Vagnoni S, Olivier L, Krumsick C, Palmirani M, 'CaseConnect: Enhancing Access to Justice through Cross-Jurisdictional Legal Case Retrieval' (2025) Special Issue 2 EJPLT.
https://doi.org/10.57230/EJPLT25SIJSVLOCKMP
Corresponding author. E-mail address: simone.vagnoni3@unibo.it (S. Vagnoni)
Zamboni street no. 27/29, Bologna (IT)
EJPLT
ISSN 2704-8012



CaseConnect: Enhancing Access to Justice through Cross-Jurisdictional Legal Case Retrieval

CaseConnect: migliorare l'accesso alla giustizia attraverso la ricerca trans-giurisprudenziale di casi giuridici

SIMONE VAGNONI
Ph.D. (c) in Artificial Intelligence
Universidad Politécnica de Madrid
Ph.D. (c) in Law, Science and Technology
Università di Bologna

LOU-ANNE OLIVIER
LL.M. Candidate in Legal Studies
Università di Bologna

COCO KRUMSICK
LL.M. Candidate in Legal Studies
Università di Bologna

MONICA PALMIRANI
Full Professor of Legal Informatics and ICT Law
Università di Bologna

Abstract

Access to justice across jurisdictions remains a critical challenge for legal professionals navigating heterogeneous legal systems and linguistic diversity. This article presents CaseConnect, a legal information retrieval system designed to facilitate cross-jurisdictional case comparison while addressing fundamental concerns of data protection, AI regulation, and access to justice. We demonstrate how a hybrid methodology integrating manual legal expertise with automated text processing can preserve the structural and semantic integrity of legal texts while complying with the General Data Protection Regulation (GDPR) and the EU Artificial Intelligence Act. By implementing legally-supervised document sectioning and multilingual semantic embeddings, the system enables efficient identification of factually similar cases across different legal traditions. This approach can contribute to access to justice by democratising comparative legal research capabilities, potentially supporting consistency in human rights adjudication, and reducing barriers to cross-jurisdictional legal knowledge.

We thank the Council of Europe for the opportunity to develop this project during the Digital Future of Justice 2.0 Hackathon, held 15-17 November 2024 in Bologna, where CaseConnect was awarded second place. We are grateful to Professor Monica Palmirani, Salvatore Sapienza, and Michele Corazza for their guidance throughout the competition, and to Leonardo Zilli for contributing to the creation of the prototype.

© The author(s) 2025, published by Suor Orsola Benincasa Università Editrice.
This contribution is licensed under a Creative Common Attribution 4.0 International Licence CC-BY-NC-ND, all the details on the license are available at:
https://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract
L'accesso alla giustizia tra diverse giurisdizioni rimane una sfida critica per i professionisti del diritto che si confrontano con sistemi giuridici eterogenei e con la diversità linguistica. Il presente articolo introduce CaseConnect, un sistema di recupero di informazioni giuridiche progettato per facilitare il confronto tra casi giurisprudenziali transgiurisdizionali, affrontando al contempo questioni fondamentali relative alla protezione dei dati, alla regolamentazione dell'intelligenza artificiale e all'accesso alla giustizia. Dimostriamo come una metodologia ibrida che integra competenze giuridiche manuali e tecniche automatizzate di elaborazione del testo possa preservare l'integrità strutturale e semantica dei testi giuridici, nel rispetto del Regolamento generale sulla protezione dei dati (GDPR) e dell'AI Act dell'Unione Europea. Attraverso l'implementazione di una segmentazione documentale giuridicamente supervisionata e di embedding semantici multilingue, il sistema consente l'identificazione efficiente di casi fattualmente simili tra differenti tradizioni giuridiche. Questo approccio può contribuire al miglioramento dell'accesso alla giustizia, democratizzando le capacità di ricerca giuridica comparata, potenzialmente sostenendo la coerenza nella giurisprudenza in materia di diritti umani e riducendo le barriere all'accesso alla conoscenza giuridica transgiurisdizionale.

