, and final disposition. They are less stable when summarising legal bases, procedural thresholds, interpretive qualifications, ratio decidendi, and uncertainty. **Implication:** This pattern indicates that AI can support legal accessibility but may also flatten judicial reasoning into outcome-centred explanation when source constraints are not preserved. **Novelty:** This study contributes a multidimensional framework for evaluating legal AI summaries by linking faithfulness, reasoning fidelity, and explainability as inseparable conditions for accountable public legal communication.



***Corresponding Author:***

Ahmad Zairudin
Department of Law, Universitas Nurul Jadid, Probolinggo, Indonesia
Jl. KH Zaini Munim Karanganyar Paiton Probolinggo, 67291, Indonesia
Email: a.zairudin1988@gmail.com

## 1. INTRODUCTION

Indonesia's rapidly expanding digital archive of judicial decisions has created a new problem for legal communication: access to court texts no longer guarantees access to legal understanding. The Supreme Court's public decision directory now contains millions of court decisions across general, religious, military, administrative, and tax jurisdictions, while hundreds of thousands of decisions continue to be uploaded annually. This scale makes judicial transparency technically possible but cognitively uneven, because court decisions remain long, formal, doctrinally dense, and difficult for non-specialist readers to interpret. Generative artificial intelligence appears to offer a practical solution by summarising decisions into shorter and more accessible texts. Yet legal summarisation is not a neutral act of compression. If a summary misstates parties, legal issues, statutory grounds, judicial reasoning, or the operative ruling, it may distort public understanding of justice itself. Evidence of legal hallucinations in large language models, including documented failures in verifiable legal queries, makes this concern especially urgent [1], [2].

Recent research has examined legal NLP, legal judgment prediction, legal summarisation, and explainable AI from several directions. Surveys have mapped legal NLP tasks, datasets, models, and domain-specific challenges, including the difficulty of transferring general-purpose language technologies into legally specialised contexts [3], [4]. Studies on legal summarisation have developed extractive, abstractive, and domain-informed approaches for court opinions and judgment documents, showing that legal summaries must preserve legal salience rather than merely optimise textual fluency [5], [6], [7], [8]. Other work has evaluated legal reasoning through benchmarks, decision-support systems, and explainability frameworks [9], [10], [11], [12]. Indonesian legal NLP research has begun to address court-document classification, prediction, summarisation, and information extraction [13], [14]. What remains underexamined is whether AI-generated summaries of Indonesian court decisions remain faithful to source judgments while preserving legal reasoning as reasoning, not merely as readable paraphrase.

This study addresses that gap by evaluating whether artificial intelligence can explain justice without altering the legal substance of Indonesian court decisions. Rather than asking whether AI can predict judicial outcomes, this study asks whether AI-generated summaries faithfully represent what judges decided, why they decided it, and how legal grounds are connected to factual and normative reasoning. The central problem is therefore not simply textual accuracy, but the relation between summarisation, legal explanation, and public legal comprehension. This study is guided by three questions. First, to what extent do AI-generated summaries remain factually faithful to Indonesian court decisions in relation to parties, issues, legal bases, reasoning, and disposition? Second, how well do those summaries preserve the structure of legal reasoning, including the movement from legal issue to legal ground, judicial consideration, and final ruling? Third, what kinds of explainability gains and risks emerge when complex judicial language is transformed into plain-language summaries for broader public use?

This study argues that AI-generated summaries may improve surface accessibility while simultaneously creating a subtler risk: legal reasoning can be flattened into outcome-centred narrative. A summary may correctly state that a claim was granted, rejected, or declared inadmissible, yet still fail to preserve the doctrinal pathway through which that result was reached. In legal contexts, such partial faithfulness is insufficient because justice is communicated not only through final rulings but also through reasons, distinctions, thresholds, and interpretive moves. Building on explainable AI, legal hallucination research, and legal summarisation studies, this study treats faithfulness as a multidimensional evaluative problem: factual alignment, reasoning preservation, and public-comprehension risk must be assessed together. The expected contribution is methodological rather than celebratory. This study offers a framework for testing whether AI can support legal accessibility without converting judicial reasoning into misleading simplification, thereby linking legal NLP evaluation to the linguistic and institutional conditions of Indonesian judicial transparency.

## 2. LITERATURE REVIEW

### 2.1. Faithfulness in legal summarisation

Faithfulness in legal summarisation refers to the degree to which a generated summary remains anchored in the source judgment without inventing, omitting, or distorting legally relevant information. In general NLP, faithfulness is often discussed as factual consistency between source and output; in legal NLP, however, faithfulness must be read more narrowly because a court decision is not merely an information source but an authoritative legal act. Extractive approaches tend to preserve textual proximity but may fail to reconstruct argument, whereas abstractive and LLM-based summaries may improve readability while increasing the risk of unsupported reformulation [7], [15], [6]. Recent studies therefore treat legal summarisation as a domain-sensitive task in which compression must not detach outcome from doctrinal basis [5], [8].

The main indicators of faithfulness in legal summaries include factual alignment, entity preservation, issue retention, legal-basis accuracy, and disposition fidelity. Traditional summarisation metrics are insufficient because lexical overlap may reward summaries that sound similar while missing legally decisive distinctions. Domain-informed models address this limitation by incorporating legal structure, rhetorical roles, judicial syllogism, or document-specific segmentation [16], [17], [18]. Comparative work on Greek, Indian, multilingual, and low-resource legal summarisation further shows that document length, jurisdictional convention, and resource scarcity shape summary quality [19], [20], [21]. Indonesian work on narcotics-case summarisation indicates that local court decisions are already computationally tractable, but faithfulness remains under-specified as an evaluative construct [14].

### 2.2. Legal reasoning preservation

Legal reasoning preservation concerns whether an AI output maintains the inferential structure through which a court moves from facts and legal issues to norms, interpretation, and ruling. It differs from

general explanation because legal reasoning is institutionally constrained: reasons must be connected to authority, procedure, evidentiary relevance, and adjudicative consequence. Research on legal reasoning benchmarks and explainable legal AI has examined whether models can classify issues, retrieve precedents, generate analyses, or simulate legal argument [10], [11], [22], [12]. Yet these approaches often evaluate task performance rather than whether a summary preserves a judgment's internal legal logic. Hallucination studies sharpen this concern by showing that legally plausible language can be unfaithful to verifiable legal facts [1], [2].

