When AI Should Refuse to Decide: Formal Irresolution and the Ethics of Legal Indeterminacy in LLM-Based Reasoning for Biodiversity Law

Rohan Nanda[1,2,*], Henrique Marcos[3] and Julia Schutz Veiga[4]

[1]Law and Tech Lab, Maastricht University, The Netherlands
[2]Department of Advanced Computing Sciences, Maastricht University, The Netherlands
[3]Faculty of Law, Maastricht University, The Netherlands
[4]NOVA School of Law; Ocean Voices, Nippon Foundation, University of Edinburgh

Abstract

Large language models asked a legal question will always produce an answer. In fragmented treaty regimes, where overlapping international instruments create genuinely contested obligations, this is an ethical problem rather than a technical achievement. This paper argues that the capacity for formal irresolution—the ability of a system to identify and communicate that the applicable law does not yield a determinate answer—is an ethical requirement for AI systems operating in high-stakes legal domains. We ground this argument in empirical findings from LLM-based legal information retrieval over a Hohfeld-Structured Normative Knowledge Base of 377 provisions from three marine biodiversity treaties, where LLMs achieve modest precision and middling recall, systematically missing dispersed obligations. We distinguish formal irresolution from technical approaches to epistemic uncertainty in machine learning, arguing that legal indeterminacy is a structural property of fragmented regimes rather than an epistemic gap that better models could close. We illustrate the proposal through a worked scenario demonstrating three distinct output types: determinate answers, structured presentations of competing obligations, and acknowledgements of incompleteness. Formal reasoning frameworks grounded in legal theory can distinguish between settled and contested legal questions, preserving the interpretive authority of human legal actors and preventing what we characterise as a form of hermeneutical injustice through technical architecture.

Keywords

Legal AI ethics, formal irresolution, legal indeterminacy, treaty fragmentation, Hohfeldian analysis, LLM reasoning

1. Introduction

The international legal framework for biodiversity is fragmented across multiple treaty regimes, each establishing its own set of obligations for states and other actors [1, 2]. Three treaties govern the access and use of marine genetic resources across overlapping terrain, each addressing distinct actors and establishing separate procedural obligations: the Biodiversity Beyond National Jurisdiction (BBNJ) Agreement, the Convention on Biological Diversity (CBD), and the Nagoya Protocol [3]. The obligations overlap, the terminology is inconsistent across instruments, and in some scenarios the treaties point in different directions. A lawyer who is asked which rules apply to a given marine bioprospecting scenario would say: it depends, and then explain on what. A large language model asked the same question will produce a fluent, confident answer that picks one reading—delivering conclusions without arguments [4].

But too much confidence can be a problem. Recent empirical work has shown that LLMs applied to marine policy systematically generate outputs that favour the perspectives of developed-country negotiating positions while underweighting viewpoints from developing states and Small Island Developing

Joint Workshop on Statistics and Knowledge Integration for Logic, Learning, Ethical Decisions, and LLMs (SKILLED-LLMs 2026), co-located with FLoC 2026, Lisbon, Portugal
r.nanda@maastrichtuniversity.nl (R. Nanda); h.jbmarcos@maastrichtuniversity.nl (H. Marcos); julia.schutz@novalaw.unl.pt (J. S. Veiga)
https://www.maastrichtuniversity.nl/nl/r-nanda (R. Nanda); https://henriquemarcos.com/ (H. Marcos); https://orcid.org/0000-0002-2504-7353 (J. S. Veiga)
0000-0001-7124-4799 (R. Nanda); 0000-0003-0259-5530 (H. Marcos); 0000-0002-2504-7353 (J. S. Veiga)

States [5]. When such systems resolve genuinely contested legal questions without signalling that the resolution involves an interpretive choice, they do not merely provide imperfect information. They perform a form of what Fricker [6] has called epistemic injustice: they exclude certain actors from the interpretive process by rendering alternative readings invisible. The capacity to formally recognise and communicate legal indeterminacy is therefore not a limitation of an AI system but an ethical obligation, particularly when the system's outputs influence the legal positions of actors who lack the resources to independently verify those outputs [7, 8].