Keywords: Legal Information Retrieval; Cross-Jurisdictional Comparison; Access to Justice; GDPR; EU AI Act; Semantic Embeddings; Human Rights Adjudication

Summary: 1. Introduction. – 2. The Legal Challenge of Cross-Jurisdictional Research. – 2.1 Access to Justice and Legal Knowledge Asymmetries. – 2.2 The Structural Complexity of Legal Texts. – 2.3 The Limits of Automation in Legal Contexts. – 3. A Hybrid Approach: Integrating Legal Expertise with Technology. – 3.1 Legal Supervision of Technical Implementation. – 3.2 Multilingual Semantic Retrieval. – 4. Privacy, Data Protection and AI Regulation. – 4.1 GDPR Compliance Framework. – 4.2 EU Artificial Intelligence Act Considerations. – 5. Enhancing Access to Justice across Jurisdictions. – 5.1 Democratising Comparative Legal Research. – 5.2 Supporting Consistency in Human Rights Adjudication. – 5.3 Limitations and Professional Responsibility. – 6. Technical Implementation and Future Directions. – 6.1 System Architecture. – 6.2 Future Development Priorities. – 7. Conclusion.

1. Introduction.

The globalisation of legal practice has created an urgent need for tools that enhance access to justice across jurisdictional boundaries.[1] Legal professionals confronting human rights cases, environmental litigation, or transnational disputes must navigate a landscape characterised by linguistic diversity, varying legal traditions, and heterogeneous document structures. Traditional keyword-based search systems prove inadequate for this task, failing to capture

[1] T Lawal, 'Rethinking Jurisdictional Barriers to Practising Law Abroad: A Soft Technological Deterministic Approach' (2024) 6(1) Law Tech Hum 21.

semantic similarity across languages while neglecting the structural organisation that defines legal reasoning.[2]

The challenge of cross-jurisdictional legal comparison extends beyond technical retrieval to fundamental questions of legal practice, professional responsibility, and equality before the law. When lawyers cannot efficiently locate relevant foreign precedents, when smaller legal organisations lack access to sophisticated research tools, and when courts lack visibility into how similar factual scenarios are addressed across jurisdictions, the administration of justice suffers. These barriers, rooted in asymmetries of legal knowledge,[3,4] are exacerbated by linguistic and structural heterogeneity across jurisdictions. Digital tools can bridge these gaps, as recognised in EU policy assessments,[5] underscoring the need for advanced retrieval technologies that respect both legal complexity and regulatory frameworks. Such barriers disproportionately affect individuals and organisations with limited resources, [6, 7] creating inequalities in access to legal knowledge that undermine the rule of law.

This article presents CaseConnect, a legal information retrieval system developed to address these access-to-justice challenges whilst respecting the regulatory frameworks governing data protection and artificial intelligence in the European Union. Our approach recognises that legal texts demand precision, contextual understanding, and interpretive nuance that cannot be fully captured through automation alone, whilst simultaneously acknowledging that manual analysis cannot scale to the volume of legal documents requiring examination across multiple jurisdictions.

We make three principal contributions. First, we present a hybrid methodology that combines manual legal analysis of structural patterns with automated implementation across large datasets, demonstrating how legal expertise can guide technological development. Second, we provide a compliance framework addressing GDPR and EU AI Act requirements from the design stage, establishing a model for responsible legal AI development that prioritises transparency, data protection, and human oversight. Third, we demonstrate how such systems can contribute to access to justice by democratising comparative legal research capabilities, enabling efficient identification of similar cases across jurisdictions, and potentially supporting consistency in human rights adjudication.

---

[2] C Kim, J Lee and W Hwang, 'LegalSearchLM: Rethinking Legal Case Retrieval as Legal Elements Generation' (2025) arXiv:2505.23832.

[3] M Cappelletti and B Garth, 'Access to Justice: The Newest Wave in the Worldwide Movement to Make Rights Effective' (1978) 27 Buff L Rev 181. <https://digitalcommons.law.buffalo.edu/buffalolawreview/vol27/iss2/1/>.

[4] European Parliament, 'Effective Access to Justice' (IPOL Study 2017). <https://www.europarl.europa.eu/RegData/etudes/STUD/2017/596818/IPOL_STU(2017)596818_EN.pdf>.

[5] ibid.

[6] R Smith, 'The Future of Law: Technology, Innovation and Access to Justice in Legal Services Around the World' in SD de Souza and M Spohr (eds), *Technology, Innovation and Access to Justice: Dialogues on the Future of Law* (Edinburgh University Press 2021) 77.

[7] Lawal (n 1).