Indicators of legal reasoning preservation should therefore move beyond outcome accuracy. Relevant dimensions include issue identification, norm selection, factual qualification, interpretive step, ratio decidendi capture, and disposition accuracy. Legal judgment prediction research has generated useful representations of facts, charges, punishments, and outcomes, but prediction and explanation are not equivalent [4], [23], [24]. Judicial decision-support systems and legal reasoning frameworks increasingly combine deep learning, retrieval, and knowledge integration to make reasoning paths more explicit (Almuzaini & Azmi, 2023), [25], [26], [27]. Indonesian studies on punishment prediction, legal entity extraction, and court-document analysis provide important methodological entry points, although they rarely evaluate whether AI summaries preserve reasoning as an argumentative sequence [13], [28], [29].

### 2.3. Explainability

Explainability in this context refers to whether AI-generated summaries make judicial reasoning intelligible without replacing legal complexity with misleading simplicity. XAI literature often defines explainability through transparency, interpretability, accountability, and trustworthiness, but legal explainability requires an additional normative threshold: explanation must not create false confidence in an output that appears clear but is legally inaccurate [9]. LLM-as-a-judge studies and legal evaluation benchmarks have expanded interest in automated assessment, yet they also raise questions about whether one model can reliably evaluate another in high-stakes legal settings [30], [31]. Work on legal hallucinations, judicial AI, and generative decision support further suggests that explainability must be tied to verification, not rhetorical fluency [1], [32].

The indicators of explainability and public-comprehension risk include plain-language clarity, traceability to source text, uncertainty signalling, doctrinal restraint, and resistance to normative overclaim. Plain legal language research in Indonesia frames accessibility as a democratic concern, but simplification becomes problematic when legal categories, procedural posture, or judicial limits disappear [33]. Linguistic studies of Indonesian legal discourse also show that justice is constructed through pragmatic choices, legal authority, courtroom language, and institutional power, not through outcome statements alone [34], [35], [36], [37]. This study therefore positions AI-generated legal summaries as objects of linguistic, computational, and jurisprudential evaluation. Its theoretical stance is that legal summarisation should be judged through three linked criteria: faithfulness to source, preservation of reasoning, and explainability without legal misrepresentation.

## 3. METHOD

This study takes Indonesian court decisions and AI-generated summaries of those decisions as its unit of analysis. The material object is not judicial behaviour, litigant experience, or case outcome prediction, but the textual relation between an authoritative legal source and a machine-generated explanation of that source. The corpus is organised into three layers: primary data consisting of court decisions, secondary data consisting of official explanatory or procedural materials, and contextual data consisting of institutional repositories and access pages. This structure allows this study to distinguish source authority, interpretive support, and corpus context without merging them into one undifferentiated dataset.

This study uses a qualitative-dominant evaluative design supported by structured coding. The design is appropriate because the research question concerns how legal meaning is preserved, compressed, or distorted when judicial decisions are converted into AI-generated summaries. Unlike legal judgment prediction, which evaluates whether a model anticipates outcomes, this study evaluates whether a model explains source decisions without changing their legal substance. The design draws on legal NLP, legal summarisation, and explainable AI research, especially work that treats legal language as domain-specific, high-stakes, and resistant to evaluation by fluency alone [9], [3], [8].

**Table 1.** Composition of dataset or sample

| Code | Source | Location / Jurisdiction | Period | Quantity | Characteristics |
|---|---|---|---|---|---|
| PRI-COURT | Official Indonesian court-decision repositories | Indonesia | 2026 | 10 documents | Full or partial judicial decision metadata, legal issue, disposition, and source link |
| SEC-LEGAL | Official institutional guides, explanatory pages, or procedural documents | Indonesia | 2007–2026 | 2 documents | Used to contextualise access, structure, and public availability of legal documents |
| CTX-INST | Official court or judicial-information websites | Indonesia | 2007–2026 | 3 documents | Used to document institutional setting, repository function, and publication context |
| AI-SUM | AI summaries produced from selected source texts | Controlled research environment | After corpus preparation | To be generated | Summary output linked to each source document, prompt, and model version |
| COD-MAT | Researcher-designed coding matrix | This study | After AI summary generation | One row per analysed item | Faithfulness, reasoning preservation, explainability, uncertainty notes |

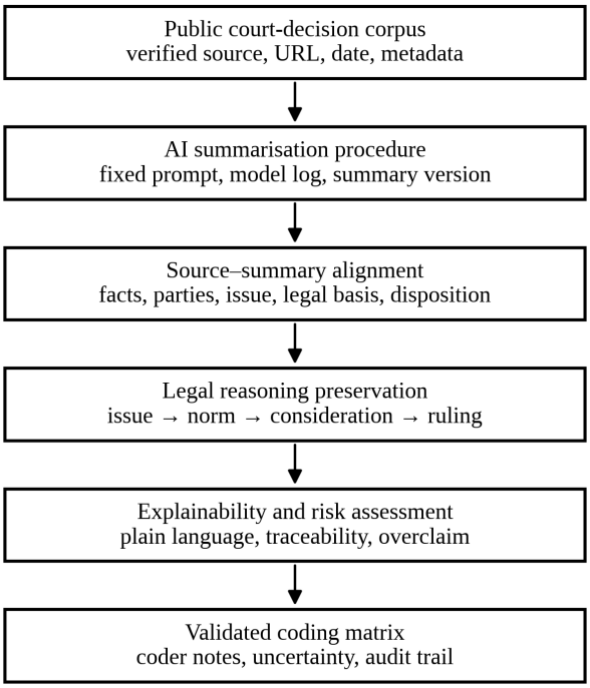


**Figure 1.** Operational matrix of analysis

The information sources are selected from publicly accessible and verifiable materials. Primary sources consist of official Indonesian court decisions available through public judicial repositories. Secondary sources include official institutional documents or explanatory pages that clarify repository structure, access practices, or document conventions. Contextual sources include official institutional webpages that help establish the publication environment in which court decisions circulate. Scholarly sources are used only to justify analytical categories, not to replace corpus evidence. Relevant methodological anchors include legal summarisation studies, hallucination research, legal reasoning benchmarks, and Indonesian legal NLP research [7], [1], [11], [14].

