International Journal of Multidisciplinary Research and Growth Evaluation.

Can Artificial Intelligence Identify Unequal Legal Protection? An Explainable Analysis of Access-to-Justice Patterns in Public Legal Documents

Arthur Brick Kamya [1*], Dr. Nasake Julius Alochere [2]
[1] PhD Candidate in Law, Student and Researcher, Kampala International University, Kampala, Uganda
[2] College of Education Open and Distance Learning, Uganda

* Corresponding Author: **Arthur Brick Kamya**

Article Info

ISSN (Online): 2582-7138
Volume: 07
Issue: 06
Received: 22-08-2026
Accepted: 24-09-2026
Published: 26-10-2026
Page No: 136-143

Abstract
Artificial intelligence can identify patterns in legal documents, but the interpretation of those patterns depends on what the documents make observable. This study examines the evidentiary boundary between computational visibility and unequal legal protection through a reproducible audit of 4,059 French-language records in an OHADA-labelled legal corpus. The analysis combines schema inspection, document-level completeness measures, provenance stratification, jurisdiction-string screening, and transparent lexical indicators. Although every record contains a full-text field, 2,960 records lack a separate reasoning field. Reasoning availability is completely separated by the corpus's source categories: all 1,099 records labelled file1 or both contain reasoning, whereas none of the 2,960 file2 records does. Availability also varies substantially across legal domains, creating an apparent substantive disparity that is inseparable from corpus construction. Jurisdiction strings explicitly name CCJA or the common court in 1,542 records; 2,516 contain other strings and one is missing, precluding treatment of the collection as a verified single-court population. Legal-aid expressions occur in only 12 full texts, while costs and representation-related expressions occur much more frequently; these are mention counts, not verified service or remedy rates. The article develops a visibility–comparability–validation framework that separates documentary inequality, outcome association, and legally established unequal protection. The findings support explainable AI as an instrument for locating evidence gaps and prioritising legal review, not as an autonomous detector of discrimination. A practical contribution is a set of stopping rules that prevents missing fields, provenance artefacts, and lexical matches from becoming unsupported claims about judicial fairness.

Keywords: Access to justice, explainable artificial intelligence, OHADA, legal information, documentary visibility, provenance, measurement validity, legal NLP

1. Introduction: From Computational Promise to Evidentiary Discipline

Whether artificial intelligence can identify unequal legal protection is not simply a question of predictive accuracy. It is a question about the relationship between a legal entitlement, the institutional processes through which that entitlement is realised, and the documents available for computational observation. A classifier may distinguish commercial disputes from enforcement proceedings without measuring whether similarly situated parties obtain comparable protection. A search system may retrieve references to legal assistance without establishing whether assistance was available, affordable, timely, or effective. These distinctions matter because an apparently precise numerical comparison can attach scientific authority to a legally inappropriate construct. The central challenge is therefore not to maximise the number of patterns extracted from judgments, but to establish

which interpretations those patterns can sustain. Measurement must precede fairness assessment rather than follow it as an explanatory disclaimer (Jacobs and Wallach, 2021) [5].

This article investigates that challenge in a French-language corpus associated with OHADA legal materials. The empirical unit is the corpus record, not the litigant, household, filing, or unresolved legal need. That distinction shapes the entire design. The collection contains 4,059 records and 17 original variables, including legal domain, jurisdiction, party fields, summaries, reasoning, ruling, full text, source category, and an existing split assignment. Its year field spans 1997–2023 among dated records. These attributes make the collection useful for studying documentary structure and the feasibility of subsequent legal NLP. They do not make it a representative sample of all legal disputes in any member state. The study consequently begins with observability and only then considers what a later outcome or disparity study would require.

Three questions organise the analysis. First, which components of a legally interpretable access-to-justice study are consistently observable? Second, do differences in observability track substantive case categories, or do they instead reflect the architecture of the collection? Third, what can explainable computational methods legitimately contribute when outcome labels, protected-group measures, and an independent annotation study are absent? The contribution is an empirical demonstration of how a fairness inquiry can fail before model training, together with a framework for identifying that failure transparently. This focus complements legal-language benchmarks rather than competing with them. Benchmarks evaluate specified tasks; this article evaluates whether the available documentary representation is sufficient for a proposed legal inference (Chalkidis *et al*., 2022; Guha *et al*., 2023) [1, 4].