The rest of the paper is organised as follows. Section 2 characterises the problem of forced resolution and distinguishes legal indeterminacy from epistemic uncertainty. Section 3 grounds the argument in empirical findings from LLM-based retrieval over a Hohfeld-Structured Normative Knowledge Base. Section 4 develops formal irresolution as an ethical obligation. Section 5 illustrates the proposal through a worked example. Section 6 considers regulatory alignment. Section 7 presents the conclusion.

2. The Problem of Forced Resolution

LLMs generate text by selecting tokens that are probable under their learned distributions [9]. This architecture has no built-in mechanism for reliably distinguishing between questions where the applicable law clearly supports one answer and questions where the law is genuinely contested. The model will produce an answer in both cases, and the answer will appear equally confident in both. Empirical studies have shown that LLMs invent false information about case law, struggle to predict when they are doing so, and produce legally incoherent reasoning with the same fluency they apply to well-supported conclusions [10, 7], a problem that persists even in retrieval-augmented legal research tools specifically designed to mitigate it [11]. More broadly, the problem of hallucination in natural language generation extends well beyond factual error to encompass unfaithful representation of source material [12].

In fragmented treaty regimes, however, the problem is more subtle than hallucination. The model may correctly identify relevant provisions and accurately state what each says, yet still produce an ethically problematic output by synthesising those provisions into a single determinate conclusion where the law does not warrant one. Legal language is frequently open-textured by design, permitting multiple plausible readings depending on how one construes the operative terms, identifies the relevant actors, or situates a provision within the broader treaty architecture [13]. This indeterminacy is not a deficiency of the legal regime but a structural feature of international law that enables flexibility in application across diverse circumstances. An AI system that collapses this indeterminacy into a single answer transforms a deliberate feature of legal design into an invisible interpretive choice.

The problem described here must be distinguished from the better-studied problem of epistemic uncertainty in machine learning, which arises when a model lacks sufficient knowledge to make a reliable prediction and can be addressed by providing the model with more data [14]. Abstaining classifiers respond to such uncertainty by allowing a model to withhold predictions on inputs where it cannot produce a reliable output [15]. A systematic evaluation of LLM abstention across 20 frontier models found that these models still struggle to abstain from answering questions that are unanswerable or ill-posed [16]. Recent work has also proposed iterative prompting strategies that estimate epistemic uncertainty directly from LLM outputs, detecting cases where the model lacks the knowledge to produce a reliable response [17]. These approaches treat uncertainty as a property of the model, arising from insufficient knowledge that can be reduced with additional data [14], detected through prompting [17], or managed by withholding unreliable predictions [15].

Legal indeterminacy is structurally different. When the CBD, the Nagoya Protocol, and the BBNJ Agreement impose overlapping but incompatible requirements on the same activity, the indeterminacy is not a gap in the model's knowledge that could be closed with better training data. It is a property of the legal regime itself. A system that treats legal indeterminacy as epistemic uncertainty mischaracterises the problem, because it frames the indeterminacy as a gap in the model's knowledge that could in principle be closed with additional information. Formal irresolution, by contrast, communicates that the law itself does not resolve the question, and that the resolution requires an interpretive choice that

properly belongs to human legal actors [18].