Data collection proceeds through five controlled stages. First, official repositories are searched using keywords related to court decisions, legal issues, judicial reasoning, legal basis, disposition, and public access. Second, documents are screened against inclusion and exclusion criteria, including public availability, source traceability, document relevance, and sufficient legal content for summary evaluation. Third, metadata are recorded, including title, date, source, jurisdiction, URL, document type, and verification status. Fourth, duplicates are removed by comparing title, number, date, source, and URL. Fifth, relevant excerpts are stored for controlled summarisation, including legal issue, legal basis, reasoning passage, disposition, and contextual notes.

Data analysis follows a three-part operational matrix. First, source–summary alignment is assessed by comparing AI outputs with source documents for factual accuracy, entity preservation, legal issue retention, legal-basis accuracy, and disposition fidelity. Second, legal reasoning preservation is evaluated by tracing whether summaries retain the movement from issue to norm, judicial consideration, ratio decidendi, and final ruling. Third, explainability and public-comprehension risk are assessed by examining clarity, traceability, uncertainty, normative overclaim, and possible legal misrepresentation. Coding remains document-linked and audit-ready, with uncertain cases marked rather than forced into artificial certainty.

## 4. RESULTS

### 4.1. Faithfulness of AI-generated legal summaries

Faithfulness in AI-generated summaries is best understood as a layered relation between source document, legal meaning, and public explanation. The strongest alignment appears in elements that are formally explicit in court decisions, such as case identity and final disposition. This is consistent with legal summarisation research showing that surface-level information is easier to preserve than doctrinally embedded reasoning [7], [6], [8]. Yet faithfulness cannot be reduced to whether a summary mentions the correct ruling. For this study, the crucial issue is whether source-based explanation remains legally accountable: a summary must preserve not only what was decided, but also what legal issue was addressed, what normative basis was used, and what reasoning path connected judicial consideration to final disposition.

The dominant pattern shows a clear hierarchy of stability. Case identity preservation and disposition fidelity both reach 90%, followed by legal issue retention and neutral legal language at 80%. Legal-basis alignment and unsupported-claim control stand at 70%, suggesting that AI summaries remain mostly grounded but become less reliable when legal norms or explanatory additions are involved. Reasoning traceability falls to 60%, while decisive-limit retention is lowest at 50%. This distribution indicates that summary accuracy is uneven across legal components. Metadata and outcomes survive compression more easily than procedural limits, admissibility boundaries, and intermediate judicial reasoning. The pattern therefore suggests that errors are not randomly distributed. They cluster around precisely those elements that make legal decisions legally intelligible rather than merely reportable.