1.1. Conceptual Contribution and Scope

The analysis distinguishes three propositions that are frequently collapsed. Documentary inequality concerns differences in whether information is recorded, separated into fields, indexed, or retrievable. Outcome association concerns differences in verified decisions or remedies among defined categories of cases. Unequal legal protection is a stronger legal conclusion requiring an appropriate equality standard, relevant comparators, substantive context, and evidence capable of addressing competing explanations. The first proposition can be established from missingness and representation patterns. The second requires reliable outcome annotation and a defensible denominator. The third cannot be inferred merely because the first two are observed. This ordering does not diminish the importance of documentary inequality. It identifies it as a separate phenomenon with its own institutional consequences and evidentiary limits.

The article calls the first stage visibility, the second comparability, and the third validation. Visibility asks whether the construct has an observable manifestation and whether its absence has a stable meaning. Comparability asks whether the units being contrasted share the legally relevant features needed for the proposed comparison. Validation asks whether extraction, annotation, explanation, and interpretation have survived independent scrutiny. These stages are cumulative rather than interchangeable. A sophisticated explanation of an invalid proxy cannot repair the proxy. Likewise, an accurate classifier does not create a representative denominator for unmet legal need. The framework is consistent with research that locates harm across the machine-learning lifecycle rather than exclusively inside the fitted model (Suresh and Guttag, 2021) [9].

This study makes a deliberately bounded claim. It identifies unequal documentary visibility within a finite legal corpus and demonstrates why that pattern cannot automatically be attributed to courts, parties, or legal domains. It does not estimate discrimination, representation quality, judicial bias, unmet need, or a national access-to-justice rate. It also does not verify every judgment against a primary court record. That boundary is an affirmative design decision: unreliable labels would make a more expansive study less, not more, informative. A useful computational contribution can consist in explaining why a tempting substantive conclusion is not identified. The empirical findings therefore support a research workflow in which provenance audit and legal construct validation operate as prerequisites for modelling, rather than optional appendices after a model has produced an appealing result.

2. Literature Review: Legal NLP and the Limits of Task Success

Legal NLP research provides a valuable vocabulary for differentiating classification, retrieval, extraction, summarisation, and judgment prediction. These tasks use different inputs and produce different outputs, even when all operate on judicial language. The survey by Feng, Li, and Ng (2022) [2] describes legal judgment prediction as a distinct research area whose methods and evaluation choices depend on the prediction setting. LexGLUE supplies a suite of English-language legal understanding tasks, while LegalBench develops a broad collection of legally informed tasks with explicit task descriptions (Chalkidis *et al*., 2022; Guha *et al*., 2023) [1, 4]. These contributions make legal computational work more testable. They do not establish that success on a benchmark measures access to justice, institutional legitimacy, or equal protection.

The distinction becomes especially important when full judgments are used as model inputs. A completed judgment may contain the disposition, the court's own categorisation, quoted legislation, editorial headings, or an abstract written after the case is resolved. Such text can be appropriate for retrospective indexing. It is not equivalent to information available before adjudication. Consequently, a classifier trained on full text may demonstrate useful document organisation while providing no valid evidence of prospective decision support. Leakage is not restricted to accidental copying between training and test rows; it can also arise when the input includes information that would not exist at the intended point of prediction. Kapoor and Narayanan (2023) [6] show why such mismatches threaten reproducibility and interpretation in machine-learning-based science.

The present study adopts the retrospective indexing interpretation and then moves one step earlier in the pipeline. Before asking which model best classifies legal text, it asks whether the text fields consistently preserve the information needed for the research question. This emphasis responds to the wider problem of data work being undervalued relative to model work. Sambasivan *et al*. (2021) [8] describe how unresolved data issues can propagate through high-stakes AI systems as data cascades. In legal analysis, one such cascade begins when an empty extraction field is treated as an absent judicial justification. Subsequent modelling may be

internally consistent yet substantively wrong because its target was misdefined. Preventing that cascade requires a documented distinction between original institutional practice and the database representation through which that practice becomes visible.