2. The Legal Challenge of Cross-Jurisdictional Research.

2.1 Access to Justice and Legal Knowledge Asymmetries.

Access to justice encompasses not merely the ability to bring claims before courts, but the capacity to meaningfully participate in legal processes with adequate knowledge and resources.[8] This broad conception, originally framed by Cappelletti and Garth, has been further developed in ECtHR jurisprudence as encompassing both formal judicial remedies and effective participation in legal processes.[9] In cross-jurisdictional contexts, this principle faces particular strain. Large international law firms maintain dedicated research departments with multilingual capabilities and access to comprehensive legal databases. Individual practitioners, small legal organisations, and civil society groups typically lack such resources, creating profound asymmetries in the ability to identify and leverage relevant foreign precedents.[10]

These asymmetries have concrete consequences for the administration of justice. [11] When legal professionals cannot efficiently discover how other jurisdictions address factually similar situations, several problems emerge: inconsistent application of human rights principles across jurisdictions; reinvention of legal arguments already developed elsewhere; inability to learn from successful litigation strategies in comparable cases; and disadvantage for parties lacking resources to conduct comprehensive comparative research.[12]

These concerns are reflected in the Council of Europe's work on the online publication and categorisation of case law. The CEPEJ Guidelines on the online publication of judicial decisions emphasise that digital dissemination of judgments should ensure fair, transparent, comprehensive and accessible access to case-law information, while preventing manipulative or overly selective filtering that could undermine equal access to judicial information and must remain consistent with fundamental rights, including data protection and the right to be forgotten under Convention 108+.[13] At the same time, Council of Europe projects such as TJENI have highlighted that traditional relevance-based classification of case law, historically performed by dedicated committees within higher courts, risks generating new forms of polarisation when it is replicated through AI-driven ranking without human control, potentially obscuring legally significant decisions that are crucial for

[8] Cappelletti and Garth (n 3); and J Gerards, 'Access to Justice in the European Convention on Human Rights: A New Strategy' (2017) 15(2) Int'l J Const L 332.
[9] Gerards (n 8) 335-340.
[10] V Ruiz Abou-Nigm, 'Iconic asymmetries of our times: "super Highways" and "jungletracks" in transnational access to justice' (2024) 20(3) Journal of Private International Law 599.
[11] N Freeman Engstrom and D Freeman Engstrom, 'The Making of the A2J Crisis' (2024) 75 Stan. L. Rev. Online 159-163.
[12] VC. Jackson, *Constitutional Engagement in a Transnational Era* (Oxford University Press, 2010) 71-73; V Ruiz Abou-Nigm, 'Iconic asymmetries of our times: "super Highways" and "jungletracks" in transnational access to justice' (2024) 20(3) Journal of Private International Law 599; N Freeman Engstrom and D Freeman Engstrom (n 12) 159-163.
[13] Council of Europe - CEPEJ, 'Guidelines on the online publication of judicial decisions' CEPEJ (2024) 9. <https://rm.coe.int/cepej-2024-9-guidelines-on-the-online-publication-of-judicial-decision/1680b2d0de>.

addressing novel societal issues.[14] In response to these challenges, the TJENI project has explored semantic approaches to the categorisation of judicial decisions that aim to preserve diversity and contextual richness in case-law retrieval.[15]

The European human rights system illustrates these challenges particularly clearly. The European Court of Human Rights (ECtHR) adjudicates cases from 46 member states, each with distinct legal traditions, languages, and procedural systems.[16] Applicants and their legal representatives must navigate this complexity whilst often lacking visibility into how the Court has previously addressed factually similar situations in different jurisdictional contexts. National courts implementing ECtHR jurisprudence face similar challenges in identifying relevant precedents across linguistic and legal boundaries. Constitutional courts drawing on the CODICES database[17] confront analogous difficulties in comparative constitutional analysis.

2.2 The Structural Complexity of Legal Texts.

Legal documents cannot be processed as generic texts. Court decisions follow structural conventions that carry substantive meaning: the organisation of facts, legal reasoning, and conclusions reflects the logical progression of legal analysis.[18] Any system for legal information retrieval must respect these structural features or risk producing technically sophisticated but legally unreliable results.

The challenge intensifies across jurisdictions. Factual sections of court decisions are introduced using varying headings and stylistic conventions that differ by language, legal tradition, and institutional practice. Belgian courts may use 'Objet du recours et procédure', the Court of Justice of the European Union employs 'The main proceedings', whilst the ECtHR favours 'Les circonstances de l'espèce'. These differences are not merely cosmetic; they reflect different conceptions of how facts should be presented and what information merits emphasis.[19]

Moreover, the temporal dimension adds further complexity. Several jurisdictions have reformed their drafting practices over time, creating substantial changes in terminology and section placement. France reformed its