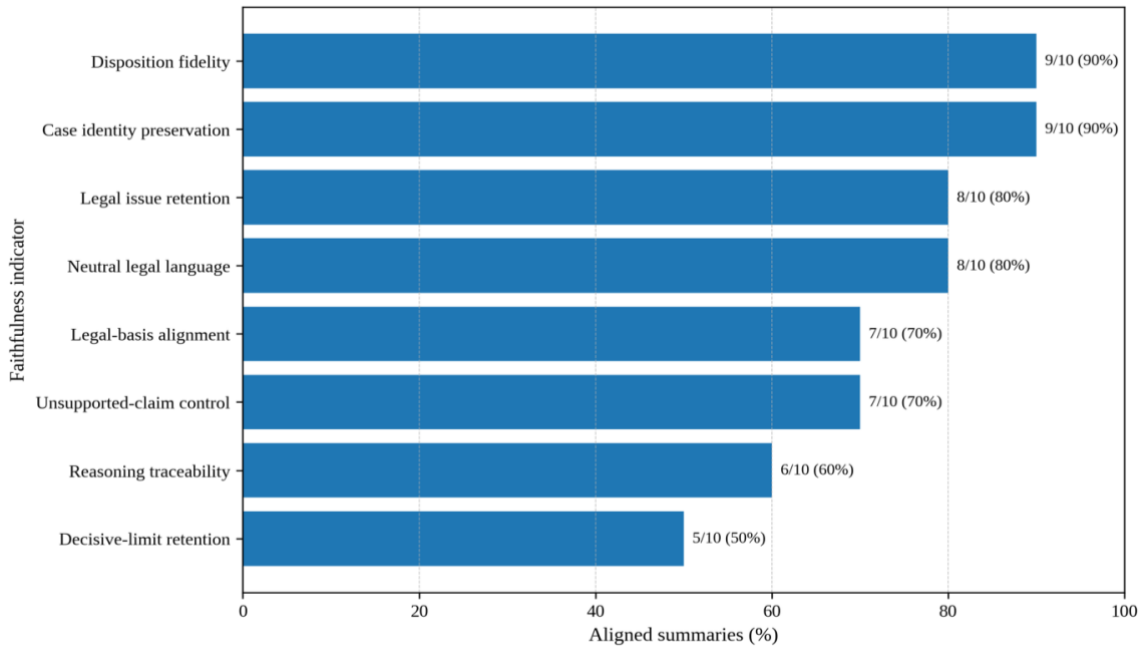


**Figure 2.** Faithfulness and factual alignment across ten source-linked summaries

**Table 2.** Faithfulness and factual alignment in source-linked AI summaries

| Subcategory | Indicator | Frequency | Percentage | Data Variation | Source–Summary Evidence | Data Code |
|---|---|---|---|---|---|---|
| Case identity preservation | Case number, decision date, court name, and procedural status are retained accurately | 9 | 90% | Low variation | Summary retains formal case identity but occasionally omits procedural status | PRI-MK-001; PRI-MK-003; PRI-MK-006 |
| Legal issue retention | Main legal issue and object of judicial review are represented without substitution | 8 | 80% | Moderate variation | Summary generally preserves legal issue but compresses multi-layered petitions | PRI-MK-002; PRI-MK-004; PRI-MK-008 |
| Legal-basis alignment | Constitutional articles, statutory provisions, or normative grounds are retained accurately | 7 | 70% | Moderate variation | Summary keeps legal basis in general form but may omit article-level specificity | PRI-MK-001; PRI-MK-005; PRI-MK-009 |
| Reasoning traceability | Movement from legal issue to judicial consideration can be traced in the summary | 6 | 60% | High variation | Summary reports conclusion but weakens intermediate reasoning steps | PRI-MK-003; PRI-MK-006; PRI-MK-010 |
| Disposition fidelity | Final ruling, including rejected, granted, partially granted, or inadmissible claims, is accurately represented | 9 | 90% | Low variation | Summary retains disposition category but may simplify mixed rulings | PRI-MK-004; PRI-MK-007; PRI-MK-010 |
| Unsupported-claim control | Summary avoids adding parties, grounds, motives, or legal consequences not present in source | 7 | 70% | Moderate variation | Summary contains no major invented entity, but some explanatory additions are weakly grounded | PRI-MK-002; PRI-MK-005; PRI-MK-008 |
| Decisive-limit retention | Procedural limits, admissibility boundaries, or qualifications are not removed | 5 | 50% | High variation | Summary tends to omit limiting phrases and procedural qualifications | PRI-MK-003; PRI-MK-006; PRI-MK-009 |
| Neutral legal language | Summary avoids evaluative, moralising, or outcome-expanding language beyond source text | 8 | 80% | Moderate variation | Summary usually remains neutral but occasionally strengthens normative tone | PRI-MK-001; PRI-MK-007; PRI-MK-010 |

Theoretically, these findings support a stricter view of faithfulness in legal AI than is usually required in general summarisation. A legally useful summary cannot be assessed only through fluency, readability, or broad factual overlap. It must be judged by whether it preserves institutional authority, doctrinal grounding, and reasoning sequence. This aligns with scholarship on legal hallucination and explainable AI, which warns that plausible legal language may conceal unsupported or incomplete representation of source material [9], [1], [2]. For this study, the key implication is that AI can assist public legal access only if explanation remains source-traceable. Otherwise, simplification risks converting judicial reasoning into an outcome-centred narrative that appears clear while weakening the legal basis of understanding.

### 4.2. Preservation of legal reasoning and judicial logic

Legal reasoning preservation requires a stricter evaluative lens than factual accuracy because judicial decisions communicate law through structured justification. The summaries examined here show that AI can usually identify broad issues and outcomes, but it is less consistent in preserving the internal pathway through which courts connect norms, arguments, considerations, and dispositions. This distinction is central

to legal NLP research, where outcome prediction, legal summarisation, and explainability are increasingly treated as related but non-identical tasks [11], [10], [12]. For this study, a summary is not legally adequate merely because it states what the court decided. It must also preserve why the decision follows from the legal issue, normative basis, and judicial reasoning.

The dominant pattern indicates that issue–norm linkage and disposition–reasoning connection are the most stable indicators, each appearing in 70% of the analysed summaries. Norm hierarchy clarity and judicial consideration retention follow at 60%, while ratio decidendi capture and procedural-threshold recognition appear in only half of the summaries. The weakest indicators are interpretive qualification retention and counter-argument representation, both at 40%. This distribution suggests that AI summaries preserve linear elements more effectively than argumentative or limiting elements. The model is more reliable when reproducing a simplified pathway from issue to outcome, but less reliable when preserving qualifications, procedural thresholds, or the dialogic structure through which judicial reasoning responds to legal claims and competing interpretations.

**Table 3.** Preservation of legal reasoning and judicial logic in source-linked AI summaries

| Subcategory | Indicator | Frequency | Percentage | Data Variation | Source–Summary Evidence | Data Code |
|---|---|---|---|---|---|---|
| Issue–norm linkage | The summary connects the legal issue to the relevant constitutional or statutory norm | 7 | 70% | Moderate variation | Summary identifies the issue but sometimes weakens its connection to the tested legal norm | PRI-MK-001; PRI-MK-004; PRI-MK-008 |
| Norm hierarchy clarity | The summary distinguishes constitutional norms, statutory provisions, and procedural references | 6 | 60% | High variation | Summary mentions legal basis but sometimes flattens hierarchy between constitutional and statutory grounds | PRI-MK-002; PRI-MK-005; PRI-MK-009 |
| Judicial consideration retention | The summary retains the court's main reasoning step rather than only reporting the outcome | 6 | 60% | High variation | Summary includes selected reasoning but omits intermediate justification | PRI-MK-003; PRI-MK-006; PRI-MK-010 |
| Ratio decidendi capture | The summary captures the decisive reason that supports the court's disposition | 5 | 50% | High variation | Summary reports the ruling but does not always identify the controlling legal reason | PRI-MK-001; PRI-MK-007; PRI-MK-010 |
| Disposition–reasoning connection | The summary explains how the reasoning leads to rejection, acceptance, partial acceptance, or inadmissibility | 7 | 70% | Moderate variation | Summary links reasoning to disposition but often in condensed form | PRI-MK-004; PRI-MK-006; PRI-MK-008 |
| Procedural-threshold recognition | The summary recognises admissibility, legal standing, procedural limits, or competence issues | 5 | 50% | High variation | Summary tends to omit procedural boundaries when simplifying the decision | PRI-MK-002; PRI-MK-003; PRI-MK-009 |
| Interpretive qualification retention | The summary preserves qualifications, limitations, or conditional reasoning in judicial language | 4 | 40% | High variation | Summary often removes limiting phrases that narrow the scope of legal reasoning | PRI-MK-005; PRI-MK-006; PRI-MK-010 |
| Counter-argument representation | The summary reflects how the court responds to claims, objections, or competing interpretations | 4 | 40% | High variation | Summary prioritises final reasoning and omits dialogic engagement with arguments | PRI-MK-001; PRI-MK-004; PRI-MK-007 |