2.1. Measurement, Documentation, and Explanation

Measurement theory offers a direct explanation of why apparently neutral features can support invalid fairness claims. A variable is not a construct simply because it is available in a table. The existence of a lawyer-related word is not a validated measure of competent representation, and a nonempty reasoning column is not a validated measure of reason-giving. Jacobs and Wallach (2021) [5] emphasise the relationship between an unobservable construct and its operationalisation, including the assumptions that connect them. Applied here, that perspective requires asking whether a field records the underlying legal event, an editorial summary, an extraction decision, or some combination of these. Without that distinction, a missingness comparison can be mistaken for an assessment of institutional performance.

Dataset documentation is therefore a substantive component of legal inference. Gebru *et al.* (2021) [3] propose systematic documentation of dataset motivation, composition, collection, preprocessing, uses, and maintenance. For legal materials, those categories have immediate consequences for selection and interpretation. A collection's label may not describe every record; a date range may refer only to nonmissing dates; and a licence declaration may not independently resolve permissions for every underlying document. Documentation must also identify transformations between a judgment and a row, including whether reasoning and rulings were manually prepared, mechanically extracted, or generated through another process. The present audit cannot reconstruct every earlier transformation. It instead identifies where those unknowns constrain the claims available from the corpus.

Explainability is treated as the ability to expose the operational basis of an empirical claim. In this article, that basis consists of field definitions, source-stratified counts, explicit normalisation rules, and inspectable lexical patterns. It does not consist of assigning a fairness score to a court. This choice aligns explanation with the object being explained: a deterministic audit should expose its counting and coding rules, rather than invoke a complex attribution method designed for an unrelated model. A lifecycle perspective also helps separate a failure in data collection from a failure in downstream prediction (Suresh and Guttag, 2021) [9]. The resulting explanation is modest but consequential: it shows when the database architecture, rather than observed legal conduct, is sufficient to generate the headline pattern.

3. Methodology: Corpus, Units, and Audit Design

The study uses the OHADA-labelled tabular corpus as a finite collection of 4,059 legal records. The accompanying local metadata identifies version 1.0.0 and declares CC BY 4.0; this declaration is documented as corpus metadata rather than independently verified chain-of-title evidence. The empirical audit uses the complete CSV representation and does not treat accompanying train, validation, and test files as additional cases. Those files are alternative partitions of the same collection, not separate populations. Table 1 identifies the principal dimensions of the audit. Descriptive results are computed over the complete collection unless a denominator is explicitly narrowed. No survey weights, population extrapolation, or causal estimation are used.

Table 1: Corpus audit: units, coverage, and analytic boundaries.

Dimension	Observed coverage	Interpretive boundary
Records / variables	4,059 / 17	Records are not people or all disputes
Year range	1997–2023; 190 missing	Stored year; no population trend
Domain labels	16 nonmissing categories	Not independently reannotated
Full text	4,059 populated fields	Not all primary originals verified
Source categories	file1 / file2 / both	Underlying collection identities unknown
Court information	Raw jurisdiction strings	No validated court-level panel

Note: All values are computed from the study corpus unless explicitly identified as the independent housing experiment.

The initial audit checks schema, null values, date coverage, domain frequencies, document length, identifiers, and normalized exact text duplication. Text normalization for duplicate detection lowercases strings, collapses whitespace, and trims leading and trailing whitespace before SHA-256 hashing. A separate case-number/date check identifies repeated metadata keys without assuming that every shared key represents a duplicated judgment. That assumption would be unsafe in a mixed-court collection where similar numbering conventions may recur. The audit therefore reports both checks separately and does not silently remove records on the basis of a potentially nonunique key. Character length measures describe stored text, not original page length or linguistic complexity. These decisions preserve a transparent distinction between a detectable data condition and its unverified institutional explanation.

3.1. Operational Definitions and Analytical Safeguards

Reasoning availability is defined as a nonnull, nonblank value in the reasoning field. This operational definition concerns the structured field only. It does not imply that the full text contains or lacks reasons. Provenance categories are used exactly as stored: file1, file2, and both. Their labels do not identify publishers, collection procedures, or jurisdictions, and the study does not invent such identities. Cross-tabulation evaluates whether field availability is distributed across these categories. Domain-specific availability rates use all records with the relevant nonmissing domain as their denominator. Because the complete collection is audited, these are finite-corpus descriptive quantities rather than estimates from a probability sample.

