ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL

Vol.9. No.1.2026

Computational Models of Legal Reasoning: A Research Study of the Theoretical Foundations and Practical Limits of Machine Decision-Making

1. Asma Rehman
Advocate, Lahore High Court, Lahore; Visiting Faculty Member, Law Department, Punjab University, Gujranwala Campus, Pakistan.
Email: Asmarehman014@gmail.com
2. Gulshan Rehman
Advocate High Court & Tax Consultant, Irfan & Farhad Law Associates, Lahore, Pakistan.
Email: Grehman136@gmail.com
3. Dr. Hafiz Muhammad Abrar Awan
Assistant Professor, Department of Urdu Encyclopedia of Islam, University of the Punjab, Lahore, Pakistan.
Email: abrara.uei@pu.edu.pk
ORCID: https://orcid.org/0000-0002-5345-8424

Abstract

Computational legal reasoning has developed from early rule-based expert systems into a diverse field encompassing defeasible logic, case-based reasoning, formal argumentation, value-sensitive models, statistical prediction, machine learning, and large language models. These approaches seek to represent, reconstruct, predict, or assist legal reasoning by translating selected features of legal judgment into computationally processable structures. Their development has generated an important jurisprudential question: which parts of legal reasoning can be formalised, and which depend upon interpretation, institutional authority, contextual judgment, contested values, or forms of practical reasoning that resist complete computational representation? This study examines the principal theoretical foundations of computational legal reasoning and evaluates the practical limits of machine decision-making in legal contexts. It analyses rule-based and defeasible reasoning, precedent-based models, argumentation frameworks, value-sensitive approaches, data-driven prediction, and contemporary language-model systems. Particular attention is given to open-textured legal concepts, analogical reasoning, conflicts among values, uncertainty in facts, source hierarchy, model hallucination, benchmark limitations, and the institutional legitimacy of automated legal decisions. The study argues that computational models are most effective when they represent bounded aspects of legal reasoning rather than when they are treated as substitutes for the entire judicial process. Formal systems can improve consistency, retrieval, argument mapping, and structured comparison, while statistical systems can detect patterns and assist prediction. However, neither approach eliminates the need for legal interpretation, evidentiary judgment, normative evaluation, and authoritative human responsibility. The article concludes that the most defensible future lies in hybrid systems in which computation supports legal reasoning while final normative responsibility remains embedded in legally accountable institutions.

Keywords: Computational Legal Reasoning, Artificial Intelligence and Law, Defeasible Logic, Case-Based Reasoning, Legal Argumentation, Machine Decision-Making, Open Texture, Large Language Models, Judicial Reasoning

Introduction

The attempt to represent legal reasoning computationally is older than the contemporary wave of generative artificial intelligence. For decades, researchers in artificial intelligence and law have investigated whether legal rules, precedents, arguments, priorities, exceptions,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

analogies, and normative values can be represented through formal systems capable of producing legally relevant conclusions.

The project is intellectually demanding because legal reasoning differs from simple calculation. Courts frequently reason under incomplete information, conflicting authorities, exceptions, uncertain facts, competing principles, and language that deliberately leaves room for contextual judgment. A legal rule may appear simple in abstract form yet become difficult to apply once its concepts interact with unforeseen facts. Precedent adds another layer: prior decisions constrain later reasoning, but similarity itself requires interpretation.

Computational models therefore do not merely ask whether a machine can "reach the correct answer." They force jurisprudence to identify the internal components of legal reasoning. What is a rule? When may it be defeated by an exception? Why is one precedent more relevant than another? How do judges distinguish cases? What makes one argument stronger when both are legally plausible? Which values justify competing interpretations? These questions have driven the development of symbolic AI, case-based systems, argumentation theory, and more recently large-scale statistical models.

Contemporary machine-learning systems have substantially expanded the field. Rather than requiring every legal rule to be manually encoded, statistical systems can learn patterns from large corpora of cases, statutes, contracts, and other legal documents. Large language models can generate legal analysis, classify textual materials, summarise cases, draft arguments, and respond to legal questions in natural language. Yet this increased flexibility introduces new problems. Systems may generate convincing but incorrect propositions, obscure their reasoning path, confuse authoritative and non-authoritative sources, or treat jurisprudential disagreement as though it were merely a prediction problem.

This study therefore distinguishes between computational assistance and legal authority. The central research question is not simply whether machines can reproduce certain outputs associated with legal reasoning, but whether the internal structure, uncertainty, value judgments, and institutional legitimacy of law can be adequately captured by computational models.