[14] Council of Europe, 'TJENI project outputs'. <https://www.coe.int/en/web/implementation/tjeni-project-outputs>.
[15] Council of Europe, 'Report on semantic categorisation of judicial decisions'. <https://rm.coe.int/report-on-semantic-categorisation-of-judicial-decisions/1680aeb729>.
[16] European Court of Human Rights, 'HUDOC Database' <https://hudoc.echr.coe.int>.
[17] Council of Europe, 'CODICES Database' <https://codices.coe.int>.
[18] KT Maxwell, B Schafer, 'Concept and Context in Legal Information Retrieval' in Proceedings of JURIX 2008: The Twenty-First Annual Conference on Legal Knowledge and Information Systems (IOS Press 2008) 63 <https://dl.acm.org/doi/10.5555/1564008.1564016>; J Šavelka, KD Ashley, 'Segmenting US Court Decisions into Functional and Issue Specific Parts' in Proceedings of JURIX 2018: The 31st International Conference on Legal Knowledge and Information Systems (IOS Press 2018) 111.
[19] P Curtoni, MA Biasiotti, S Faro, 'Semantic Access to Multilingual Legal Information' in Proceedings of the Workshop 'Free EU Information on the Web: The Future Beyond the New EUR-Lex' (JURIX 2005 Satellite Workshop 2005).

judgment formatting in 2019,[20] necessitating systems that can accommodate both historical and contemporary conventions. Any technological solution must account for these variations whilst maintaining consistency in how documents are processed and compared. Early attempts at semantic access to multilingual legal information[21] and cross-lingual legal information retrieval using ontological approaches[22] demonstrated the persistent challenges of harmonising structural heterogeneity across jurisdictions.

2.3 The Limits of Automation in Legal Contexts.

Artificial intelligence techniques offer powerful capabilities for processing large volumes of text, but their application to legal materials requires careful consideration of domain-specific requirements.[23] Semantic similarity does not automatically indicate legal relevance.[24] The same factual scenario may have different legal significance depending on jurisdictional context, applicable legal frameworks, procedural posture, and temporal factors. A case involving surveillance in the context of national security raises different legal questions than surveillance for criminal investigation, even if the factual scenarios appear similar.

Recent approaches to legal case retrieval, such as those framing the task as legal elements generation,[25] highlight the importance of structured legal analysis beyond pure text similarity. Systems focused on mapping recitals and normative provisions[26] demonstrate progress in multilingual legal IR but operate on statutory rather than case law materials, presenting different retrieval challenges. Traditional information extraction from case law,[27] whilst foundational, often struggled with the interpretive complexity required for cross-jurisdictional comparison.

Furthermore, the absence of precedent carries meaning in legal reasoning. From a doctrinal perspective, if a search system fails to locate relevant cases, legal professionals must know whether this reflects genuine absence of precedent or limitations of the search methodology. Pure automation cannot make such determinations reliably; human legal expertise remains essential for

[20] Cour de cassation, 'Mode de rédaction et motivation des décisions de la Cour de cassation' <https://www.courdecassation.fr/toutes-les-actualites/comprendre-une-decision-de-la-cour/mode-de-redaction-et-motivation-des>.
[21] Curtoni, Biasiotti, Faro (n 19).
[22] L Dini and others, 'Cross-Lingual Legal Information Retrieval Using a WordNet Architecture' in Proceedings of the 10th International Conference on Artificial Intelligence and Law (ACM 2005) 163.
[23] C Braun and others, 'The Hidden Structure- Improving Legal Document Understanding Through Explicit Text Formatting' (2025) arXiv:2505.12837v1 2.
[24] P Tritto, H Ponce, 'Causal Artificial Intelligence in Legal Language Processing:ASystematic Review' (2025) 27 Entropy, 3-5.
[25] Kim, Lee and Hwang (n 2).
[26] R Nanda, L Di Caro and G Boella, 'A Multilingual Legal Information Retrieval System for Mapping Recitals and Normative Provisions in EU Legislation' in *Proceedings of JURIX 2020: The Thirty-Third Annual Conference on Legal Knowledge and Information Systems* (IOS Press 2020) 171.
[27] P Jackson and others, 'Information Extraction from Case Law and Retrieval of Prior Cases by Partial Parsing and Query Generation' (2003) 31 Artif Intell 233.

validating results and contextualising findings within appropriate legal frameworks.

These considerations point towards hybrid approaches that integrate technological capabilities with human legal expertise. The question is not whether to use AI in legal contexts, but how to deploy it responsibly whilst preserving the precision, contextual understanding, and interpretive nuance that legal analysis requires.

3. A Hybrid Approach: Integrating Legal Expertise with Technology.

3.1 Legal Supervision of Technical Implementation.

CaseConnect's methodology recognises that effective legal AI requires careful integration of domain expertise with technical capabilities. We adopted a hybrid approach in which legal experts within our team conducted manual analysis of representative judicial decisions to identify structural patterns and common terminology for presenting factual content. This analysis served as the foundation for broader technical implementation, with insights systematically documented and implemented by the technical team across the entire corpus.