3. Empirical Grounding

Recent work on LLM-based legal information retrieval provides empirical evidence for this concern. A Hohfeld-Structured Normative Knowledge Base (HSNKB) was constructed from 377 expert-annotated provisions across the CBD, the Nagoya Protocol, and the BBNJ Agreement [19]. Hohfeld's analytical framework [20] disaggregates vague terms like "rights" and "duties" into eight precise relational categories (claim-right, duty, liberty, no-right, power, liability, immunity, disability), each defined as a directed relation between two actors. Each provision in the HSNKB was mapped into a relational triple of actor-holder, Hohfeldian legal position, and actor-affected, making normative relationships machine-readable while preserving the legal precision that the domain requires. To illustrate: BBNJ Agreement Article 14(3), which requires that marine genetic resources and digital sequence information subject to utilisation be deposited in publicly accessible repositories, is encoded as a duty held by Contracting Parties, with stakeholders as the affected actors holding a correlative claim-right. This encoding makes explicit the relational structure that natural-language treaty text leaves implicit: who bears the obligation, against whom it holds, and what legal position the counterparty occupies.

Four LLMs were evaluated on retrieving relevant HSNKB rows to answer 15 fact-pattern legal questions. The best-performing model achieved precision of 0.77 and recall of 0.53 [19]. Two observations carry ethical significance. The Hohfeldian structure constrains the output space to genuine normative content, ensuring that every retrieved provision corresponds to an actual treaty entry in the knowledge base. However, the middling recall means the system systematically missed relevant provisions, particularly conditional and multi-actor duties dispersed across treaties. When a system informs a state actor that five provisions apply but silently omits five others, the actor may fail to comply with obligations it was never made aware of.

The error analysis revealed that false negatives mainly include benefit-sharing obligations triggered by overlapping conditions involving access to resources, traditional knowledge, and digital sequence information. These are precisely the provisions where the treaties interact most complexly, and where the question of which obligations apply depends on interpretive choices about jurisdictional scope, resource classification, and the relationship between instruments.

Treaty provisions frequently contain explicit cross-references to other articles, both within the same instrument and across treaties. For instance, benefit-sharing obligations under the Nagoya Protocol reference specific articles of the CBD, and monetary benefit-sharing provisions under the BBNJ Agreement reference the article establishing the special fund. In future, the HSNKB could be supplemented with these cross-reference links, so that when the system retrieves a provision that references another, it can check whether the referenced provision was also retrieved. If it was not, the system has a structural signal of potential incompleteness. This does not guarantee full coverage, but it provides a principled basis for flagging gaps rather than presenting partial retrieval as a complete answer.

4. Formal Irresolution as Ethical Obligation

We propose that AI systems operating in fragmented legal domains should be designed to produce three distinct types of output rather than collapsing all queries into determinate answers. The first is a *determinate answer*, appropriate when the retrieved provisions clearly and consistently support a single legal conclusion. The second is a *structured presentation of competing obligations*, appropriate when provisions from different treaties impose different requirements on the same actor for the same activity, accompanied by an explicit indication that the law is contested. The third is an *acknowledgement of incompleteness*, appropriate when the system's recall limitations mean that relevant provisions may have been missed.

Structured knowledge representations grounded in legal theory make this distinction possible. The Hohfeldian relational structure of the HSNKB allows the system to identify where obligations overlap (the same actor bears duties under multiple treaties), where they conflict (the duties impose incompatible requirements), and where they leave gaps (no treaty addresses a particular aspect of the scenario). When competing interpretations are preserved in the structured representation rather than collapsed into a single narrative, human legal actors retain the ability to examine the alternatives and exercise their own interpretive judgment.

This argument carries particular force when considered alongside the distributional consequences of opaque legal AI. The governance complexity of marine genetic resources falls hardest on actors from biodiversity-rich regions who lack the institutional capacity to navigate overlapping treaty obligations independently [5, 2]. Small Island Developing States and Indigenous Peoples and Local Communities are directly affected by marine biodiversity governance yet often lack the specialised legal resources to assess whether an AI-generated legal answer reflects settled law or a contestable interpretation. For these actors, a system that silently resolves ambiguity forecloses interpretive possibilities that might serve their interests, without making that foreclosure visible. This constitutes what Fricker [6] terms hermeneutical injustice: a structural gap in shared interpretive resources that prevents certain forms of knowledge or experience from being made intelligible within a given framework. Hermeneutical injustice differs from testimonial injustice (where a speaker's credibility is deflated) in that it operates at an earlier stage: before an alternative legal reading can even be evaluated, the interpretive categories required to make it intelligible are absent. Critically, hermeneutical injustice involves no individual culprit—it is a purely structural phenomenon produced by the framework itself. When an LLM collapses competing treaty interpretations into a single answer, the system's output categories lack the resources to make alternative readings visible. Actors whose interests would be served by those readings are excluded not from information as such, but from the interpretive practices through which legal meaning is constructed [21].