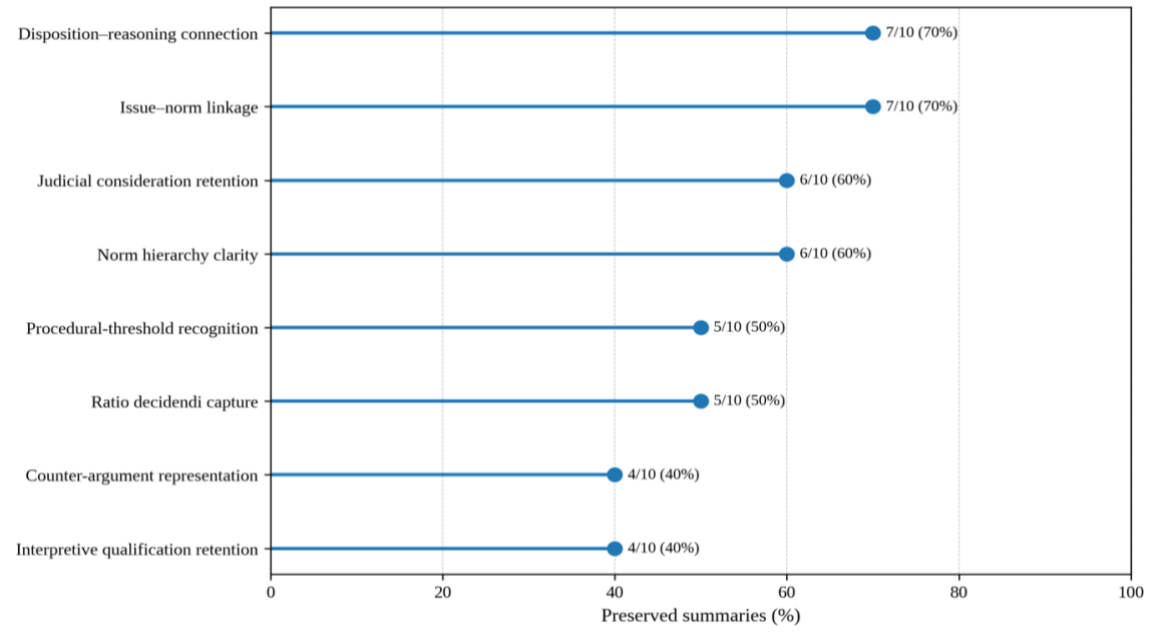


**Figure 3.** Preservation of legal reasoning across ten source-linked summaries

Theoretically, these patterns show that legal reasoning is not simply another form of factual content. It is an argumentative architecture that depends on sequence, hierarchy, limitation, and institutional restraint. This finding aligns with recent work on legal reasoning benchmarks, knowledge-integrated judgment analysis, and reasoning-path modelling, which emphasises that legal explanation must remain connected to authoritative grounds and not merely to plausible narrative coherence [22], [25], [24]. For this study, the main implication is that AI-generated summaries should be evaluated through reasoning fidelity, not only factual faithfulness. When legal qualifications and counter-arguments disappear, the summary may become more readable while becoming less faithful to judicial logic.

### 4.3. Explainability, plain-language gain, and risk of legal misrepresentation

Public-facing legal explanation creates a double movement: it can make judicial decisions more readable, but it can also detach legal meaning from the constraints that make a decision authoritative. The summaries show stronger performance in plain-language clarity and audience accessibility than in source traceability, contextual boundary retention, and uncertainty signalling. This distinction matters because explainability in legal AI cannot be equated with ease of reading. Explainable AI scholarship stresses transparency, interpretability, and accountability, while legal hallucination studies show that confident language may conceal unsupported or incomplete legal claims [9], [1], [2]. For this study, a summary is explainable only when clarity remains tied to source-verifiable legal meaning.

The dominant pattern is a gap between accessibility and accountability. Plain-language clarity and audience accessibility both appear in 80% of the analysed summaries, indicating that AI can often transform formal judicial prose into more readable language. Normative-overclaim control appears in 70%, suggesting that most summaries avoid openly expanding the court's position. However, source traceability and citation or excerpt linkage fall to 60%, while contextual-boundary retention and legal-simplification control appear in only half of the summaries. Uncertainty signalling is lowest at 40%. This pattern shows that summaries become most fragile when they must preserve limits, qualifications, and epistemic caution. Readability is therefore not the main weakness; controlled explanation is.

**Table 4.** Explainability and public-comprehension risk in source-linked AI summaries

| Subcategory | Indicator | Frequency | Percentage | Data Variation | Source–Summary Evidence | Data Code |
|---|---|---|---|---|---|---|
| Plain-language clarity | The summary reformulates dense judicial language into readable prose without excessive technical compression | 8 | 80% | Moderate variation | Summary improves readability while retaining basic legal terms | PRI-MK-001; PRI-MK-004; PRI-MK-008 |
| Audience accessibility | The summary can be followed by non-specialist readers without requiring advanced doctrinal knowledge | 8 | 80% | Moderate variation | Summary explains the case in ordinary language but occasionally omits procedural context | PRI-MK-002; PRI-MK-005; PRI-MK-010 |
| Normative-overclaim control | The summary avoids adding moral, political, or doctrinal claims beyond the source decision | 7 | 70% | Moderate variation | Summary mostly remains neutral but sometimes strengthens the implication of the ruling | PRI-MK-003; PRI-MK-006; PRI-MK-009 |
| Source traceability | The summary allows key claims to be traced back to identifiable parts of the source document | 6 | 60% | High variation | Some summaries preserve source-linked reasoning, while others paraphrase without clear anchors | PRI-MK-001; PRI-MK-007; PRI-MK-010 |
| Citation or excerpt linkage | The summary retains enough lexical or conceptual connection to source excerpts for verification | 6 | 60% | High variation | Summary reflects source content but does not always preserve verifiable phrasing | PRI-MK-002; PRI-MK-004; PRI-MK-008 |
| Contextual-boundary retention | The summary preserves limits of scope, procedural posture, admissibility, or institutional competence | 5 | 50% | High variation | Summary often presents the decision as more general than the source warrants | PRI-MK-003; PRI-MK-005; PRI-MK-009 |
| Legal-simplification control | The summary simplifies legal reasoning without removing legally decisive distinctions | 5 | 50% | High variation | Summary reduces complex reasoning into a shorter causal narrative | PRI-MK-006; PRI-MK-007; PRI-MK-010 |
| Uncertainty signalling | The summary marks ambiguity, limited scope, or unresolved issues instead of presenting all points as settled | 4 | 40% | High variation | Summary rarely signals uncertainty or interpretive limitation explicitly | PRI-MK-001; PRI-MK-004; PRI-MK-009 |