This legally-supervised sectioning framework [28] proved essential for addressing the substantial heterogeneity of legal documents across multiple jurisdictions. Rather than attempting to derive structural patterns automatically through machine learning, we grounded our system in recognition of how legal professionals actually organise and present information. The technical team could then implement these insights consistently whilst the legal experts validated results and refined the approach iteratively.

This methodology offers several advantages for building trustworthy legal AI systems. First, it produces results that legal professionals can understand and verify, as the structural analysis follows recognisable legal patterns. Second, it accommodates the variations across jurisdictions and temporal periods that pure automation might miss. Third, it establishes clear lines of accountability, with legal experts responsible for the analytical framework and technical experts responsible for scalable implementation.

3.2 Multilingual Semantic Retrieval.

To enable cross-lingual comparison whilst respecting document structure, we implemented semantic embeddings using the BGE-M3 model.[29] This model provides native handling of over 100 languages without requiring translation, crucial for European legal documents spanning numerous languages. The

[28] Šavelka, Ashley (n 18).

[29] J Chen and others, 'BGE M3-Embedding: Multi-Lingual, Multi-Functionality, Multi-Granularity Text Embeddings through Self-Knowledge Distillation 'in *Findings of the Association for Computational Linguistics: ACL 2024* (ACL Anthology 2024) <https://aclanthology.org/2024.findings-acl.137/>; arXiv:2402.03216 <https://arxiv.org/abs/2402.03216>.

8,192-token context window accommodates lengthy legal texts without aggressive truncation that might lose important contextual information.

The system generates embeddings for complete documents, maintaining contextual integrity across the entire case rather than fragmenting texts into disconnected segments. We opted for a single document-level embedding to preserve global factual and legal context, rather than chunk-level embeddings; a systematic comparison with multi-chunk strategies is left for future work. Whilst the preprocessing pipeline identifies and extracts distinct sections (facts, reasoning, conclusions) for metadata purposes, we generate a single embedding per document to preserve the full context in which factual scenarios arise. This design choice reflects the legal reality that facts cannot be fully understood in isolation from the legal reasoning they support.

Recent work on cross-lingual legal paragraph retrieval[30] and query-driven relevant paragraph extraction [31] demonstrates alternative granularity approaches. Cross-lingual statutory article retrieval for specific jurisdictions[32] and advances in Vietnamese-English legal CLIR [33] illustrate the growing sophistication of multilingual legal IR, though these typically focus on normative texts rather than case law. Historical dictionary-based approaches[34] and comprehensive surveys of translation techniques in CLIR [35] inform our decision to leverage multilingual embeddings rather than translation-based pipelines.

Retrieval operates through dot product similarity between query and document embeddings, enabling efficient comparison across thousands of documents. Users can filter results by jurisdiction, date range, and similarity threshold, allowing refined searches that respect the boundaries and characteristics of different legal systems. The system returns ranked results that surface the most relevant cases regardless of originating jurisdiction, whilst preserving metadata that allows users to contextualise findings appropriately.

---

[30] A Upadhya, KC Santosh, 'LexCLIPR: Cross-Lingual Paragraph Retrieval for Legal Applications ' in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (ACL 2025 forthcoming).

[31] J Tyss and others, 'Query-Driven Relevant Paragraph Extraction from Legal Judgments 'in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING 2024).

[32] CH Wang and others, 'Cross-Lingual Statutory Article Retrieval for Taiwan Legal System '(2024) arXiv:2410.11450.

[33] HD Nguyen, NC Dang, NT Son, 'Improving Cross-Lingual Information Retrieval on Low-Resource Languages via Optimal Transport Distillation 'in *Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (NAACL 2025 forthcoming).

[34] DA Hull, G Grefenstette, 'Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval' in *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM 1996) 49.

[35] D Zhou and others, 'Translation Techniques in Cross-Language Information Retrieval' (2012) 45(1) ACM Comput Surv, Article 1.

4. Privacy, Data Protection and AI Regulation.

4.1 GDPR Compliance Framework.

In developing CaseConnect, we designed the system with awareness of the General Data Protection Regulation[36] from the initial stages, implementing data protection by design and by default as required by Article 25. This approach addresses the fundamental principles of Article 5: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, and accountability.