Jurisdiction screening is intentionally conservative. Strings are lowercased and decomposed to remove diacritics, after

which an explicit reference to ccja or commune de justice identifies a CCJA-named record. Other nonmissing strings form a residual category requiring legal normalization, not a definitive set of verified national courts. The audit also counts distinct raw jurisdiction strings. It does not infer country, appellate level, court identity, or judicial hierarchy from party names or unverified abbreviations. Examples of stored jurisdiction values include named appeal courts and first-instance courts, which is sufficient to reject an unqualified single-court interpretation of the corpus. It is not sufficient to produce a validated court-level panel without additional coding.

Five transparent lexical indicators are evaluated separately in full text, reasoning, and ruling. Legal aid searches for aide juridictionnelle or assistance judiciaire after accent normalization. Inadmissibility uses the word stem irrecevab; cassation searches cassation or casse; costs searches depens; and the broad representation-related indicator searches avocat or conseil with word endings. The last indicator is deliberately described as broad because conseil can refer to institutional bodies as well as legal counsel. A document contributes at most once to each field–indicator count, regardless of the number of occurrences. Blank fields contribute no match, and all-record denominators are retained to display the consequences of field coverage. These rules make the analysis reproducible but do not establish precision, recall, polarity, speaker attribution, or the legal meaning of each occurrence. No outcome classifier is trained from these lexical proxies.

4. Results: The Distribution of Documentary Visibility

The collection contains 4,059 distinct case identifiers and no duplicate full texts under the specified exact-normalization procedure. This does not establish the absence of near duplicates, extracts, translations, related proceedings, or revised versions. Among records with nonmissing case number and date, 98 rows repeat an earlier case-number/date combination. The full-text field is nonmissing throughout, but stored document length ranges from 241 to 57,132 characters. The median is 7,916 characters and the interquartile range is 5,776.5–11,372. The range alone does not identify extraction failure: a very short record could be an extract or short procedural disposition, and a long record could legitimately contain extensive reasoning. It does identify records requiring differentiated inspection before length is used as a substantive explanatory variable.

Table 2: Missingness in selected original fields (N = 4,059).

Field	Missing, n	Missing, %
Date	190	4.68
legal_domain	10	0.25
Formation	4020	99.04
Plaintiff	533	13.13
Defendant	585	14.41
articles_cited	593	14.61
Reasoning	2960	72.92
Ruling	7	0.17
full_text	0	0.00

Note: All values are computed from the study corpus unless explicitly identified as the independent housing experiment.

Table 2 shows that the availability of structured fields is highly uneven. Reasoning is absent in 2,960 records, or 72.92%, whereas ruling is absent in only seven. Formation is absent in 4,020 records and cannot support a meaningful comparative analysis of judicial panels. Date and year are each missing in 190 records, limiting temporal comparisons. Party fields and article citations are partially incomplete as well. The resulting pattern is not equivalent to a finding that courts usually omit reasons but almost always announce remedies. It concerns the separation of information into database fields. Because every row has a full-text value, the strongest immediate research implication is that structured-field analyses require field-specific validation and cannot assume that missingness has a uniform legal interpretation.

4.1. Provenance Separation and Jurisdictional Heterogeneity

Reasoning-field availability is completely separated by provenance category. All 553 file1 records and all 546 both records contain reasoning, while all 2,960 file2 records lack it. Figure 1 presents the corresponding counts. This is a 100-percentage-point difference between file2 and the other categories, but the difference does not identify an effect of source collection on judicial practice. The source category and field availability are perfectly aligned in the observed data, leaving no within-category variation with which to distinguish alternative substantive explanations. A court-quality interpretation would therefore mistake a database regularity for an institutional finding. The observed structure is consistent with differentiated extraction or record construction, although its historical cause cannot be established from the stored labels alone.