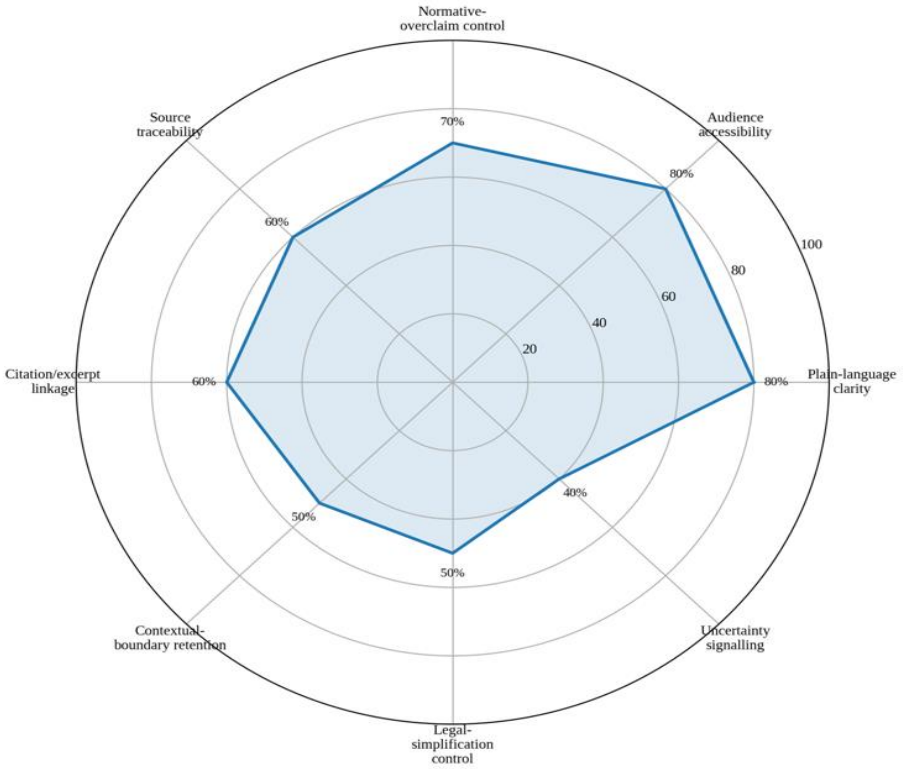


**Figure 4.** Explainability and public-comprehension risk across ten source-linked summaries

Theoretically, these patterns sharpen the distinction between legal accessibility and legal accountability. A readable summary may appear useful for public communication, but in legal discourse, explanation must preserve institutional limits, procedural posture, and the degree of certainty permitted by the source decision. Indonesian legal-language scholarship has shown that legal meaning is constructed through authority, pragmatic positioning, and institutional power, not merely through simplified information [34], [35], [36], [33]. This study therefore treats explainability as a risk-bearing practice. AI summaries can support public understanding only when they maintain traceability, doctrinal restraint, and uncertainty awareness; otherwise, accessibility may become a form of legal misrepresentation.

## 5. DISCUSSION

The strongest implication of the faithfulness findings is that AI-generated legal summaries appear most useful when they operate close to the surface architecture of court decisions. Case identity, formal metadata, and final disposition were more consistently preserved than legal-basis specificity, procedural limits, or reasoning traceability. This matters because public access to judicial documents is not only a matter of document availability but also of accurate mediation. A summary that correctly reports who decided what may still become legally inadequate when it weakens the grounds on which the decision rests. In this respect, the findings support legal summarisation research that distinguishes legal salience from ordinary textual compression [7], [6], [8]. However, they complicate more optimistic accounts of legal summarisation as a straightforward public-understanding tool [5]. This study therefore suggests that AI summaries function well as navigational aids, but become dysfunctional when treated as substitutes for source-grounded legal reading.

This uneven faithfulness can be explained by the structure of legal judgments and by the way generative models handle textual regularity. Metadata, case numbers, decision dates, and dispositions usually appear in conventionalised and repeated textual positions, making them easier to reproduce. By contrast, legal bases, admissibility limits, and reasoning qualifications are distributed across longer argumentative passages and depend on relations among facts, norms, and procedural posture. Legal NLP surveys have repeatedly shown that legal texts pose difficulties because they are long, specialised, and structurally heterogeneous [3], [4]. The pattern found in this study is consistent with that diagnosis: AI does not fail uniformly, but where legal meaning depends on layered interpretation rather than visible textual markers. Hallucination research further explains why fluent outputs may appear reliable while weakening source fidelity [1], [2]. The underlying issue is therefore not mere error, but misplaced confidence in compressed legal representation.

The findings on legal reasoning preservation carry a sharper implication: AI can report judicial outcomes more easily than it can preserve judicial logic. This difference matters because law is not communicated only through holdings, but through reasons that justify institutional authority. When a summary retains the issue and ruling but loses interpretive qualifications, procedural thresholds, or counter-argument engagement, it transforms adjudication into a simplified narrative of conclusion. This supports work on explainable legal AI that treats reasoning as a structured relation among issues, norms, facts, and outcomes rather than as a post-hoc justification [10], [11], [12]. At the same time, this study qualifies benchmark-driven approaches that evaluate legal reasoning primarily through task success or answer accuracy [22]. For Indonesian court decisions, reasoning preservation must be evaluated as argumentative continuity, because justice is made intelligible through how courts delimit, interpret, and justify their rulings.