- Data Minimisation and Purpose Limitation.
  The system processes publicly available court decisions from authoritative databases (HUDOC,[37] CJEU, CODICES[38]). We process only the legal content necessary for retrieval functionality. Personal data appearing in published judicial decisions is not extracted or used beyond what is already publicly disclosed in the original court documents. This approach respects the principle that whilst court decisions are public documents, processing of personal data within them must still comply with data protection principles.
- Transparency and Information.
  The system operates as a research tool with clear documentation of its AI-driven nature and limitations. Users interact with a transparent interface that presents similarity scores and allows verification of results against source documents. This transparency enables users to exercise informed judgment about the relevance and applicability of retrieved materials, rather than treating system outputs as authoritative.
- Technical and Organisational Measures.
  All components use verified permissive licences that facilitate compliance with data protection standards. The system architecture separates preprocessing from runtime operations, minimising data processing during user interactions. Future operational deployment would require comprehensive security assessments, documented processing agreements with data providers, and ongoing compliance monitoring mechanisms.

As a research prototype developed during the Digital Future of Justice Hackathon 2.0,[39] CaseConnect has not yet undergone formal data protection impact assessment as would be required under Article 35 for production systems processing personal data at scale. Such assessment would be essential before operational deployment, particularly given the system's processing of judicial decisions that may contain sensitive personal information.

[36] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [2016] OJ L119/1 <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
[37] European Court of Human Rights (n 11).
[38] Council of Europe (n 12).
[39] The Digital Future of Justice 2.0 Hackathon was organised by the Council of Europe and held 15-17 November 2024 in Bologna, Italy https://digitaljustice.tech .

4.2 EU Artificial Intelligence Act Considerations.

The EU Artificial Intelligence Act[40] establishes a risk-based framework for AI systems, with particular requirements for 'high-risk 'applications. We conducted an initial assessment to determine whether CaseConnect qualifies as high-risk under Article 6 and Annex III. As a legal information retrieval research tool that supports rather than replaces human decision-making, and does not directly produce legal effects or significantly affect individual rights, we provisionally consider that, in its current research-prototype form and absent deployment in decision-making contexts, CaseConnect would likely fall outside the 'high-risk 'categories of Article 6 and Annex III; a formal legal assessment would nevertheless be required before production deployment.

Nonetheless, we incorporated key regulatory principles into the system design from inception, anticipating that regulatory frameworks will continue to evolve and that responsible AI development should exceed minimum compliance requirements.

- Data Quality and Governance (Article 10):
  The system uses datasets from authoritative official sources that are publicly available and regularly maintained by their respective institutions. This reliance on institutional sources provides reasonable assurance of data quality, though we acknowledge the need for ongoing evaluation of potential biases in how legal texts are selected, published, and maintained across different jurisdictions.
- Transparency and Information Provision (Article 13):
  We maintain comprehensive documentation of the system's AI functionalities, embedding model characteristics, and retrieval mechanisms. The interface clearly presents similarity scores and allows users to verify results against source documents. This transparency enables users to understand how results are generated and to exercise critical judgment in evaluating retrieved materials.
- Human Oversight and Professional Responsibility (Article 14):
  CaseConnect is explicitly designed as a tool supporting legal professionals ' judgment rather than automating legal decision-making. The system facilitates discovery of potentially relevant cases but does not assess legal precedential value, evaluate jurisdictional applicability, or account for doctrinal nuances that affect case interpretation. Legal professionals maintain full agency in evaluating the relevance and applicability of retrieved results, preserving professional responsibility for legal analysis and advice.
- Record-Keeping and Accountability (Article 27):
  For a production system, comprehensive logging of system interactions, model versions, and retrieval performance would be essential. Such records would enable accountability, facilitate ongoing evaluation of system performance, and support investigation of potential issues or errors.

[40] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L184/1 https://eur-lex.europa.eu/eli/reg/2024/1689/oj .

It bears emphasis that as a hackathon prototype, CaseConnect has not undergone the formal conformity assessment procedures that would be required for regulated deployment. The regulatory considerations described here represent design principles incorporated during development rather than certified compliance. Future work towards production deployment would require comprehensive legal review, formal risk assessment under the AI Act's final implementing regulations, and ongoing compliance monitoring as the regulatory landscape evolves. The AI Act does not emerge in a normative vacuum. Its risk-based approach and emphasis on human oversight, transparency and robustness are broadly consistent with the OECD AI Principles, which promote human-centred values, fairness, accountability and reliable AI systems.[41] The Act also aligns with emerging international guidance on the use of AI in judicial settings, including UNESCO's recent guidelines for the judiciary on AI in the courtroom, which stress that automated tools must remain subordinate to judicial independence and fundamental rights. [42] Positioning CaseConnect within this broader international framework is particularly important in the Council of Europe context, where 46 member states – including non-EU countries – are concerned.