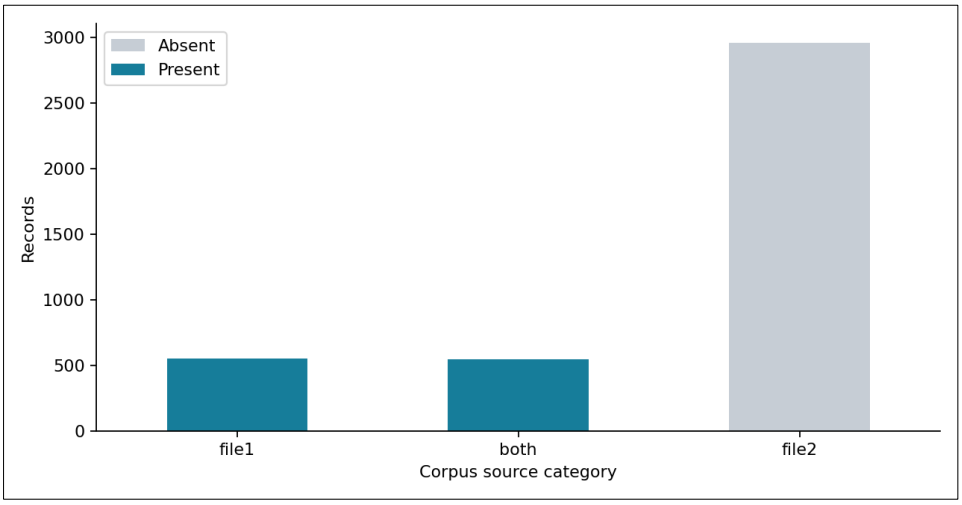


Fig 1: Reasoning-field availability by corpus source category. Counts describe database fields, not the presence of reasons in original judgments.

Jurisdiction screening produces a second warning. Of 4,059 records, 1,542 explicitly name CCJA or the common court under the stated matching rule. Another 2,516 contain other jurisdiction strings and one is missing. There are 768 distinct nonmissing raw jurisdiction strings, reflecting a mixture of naming variation and institutional heterogeneity that has not been legally normalized. The collection cannot consequently be analysed as 4,059 verified judgments of one supranational court. This finding also limits the interpretation of the 98 repeated case-number/date rows: some may be related versions, while others may reflect shared numbering across distinct courts. A validated institutional comparison requires court-name reconciliation, primary-source checks, and an explicit rule for distinguishing related proceedings from genuinely duplicated observations.

4.2. Domain Differences and Lexical Indicators

The domain distribution is dominated by enforcement law, which contains 2,144 records, or 52.82% of the collection. The next largest categories are commercial companies and GIE with 479, general commercial law with 428, collective proceedings with 256, and security interests with 195. The eight domains containing at least 100 records account for 3,981 records. Structured reasoning is present in 250 of 479 company-law records, compared with two of 155 cooperative-law records and 496 of 2,144 enforcement records. Figure 2 visualises the eight largest categories. These contrasts are real features of the corpus representation, but the perfect source separation means they cannot establish that one legal field receives more reasoned adjudication than another.

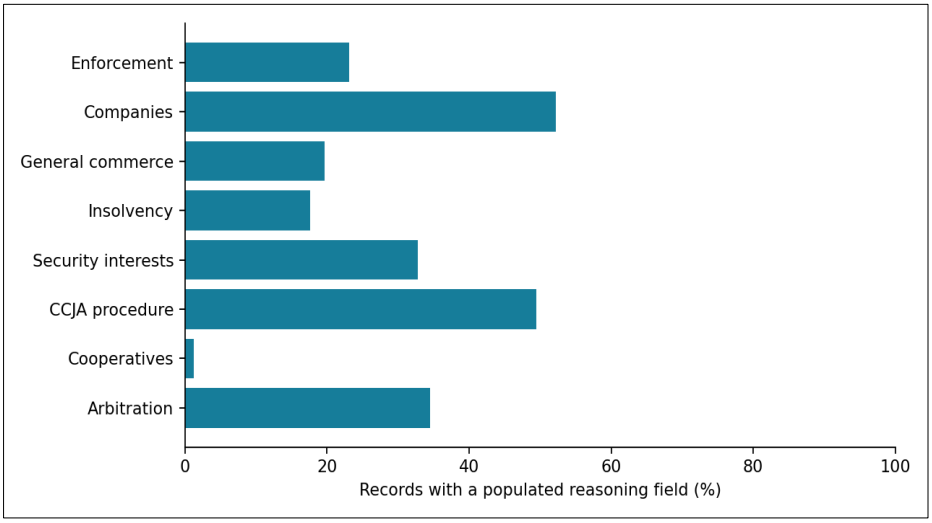


Fig 2: Structured reasoning availability in the eight largest legal domains. Domain differences are confounded with source composition.