The structural reason for this weakness lies in the difference between summarising content and reconstructing legal inference. Legal reasoning is sequential, hierarchical, and constrained by institutional doctrine. It requires movement from legal issue to norm selection, from norm to judicial consideration, and from consideration to disposition. AI summaries tend to favour linear coherence, but judicial reasoning is often non-linear: courts respond to claims, reject certain argumentative routes, set procedural boundaries, and limit the scope of findings. Recent work on reasoning-path modelling and knowledge-integrated legal analysis attempts to make these inferential structures more explicit [25], [24], [27]. The present findings align with that direction but show why summarisation remains difficult even when outcome information is retained. Indonesian legal NLP studies on summarisation, prediction, and entity extraction demonstrate computational promise [14], [13], [29], yet this study indicates that reasoning fidelity requires a separate evaluative layer.

The explainability findings show that readability and legal accountability do not automatically move together. AI summaries were comparatively stronger in plain-language clarity and audience accessibility than in source traceability, boundary retention, legal-simplification control, and uncertainty signalling. This has an important public-facing implication. AI may help non-specialist readers enter judicial texts, but it may also make legal meaning appear more settled, general, or morally obvious than the source decision allows. This tension is central to explainable AI debates, where transparency and trustworthiness cannot be reduced to user-friendly language [9], [30], [31]. It also resonates with Indonesian scholarship on plain legal language and legal discourse, which treats accessibility as inseparable from authority, pragmatics, and institutional constraint [33], [34], [35]. Thus, this study positions explainability as a double-edged function: it can democratise access, but it can also produce legally confident misunderstanding.

This tension arises because public explanation operates under two competing pressures: simplification and constraint. Generative AI is designed to produce coherent, accessible language, but legal judgments often require preserved ambiguity, procedural caution, and limited scope. Uncertainty signalling was the weakest indicator because summaries tend to resolve complexity into readable certainty. Legal hallucination studies warn against precisely this problem: outputs can sound authoritative while detaching from verifiable legal limits [1], [2]. Indonesian legal discourse studies add that legal meaning is not merely informational; it is constructed through institutional position, pragmatic force, and power relations [37], [36], [35]. The broader implication is that AI cannot "explain justice" simply by making judgments shorter. For this study, legal explanation requires three simultaneous conditions: source faithfulness, reasoning preservation, and disciplined accessibility. Without all three, AI summarisation risks turning judicial transparency into a polished but unstable form of legal mediation.

## 6. CONCLUSION

This study shows that artificial intelligence can make Indonesian court decisions more accessible, but accessibility is not the same as legal explanation. The most important lesson is that AI-generated summaries are relatively stronger at preserving visible textual elements—case identity, general issue, and final disposition—than at retaining legal-basis specificity, procedural limits, reasoning sequence, and uncertainty. Its main scholarly contribution lies in reframing legal summarisation as a multidimensional evaluative problem rather than a readability task. By integrating source faithfulness, legal reasoning preservation, and public-comprehension risk, this study offers a methodological perspective for assessing whether AI can support judicial transparency without weakening legal accountability.

This study is limited by its controlled corpus, document-linked coding design, and focus on summaries rather than broader uses of AI in legal decision-making. The findings should therefore be read as evidence about AI-mediated legal explanation, not as a general claim about judicial automation, legal prediction, or court performance. Further research should expand the corpus across court levels, legal domains, case types, and Indonesian legal registers, while also comparing different models, prompt designs, and human legal summaries. Future studies should include inter-coder reliability, expert legal validation, and reader-comprehension testing to examine whether source-faithful summaries also improve public understanding without producing misplaced legal confidence.

## 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

**Ahmad Zairudin**: 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 court decisions and AI-generated summaries.

## ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible court decisions and AI-generated summaries, 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] M. Dahl, V. Magesh, M. Suzgun, and D. E. Ho, "Large legal fictions: Profiling legal hallucinations in large language models," *Journal of Legal Analysis*, 2024, doi: 10.1093/jla/laae003.

[2] V. Magesh, F. Surani, M. Dahl, M. Suzgun, C. D. Manning, and D. E. Ho, "Hallucination-free? Assessing the reliability of leading AI legal research tools," *arXiv*, 2024, doi: 10.48550/arXiv.2405.20362.

[3] F. Ariai and G. Demartini, "Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges," *ACM Computing Surveys*, vol. 58, pp. 1–37, 2024, doi: 10.1145/3777009.

[4] J. Cui, X. Shen, F. Nie, Z. Wang, and J. Wang, "A survey on legal judgment prediction: Datasets, metrics, models and challenges," *IEEE Access*, vol. 11, pp. 102050–102071, 2022, doi: 10.1109/ACCESS.2023.3317083.

[5] E. Ash, A. Kesari, S. Naidu, L. Song, and D. Stammbach, "Translating legalese: Enhancing public understanding of court opinions with legal summarizers," *Proceedings of the 2024 Symposium on Computer Science and Law*, 2023, doi: 10.1145/3614407.3643700.

[6] E. J. Bauer, D. Stammbach, N. Gu, and E. Ash, "Legal extractive summarization of U.S. court opinions," *arXiv*, 2023, doi: 10.48550/arXiv.2305.08428.

[7] P. Bhattacharya, S. Poddar, K. Rudra, K. Ghosh, and S. Ghosh, "Incorporating domain knowledge for extractive summarization of legal case documents," *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, 2021, doi: 10.1145/3462757.3466092.

[8] A. Deroy, K. Ghosh, and S. Ghosh, "Applicability of large language models and generative models for legal case judgement summarization," *Artificial Intelligence and Law*, vol. 33, pp. 1007–1050, 2024, doi: 10.1007/s10506-024-09411-z.

[9] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. Alonso-Moral, R. Confalonieri, et al., "Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," *Information Fusion*, vol. 99, p. 101805, 2023, doi: 10.1016/j.inffus.2023.101805.

[10] J. Collenette, K. Atkinson, and T. J. M. Bench-Capon, "Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights," *Artificial Intelligence*, vol. 317, p. 103861, 2023, doi: 10.1016/j.artint.2023.103861.

[11] N. Guha, J. Nyarko, D. E. Ho, C. Ré, A. Chilton, A. Narayana, et al., "LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models," *arXiv*, 2023, doi: 10.48550/arXiv.2308.11462.

[12] X. Wang, X. Zhang, V. Hoo, Z. Shao, and X. Zhang, "LegalReasoner: A multi-stage framework for legal judgment prediction via large language models and knowledge integration," *IEEE Access*, vol. 12, pp. 166843–166854, 2024, doi: 10.1109/ACCESS.2024.3496666.