5. Enhancing Access to Justice across Jurisdictions.

5.1 Democratising Comparative Legal Research.

CaseConnect can contribute to access to justice by reducing barriers to cross-jurisdictional legal research that have historically favoured well-resourced organisations. Smaller legal practices, civil society organisations, and individual practitioners gain access to research capabilities previously available only to international law firms with dedicated research departments and multilingual staff. By reducing barriers to cross-jurisdictional legal research that have historically favoured well-resourced organisations, CaseConnect addresses digital divides in access to justice, supporting smaller practices and civil society organisations whilst aligning with emerging EU frameworks for technology-driven equity in legal services.[43] [44] This democratisation of legal knowledge supports more equitable participation in transnational legal processes.

The practical implications extend across several dimensions of legal practice. Lawyers representing clients in human rights cases can efficiently identify how other jurisdictions have addressed factually similar scenarios, strengthening their arguments through comparative analysis. Legal aid organisations can

[41] OECD, 'OECD Principles on Artificial Intelligence'. <https://oecd.ai/en/ai-principles>.
[42] UNESCO, 'AI in the courtroom: UNESCO's new guidelines for the judiciary', <https://www.unesco.org/en/articles/ai-courtroom-unescos-new-guidelines-judiciary>.
[43] SF Nicolosi, 'Enhancing Access to Justice in European Shared Administration through Principles of Good Administration' (2025) Eur Papers (forthcoming).
[44] OECD, 'AI in Justice Administration and Access to Justice: Governing with Artificial Intelligence' (2025) <https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report/ai-in-justice-administration-and-access-to-justice.html>.

leverage foreign precedents without requiring extensive multilingual research capabilities. Academic researchers can conduct systematic comparative studies that would be prohibitively resource-intensive through manual methods.

Importantly, the system's focus on factual similarity rather than pure legal doctrine enables identification of patterns that might otherwise remain invisible. Systematic analysis of how different jurisdictions address similar factual scenarios can reveal inconsistencies in human rights protection, inform policy development, and support arguments for harmonisation where appropriate. This capability serves not merely individual litigants but the broader project of developing coherent human rights jurisprudence across jurisdictional boundaries.

5.2 Supporting Consistency in Human Rights Adjudication.

The European human rights system aspires to consistent protection of fundamental rights across member states, but achieving this consistency requires visibility into how courts across jurisdictions address comparable situations. CaseConnect may support this goal by enabling efficient identification of relevant precedents across linguistic and jurisdictional boundaries.

National courts implementing ECtHR jurisprudence benefit from understanding how peer jurisdictions have applied similar principles in comparable factual contexts. The ECtHR itself faces the challenge of maintaining consistency across its own case law spanning 46 jurisdictions and multiple languages. Tools that facilitate identification of factually similar cases can support more systematic consideration of how different factual scenarios have been addressed, potentially reducing inconsistencies and strengthening the coherence of human rights protection.

This contribution to consistency should not be overstated. Legal systems appropriately account for contextual differences, and mechanical application of foreign precedents without regard for jurisdictional specifics would be legally problematic. CaseConnect facilitates discovery and comparison but cannot and should not replace the professional judgment required to determine when foreign precedents are genuinely relevant and when contextual differences counsel different approaches.

5.3 Limitations and Professional Responsibility.

We acknowledge significant limitations that must inform responsible use of the system. Retrieval based on factual similarity does not capture differences in legal frameworks, procedural contexts, or interpretive approaches that may render factually similar cases legally distinguishable. The system cannot assess the precedential value of cases, account for how jurisprudence has evolved since a decision was issued, or evaluate jurisdictional applicability of foreign precedents.

These limitations place particular responsibility on legal professionals using such tools. They must exercise independent judgment in evaluating the relevance and applicability of retrieved cases to their specific situations. They

must understand that similarity scores reflect factual resemblance detected by the embedding model, not legal relevance determined through proper legal analysis. They must contextualise findings within appropriate legal frameworks and account for doctrinal nuances that affect interpretation.

Moreover, the system's effectiveness depends on the quality and representativeness of the underlying corpus. Gaps in coverage, biases in case selection, or inconsistencies in document formatting can affect retrieval quality. We prioritise using authoritative official sources, but even official databases may have limitations that users should consider when evaluating results.

Professional responsibility in the age of legal AI requires that practitioners understand both the capabilities and limitations of the tools they employ. CaseConnect should be understood as facilitating discovery rather than providing answers, supporting professional judgment rather than replacing it.