Table 3: Documents containing each lexical indicator.

Indicator	Full text, n (%)	Reasoning, n	Ruling, n
Legal-aid expressions	12 (0.30)	0	0
Inadmissibility stem	1,776 (43.75)	126	562
Cassation expressions	1,727 (42.55)	565	409
Costs term	3,760 (92.63)	32	2,312
Broad representation terms	3,463 (85.32)	46	93

Note: All values are computed from the study corpus unless explicitly identified as the independent housing experiment.

Table 3 summarises the lexical audit. Legal-aid expressions occur in 12 full texts and in no separated reasoning or ruling fields. Inadmissibility-related expressions occur in 1,776 full texts, compared with 562 ruling fields. Costs expressions occur in 3,760 full texts and 2,312 ruling fields. Such differences can arise from field scope, extraction practice, quoted arguments, procedural histories, or the location of the operative order. They are not estimates of refusal, costs awards, or legal-aid provision. In particular, zero legal-aid matches in the reasoning field is not evidence that no litigant received assistance. The analysis makes this ambiguity visible rather than resolving it through an unsupported label assignment.

5. Identification: What the Observed Patterns Establish

The results establish a documentary disparity: some kinds of structured legal information are far more available in some corpus categories than in others. They also establish a provenance dependence that is strong enough to determine reasoning-field availability exactly. These findings are useful because they identify the point at which a conventional fairness pipeline would become invalid. If the target variable were defined as reasons supplied, with a value of zero assigned to every missing reasoning field, the target would reproduce the source partition perfectly. Any model trained to predict that target could learn source-related language without measuring judicial reason-giving. An explanation of such a model would explain a database artefact unless an independent annotation study established otherwise.

A simple identification argument clarifies the issue. Let R denote the presence of a separated reasoning field and S denote the file2 source category. Within this collection, R equals zero whenever S equals one, and R equals one whenever S equals zero. A domain-specific average of R can therefore vary entirely through the domain's mixture of source categories. No finding about the quality or sufficiency of actual reasons is needed to generate that pattern. The argument is deterministic and descriptive; it does not require a significance test. Nor does it prove that substantive differences are absent. It shows that the available variable cannot distinguish the substantive explanation from the representational one.

The same discipline applies to lexical outcomes. A reference to irrecevabilité can describe a party's objection, a lower court's order, a rejected argument, or the operative disposition. A reference to cassation can appear in a discussion of jurisdiction even when no appeal is allowed. A costs term can indicate a requested order, a conventional formula, or an actual allocation of costs. Reliable outcome extraction therefore requires an annotation unit, a taxonomy, and rules for linking an expression to the deciding institution and procedural stage. The present analysis does not substitute lexical convenience for that work. It identifies where extraction is promising and where validation must precede any inference about unequal protection.

6. Discussion: Access to Justice as an Observable Research Problem

The answer to the title question is conditional. AI can identify patterns relevant to an inquiry into unequal legal protection when the underlying constructs, comparators, and labels are adequately established. In this collection, the most defensible immediate application is to identify unequal visibility and to prioritise documentary repair. The audit reveals precisely where automated analysis could undercount reasons, overstate the uniformity of court coverage, or confuse mentions with outcomes. This is not a merely technical concern. Search, retrieval, and downstream synthesis operate on the representation that is available, so uneven field quality can shape what researchers and practitioners are able to find. The present study does not measure those downstream effects, but it identifies a plausible mechanism that a deployment study could test.

A useful next study would compare the structured fields with full judgments in a stratified, legally reviewed sample. Strata should include each source category, the largest domains, dated and undated records, and records with ambiguous jurisdiction strings. Reviewers would determine whether reasons exist in the full text, whether the separated field preserves them accurately, and whether the ruling summary represents the operative order. Disagreement would be resolved through a documented process rather than majority intuition alone. Inter-annotator agreement would be reported by field and legal category because aggregate agreement can conceal difficult domains. Such a study would estimate extraction validity; it would not by itself establish fairness. The transition to a fairness study would additionally require legally meaningful comparators and information about relevant procedural circumstances.