Section One: Theoretical Foundations of Computational Legal Reasoning

Computational legal reasoning emerged from attempts to formalise legal knowledge in ways that permit systematic inference. Early work frequently adopted logical and rule-based approaches, while later research incorporated exceptions, precedent, dialectical argument, and social values. These models remain important because they make their assumptions explicit and allow researchers to identify exactly which elements of legal reasoning can be represented formally.

1. Rule-Based, Non-Monotonic and Defeasible Models of Legal Reasoning

Classical logical systems initially offered an attractive model for legal reasoning because legal norms often possess an apparent conditional structure: if legally relevant facts are established, a specified legal consequence follows. However, real legal systems are saturated with exceptions, qualifications, priorities, presumptions, and rules whose application can be displaced by stronger considerations. Consequently, researchers increasingly turned from strictly deductive models toward defeasible and non-monotonic reasoning.

Phan Minh Dung described the purpose of his influential argumentation framework as:"The purpose of this paper is to study the fundamental mechanism humans use in argumentation."[1]

[1] Phan Minh Dung, "On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games," *Artificial Intelligence* 77, no. 2 (1995): 321–357.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

This statement captures a major conceptual shift. Instead of assuming that legal reasoning consists primarily of deduction from fixed premises, argumentation models focus upon conflict among competing reasons. A legal conclusion may be justified not because it is deductively unavoidable, but because the arguments supporting it survive relevant attacks. This is particularly appropriate for law, where parties routinely present incompatible constructions of statutes, facts, precedents, and principles. Dung's framework does not itself constitute a complete theory of law; rather, it provides an abstract structure for representing attack and acceptability among arguments. Its enduring importance lies in moving computational reasoning from the simple application of rules toward the evaluation of argumentative conflict.

Prakken and Sartor later developed systems in which:

> "Conflicts between arguments are decided with the help of priorities on the rules."[2]

The significance of this model lies in its recognition that legal norms rarely operate on a perfectly equal plane. A specific rule may override a general one; later legislation may displace earlier legislation; constitutional norms may control ordinary legislation; exceptional circumstances may defeat an otherwise applicable rule. Yet the authors' approach is especially important because the priorities themselves need not always be fixed. They can become part of the argumentative dispute. This reflects legal practice more accurately than a rigid hierarchy in which every conflict is predetermined. Lawyers often disagree not only about which rules apply, but also about which rule should prevail under the circumstances and why. Computational legal reasoning therefore becomes partly a model of contestable priority rather than mere logical execution.

Defeasible logic is particularly useful in areas where legal conclusions remain provisional. A prima facie entitlement may exist unless an exception applies; an evidentiary presumption may operate until rebutted; one statutory provision may govern unless a more specific provision controls. These structures can be represented more realistically through non-monotonic reasoning than through classical logic, because the addition of new information may legitimately change the conclusion.

The theoretical limitation, however, appears when the system must determine the content of the rules themselves. Formal logic can model the consequence of a rule once its meaning has been specified, but disputes about statutory interpretation may concern the very formulation of the operative rule. At that point, computation requires a prior interpretive choice.

2. Case-Based Reasoning and the Computational Representation of Precedent

Common-law reasoning cannot be reduced entirely to statutes or explicit rules. Lawyers and judges reason through prior cases, compare factual configurations, identify relevant similarities, distinguish adverse precedents, and formulate rules at different levels of generality. This led to the development of case-based reasoning as one of the most important traditions in AI and law.

Ashley explains the basic mechanism in direct terms:

> "Reasoners compare problems to prior cases to draw conclusions about a problem and guide decision making."[3]

This formulation captures the central logic of precedent without pretending that precedent operates mechanically. A prior case is not relevant simply because it shares many facts with

---

[2] Henry Prakken and Giovanni Sartor, "Argument-Based Extended Logic Programming with Defeasible Priorities," *Journal of Applied Non-Classical Logics* 7, nos. 1–2 (1997): 25–75.

[3] Kevin D. Ashley, "Case-Based Reasoning and Its Implications for Legal Expert Systems," *Artificial Intelligence and Law* 1 (1992): 113–208.

the present dispute. The legally significant question is which similarities and differences matter in light of the legal issue. A computational model must therefore represent not only factual resemblance but legally meaningful dimensions. This is why early case-based systems such as HYPO became important: they treated precedent as structured comparison rather than merely textual retrieval. The theoretical contribution of such systems was to show that legal analogy can be partially represented through dimensions, factors, competing cases, and arguments for distinguishing or following precedent.

Ashley and Rissland later characterised HYPO-style systems as adopting:

> "a 'lazy' approach"[4]

to generalisation.