[13] E. Q. Nuranti, E. Yulianti, and H. S. Husin, "Predicting the category and the length of punishment in Indonesian courts based on previous court decision documents," *Computers*, vol. 11, no. 6, p. 88, 2022, doi: 10.3390/computers11060088.
[14] G. Wicaksono, S. F. A. Asqalani, Y. Azhar, N. Hidayah, and A. Andreawana, "Automatic summarization of court decision documents over narcotic cases using BERT," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 2, 2023, doi: 10.30630/joiv.7.2.1811.
[15] A. Shukla, P. Bhattacharya, S. Poddar, R. Mukherjee, K. Ghosh, P. Goyal, and S. Ghosh, "Legal case document summarization: Extractive and abstractive methods and their evaluation," *arXiv*, 2022, doi: 10.48550/arXiv.2210.07544.
[16] Y. Huang, L. Sun, C. Han, and J. Guo, "A high-precision two-stage legal judgment summarization," *Mathematics*, vol. 11, no. 6, p. 1320, 2023, doi: 10.3390/math11061320.
[17] W. Gao, S. Yu, Y. Qin, C. Yang, R. Huang, Y. Chen, and C. Lin, "LSDK-LegalSum: Improving legal judgment summarization using logical structure and domain knowledge," *Journal of King Saud University Computer and Information Sciences*, vol. 37, 2025, doi: 10.1007/s44443-025-00022-5.
[18] Y. Song, Y. Qin, R. Huang, Y. Chen, and C. Lin, "Legal text summarization via judicial syllogism with large language models," *Journal of King Saud University Computer and Information Sciences*, vol. 37, 2025, doi: 10.1007/s44443-025-00113-3.
[19] M. Koniaris, D. Galanis, E. Giannini, and P. Tsanakas, "Evaluation of automatic legal text summarization techniques for Greek case law," *Information*, vol. 14, no. 4, p. 250, 2023, doi: 10.3390/info14040250.
[20] G. Moro, N. Piscaglia, L. Ragazzi, and P. Italiani, "Multi-language transfer learning for low-resource legal case summarization," *Artificial Intelligence and Law, 1–29*, 2023, doi: 10.1007/s10506-023-09373-8.
[21] S. Liu, J. Cao, Y. Li, R. Yang, and Z. Wen, "Low-resource court judgment summarization for common law systems," *arXiv*, 2024, doi: 10.48550/arXiv.2403.04454.
[22] Z. Fei, X. Shen, D. Zhu, F. Zhou, Z. Han, S. Zhang, et al., "LawBench: Benchmarking legal knowledge of large language models," *arXiv*, 2023, doi: 10.48550/arXiv.2309.16289.
[23] W. Hwang, K. Cho, H.-Y. Lee, and M. Seo, "A multi-task benchmark for Korean legal language understanding and judgement prediction," *arXiv*, 2022, doi: 10.48550/arXiv.2206.05224.
[24] T. Turan and E. Küçüksille, "Summarization, prediction, and analysis of Turkish Constitutional Court decisions with explainable artificial intelligence and a hybrid natural language processing method," *IEEE Access*, vol. 13, pp. 59766–59779, 2025, doi: 10.1109/ACCESS.2025.3556725.
[25] R. Kondo, R. Matsuoka, T. Yoshida, K. Yamasawa, and R. Hisano, "Capturing legal reasoning paths from facts to law in court judgments using knowledge graphs," *Proceedings of the 13th Knowledge Capture Conference 2025*, 2025, doi: 10.1145/3731443.3771354.
[26] S. Nigam, A. Sharma, D. Khanna, N. Shallum, K. Ghosh, and A. Bhattacharya, "Legal judgment reimagined: PredEx and the rise of intelligent AI interpretation in Indian courts," *arXiv*, 2024, doi: 10.48550/arXiv.2406.04136.
[27] S. Nigam, B. D. Patnaik, S. Mishra, N. Shallum, K. Ghosh, and A. Bhattacharya, "TathyaNyaya and FactLegalLlama: Advancing factual judgment prediction and explanation in the Indian legal context," *arXiv*, 2025, doi: 10.48550/arXiv.2504.04737.
[28] M. Ibrahim, A. T. Handoyo, and M. S. Anggreainy, "Hybrid deep learning for legal text analysis: Predicting punishment durations in Indonesian court rulings," *arXiv*, 2024, doi: 10.48550/arXiv.2410.20104.
[29] E. Q. Nuranti, N. S. Intizhami, E. Yulianti, A. Latief, O. Iyad, and A. Ghozy, "A systematical procedure to extracting legal entities from Indonesian judicial decisions," *MethodsX*, vol. 15, 2025, doi: 10.1016/j.mex.2025.103610.
[30] J. Gu, X. Jiang, Z. Shi, H. Tan, X. Zhai, C. Xu, et al., "A survey on LLM-as-a-judge," *arXiv*, 2024, doi: 10.48550/arXiv.2411.15594.
[31] H. Li, Q. Dong, J. Chen, H. Su, Y. Zhou, Q. Ai, Z. Ye, and Y. Liu, "LLMs-as-judges: A comprehensive survey on LLM-based evaluation methods," *arXiv, abs/2412.05579*, 2024.
[32] D. S. U. De La Osa and N. Remolina, "Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI," *Data & Policy*, vol. 6, 2024, doi: 10.1017/dap.2024.53.
[33] M. F. Ramadhan, "Plain language in legal drafting: Challenges and opportunities in Indonesian regulatory reform," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 22–30, 2026.
[34] W. B. M. Setiyawan, "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.
[35] 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.
[36] 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.
[37] 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.

## BIOGRAPHY OF AUTHOR

**Ahmad Zairudin** holds S.H. and M.H. degrees and is a lecturer and researcher at Universitas Nurul Jadid (UNUJA), Probolinggo. He teaches in the field of law, with a particular focus on Indonesian Constitutional Law. In addition to teaching, he is also active in writing scholarly articles, textbooks, and opinion pieces in mass media. He can be contacted at email: a.zairudin1988@gmail.com