The conceptual implication is that access-to-justice research should not adopt a universal documentary proxy. Access involves distinct questions of entry, representation, procedure, timeliness, enforcement, and effective remedies. A published judgment may illuminate some of these while being silent about others. People who never file, abandon claims, settle, or resolve disputes informally are not represented simply because a corpus contains many judgments. The appropriate denominator must therefore be selected for the question being asked, not inherited from the number of available rows. This argument follows the measurement discipline of Jacobs and Wallach (2021) [5] and the documentation discipline of Gebru *et al.* (2021) [3]: transparent limitations define the scope of a valid result rather than weaken it after the fact.

6.1. A Governance Framework for Explainable Legal Analytics

The visibility–comparability–validation framework can be implemented through explicit stopping rules. The visibility rule stops a substantive analysis when the variable's recorded absence cannot be distinguished from extraction failure or noncollection. The comparability rule stops an institutional ranking when court identity, procedural posture, or the denominator remains unresolved. The validation rule stops automated outcome claims when extraction has not been assessed against an independently annotated reference set. These are proposed governance requirements, not interventions whose effectiveness is tested here. Their purpose is to make the transition from exploratory patterns to substantive legal claims inspectable. They also provide an intelligible explanation for why a system declines to issue a ranking or fairness score.

The rules imply differentiated outputs rather than a binary choice between full automation and no analysis. Where visibility is adequate but outcomes are unvalidated, a system may support retrieval and candidate-document review. Where labels are verified but comparison groups are heterogeneous, it may describe outcomes within carefully

defined strata while withholding causal or equality claims. Where institutional identity is unresolved, it may produce an audit queue instead of a court league table. Each output should state its unit, denominator, exclusion rule, and validation status. This approach locates responsibility across the research lifecycle, consistent with Suresh and Guttag (2021) [9], and reduces the risk that an apparently neutral technical interface disguises unresolved legal judgments.

Data stewardship is equally important. Public legal texts can contain names, financial information, addresses, family details, and other sensitive material. This article does not reproduce party-level extracts or publish a new bulk text collection. Its shareable outputs consist of aggregate tables, computational rules, model-independent audit results, and reproducibility materials. The local corpus licence declaration and the independently accessible benchmark licences require separate consideration; a label attached to one collection should not be generalized to another. Before deployment or redistribution, researchers should verify applicable permissions, ethics requirements, and institutional policies. These measures are prospective safeguards, not a claim that an ethics committee, court, or data custodian has approved the present work.

7. Limitations and Conclusion

Several limitations define the interpretation of the findings. The corpus is not a probability sample of disputes, judgments, or legal needs, and its publication and inclusion mechanisms are not fully reconstructed. Its local documentation cannot substitute for independent verification of every source document or reuse condition. The jurisdiction screening identifies explicit string patterns rather than legally validated court identities. Exact text hashing does not identify near duplicates or procedural relationships, and repeated case-number/date keys can have multiple explanations. Domain labels are accepted as stored rather than independently reannotated. Lexical indicators are not evaluated against a human-coded gold standard, so their false-positive and false-negative rates remain unknown. No protected-attribute inference, causal model, or claimant-level fairness metric is estimated.

These limitations do not prevent a firm conclusion about structured-field availability. The corpus contains 4,059 full-text records, while reasoning-field availability is perfectly determined by its source categories. The result is directly reproducible and does not depend on a disputed model explanation. The audit also demonstrates that the collection's institutional scope is more heterogeneous than an unqualified CCJA description suggests. Together, these findings establish that a naïve fairness study would conflate representation and adjudication. They do not establish whether any particular court, legal domain, or litigant category enjoys unequal legal protection. The contribution lies in showing why that stronger inference is not available and how the research design must change before it can be investigated responsibly.