The expression is conceptually significant. Rather than constructing a universal legal rule in advance, the system postpones generalisation until a new factual problem appears. It then positions the new dispute among earlier cases and reasons from the pattern of similarities and differences. This resembles an important feature of common-law adjudication: courts often do not formulate the broadest possible rule before they know the factual dispute requiring resolution. The meaning of precedent evolves through successive applications, distinctions, and reformulations. Computational case-based reasoning thus captures something that fixed-rule systems miss: legal categories can emerge through iterative comparison rather than complete ex ante specification.

Bench-Capon and Sartor extend the theory further by presenting reasoning with cases as:

> "a process of constructing, evaluating and applying a theory."[5]

This view moves beyond the idea that precedent is simply a database of similar decisions. Cases support competing theories about which factual features matter and which values justify the legal rule. A lawyer may therefore construct one account of the precedents while an opponent constructs another. The computational task becomes the evaluation of rival explanatory structures rather than the identification of the numerically closest case. This is a major jurisprudential insight because legal analogy is rarely reducible to similarity alone. The decisive similarity depends upon an underlying theory of the field: why a feature matters, what legal value it advances, and how it relates to prior outcomes.

The practical strength of case-based systems is their capacity to support argument construction, legal education, and precedent retrieval. Their limitation is that the representation of legally significant factors still requires a theory of relevance. Machines can compare encoded features, but deciding which features should be encoded may itself involve contested legal judgment.

3. Argumentation, Competing Values and the Normative Structure of Legal Choice

Many difficult legal disputes cannot be resolved solely by determining which rule technically applies. Competing legal interpretations may each be doctrinally plausible while advancing different constitutional, social, economic, or institutional values. Legal decision-making therefore includes practical reasoning about what the law is seeking to protect and how competing values should be balanced.

Bench-Capon argues that:

> "the strength of an argument depends on the social values that it advances."[6]

---

[4] Kevin D. Ashley and Edwina L. Rissland, "Law, Learning and Representation," *Artificial Intelligence* 150, nos. 1–2 (2003): 17–58.

[5] Trevor J. M. Bench-Capon and Giovanni Sartor, "A Model of Legal Reasoning with Cases Incorporating Theories and Values," *Artificial Intelligence* 150, nos. 1–2 (2003): 97–143.

[6] Trevor J. M. Bench-Capon, "Persuasion in Practical Argument Using Value-Based Argumentation Frameworks," *Journal of Logic and Computation* 13, no. 3 (2003): 429–448.

This proposition identifies a fundamental limit of purely syntactic legal models. Arguments often derive persuasive force not simply from formal structure but from the values they protect. In one dispute, legal certainty may compete with substantive fairness; in another, privacy may conflict with public safety; elsewhere, freedom of contract may confront consumer protection. A computational model that represents arguments without representing their normative foundations may reproduce the structure of disagreement while missing why one interpretation is preferred. Value-based argumentation frameworks seek to address this problem by linking arguments to values and examining how different orderings of values affect acceptability.

The advantage of value-based models is not that they automatically determine the correct ordering of legal values. Rather, they make the underlying normative disagreement explicit. Two decision-makers may agree on all relevant facts and rules yet disagree because they assign different weight to competing principles.

This is especially important in constitutional and administrative law, where proportionality, reasonableness, necessity, fairness, and public interest cannot always be reduced to binary variables. A computational system can assist by mapping the competing arguments and identifying consequences of different value priorities. It cannot, however, derive the authoritative priority of values from logic alone.

Legal authority also matters. Even a formally coherent value ordering does not become law merely because a model prefers it. Constitutional texts, statutes, precedent, institutional competence, and judicial legitimacy determine which actors are authorised to resolve normative conflict. Computational legal reasoning must therefore distinguish between modelling a normative choice and possessing legal authority to make that choice.

Section Two: Machine Decision-Making and the Limits of Formalisation

The transition from symbolic legal systems to machine learning has significantly expanded computational capabilities. Modern models can work with unstructured legal language and large corpora without requiring every concept to be manually formalised. Yet this flexibility changes rather than eliminates the theoretical problems. Statistical systems can recognise patterns while remaining vulnerable to semantic ambiguity, unstable authority, factual uncertainty, and fabricated content.

1. Statistical Learning, Large Language Models and the Expansion of Legal Reasoning Tasks

Contemporary legal AI increasingly relies upon statistical learning rather than manually encoded legal rules. Large language models generate outputs from probabilistic relationships learned across extensive textual corpora. They can therefore perform tasks that earlier symbolic systems could not easily approach, including long-form drafting, summarisation, classification, and natural-language interaction.