5. Worked Example: Three Outputs for One Scenario

To illustrate how formal irresolution operates in practice, consider a simplified scenario drawn from cross-treaty analysis using the HSNKB. A research consortium collects marine genetic resources from within a coastal state's exclusive economic zone and from an adjacent area beyond national jurisdiction, relying on traditional knowledge provided by an Indigenous community. The resulting digital sequence information is deposited in open-access repositories. A commercial entity later accesses the data and develops a product.

A conventional LLM asked which benefit-sharing obligations apply would synthesise the applicable provisions into a narrative answer, presenting a single reading as if the law were settled. The HSNKB enables three distinct outputs corresponding to the typology proposed in Section 4.

5.1. Determinate answer

The BBNJ Agreement imposes clear duties on Contracting Parties to notify the Clearing-House Mechanism before in situ collection (Article 12(1)–(2)) and to ensure that marine genetic resources and digital sequence information subject to utilisation are deposited in publicly accessible repositories (Article 14(3)). The Hohfeldian encoding identifies these as duties held by Contracting Parties with correlative claim-rights held by stakeholders. There is no competing interpretation across instruments; the system can deliver a determinate answer.

5.2. Structured presentation of competing obligations

The Nagoya Protocol requires that benefits from the utilisation of traditional knowledge be shared with the communities holding that knowledge (Article 5(5)), but the territorial scope of both the CBD and the Nagoya Protocol excludes traditional knowledge related to marine genetic resources from areas

beyond national jurisdiction [19]. The BBNJ Agreement addresses areas beyond national jurisdiction but leaves the relationship between its benefit-sharing provisions and the Nagoya Protocol's traditional knowledge provisions unresolved. The HSNKB surfaces both sets of obligations as applicable, with an explicit indication that their interaction is contested. The system presents competing obligations rather than selecting one reading.

5.3. Acknowledgement of incompleteness

The retrieval experiment showed recall of 0.53, with false negatives clustering around conditional and multi-actor duties dispersed across treaties [19]. In a production setting, a system cannot compute its own recall without access to expert gold answers. However, two design choices can address this limitation. First, where treaty provisions contain explicit cross-references to other articles, the system can check whether the referenced provisions were also retrieved and flag missing links as potential gaps. Second, and more fundamentally, the system can be designed to present its output as a set of retrieved provisions rather than as a definitive account of the applicable law, communicating by default that retrieval in fragmented treaty regimes is inherently partial.

This tripartite output structure preserves the interpretive authority of human legal actors at each stage. It does not require the system to resolve what the law does not resolve, only to communicate that its retrieval may not be exhaustive.

6. Regulatory Alignment

The EU AI Act imposes transparency requirements (Article 13) and human oversight obligations (Article 14) on high-risk AI systems [22]. Formal irresolution satisfies both requirements in a substantive rather than merely procedural sense. Transparency is served because the system makes explicit which provisions it retrieved, how they relate to each other, and where they conflict. Human oversight becomes meaningful because the system identifies the specific cases that require human interpretive judgment, rather than presenting all outputs with equal confidence. An AI system that always produces a determinate legal answer renders human oversight procedurally available but substantively hollow, because the human reviewer has no principled basis for distinguishing outputs that warrant scrutiny from those that do not. The requirement for formal irresolution also aligns with the principle that the application of legal rules must remain contestable rather than rendered opaque by algorithmic processes [23, 24]. As Hildebrandt has argued, smart technologies must be designed to enable legal protection rather than foreclose it [24]. A system that always produces a determinate answer removes the visibility of the interpretive choice that would allow affected parties to challenge its outputs.