6. Technical Implementation and Future Directions.

6.1 System Architecture.

CaseConnect implements a client-server architecture with clear separation between offline preprocessing and runtime retrieval operations. This design addresses the computational intensity of processing large legal document corpora whilst maintaining responsive user experience. The backend provides a REST API built on Flask, with endpoints for system health monitoring and case retrieval. The frontend implements a React single-page application with accessible interface components for search, filtering, and result display.

The offline preprocessing pipeline handles document processing during dataset preparation: extracting text from diverse formats (HTML, PDF, DOCX), applying legally-supervised sectioning rules to identify factual content, generating semantic embeddings using the BGE-M3 model, and capturing metadata for filtering and contextualisation. This architecture separates computationally expensive processing from latency-sensitive retrieval, enabling fast response times during user interactions.

6.2 Future Development Priorities.

Several directions for future development would enhance the system's utility for legal professionals whilst maintaining its foundational commitments to transparency and human oversight.

- Comparative Analysis Features:
- Enhanced functionality enabling users to view and contrast how similar facts are treated across different jurisdictions would support deeper comparative legal reasoning. Such features might allow side-by-side comparison of retrieved cases, highlighting similarities and differences in how courts approach factually analogous scenarios.

- Integration of Akoma Ntoso:
  The Akoma Ntoso standard for legal document representation[45] offers standardised structure, semantic markup, enhanced interoperability, and citation network analysis capabilities. Integration would enable more sophisticated structural analysis and leverage of standardised legal metadata across jurisdictions.
- Real-Time Document Processing:
  The current system relies on pre-computed embeddings, creating latency between case publication and system availability. Developing capabilities for incremental document addition and real-time processing would make the system more suitable for practitioners requiring access to the most recent jurisprudence.
- Rigorous Evaluation:
  Quantitative evaluation with standard information retrieval metrics, expert annotations, and controlled user studies remains a priority. Creating benchmark datasets for cross-jurisdictional legal retrieval presents challenges, as relevance judgments require deep legal expertise and may vary among practitioners.[46] Nonetheless, systematic evaluation would strengthen confidence in the system's reliability and identify areas requiring improvement.

7. Conclusion.

This article has presented CaseConnect as a model for developing legal AI systems that can contribute to access to justice whilst respecting data protection principles and regulatory requirements. Our hybrid methodology demonstrates that effective legal AI requires careful integration of legal expertise with technical capabilities, rather than pure automation that neglects domain-specific requirements. The compliance framework addressing GDPR and EU AI Act requirements illustrates how responsible AI development can prioritise transparency, accountability, and human oversight from the design stage.

CaseConnect can contribute to access to justice by democratising comparative legal research capabilities, enabling smaller legal organisations and individual practitioners to access research tools previously available only to well-resourced international firms. AI-driven systems like CaseConnect align with emerging international frameworks for technology-enhanced access to justice,[47] demonstrating that responsible deployment of advanced retrieval tools can support equity in legal knowledge whilst maintaining professional oversight and regulatory compliance. By facilitating efficient identification of factually similar cases across jurisdictions, the system may support consistency

[45] M Palmirani and others, 'Akoma Ntoso for Legal Documents 'in Giovanni Sartor and others (eds), *Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management* (Springer 2011) 75.
[46] J Rabelo and others, 'COLIEE 2021: Methods for Legal Document Retrieval and Entailment ' (2022) 16 *Review of Socionetwork Strategies*.
[47] OECD (n 44).

in human rights adjudication and enables legal professionals to learn from how different courts approach comparable scenarios.

However, we have emphasised that such systems must be understood as supporting rather than replacing professional judgment. Semantic similarity does not automatically indicate legal relevance; contextual differences may render factually similar cases legally distinguishable. Legal professionals using AI-powered tools bear responsibility for exercising independent judgment, validating results, and accounting for jurisdictional and doctrinal nuances that affect interpretation.

The development of responsible legal AI requires ongoing attention to both technical capabilities and legal considerations. As the EU regulatory framework for AI enters into force and implementing regulations are established, systems like CaseConnect will need to demonstrate not merely technical sophistication but genuine commitment to transparency, accountability, and enhancement of – rather than substitution for – human professional judgment.

Through tools that respect the complexity of legal texts, account for regulatory requirements, and maintain appropriate human oversight, we can work towards a future where legal professionals can efficiently navigate multi-jurisdictional legal landscapes. This supports more consistent, informed, and equitable administration of justice across borders, ultimately serving the fundamental principle that access to justice should not depend on the resources available to parties or their legal representatives.