LegalBench describes itself as:

> "a collaboratively constructed legal reasoning benchmark consisting of 162 tasks."[7]

The scale and diversity of this benchmark demonstrate that "legal reasoning" is not a single capability. The benchmark distinguishes among multiple forms of reasoning and includes tasks constructed by legal professionals. This is theoretically important because a model may perform strongly on one type of legal task while failing on another. Statutory classification, issue spotting, doctrinal application, rule recall, and interpretation should not be collapsed

[7] Neel Guha et al., "LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models," *Advances in Neural Information Processing Systems* 36 (2023): 44123–44279.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

into a single measure of legal intelligence. Evaluation must therefore be task-specific. A system that predicts outcomes successfully may still generate poor reasons; a system skilled at textual entailment may not handle analogy or contested interpretation well.

The LegalBench project further describes its purpose as creating:

> "a common vocabulary"[8]

for lawyers and model developers.

This objective addresses an important interdisciplinary problem. Computer scientists may evaluate systems through accuracy scores while lawyers distinguish among doctrinal interpretation, precedent, evidentiary reasoning, analogical reasoning, and argument construction. Without a shared taxonomy, strong performance on a narrow benchmark can be mistaken for broad legal competence. A common vocabulary encourages more precise evaluation by asking which form of reasoning has actually been tested. It also makes limitations more visible: a model should not be described as capable of "legal reasoning" generally when the evidence supports competence only on selected tasks under controlled conditions.

Statistical models possess an important advantage over purely symbolic approaches: they can process legal language without requiring full formal translation. Yet their internal operation is generally not equivalent to doctrinal reasoning. A model may produce an answer that resembles a legally reasoned conclusion without having represented authority, precedent, burden of proof, or institutional hierarchy in the same manner as a lawyer.

Accordingly, output similarity must not be confused with reasoning identity. A model can arrive at a correct conclusion through statistical pattern recognition while failing to reproduce the legal path by which that conclusion is justified.

2. Open Texture, Semantic Indeterminacy and Context-Dependent Norms

One of the oldest jurisprudential difficulties for computational law is open texture. Legal systems frequently employ expressions such as "reasonable," "substantial," "appropriate," "good faith," "undue," or "proportionate." Such terms are deliberately capable of application across changing circumstances and cannot always be reduced to exhaustive factual conditions in advance.

Recent research defines the difficulty directly:

> "An open-textured norm requires legal decision makers to apply one or more underlying principles."[9]

This distinction separates open-textured standards from more determinate rules. A numerical speed limit can usually be represented through explicit conditions; a duty to drive "reasonably" requires contextual assessment. The latter may depend upon weather, traffic, visibility, road conditions, emergency circumstances, and accepted social expectations. A computational system therefore requires more than lexical interpretation. It must identify which contextual facts matter and how the underlying legal purpose should influence application. Open texture is not merely defective drafting; it often performs an intentional legal function by allowing norms to operate across unforeseen situations.

A later study notes that open-textured terms:

> "are an obstacle to the automatic processing of regulation by computers."[10]

---

[8] Guha et al., "LegalBench"; see also the LegalBench project description.

[9] Clement Guitton et al., "The Challenge of Open-Texture in Law," *Artificial Intelligence and Law* 33 (2025): 405–435, published online 17 April 2024.

[10] Clement Guitton et al., "Identifying Open-Texture in Regulations Using LLMs," *Artificial Intelligence and Law*, published 6 May 2025, https://doi.org/10.1007/s10506-025-09450-0.

This observation should not be understood to mean that automation is impossible whenever legal language is vague. Rather, it identifies where additional legal information becomes necessary. A system may detect that a provision contains an open-textured concept, retrieve interpretive precedents, and present competing constructions. The limitation arises if the machine is expected to transform semantic indeterminacy into a single determinate answer without importing a theory of interpretation. At that point, uncertainty has not been solved; it has merely been hidden inside the model. The most responsible computational system may therefore be one capable of identifying interpretive uncertainty rather than suppressing it.

The European Union AI Act recognises the institutional importance of this problem in judicial contexts, observing that systems assisting courts may affect:

> "the right to an effective remedy and to a fair trial."[11]

The legal significance of this classification extends beyond technical risk. Judicial reasoning is institutionally authoritative because courts are empowered to interpret legal texts, assess facts, distinguish precedents, and justify decisions publicly. Where AI assists those functions, errors or opacity can affect not merely administrative efficiency but the legitimacy of adjudication itself. This explains why judicial AI cannot be assessed solely through accuracy metrics. The legal process includes reason-giving, party participation, appellate review, and institutional responsibility.