7. Conclusion

Although we have developed the argument through marine biodiversity law, the principle of formal irresolution holds wherever law is fragmented across overlapping instruments. International law is the clearest case. The international legal order is a system whose rulesets are frequently inconsistent, yet from which legal subjects are nonetheless expected to derive a consistent set of legal consequences for a given case [25]. Moving from an inconsistent body of rules to a consistent set of consequences is interpretive work, accomplished through reasoning with principles, and it belongs to legal actors rather than to the instruments themselves [26]. A system that resolves competing treaty obligations into a single answer performs that interpretive work silently and presents the result as though the rules had decided the question. Formal irresolution keeps the inconsistency of the rules visible, so that the interpretive step through which consistency is reached remains open to the actors who are entitled to take it. The same structure recurs across other fragmented domains of international and transnational law, including the law of investment protection and the interaction of trade and environmental regimes, wherever obligations drawn from separate instruments bear on a single course of conduct.

The property that licenses formal irresolution belongs to the represented domain. The indeterminacy is structural. It is a feature of the body of norms being reasoned over, of the kind we distinguished in Section 2 from the epistemic uncertainty. Any knowledge base that encodes content with this property raises the same demand. A representation of conflicting clinical guidelines, of safety standards that diverge across jurisdictions, or of expert annotations over which qualified annotators legitimately disagree does not resolve to a single answer. A reasoning system built over such a representation can surface the conflict or conceal it behind a fluent resolution. The Hohfeldian knowledge base studied here is one instance of a general problem for knowledge representation and reasoning. Where indeterminacy is epistemic, the techniques discussed in Section 2 can in principle address the problem. Where it is structural, the system must be able to communicate that the knowledge it represents does not decide the question.

The capacity for formal irresolution is an ethical requirement for AI systems operating in fragmented legal domains, and the same requirement follows wherever a reasoning system is built over a contested body of knowledge. LLMs on their own struggle to distinguish answerable from unanswerable questions and frequently fail to abstain [16]. Formal irresolution does not attempt to solve this problem within the model itself. Instead, it imposes external structure through formal reasoning frameworks grounded in legal theory, such as Hohfeldian relational analysis applied to structured normative knowledge bases [20, 19]. These frameworks can surface competing obligations and signal incomplete retrieval regardless of the model's own confidence. This preserves the interpretive authority of human legal actors and ensures that the contestability of international law remains visible to those whose rights and obligations depend on it. Designing AI systems that recognise when they cannot decide is not a concession of weakness but a commitment to epistemic honesty in domains where forced resolution carries real consequences for real actors.

Declaration on Generative AI

During preparation of this work, the authors used Claude for sentence polishing and rephrasing. After using it, the authors reviewed and edited the content and take full responsibility for the content of the publication.

References

[1] G. Ardito, G. Andreone, M. Rovere, Overlapping and fragmentation in the protection and conservation of the marine environment in areas beyond national jurisdiction, Frontiers in Marine Science Volume 9 (2023).

[2] E. Morgera, Fair and equitable benefit-sharing in a new international instrument on marine biodiversity: A principled approach towards partnership building?, Maritime Safety and Security Law Journal 5 (2018) 48–77.

[3] R. Rayfuse, A. Jaeckel, N. Klein, Chapter 1: International marine environmental law in the 21st century, in: Research Handbook on International Marine Environmental Law, Edward Elgar Publishing, Cheltenham, UK, 2023, pp. 2 – 32.

[4] H. Marcos, Can large language models apply the law?, AI & Society 40 (2025) 3605–3614. doi:10.1007/s00146-024-02105-9.

[5] M. Ziegler, S. Lothian, B. O'Neill, R. Anderson, Y. Ota, AI language models could both help and harm equity in marine policymaking, npj Ocean Sustainability 4 (2025) 32. doi:10.1038/s44183-025-00132-7.