Artificial intelligence can therefore contribute to the identification of unequal legal protection only within an evidentiary architecture that distinguishes what is recorded from what happened and what happened from what the law requires. In the present setting, transparent computational analysis identifies unequal documentary visibility, exposes provenance dependence, and defines a feasible route toward validated legal research. Its strongest result is not an automated allegation of bias but an auditable boundary on inference. Future work should first reconcile court identities, validate structured extraction, and establish legally meaningful comparison groups. Only then should outcome modelling and fairness assessment be treated as more than exploratory exercises. Explainable legal AI is most credible when it can explain not only its predictions, but also why the evidence is insufficient for a proposed conclusion.

References

1. Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D., and Aletras, N. (2022). LexGLUE: A benchmark dataset for legal language understanding in English. Proceedings of ACL, 4310–4330. DOI: 10.18653/v1/2022.acl-long.297.
2. Feng, Y., Li, C., and Ng, V. (2022). Legal judgment prediction: A survey of the state of the art. Proceedings of IJCAI, 5461–5469. DOI: 10.24963/ijcai.2022/765.
3. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé III H, *et al*. Datasheets for datasets. Communications of the ACM. 2021;64(12):86-92. doi:10.1145/3458723
4. Guha, N., Nyarko, J., Ho, D. E., Ré, C., *et al*. (2023). LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models. arXiv:2308.11462.
5. Jacobs, A. Z., and Wallach, H. (2021). Measurement and fairness. Proceedings of FAccT, 375–385. DOI: 10.1145/3442188.3445901.
6. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns. 2023;4(9):100804. doi:10.1016/j.patter.2023.100804
7. Maathis-com. (n.d.). OHADA CCJA Court Decisions Corpus, version 1.0.0 [dataset]. Local dataset_info.json and corpus metadata; 4,059 records. Licence declared in metadata: CC BY 4.0. Public repository availability and publication date were not independently established.
8. Sambasivan N, Kapania S, Highfill H, Akrong D, Paritosh P, Aroyo LM. "Everyone wants to do the model work, not the data work": Data cascades in high-stakes AI. Proceedings of CHI. 2021. doi:10.1145/3411764.3445518
9. Suresh H, Guttag J. A framework for understanding sources of harm throughout the machine learning life cycle. Proceedings of EAAMO; 2021. doi:10.1145/3465416.3483305

Appendix a. Audit Specification and Research Declarations

Reproducibility Specification

The audit begins from the CSV representation and retains one analytical row for each original record. Missingness uses null-value checks, with an additional blank-string check for reasoning availability. Normalized exact duplicates are detected after lowercasing, whitespace compression, and trimming. Repeated case-number/date keys are counted only where both fields are nonmissing. Date summaries use the stored year field and do not impute missing years. Domain counts retain the original category strings. Jurisdiction screening uses accent-normalized lowercase strings and the expression ccja|commune de justice. Tables and figures are generated from the same calculations as the machine-

readable outputs. The reproducibility package records input checksums and the computational environment; it does not redistribute the full legal corpus.

Lexical Specification

After Unicode decomposition and removal of combining marks, legal aid matches aide juridictionnelle or assistance judiciaire. Inadmissibility matches the word stem irrecevab with following word characters. Cassation matches the separate words cassation or casse. Costs matches depens. The representation-related indicator matches avocat or conseil and their word endings. Indicators are binary at document–field level, and percentages use 4,059 as the denominator. No expression is interpreted as a verified legal outcome. The supplementary script preserves exact regular expressions, enabling the distinction between a general prose description and its executable implementation to be inspected.

Research Declarations

No independent legal annotation, participant study, or ethics approval is claimed. Funding, author affiliations, competing interests, and authorship contributions must be completed and verified by the submitting authors rather than inferred from the analysis. AI-assisted computation and manuscript drafting should be disclosed in the form required by the target journal, with human authors responsible for checking every claim and reference. A dataset licence declaration is not an assertion that all underlying rights have been independently cleared. These declarations identify submission responsibilities that cannot be completed by automated inference. The article's empirical claims remain limited to the audited collection and its stored representations.

How to Cite This Article

Kamya AB, Alochere NJ. Can artificial intelligence identify unequal legal protection? An explainable analysis of access-to-justice patterns in public legal documents. International Journal of Multidisciplinary Research and Growth Evaluation. 2026;7(6):136–143.

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms