AI-Augmented Judicial Relativity Framework (JRF).

Abstract

Modern judicial systems must synthesize numerous interconnected variables — statutory provisions, factual evidence, intent, witness credibility, forensic findings, constitutional principles, and historical precedent — while remaining consistent and unbiased. This paper proposes the Judicial Relativity Framework (JRF), an AI-assisted decision-support methodology inspired by Einstein's concept of multiple frames of reference and by dimensionality-reduction principles from machine learning. JRF transforms a complex legal problem into a small number of interpretable decision dimensions — factual certainty, legal similarity, ethical impact, constitutional compatibility, and precedent consistency — without replacing judicial discretion. It combines explainable artificial intelligence, legal knowledge graphs, semantic precedent retrieval, causal inference, and evidence weighting to produce a transparent judicial-assistance report; its dimensionality-reduction, semantic-embedding, contradiction-detection, precedent-retrieval, and explainable-attribution components are given a full mathematical formalization, so every quantitative claim is precisely defined rather than merely illustrated. The resulting judicial dimensions are exact, traceable linear combinations of the underlying evidentiary features, and their contributing evidence can be recovered through an exact Shapley-value decomposition. Two illustrative case studies — a property-ownership dispute and a fatal road accident — show that the same architecture and mathematics generalize across areas of law, while judges retain complete authority over the final verdict. By reframing legal disputes as a small set of interpretable, mathematically grounded dimensions rather than a mass of correlated variables, JRF offers augmented rather than automated justice, improving consistency and transparency subject to fairness, explainability, and due-process safeguards. The framework remains a conceptual and mathematical proposal; empirical validation is identified as necessary future work.

Keywords judicial decision support · explainable AI · dimensionality reduction · legal natural language processing · precedent retrieval · human-in-the-loop AI

1 Introduction

1.1 The Multidimensional Nature of Legal Disputes

This section motivates the framework: it examines why legal disputes are inherently multidimensional (Section 1.1), draws an analogy to dimensionality reduction in machine learning (Section 1.2), positions the framework as explainable augmentation rather than opaque prediction (Section 1.3), and summarizes the paper's contributions (Section 1.4).

Justice has traditionally been viewed as the application of statutory law to established facts. In practice, however, real-world legal disputes rarely exist in a single dimension; every case involves multiple perspectives that interact simultaneously (Ashley 2017). Representative dimensions include: legal provisions; factual evidence; witness reliability; forensic reports; intent; economic conditions; constitutional rights; ethical considerations; previous judicial precedents; and societal consequences.

Although two criminal cases may invoke the same legal section, their surrounding circumstances often differ substantially, producing different judicial outcomes. This reflects an implicit principle of judicial relativity: the same law may yield different results because the contextual variables differ. The observation is consistent with a large body of work in legal analytics showing that judicial interpretation depends on factual context and precedent in addition to statutory text, and that these dependencies are learnable from case documents (Surden 2019; Aletras et al. 2016; Katz et al. 2017).

1.2 Dimensionality Reduction as a Conceptual Analogy

In modern data science, high-dimensional datasets are frequently simplified through dimensionality-reduction methods such as Principal Component Analysis (PCA). Instead of examining hundreds of correlated variables independently, PCA identifies a smaller number of components that preserve most of the meaningful variation in the data (Jolliffe and Cadima 2016). Inspired by this idea, the present paper proposes an analogous — but deliberately interpretable — organization of legal reasoning. Rather than asking judges to weigh hundreds of interconnected variables manually, the framework organizes evidence into a small number of interpretable judicial

dimensions such as: Evidence Strength; Intent Probability; Constitutional Compatibility; Precedent Similarity; and Social Harm Index.

These dimensions do not determine guilt or innocence. Instead, they provide judges with a simplified, auditable representation of an otherwise highly complex legal problem. It is important to stress that the dimensionality-reduction step is used here as a conceptual analogy and as an interpretability aid, not as a literal claim that latent statistical components equal legal categories. Section 4.5.1 makes this analogy mathematically precise: it gives the exact linear-algebraic relationship between the underlying evidentiary variables and each judicial dimension, so that the analogy to PCA is not merely rhetorical.

1.3 From Opaque Prediction to Explainable Augmentation

Unlike opaque predictive systems that output a single risk score or verdict, the proposed framework emphasizes explainability. Concerns about black-box risk-assessment tools in the justice system — including questions of accuracy, contestability, and disparate impact across demographic groups — are now well documented, and they motivate a design in which every recommendation is traceable (Dressel and Farid 2018; Angwin et al. 2016; Barredo Arrieta et al. 2020). Accordingly, each recommendation produced by the JRF must identify its supporting evidence, contradictory evidence, similar decided cases, confidence level, the legal provisions involved, and the reasoning chain that links them. Consequently, AI functions as an intelligent legal analyst rather than an autonomous judge.

1.4 Contributions

This paper makes the following contributions, which are developed in the sections that follow:

1. It formalizes judicial relativity as a design principle for multi-dimensional legal-evidence modeling (Sections 1 and 3).

2. It proposes an AI-assisted dimensional-reduction view of judicial interpretation, formalized mathematically via principal component analysis and, for nonlinear evidentiary interactions, autoencoders (Section 4.5.1), and used as a conceptual analogy to PCA rather than a literal application (Sections 3 and 4).

3. It specifies an explainable-AI layer that yields traceable, additively decomposable reasoning — formalized through Shapley-value attribution (Section 4.8) — rather than opaque prediction (Sections 3, 4.8, and 7).

4. It integrates semantic retrieval of precedent cases based on factual similarity, formalized as approximate nearest-neighbor search over case embeddings (Section 4.7).

5. It embeds a human-in-the-loop protocol in which AI supports, but never replaces, the judiciary, together with the accompanying ethical and constitutional safeguards (Sections 6–8).

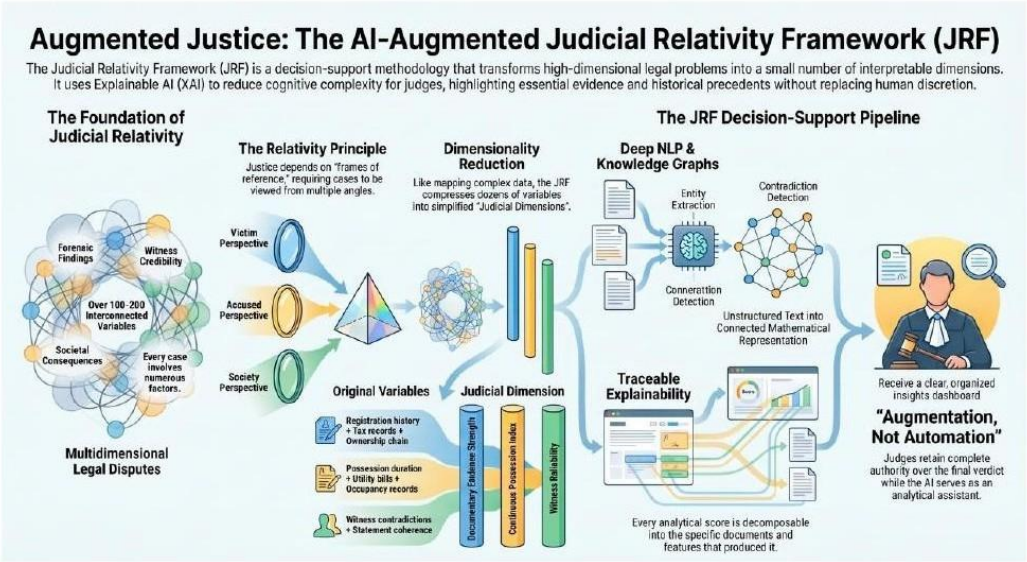

Fig. 1 Overview of the AI-Augmented Judicial Relativity Framework (JRF) pipeline: multidimensional legal disputes are organized via the relativity principle and dimensionality reduction into interpretable judicial dimensions; an NLP and knowledge-graph pipeline converts unstructured text into a connected mathematical representation; a traceable explainability layer decomposes every score into its contributing evidence; and the judge receives an organized decision-support dashboard while retaining full authority over the verdict ("Augmentation, not Automation").

2 Related Work and Background

The JRF draws on five strands of prior research: predictive modeling of judicial decisions, natural language processing (NLP) for legal text, dimensionality reduction and representation learning, explainable AI, and the critical literature on algorithmic risk assessment. This section situates the framework within each strand and identifies the gap it addresses.

2.1 Predictive Modeling of Judicial Decisions

A substantial line of computational-legal research has shown that features extracted from case documents carry signal about outcomes. Aletras et al. (2016) demonstrated that the outcomes of European Court of Human Rights cases can be predicted from textual content with strong accuracy, while Katz et al. (2017) built a time-evolving classifier that predicts the votes and case outcomes of the U.S. Supreme Court across nearly two centuries. Neural approaches to legal-judgment prediction have since extended these results (Chalkidis et al. 2019), and broad surveys catalogue how NLP is being applied across the legal system (Zhong et al. 2020). The JRF departs from this tradition in a crucial respect: its goal is not to predict a verdict but to organize and explain evidence so that a human judge can reason more effectively.

2.2 NLP and Language Models for Legal Text

Transformer-based language models underpin most current legal-NLP systems (Vaswani et al. 2017). General-purpose encoders such as BERT (Devlin et al. 2019) and their legal-domain adaptations, notably LEGAL-BERT (Chalkidis et al. 2020), produce contextual representations that support classification, retrieval, and entity extraction on legal corpora. The JRF's text pipeline relies on established components: named entity recognition to locate legal entities (Nadeau and Sekine 2007), relation extraction to populate a legal knowledge graph (Hogan et al. 2021), sentence-level embeddings for semantic comparison (Reimers and Gurevych 2019), and natural language inference to detect contradictions between statements (Bowman et al. 2015).

2.3 Dimensionality Reduction and Representation Learning

Reducing many correlated variables to a few interpretable factors is a mature area of machine learning. PCA remains the canonical linear method (Jolliffe and Cadima 2016); autoencoders learn nonlinear compressed representations (Hinton and Salakhutdinov 2006); t-SNE (van der Maaten and Hinton 2008) and UMAP (McInnes et al. 2018) are widely used for visualizing high-dimensional structure; and non-negative matrix factorization yields additive, and therefore often more interpretable, components (Lee and Seung 1999). The JRF treats these methods as candidate tools for compressing a legal feature matrix, with the choice governed by the interpretability requirements discussed in Section 4.5, and gives the linear (PCA) and nonlinear (autoencoder) cases an explicit mathematical treatment in Section 4.5.1.

2.4 Explainable AI in High-Stakes Decision-Making

Because judicial decisions affect fundamental rights, the framework must expose its reasoning. Model-agnostic explanation methods such as LIME (Ribeiro et al. 2016) and SHAP (Lundberg and Lee 2017) attribute a prediction to its contributing features, and a broader methodological literature argues for rigorous, human-centered notions of interpretability (Doshi-Velez and Kim 2017) and surveys the concepts and open challenges of explainable AI (Barredo Arrieta et al. 2020). The JRF adopts these principles at the level of each judicial dimension, requiring every score to be decomposable into the evidence that produced it — a requirement given exact mathematical content in Section 4.8 via the Shapley-value efficiency property.

2.5 Algorithmic Risk Assessment and Its Critiques

The deployment of algorithmic risk-assessment tools in criminal justice has attracted sustained scrutiny. Investigative reporting raised concerns that a widely used recidivism tool produced racially disparate error rates (Angwin et al. 2016); subsequent work showed that the same tool was no more accurate than untrained humans and could be matched by a two-feature linear model (Dressel and Farid 2018); and formal analyses established that competing fairness criteria cannot generally be satisfied simultaneously (Chouldechova 2017; Barocas and Selbst 2016). These findings directly shape the JRF's design: the framework is explicitly positioned as decision support under human control, with auditable dimensions and explicit safeguards (Section 7), rather than as an automated predictor of guilt or sentence.

3 The Judicial Relativity Framework

3.1 Design Principles

This section presents the Judicial Relativity Framework (JRF) itself. It states the design principles that govern the framework (Section 3.1), describes how case evidence is organized into a small set of interpretable judicial dimensions (Section 3.2), and outlines the end-to-end system architecture (Section 3.3), which the two case studies in Sections 4 and 5 then instantiate.

Four principles govern the framework. First, augmentation over automation: the system organizes and explains evidence but never issues a verdict. Second, traceability: every quantitative output must be decomposable into the underlying documents and features that produced it. Third,

interpretability by construction: reduced dimensions are mapped onto legally meaningful concepts rather than presented as anonymous statistical factors. Fourth, contestability: contradictory evidence and confidence estimates are surfaced so that parties and judges can challenge the analysis.

3.2 Judicial Dimensions

The framework compresses hundreds of case variables into a small set of interpretable judicial dimensions — for example, documentary evidence strength, continuous-possession index, legal-ownership consistency, witness reliability, and precedent similarity in a property dispute, or driver-negligence probability, mechanical-failure contribution, and environmental contribution in an accident case. Each dimension is derived from many underlying features and must remain explainable by tracing its score back to the contributing evidence.

3.3 System Architecture Overview

At a high level, the JRF processes a case through seven stages: (i) a data-collection layer that ingests unstructured documents and structured records; (ii) an NLP pipeline that extracts entities, relations, embeddings, and contradictions; (iii) structured feature engineering; (iv) feature fusion into a unified matrix; (v) dimensionality reduction into judicial components; (vi) semantic retrieval of comparable precedents; and (vii) an explainable-AI layer that renders the results as an auditable judicial dashboard. The next two sections instantiate this architecture on two illustrative case studies, and Sections 4.5.1, 4.7, and 4.8 give stages (v), (vi), and (vii) a complete mathematical formulation.

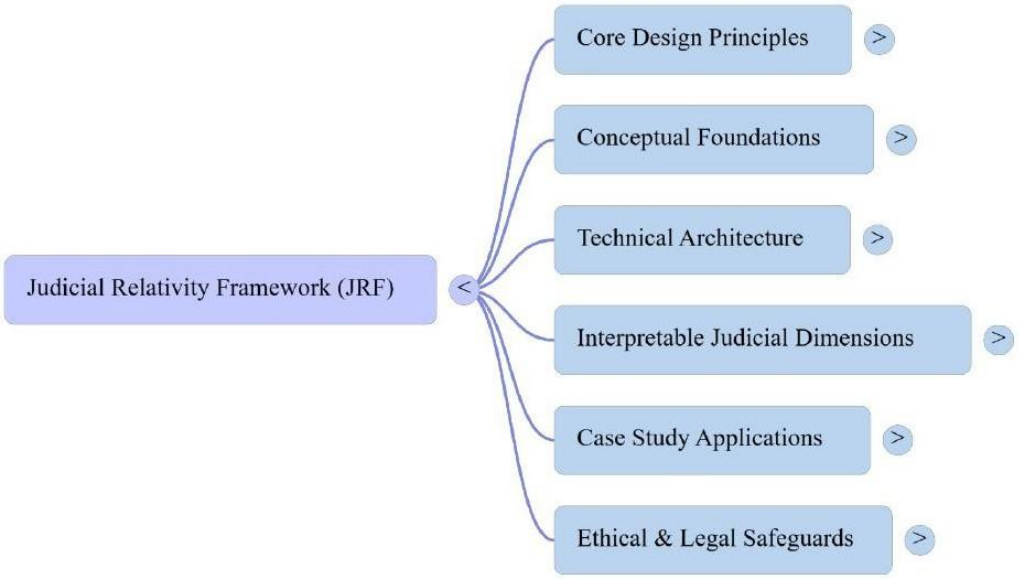

Fig. 2 Conceptual map of the six components of the Judicial Relativity Framework: core design principles, conceptual foundations, technical architecture, interpretable judicial dimensions, case-study applications, and ethical and legal safeguards.

4 Case Study I: Property Ownership Dispute

This case study provides a technical interpretation of the NLP and dimensionality-reduction stages of the framework applied to a contested claim of property ownership.

4.1 Data Collection Layer

The case contains both unstructured and structured data.

4.1.1 Unstructured Textual Data

The unstructured record typically includes sale deeds, family-settlement agreements, revenue orders, witness statements, court judgments, legal notices, cross-examination transcripts, affidavits, and mutation orders. These documents are processed using natural language processing.

4.1.2 Structured Numerical Data

Representative structured variables are summarized in Table 1. Together with the textual signals, a single case may yield over 100–200 measurable variables.

Table 1 Representative structured numerical variables

Variable	Type
Number of ownership transfers	Integer
Years of uninterrupted possession	Numeric
Number of tax payments	Numeric
Number of witnesses	Integer
Signature verification confidence	Percentage
Number of contradictions	Integer
Similar precedent score	Float
Document age	Numeric
Timeline gaps	Numeric

4.2 NLP Processing Pipeline

The textual documents first undergo preprocessing. The full pipeline proceeds as a sequence of stages:

Raw documents → OCR (if scanned) → Sentence segmentation → Tokenization → Stop-word removal → Lemmatization → Named entity recognition → Relation extraction → Semantic embedding

4.2.1 Named Entity Recognition (NER)

An NER model identifies legal entities such as person, property number, survey number, village, revenue office, judge, court, date, witness, document ID, and government department (Nadeau and Sekine 2007). For example, the sentence "Ram Kumar transferred Survey No. 144 to Mohan Kumar on 12 June 1998" is extracted as: Seller = Ram Kumar; Buyer = Mohan Kumar; Survey No. = 144; Date = 12-06-1998; Transaction = Sale. These extracted entities populate a legal knowledge graph.

4.2.2 Relation Extraction

The system next identifies relationships between entities — for instance, that Person A owns Property X, that Person B claims Property X, and that Witness C supports Person A. Instead of isolated text, the system thereby constructs a connected legal knowledge graph in which such relations are first-class objects (Hogan et al. 2021).

4.2.3 Semantic Embeddings

Each document is converted into a dense vector using transformer-based language models such as BERT, LEGAL-BERT, or comparable legal-domain encoders (Devlin et al. 2019; Chalkidis et al. 2020; Vaswani et al. 2017). A statement such as "This land was inherited…" becomes a high-dimensional vector (often 768 or more dimensions), enabling semantic comparison beyond exact wording (Reimers and Gurevych 2019). This allows the system to recognize that "The property was inherited" and "The land came through succession" express the same fact even though the wording differs.

Formally, an encoder maps each statement or document segment to a dense vector $e \in \mathbb{R}^{768}$ (or another fixed dimensionality determined by the chosen transformer). Two embeddings *e_u* and *e_v* are compared using cosine similarity:

$$\mathrm{sim}(e_u, e_v) = \frac{e_u \cdot e_v}{\|e_u\| \, \|e_v\|} \in [-1, 1] \tag{1}$$

with similarity close to 1 indicating that the two statements express the same underlying fact regardless of surface wording. This score itself becomes an input feature to the Precedent Similarity dimension (Section 4.6) and to the case-retrieval stage (Section 4.7).

4.2.4 Contradiction Detection

Natural language inference (NLI) models compare pairs of statements and label them as entailment, neutrality, or contradiction (Bowman et al. 2015). For example, "The father never sold the property" and "The property was sold in 1998" are flagged as contradictory, and the system assigns a contradiction-confidence score that later feeds the witness-reliability dimension.

Formally, the classifier maps a premise–hypothesis pair (p, h) to a contextual representation and then to class logits $\ell \in \mathbb{R}^3$, one per label. A softmax converts logits to a probability distribution over the three labels:

$$P(\mathrm{class} = k \mid p, h) = \frac{\exp(\ell_k)}{\sum_{j \in \{ent, neu, con\}} \exp(\ell_j)} \tag{2}$$

for $k \in$ {entailment, neutral, contradiction}. The contradiction-confidence score referenced above is P(contradiction | p, h); it is one of the raw features later compressed into the Witness Reliability dimension (Table 3).

4.3 Structured Feature Engineering

Numerical variables are extracted and normalized into a feature record. An example feature vector for a case is shown in Table 2. Every case eventually becomes a numerical feature vector.

Table 2 Example structured feature vector for a single case

Feature	Value
Years of possession	29
Number of tax receipts	31
Ownership transfers	2
Mutation approvals	4
Missing records	1
Witness consistency	0.82
Signature authenticity	0.94
Timeline completeness	0.91

4.4 Feature Fusion

Textual and numerical information are then merged into a single representation:

Semantic embeddings + Numerical features + Knowledge-graph features + Legal-precedent similarity → Unified feature matrix

For instance, Case 101 might be represented as the fused vector [0.84, 0.33, 0.77, …, 29, 31, 4, 0.91, 0.82, 0.95, …]. Every case can now be represented in a common mathematical feature space.

4.5 Dimensionality Reduction

Instead of presenting roughly 120 correlated variables to the judge, dimensionality reduction identifies latent structures that explain most of the variation in the data:

≈120 variables → principal components → 5 judicial dimensions

Candidate techniques include PCA for linear feature compression (Jolliffe and Cadima 2016), autoencoders for nonlinear representation learning (Hinton and Salakhutdinov 2006), UMAP or t-SNE for exploratory visualization of case similarity (van der Maaten and Hinton 2008; McInnes et al. 2018), and non-negative matrix factorization when interpretable additive components are desired (Lee and Seung 1999). The exact algorithm should be selected according to the data characteristics and the interpretability requirements. For example, dozens of variables related to document quality, chronology, and registration may collectively contribute to a latent component interpreted as documentary evidence strength. Section 4.5.1 formalizes this reduction step mathematically.

4.5.1 Mathematical Formulation of the Dimensionality-Reduction Step

Let each case i be represented, after feature fusion (Section 4.4), by a vector $x\_i \in \mathbb{R}^d$, where $d \approx 120$ for the present case study. Stacking n cases (the current case together with the retrieved precedent corpus) yields a data matrix $X \in \mathbb{R}^{n \times d}$. Writing the column-wise mean as $\mu = (1/n) \Sigma_i x\_i \in \mathbb{R}^d$ and the mean-centered matrix as $\tilde{X} = X - 1_n \mu^T$, the sample covariance matrix of the fused evidentiary features is

$$\Sigma = \frac{1}{n-1} \tilde{X}^T \tilde{X} \tag{3}$$

Principal Component Analysis diagonalizes Σ:

$$\Sigma = W \Lambda W^T \tag{4}$$

where $W = [w_1, \ldots, w\_d]$ is an orthogonal matrix of eigenvectors (the principal directions) and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda\_d)$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda\_d \geq 0$, are the corresponding eigenvalues. Retaining the top k eigenvectors, $W\_k = [w_1, \ldots, w\_k] \in \mathbb{R}^{d \times k}$, gives the reduced representation of case i — its vector of judicial-dimension scores:

$$z_i = W_k^T (x_i - \mu) \tag{5}$$

Equation (5) is the formal counterpart of the arrow diagram '≈120 variables → principal components → 5 judicial dimensions' above: each of the five entries of z_i is one judicial dimension (Documentary Evidence Strength, Continuous Possession Index, etc.; Table 3).

Because *W_k* is fixed once estimated, every judicial-dimension score is, by construction, a weighted linear combination of the underlying evidentiary features:

$$z_{i,j} = \sum_{l=1}^{d} w_{l,j}(x_{i,l} - \mu_l) \tag{6}$$

The term *c_l,j = w_l,j(x_i,l − μ_l)* is the exact contribution of raw feature *l* (e.g., 'years of possession') to judicial dimension *j* (e.g., 'Continuous Possession Index'), and summing *c_l,j* over all *l* reproduces *z_i,j* exactly. Equation (6) is therefore the mathematical basis of the traceability principle in Section 3.1: because PCA is linear, the decomposition of a judicial-dimension score into its contributing evidence (Section 4.8) is exact rather than approximate.

Choosing the number of retained dimensions *k* is not arbitrary. The proportion of total evidentiary variation captured by the *j*-th component is its explained-variance ratio,

$$\mathrm{EVR}_j = \frac{\lambda_j}{\sum_{m=1}^{d} \lambda_m} \tag{7}$$

and *k* is chosen as the smallest integer such that the cumulative explained variance exceeds a pre-registered threshold τ (e.g., $\tau = 0.85$):

$$\mathrm{CEV}_k = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{m=1}^{d} \lambda_m} \geq \tau \tag{8}$$

This gives the reduction to five judicial dimensions a quantitative justification rather than treating five as a fixed, aesthetically convenient number: *k* is only fixed at 5 once Equations (7)–(8) show that five components already capture the large majority of the variation across the ≈120 fused features.

Illustrative example. Suppose two standardized evidentiary indicators contributing to Documentary Evidence Strength — a normalized count of tax receipts, x_1, and a normalized

registration-continuity score, x_2 — are found, across the precedent corpus, to have sample covariance

$$\Sigma = [\ 1.00\ \ 0.80\ ;\ \ 0.80\ \ 1.00\ ]$$

(both features standardized to unit variance and correlated at $\rho = 0.8$, since continuous tax payment and continuous registration typically move together; rows are read left-to-right and separated by a semicolon). Diagonalizing this covariance matrix gives eigenvalues

$$\lambda_1 = 1 + \rho = 1.8, \quad \lambda_2 = 1 - \rho = 0.2 \tag{10}$$

with eigenvectors $w_1 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$ and $w_2 = (1/\sqrt{2}, -1/\sqrt{2})^{\mathrm{T}}$. By Equation (7), the first component alone explains

$$\mathrm{EVR}_1 = \frac{1.8}{2.0} = 0.90 = 90\% \tag{11}$$

of the joint variation in these two indicators, so a single judicial dimension

$$z = \frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_2 \tag{12}$$

— essentially the average of the two standardized indicators — is a near-lossless summary of both. This is the precise sense in which 'tax records' and 'registration continuity' collapse into one entry of Table 3.

Nonlinear alternative. When evidentiary interactions are not well approximated by a linear combination — for example, if additional witnesses beyond some threshold add diminishing credibility — the framework may instead use an autoencoder (Section 2.3; Hinton and Salakhutdinov 2006). An encoder f_θ: $\mathbb{R}^{\mathrm{d}} \to \mathbb{R}^{\mathrm{k}}$ and decoder g_φ: $\mathbb{R}^{\mathrm{k}} \to \mathbb{R}^{\mathrm{d}}$ are trained jointly to minimize reconstruction error,

$$L(\theta, \varphi) = \frac{1}{n} \sum_i \| x_i - g_\varphi(f_\theta(x_i)) \|^2 \tag{13}$$

with the latent code $z\_i = f\_\theta(x\_i) \in \mathbb{R}^{\mathrm{k}}$ again serving as the judicial-dimension vector. In the special case where f_θ and g_φ are linear and the loss is squared error, the optimal solution spans the same subspace as the top-k PCA eigenvectors of Equation (4) (Hinton and Salakhutdinov

2006); the nonlinear case generalizes this to curved manifolds in evidence space, at some cost to the exact linear decomposition of Equation (6), which the explainable-AI layer (Section 4.8) then recovers approximately rather than exactly.

4.6 Formation of Judicial Components

Rather than displaying every original variable, the reduced representation of Equation (5) is mapped onto interpretable legal dimensions, as summarized in Table 3. These judicial dimensions are derived from multiple underlying features (Equation (6)) and should remain explainable by tracing each score back to its contributing evidence (Section 4.8).

Table 3 Mapping of underlying variables to interpretable judicial dimensions

Original Variables	Judicial Dimension
Registration history + tax records + mutation chain + ownership chronology	Documentary Evidence Strength
Possession duration + electricity bills + tax payments + occupancy records	Continuous Possession Index
Previous judgments + applicable statutes + similar legal reasoning	Legal Ownership Consistency
Witness contradiction score + cross-examination consistency + statement coherence	Witness Reliability
Semantic similarity to prior cases + outcome similarity	Precedent Similarity

4.7 Retrieval of Similar Cases

Each historical judgment is also represented as an embedding, so the current case can be matched against a precedent corpus through a progressive narrowing:

Current case → vector search → top 100 similar → top 10 relevant → top 3 nearly identical

Formally, let $q \in \mathbb{R}^p$ denote the embedding of the current case and $v_1, \ldots, v\_N \in \mathbb{R}^p$ the embeddings of the N precedents in the corpus, compared under the cosine distance $d(q, v) = 1 - \text{sim}(q, v)$ of Equation (1). Exact k-nearest-neighbor retrieval returns the subset of size K that minimizes total distance to q:

$$\mathrm{kNN}_K(q) = \underset{S \subseteq \text{Corpus},\ |S| = K}{\arg\min} \sum_{v \in S} d(q, v) \tag{14}$$

which costs O(*Np*) per query under brute-force search. At the corpus sizes relevant to a national precedent database this cost is prohibitive, so the system instead uses a graph-based approximate nearest-neighbor index — hierarchical navigable small-world (HNSW) graphs (Malkov and Yashunin 2020) — that answers each query in approximately O(log *N*) time while empirically returning a result set close to the true *k*-NN of Equation (14). This is what supports the progressive narrowing 'top 100 → top 10 → top 3' described above at national-precedent-database scale. Approximate nearest-neighbor search over legal embeddings retrieves factually similar precedents efficiently, even at large scale (Malkov and Yashunin 2020; Johnson et al. 2021). The system then explains which factual and legal features contributed to each match, so that similarity is itself auditable rather than asserted.

4.8 Explainable AI Layer

The system provides transparent reasoning rather than only numerical scores (Ribeiro et al. 2016; Lundberg and Lee 2017). For example, a documentary-evidence-strength score of 92% might be attributed to a verified registered sale deed, continuous tax payments over thirty years, the absence of registration conflict, consistent revenue records, and 94% signature authenticity. A witness-reliability score of 76% might be attributed to two witnesses contradicting earlier affidavits, three witnesses remaining consistent across examinations, and one witness lacking documentary support. This traceability lets judges inspect the reasoning rather than relying on a black-box output.

These attributions have a precise mathematical form. For a judicial-dimension score $y = h(x)$ computed from underlying features $F = \{1, \ldots, d\}$ (Equation (6), or a downstream nonlinear aggregation such as Equation (13)), the Shapley value of feature l is

$$\varphi_l = \sum_{S \subseteq F \setminus \{l\}} \frac{|S|!\ (|F| - |S| - 1)!}{|F|!} \cdot \left[h(S \cup \{l\}) - h(S) \right] \tag{15}$$

where $h(S)$ denotes the model's expected output when only the features in subset S are observed and the rest are marginalized out (Lundberg and Lee 2017). Shapley values satisfy the efficiency

property — they sum exactly to the gap between the full-information and no-information predictions — so every dimension score decomposes exactly as

$$y = \varphi_0 + \sum_{l \in F} \varphi_l \tag{16}$$

with baseline $\varphi_0 = h(\varnothing)$ (e.g., the corpus-average score). Equation (16) is the formal statement of 'decomposable into the specific documents and features that produced it' (Fig. 1; Section 3.1). For the linear PCA dimensions of Section 4.5.1, this decomposition coincides exactly with the loadings of Equation (6) — $\varphi_l = w_{l,j}(x_l - \mu_l)$ — so no approximation is needed. For nonlinear components — the autoencoder of Equation (13), or any downstream classifier — exact evaluation of Equation (15) is combinatorial in $|F|$ and is instead estimated via sampling or the KernelSHAP approximation (Lundberg and Lee 2017).

4.9 Final Judicial Dashboard

Instead of reviewing hundreds of variables independently, the judge receives an evidence-centric summary, illustrated in Table 4. The dashboard also links to the underlying documents, explains how each score was computed via Equations (6) and (15)–(16), identifies contradictory evidence, and presents confidence estimates. The judge remains the ultimate decision-maker; the AI serves as an explainable analytical assistant that organizes complex evidence into a form that is easier to review while preserving transparency and accountability.

Table 4 Illustrative judicial dashboard for the property-ownership case

Judicial Dimension	Score	Key Evidence
Documentary Evidence Strength	92%	Registered deeds, mutation records, tax receipts
Continuous Possession Index	87%	Occupancy history, utility records, tax continuity
Legal Ownership Consistency	95%	Statutory compliance, inheritance rules, ownership chain
Witness Reliability	76%	Statement consistency and contradiction analysis
Precedent Similarity	91%	Closely matching Supreme Court and High Court cases

5 Case Study II: Road Accident Causing Death

5.1 Traditional Analysis

This second case study applies the architecture introduced in Section 3 to a road-accident dispute, illustrating that the framework generalizes across areas of law. It first reviews how liability is traditionally assessed (Section 5.1), then shows how the framework integrates the relevant evidence (Section 5.2) and reduces it to an interpretable decision profile (Section 5.3).

Consider a collision that results in one fatality. Liability turns on many entangled questions: overspeeding, weather, brake failure, intoxication, pedestrian negligence, CCTV quality, road condition, vehicle maintenance, driver reaction time, and emergency-response delay. Each factor influences liability differently, and no single factor is dispositive on its own.

5.2 AI-Assisted Evidence Integration

The framework integrates CCTV footage, GPS trajectory, vehicle telematics, weather history, road-engineering data, forensic reports, mobile-phone records, alcohol tests, witness testimony, and prior judgments involving similar circumstances. More than 200 variables may be examined simultaneously. Because the central legal question concerns causation rather than mere correlation, the analysis draws on causal-inference methods to distinguish contributing causes from incidental factors (Pearl 2009).

5.3 Reduced Judicial Representation

Instead of overwhelming the court with hundreds of parameters, the framework generates an interpretable decision profile — the vector z_i of Equation (5), here computed over the ≈200 accident-related variables and read off in qualitative bands rather than raw scores — shown in Table 5. The report further explains which evidence increased the negligence probability, which evidence reduced certainty, why comparable precedents produced different outcomes, the relevant constitutional considerations, and the applicable statutory provisions. The judge therefore receives an explainable reasoning pathway rather than a black-box prediction.

Table 5 Reduced judicial representation for the road-accident case

Judicial Dimension	Result
Driver Negligence Probability	High

Judicial Dimension	Result
Mechanical Failure Contribution	Low
Environmental Contribution	Moderate
Victim Contribution	Minimal
Similar Historical Case Match	94%

6 Discussion

6.1 Research Contributions

This section discusses the framework's contributions and how it should be evaluated. Section 6.1 summarizes the paper's core innovations, and Section 6.2 outlines the evaluation axes — retrieval quality, explanation faithfulness, fairness, and human-subjects studies — along which the framework should be assessed in future work.

The Judicial Relativity Framework introduces five core innovations: multi-dimensional legal-evidence modeling; AI-assisted dimensional reduction for judicial interpretation, now given a full linear-algebraic treatment in Section 4.5.1 (as a conceptual analogy to PCA, not a literal application); explainable AI with traceable, Shapley-decomposable reasoning rather than opaque prediction (Section 4.8); semantic retrieval of precedent cases based on factual similarity, formalized as approximate nearest-neighbor search (Section 4.7); and human-in-the-loop decision-making, in which AI supports but does not replace the judiciary. Taken together, these elements aim to improve judicial efficiency, reduce cognitive overload, and promote greater consistency while preserving judicial independence, due process, and the rule of law.

6.2 Evaluation Considerations

A framework of this kind should be evaluated along several axes rather than by predictive accuracy alone. Retrieval quality can be measured with standard information-retrieval metrics (for example, precision@k and mean reciprocal rank) against expert-annotated precedents, using the distance function of Equation (14). The faithfulness of the explanations — whether the Shapley attributions of Equation (15) actually drive each dimension — should be assessed separately from raw performance, since a plausible-sounding explanation is not necessarily a faithful one (Doshi-Velez and Kim 2017). Fairness should be audited across protected groups and case types, given the

documented risk that error rates diverge across populations (Dressel and Farid 2018; Angwin et al. 2016; Chouldechova 2017). Finally, because the ultimate goal is augmentation, the most important evaluation is a human-subjects study measuring whether judges reach more consistent and better-justified decisions with the tool than without it, while guarding against over-reliance.

7 Limitations, Risks, and Ethical Safeguards

7.1 Algorithmic Bias and Fairness

Deploying AI in judicial settings carries real risks that must be addressed directly. This section discusses algorithmic bias and fairness (Section 7.1), explainability and automation bias (Section 7.2), data quality, provenance, and security (Section 7.3), and legal and constitutional safeguards (Section 7.4).

Models trained on historical judgments can absorb and reproduce historical bias, and formal results show that several intuitive fairness criteria cannot be satisfied simultaneously once base rates differ across groups (Chouldechova 2017; Barocas and Selbst 2016). The JRF mitigates — but cannot eliminate — this risk by exposing per-dimension evidence via Equations (6) and (15)–(16), enabling group-wise error audits, and never emitting a verdict. Deployment should therefore be accompanied by ongoing, independent fairness monitoring rather than a one-time certification.

7.2 Explainability and Automation Bias

Transparent scores can paradoxically increase automation bias, in which a decision-maker defers to a confident-looking system. Because explanation methods such as LIME and SHAP approximate model behavior locally — and, for the nonlinear components of Section 4.5.1, Equation (15) itself is only estimated rather than computed exactly — their outputs must be presented as evidence for inspection rather than as ground truth (Ribeiro et al. 2016; Lundberg and Lee 2017; Doshi-Velez and Kim 2017). Interfaces should highlight uncertainty and contradictory evidence prominently so that the judge is prompted to interrogate, not merely accept, the analysis.

7.3 Data Quality, Provenance, and Security

The framework's outputs are only as reliable as its inputs. OCR errors, incomplete registries, forged documents, and gaps in the precedent corpus can all distort the fused feature matrix X of Section 4.5.1. Every ingested document should therefore carry provenance metadata, and the system

should flag low-confidence extractions instead of silently propagating them. Sensitive personal data demands strict access control, auditability, and compliance with applicable data-protection law.

7.4 Legal and Constitutional Safeguards

Automated analysis in adjudication raises due-process questions, including the right to contest and to understand an adverse process. Existing scholarship cautions that a general 'right to explanation' may be weaker in law than is often assumed, which makes voluntary, robust explainability all the more important (Wachter et al. 2017), and that introducing artificial intelligence into justice systems requires careful institutional design to preserve legitimacy and accountability (Re and Solow-Niederman 2019). The JRF is accordingly constrained to an advisory role: parties must be able to examine and challenge the evidence behind every dimension, the human judge retains full authority over the verdict, and the system's analysis forms part of the record rather than replacing judicial reasoning.

8 Conclusion

This paper has proposed the Judicial Relativity Framework, an AI-augmented decision-support methodology that reframes a legal dispute as a small set of interpretable judicial dimensions rather than a mass of correlated variables. Combining legal NLP, knowledge graphs, precedent retrieval, dimensionality reduction, and explainable AI, the framework produces a transparent, evidence-centric analysis while leaving the verdict entirely to the judge. The machine-learning core of the framework is made mathematically explicit: Section 4.5.1 derives the judicial dimensions as principal-component projections (or, in the nonlinear case, autoencoder latent codes) of the fused evidentiary feature vector; Section 4.7 formalizes precedent retrieval as approximate nearest-neighbor search under cosine distance; and Section 4.8 shows that every dimension score admits an exact Shapley-value decomposition into its contributing evidence. Two illustrative case studies — a property-ownership dispute and a fatal road accident — show how the same architecture, and the same mathematics, applies across very different areas of law.

The framework's value proposition is augmented justice, not automated justice: reducing cognitive complexity, surfacing overlooked and contradictory evidence, and improving consistency and transparency, all under human control and subject to the fairness, explainability, and due-process

safeguards set out above. Future work should validate the framework empirically through retrieval-quality benchmarks, explanation-faithfulness studies, fairness audits, and human-subjects trials with practicing judges.

Statements and Declarations

Competing Interests The author declares no competing interests.

Funding No funding was received to assist with the preparation of this manuscript.

Data Availability This is a conceptual and mathematical framework paper. No empirical datasets were generated or analyzed during this study; the illustrative case data used in Sections 4 and 5 are hypothetical and are presented in full within the manuscript (Tables 1–5).

Author Contribution Prabhat Kumar is the sole author and is responsible for the conception, mathematical formalization, and writing of this work.

References

Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V (2016) Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. PeerJ Comput Sci 2:e93

Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias. ProPublica, 23 May 2016

Ashley KD (2017) Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge University Press, Cambridge

Barocas S, Selbst AD (2016) Big data's disparate impact. Calif Law Rev 104(3):671–732

Barredo Arrieta A et al (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Bowman SR, Angeli G, Potts C, Manning CD (2015) A large annotated corpus for learning natural language inference. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 632–642

Chalkidis I, Androutsopoulos I, Aletras N (2019) Neural legal judgment prediction in English. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pp 4317–4323

Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I (2020) LEGAL-BERT: the muppets straight out of law school. In: Findings of the Association for Computational Linguistics: EMNLP 2020, pp 2898–2904

Chouldechova A (2017) Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data 5(2):153–163

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT 2019, pp 4171–4186

Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. arXiv:1702.08608

Dressel J, Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. Sci Adv 4(1):eaao5580

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507

Hogan A et al (2021) Knowledge graphs. ACM Comput Surv 54(4):1–37

Johnson J, Douze M, Jégou H (2021) Billion-scale similarity search with GPUs. IEEE Trans Big Data 7(3):535–547

Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A 374(2065):20150202

Katz DM, Bommarito MJ, Blackman J (2017) A general approach for predicting the behavior of the Supreme Court of the United States. PLOS ONE 12(4):e0174698

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS), vol 30, pp 4765–4774

Malkov YA, Yashunin DA (2020) Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. IEEE Trans Pattern Anal Mach Intell 42(4):824–836

McInnes L, Healy J, Saul N, Melville J (2018) UMAP: uniform manifold approximation and projection for dimension reduction. J Open Source Softw 3(29):861

Nadeau D, Sekine S (2007) A survey of named entity recognition and classification. Lingvisticae Investig 30(1):3–26

Pearl J (2009) Causality: models, reasoning, and inference, 2nd edn. Cambridge University Press, Cambridge

Re RM, Solow-Niederman A (2019) Developing artificially intelligent justice. Stanford Technol Law Rev 22:242–289

Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: Proceedings of EMNLP-IJCNLP 2019, pp 3982–3992

Ribeiro MT, Singh S, Guestrin C (2016) "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1135–1144

Surden H (2019) Artificial intelligence and law: an overview. Ga State Univ Law Rev 35(4):1305–1337

van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

Vaswani A et al (2017) Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS), vol 30, pp 5998–6008

Wachter S, Mittelstadt B, Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. Int Data Priv Law 7(2):76–99

Zhong H, Xiao C, Tu C, Zhang T, Liu Z, Sun M (2020) How does NLP benefit legal system: a summary of legal artificial intelligence. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pp 5218–5230