Open texture therefore marks a boundary between computation and legal authority. Machines may organise relevant materials, detect patterns, and propose interpretations. The authoritative resolution of genuinely contestable legal meaning remains an institutional act.

3. Factual Uncertainty, Legal Authority and the Problem of Hallucination

Legal reasoning depends not only upon norms but upon accurate facts and valid authorities. A persuasive legal argument built upon a nonexistent case, misquoted statute, or inaccurate procedural rule is defective regardless of how sophisticated its language may appear.

A large empirical study of specialised legal research tools concluded:

> "the providers' claims are overstated."[12]

The study is significant because it evaluated products specifically designed for professional legal research rather than general-purpose chatbots. Retrieval-augmented systems substantially reduced some errors, yet the researchers still found nontrivial rates of hallucination. This shows that access to a legal database does not automatically solve the epistemic problem. Errors may arise from retrieval, interpretation, synthesis, or unsupported inference. For legal practice, the relevant standard cannot therefore be whether a specialised model performs better than a general model; the question is whether its remaining error rate is compatible with the stakes of the task and the verification procedures surrounding its use.

The same research reports that some evaluated systems:

> "hallucinate more than 17% of the time."[13]

This figure should not be universalised to every legal AI product or every type of query. It arose from a specific preregistered evaluation and should be interpreted within that methodological context. Its broader significance is nevertheless considerable: professional branding and retrieval architecture do not justify treating model output as self-authenticating.

---

[11] Regulation (EU) 2024/1689, Artificial Intelligence Act, Recital 61, concerning AI systems used in the administration of justice.

[12] Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho, "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools," *Journal of Empirical Legal Studies* 22, no. 2 (2025): 216–242.

[13] Magesh et al., "Hallucination-Free?," empirical results concerning specialised legal research tools.

In legal reasoning, the reliability of citations is structurally important because precedent operates through an authority hierarchy. A fabricated citation is not merely an informational mistake; it corrupts the chain of legal justification upon which a conclusion is presented. Verification must therefore be built into professional workflow rather than treated as an optional final check.

The United States federal judiciary's interim guidance:

> "Cautions against delegating core judicial functions to AI."[14]

This institutional position draws a critical distinction between assistance and adjudicative delegation. AI may support clerical review, document organisation, research, or procedural analysis, but core judicial functions include authoritative determination of disputes. The caution reflects more than a concern about model accuracy. Judicial decision-making involves responsibility to parties, compliance with law, reasoned justification, ethical obligations, and the exercise of constitutionally conferred authority. Even a system that became statistically more accurate than an individual judge would not automatically inherit the legal office of judging. The institutional identity of the decision-maker remains part of the law.

The hallucination problem therefore reveals a deeper point: legal reasoning is source-sensitive. An answer may be linguistically plausible but legally invalid if it relies upon the wrong jurisdiction, outdated law, dicta instead of holding, repealed legislation, or nonexistent authority.

Section Three: Designing Legally Defensible Computational Decision Systems

The practical question is not whether legal computation should be abandoned because complete automation is difficult. Computational tools can provide substantial value when their function is carefully defined. The challenge is to design systems that complement legal reasoning without concealing uncertainty, displacing authority, or creating an illusion of mechanical certainty.

1. Source Provenance, Verifiability and the Architecture of Legal Evidence

A legally defensible machine system should maintain a distinction between generated language and authoritative source material. Every material legal proposition should ideally remain traceable to the statute, regulation, judgment, record, or other source from which it is derived.

Materials prepared for the U.S. federal judiciary emphasise:

> "Without knowing what the AI was designed and programmed to do, none of these fundamental questions can begin to be answered."[15]

This statement is particularly important for legal evidence and machine-generated analysis. Reliability cannot be assessed in the abstract. A court or reviewer needs to know the task for which a system was designed, the type of data on which it was developed, the assumptions embedded in its architecture, and the relationship between the system's output and the legal issue under consideration. A model suitable for document classification may not be suitable for evaluating credibility; a system designed for one population may not generalise to

---

[14] Administrative Office of the U.S. Courts, "Developing Artificial Intelligence Policies," *Court Operations—Annual Report 2025* (Washington, DC: Administrative Office of the U.S. Courts, 2025).