[6] M. Fricker, Epistemic Injustice: Power and the Ethics of Knowing, Oxford University Press, Oxford, 2007.

[7] S. Wachter, B. Mittelstadt, C. Russell, Do large language models have a legal duty to tell the truth?, Royal Society Open Science 11 (2024) 240197.

[8] H. Westermann, K. Benyekhlef, Justicebot: A methodology for building augmented intelligence tools for laypeople to increase access to justice, in: Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law, 2023, pp. 351–360.
[9] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, J. Gao, Large language models: A survey, arXiv preprint arXiv:2402.06196 (2024).
[10] M. Dahl, V. Magesh, M. Suzgun, D. E. Ho, Large legal fictions: Profiling legal hallucinations in large language models, Journal of Legal Analysis 16 (2024) 64–93. doi:10.1093/jla/laae003.
[11] V. Magesh, F. Surani, M. Dahl, M. Suzgun, C. D. Manning, D. E. Ho, Hallucination-Free? Assessing the reliability of leading AI legal research tools, Journal of Empirical Legal Studies 22 (2025) 216–242.
[12] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung, Survey of hallucination in natural language generation, ACM Computing Surveys 55 (2023) 1–38. doi:10.1145/3571730.
[13] R. Poscher, Ambiguity and vagueness in legal interpretation, in: L. M. Solan, P. M. Tiersma (Eds.), The Oxford Handbook of Language and Law, Oxford University Press, 2012, pp. 128–145. doi:10.1093/oxfordhb/9780199572120.013.0010.
[14] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Machine learning 110 (2021) 457–506.
[15] Y. J. Choe, A. Gangrade, A. Ramdas, Counterfactually comparing abstaining classifiers, in: Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23, Curran Associates Inc., Red Hook, NY, USA, 2023.
[16] P. Kirichenko, M. Ibrahim, K. Chaudhuri, S. J. Bell, Abstentionbench: Reasoning LLMs fail on unanswerable questions, in: Advances in Neural Information Processing Systems, volume 38, Curran Associates, Inc., 2025.
[17] Y. A. Yadkori, I. Kuzborskij, A. György, C. Szepesvári, To believe or not to believe your LLM: iterative prompting for estimating epistemic uncertainty, in: Proceedings of the 38th International Conference on Neural Information Processing Systems, NIPS '24, Curran Associates Inc., Red Hook, NY, USA, 2024.
[18] J. Grimmelmann, B. Sobel, D. Stein, Generative misinterpretation, Harvard Journal on Legislation 63 (2026).
[19] R. Nanda, H. Marcos, H. Westermann, J. S. Veiga, A Hohfeldian knowledge base for LLM-assisted legal information retrieval in marine biodiversity law, in: Legal Knowledge and Information Systems, IOS Press, 2025, pp. 318–323. doi:10.3233/FAIA251604.
[20] W. N. Hohfeld, Some fundamental legal conceptions as applied in judicial reasoning, The Yale Law Journal 23 (1913) 16–59. doi:10.2307/785533.
[21] K. Whyte, Too late for indigenous climate justice: Ecological and relational tipping points, WIREs Climate Change 11 (2020) e603. doi:10.1002/wcc.603.
[22] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) no 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), 2024.
[23] K. Yeung, Algorithmic regulation: A critical interrogation, Regulation & Governance 12 (2018) 505–523. doi:10.1111/rego.12158.
[24] M. Hildebrandt, Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology, Edward Elgar Publishing, 2015. doi:10.4337/9781849808774.
[25] H. Marcos, Two kinds of systemic consistency in international law, European Journal of Legal Studies 15 (2023) 65–83.
[26] H. Marcos, From fragmented legal order to globalised legal system: Towards a framework of general principles for the consistency of international law, Athena: Critical Inquiries in Law, Philosophy and Globalization 3 (2023) 90–124. doi:10.6092/issn.2724-6299/17223.