[15] *Artificial Intelligence and the Courts: Materials for Judges*, April 2024, 20, materials prepared for the federal judiciary and included in Advisory Committee on Evidence Rules materials.

another. Legal admissibility and decision-making therefore require functional specificity rather than generic confidence in "AI."
A source-sensitive architecture should also preserve hierarchy. Constitutional provisions, statutes, binding precedent, persuasive authority, academic commentary, and party submissions do not possess equal legal status. A model that retrieves them all as semantically similar text may produce misleading synthesis unless authority is explicitly represented.
This is an area in which symbolic and statistical approaches can complement one another. Language models can retrieve and summarise large quantities of material, while symbolic layers can encode jurisdiction, court hierarchy, date, precedential status, statutory amendment, and procedural posture.
The objective should not be a system that produces the most confident answer, but one that produces an answer whose evidentiary and doctrinal route can be inspected.

2. Hybrid Human-Machine Reasoning and the Preservation of Legal Responsibility

The strengths of computational models differ. Rule-based systems offer explicit inferential structure but struggle with linguistic flexibility. Case-based systems represent analogy but depend upon factor selection. Statistical systems handle unstructured language but can lack stable doctrinal representation. Hybrid design therefore offers a plausible path forward.
The U.S. judiciary's interim guidance recommends that users:

> "review and independently verify all AI-generated content or output."[16]

Independent verification is more demanding than superficial proofreading. A legally trained reviewer must check whether cited authorities exist, remain valid, belong to the correct jurisdiction, support the proposition stated, and have been characterised accurately. The reviewer must also assess whether relevant adverse authority has been omitted. This requirement places responsibility where professional legal systems traditionally locate it: with the lawyer, judge, or official accountable for the final work. AI can reduce the cost of locating and organising material, but verification prevents computational assistance from becoming unacknowledged delegation.
The European AI Act similarly states:

> "the final decision-making must remain a human-driven activity."[17]

In judicial settings, this principle has institutional significance beyond ordinary human oversight. A judgment is not merely an answer to a legal puzzle; it is an exercise of public authority that must be attributed to a legally constituted decision-maker. Human-driven decision-making therefore requires more than clicking approval on a model-generated draft. The judge must retain substantive control over interpretation, factual assessment, legal reasoning, and the ultimate disposition. The machine may structure possibilities, identify overlooked authorities, or test alternative arguments, but legal responsibility cannot be transferred to a system that does not hold judicial office.
A strong hybrid architecture could allocate different tasks to different computational methods. Formal logic can check consistency among explicit rules. Case-based reasoning can identify comparable precedents. Language models can summarise records and formulate alternative arguments. Human legal decision-makers can evaluate normative conflict, credibility, institutional context, and authoritative interpretation.
Such a model treats artificial intelligence as cognitive infrastructure rather than an autonomous legal subject.

---

16 Administrative Office of the U.S. Courts, *Court Operations—Annual Report 2025*, "Developing Artificial Intelligence Policies," interim guidance concerning independent verification of AI-generated material.
17 Regulation (EU) 2024/1689, Artificial Intelligence Act, Recital 61.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT
JOURNAL OF APPLIED LINGUISTICS AND TESOL
Vol.9. No.1.2026

3. Benchmarking, Epistemic Limits and the Boundary of Full Automation

The evaluation of legal AI requires more than anecdotal demonstrations. Benchmarks are necessary to determine which tasks systems can perform reliably, under what conditions, and with what patterns of failure. Yet benchmarking itself has limitations.

LegalBench contains:

> "162 tasks covering six different types of legal reasoning."[18]

This diversity is valuable because it resists the idea that one accuracy score can represent general legal intelligence. At the same time, benchmark success should not be equated with professional competence. Real litigation includes incomplete records, strategic behaviour, evolving law, jurisdictional complexity, credibility disputes, ethical obligations, and procedural consequences that curated datasets may not reproduce. Benchmark tasks are therefore diagnostic instruments rather than substitutes for institutional validation. A system may perform well on controlled doctrinal questions while failing when authorities conflict, facts are contested, or procedural posture changes the applicable rule.

The open-texture literature also warns that:

> "different individuals might understand the term differently."[19]

This feature limits the idea that every disagreement can be resolved through better training data. Some legal disagreement reflects genuine contestability rather than informational deficiency. Judges may disagree over proportionality, reasonableness, public interest, or the appropriate analogy even when they possess identical factual records and authorities. Computational prediction can estimate how courts are likely to decide, but prediction does not itself establish how they legally ought to decide. The distinction between descriptive regularity and normative justification must therefore remain central to computational jurisprudence.

The federal judiciary has additionally observed that:

> "AI-generated and AI-processed evidence present fundamental questions regarding authentication, reliability, and the establishment of proper foundations."[20]

This observation extends the limits of automation beyond legal research. Machine systems may increasingly generate or transform evidence itself, creating questions about provenance, authenticity, training data, model behaviour, and admissibility. The law must therefore govern not only AI as a decision-maker but AI as a producer of inputs into decision-making. A computational legal system cannot safely reason from evidence whose authenticity is uncertain. Future automation will thus require coordinated standards for evidentiary provenance, model validation, procedural disclosure, and challenge by opposing parties.

Full automation is therefore most plausible in bounded domains where rules are relatively determinate, data are structured, and legal consequences are routine. As interpretive openness, factual contestability, normative conflict, or institutional stakes increase, the case for autonomous machine decision-making becomes substantially weaker.

Findings of the Study

This study yields several principal findings.

First, computational legal reasoning is not a single methodology. Rule-based systems, defeasible logic, case-based reasoning, argumentation frameworks, machine learning, and large language models represent different conceptions of what legal reasoning consists of.

---

[18] Guha et al., "LegalBench."

[19] Guitton et al., "The Challenge of Open-Texture in Law."

[20] Administrative Office of the U.S. Courts, "Objection! How the Federal Rules of Evidence Promote Fair Trials," 5 May 2026.

Second, classical deductive logic alone is insufficient for many legal problems because legal norms frequently include exceptions, rebuttable presumptions, competing rules, and revisable conclusions. Defeasible models better represent this provisional structure.
Third, precedent reasoning cannot be reduced to numerical similarity. Determining which features of earlier cases are legally significant depends upon theories of relevance, legal purpose, and underlying values.
Fourth, formal argumentation models provide powerful tools for representing conflict among reasons, but they do not independently determine which legal values should receive authoritative priority.
Fifth, statistical and language models expand the range of legal tasks that machines can assist with, yet linguistic fluency should not be treated as proof of doctrinal accuracy or legal justification.
Sixth, open-textured concepts establish an important boundary to complete formalisation. Terms such as reasonableness, proportionality, and good faith often require contextual and value-sensitive interpretation rather than fixed factual triggering conditions.
Seventh, legal hallucination is especially serious because law is authority-sensitive. A fabricated or mischaracterised precedent damages the normative foundation of the reasoning, not merely the factual accuracy of a sentence.
Eighth, benchmark performance should be interpreted narrowly. Controlled task success does not establish general competence across real-world legal practice.
Ninth, legal decision-making possesses an institutional dimension that cannot be reduced to outcome prediction. Judicial authority, procedural responsibility, reason-giving, and accountability remain legally significant even if computational systems become increasingly capable.
Tenth, hybrid systems offer the strongest current model for responsible legal AI because they permit different computational techniques to perform bounded functions while preserving authoritative human legal responsibility.

Recommendations

1. Computational legal systems should be evaluated according to specific legal tasks rather than through broad claims of general “legal intelligence.”
2. Rule-based systems should expressly represent exceptions, priorities, and defeasible conditions where the relevant legal domain requires them.
3. Case-based systems should explain which factual factors make a precedent similar or distinguishable rather than presenting similarity scores alone.
4. Systems used for legal research should preserve source provenance, court hierarchy, jurisdiction, date, and precedential status.
5. Legal AI interfaces should distinguish between retrieved authority, model-generated synthesis, and inferential conclusions.
6. Open-textured legal terms should trigger interpretive warnings or alternative analyses rather than automatic presentation of a single determinate answer.
7. Professional legal users should independently verify citations, quotations, holdings, statutory provisions, and procedural rules generated or summarised by AI.
8. Judicial institutions should define categories of tasks that may be computationally assisted and distinguish them from functions that must remain institutionally human.
9. Benchmarking programmes should evaluate not only final-answer accuracy but citation validity, reasoning consistency, sensitivity to jurisdiction, treatment of adverse authority, and uncertainty calibration.

10. Developers should test legal systems on false-premise questions to determine whether models challenge incorrect user assumptions rather than simply elaborate upon them.
11. Legal systems should record sufficient information about model version, source retrieval, prompts, relevant outputs, and human modifications where AI materially contributes to consequential decisions.
12. Hybrid architectures combining symbolic reasoning, authoritative retrieval, case comparison, and language models should be preferred where no single computational paradigm adequately represents the legal task.
13. Courts should maintain procedural mechanisms through which parties can challenge machine-generated or machine-processed evidence.
14. Legal education should include computational literacy so that lawyers understand both the capabilities and limitations of machine reasoning.
15. Research should distinguish predictive accuracy from normative justification; knowing what courts are likely to decide is not identical to determining what the law requires.
16. Fully automated legal decision-making should be confined, if used at all, to narrowly defined contexts where the applicable norms, factual inputs, review mechanisms, and legal consequences are sufficiently determinate.

Conclusion

Computational models of legal reasoning have substantially deepened understanding of both artificial intelligence and law. Their importance lies not only in their technological applications but in the jurisprudential questions they force legal theory to confront.

Rule-based systems demonstrate that some legal norms can be represented as structured conditions and consequences. Defeasible models reveal that legal reasoning often remains revisable in the face of exceptions and stronger competing reasons. Case-based systems show that precedent depends upon legally significant comparison rather than superficial resemblance. Argumentation frameworks capture the adversarial structure of legal justification. Value-based models reveal that many disputes turn upon normative priorities. Machine-learning systems demonstrate that patterns can be extracted from legal corpora without complete manual formalisation. Large language models extend this capacity into natural-language interaction and document generation.

Yet each advance also exposes a corresponding limitation.

Logic requires prior specification of legal meaning. Case-based reasoning requires a theory of relevant similarity. Value-sensitive reasoning cannot autonomously legitimise its value hierarchy. Statistical learning can reproduce patterns without establishing normative justification. Language models can produce fluent legal text while remaining vulnerable to hallucinated authority. Benchmarks can measure bounded capabilities without reproducing the institutional complexity of real adjudication.

The central mistake would therefore be to treat legal reasoning as a single computational problem awaiting one sufficiently powerful model. Law consists of multiple forms of reasoning operating within institutions that assign authority, create procedure, structure evidence, and impose responsibility.

Machine decision-making can assist this structure most effectively when its role is explicit and bounded. It can retrieve, classify, compare, test, map, predict, and draft. It can expose inconsistent arguments, locate analogous cases, identify possible interpretations, and organise large records. These capacities can substantially improve legal work.

What computation does not automatically supply is legal authority.

A judicial decision is not legally binding because it represents the statistically most probable answer. It is binding because it is made by an institution authorised to determine the dispute through legally prescribed procedures and to provide reasons that can be reviewed within the legal order.

The future of computational legal reasoning should therefore be understood as augmentation rather than institutional substitution. The strongest systems will not attempt to conceal legal uncertainty behind mathematical confidence. They will identify where rules are determinate, where precedents conflict, where values diverge, where sources are uncertain, and where human legal judgment must begin.

Selected Sources and References

Ashley, Kevin D. "Case-Based Reasoning and Its Implications for Legal Expert Systems." *Artificial Intelligence and Law* 1 (1992): 113–208.

Ashley, Kevin D., and Edwina L. Rissland. "Law, Learning and Representation." *Artificial Intelligence* 150, nos. 1–2 (2003): 17–58.

Bench-Capon, Trevor J. M. "Persuasion in Practical Argument Using Value-Based Argumentation Frameworks." *Journal of Logic and Computation* 13, no. 3 (2003): 429–448.

Bench-Capon, Trevor J. M., and Giovanni Sartor. "A Model of Legal Reasoning with Cases Incorporating Theories and Values." *Artificial Intelligence* 150, nos. 1–2 (2003): 97–143.

Dung, Phan Minh. "On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games." *Artificial Intelligence* 77, no. 2 (1995): 321–357.

European Parliament and Council. *Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence*.

Guha, Neel, et al. "LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models." *Advances in Neural Information Processing Systems* 36 (2023): 44123–44279.

Guitton, Clement, et al. "Identifying Open-Texture in Regulations Using LLMs." *Artificial Intelligence and Law*. Published 6 May 2025. https://doi.org/10.1007/s10506-025-09450-0.

Guitton, Clement, et al. "The Challenge of Open-Texture in Law." *Artificial Intelligence and Law* 33 (2025): 405–435.

Magesh, Varun, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning, and Daniel E. Ho. "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools." *Journal of Empirical Legal Studies* 22, no. 2 (2025): 216–242.

Prakken, Henry. *Logical Tools for Modelling Legal Argument: A Study of Defeasible Reasoning in Law*. Dordrecht: Springer, 1997.

Prakken, Henry, and Giovanni Sartor. "Argument-Based Extended Logic Programming with Defeasible Priorities." *Journal of Applied Non-Classical Logics* 7, nos. 1–2 (1997): 25–75.

United States Courts. *Court Operations—Annual Report 2025*. Washington, DC: Administrative Office of the U.S. Courts, 2025.

United States Courts. *Artificial Intelligence and the Courts: Materials for Judges*. Advisory Committee on Evidence Rules materials, April 2024.

United States Courts. "Objection! How the Federal Rules of Evidence Promote Fair Trials." 5 May 2026.