

Enhancing Access to Justice via LLMs in the Field of Asset Division: CREA2 and the Digital Journey

Marco Giacalone, Nishat Hyder-Rahman, Mattia Fonisto, Flora Amato*

Abstract

The ongoing digitalisation of both out-of-court and court services undoubtedly enhances justice by increasing speed, lowering costs and time, providing information and removing physical barriers to accessing legal services and enforcing rights. We contend that advances in large language models (LLMs) will further enhance access to justice by providing bespoke, data-driven services to citizens. In this paper, we concretise this claim by demonstrating how generative AI can be integrated into dispute resolution services through the example of our own project, CREA2, an online platform and digital tool for asset division between divorcing couples or beneficiaries of an inheritance. CREA2 was completed in June 2024, and in this paper we reflect on the finished project whilst looking ahead to the follow-up CREA3 project, which began in November 2024. This paper, moreover, seeks to address not only *why* LLMs are being implemented in dispute resolution, but also *how*. The implementation and evaluation of LLMs in legal practice is an inherently interdisciplinary endeavour. Therefore, here, drawing on both legal and technical expertise, we seek to situate the CREA project, and broader phenomena of AI-driven legal tools, within both the legal and technical discourses.

Keywords: generative AI, large language models, access to justice, asset division, online dispute resolution

* Marco Giacalone, Research Professor, Department of Private and Economic Law (PREC), Vrije Universiteit Brussel. Nishat Hyder-Rahman, Postdoctoral Research Fellow, Department of Private and Economic Law (PREC), Faculty of Law and Criminology, Vrije Universiteit Brussel. Mattia Fonisto, PhD Candidate, National PhD Program in Artificial Intelligence, University of Naples Federico II. Flora Amato, Associate Professor, Department of Electrical Engineering and Information Technology, University of Naples Federico II.

1. Introduction

It is an oft-cited fact that the law is a slow-moving beast; slow to the extent that it often appears impervious to changes and advances in society and technology. The adverse effects of the sluggish pace of formal legal processes can be perceived in the context of access to justice across international, supranational and national levels. Within the legal domain, the field of dispute resolution is an outlier in which a relatively constant stream of development is detectable. Novel out-of-court and private settlement mechanisms have emerged in alternative dispute resolution (ADR) and, more recently, in online dispute resolution (ODR). Advances in ODR are necessarily propelled by advances in technology, including large language models (LLMs). The ongoing digitalisation of both out-of-court and court services undoubtedly enhances access to justice by increasing speed, lowering costs, providing information, removing physical barriers to accessing legal services and enforcing rights. We contend that LLMs will further enhance access to justice by providing bespoke, data-driven services to citizens. We concretise this claim by demonstrating how generative AI can be integrated into dispute resolution services through the example of our own project, CREA2, an online platform and digital tool for asset division between divorcing couples or beneficiaries of an inheritance.¹ CREA2 was completed in June 2024, and this paper is an opportunity to reflect on the finished project whilst kicking off the follow-up CREA3 project, which began in November 2024.²

This paper seeks to address not only *why* LLMs are being implemented in dispute resolution, but also *how*. The implementation and evaluation of LLMs in legal practice is an inherently interdisciplinary endeavour; it is therefore essential to clearly and convincingly situate this within both the legal and technical discourses. Hence, this paper begins with an overview of the concept of access to justice as a fundamental right, and how generative AI can contribute therein (section 2). The focus then shifts to our project, namely CREA2 (section 3). After briefly introducing the project, we explain how generative AI will be implemented in the project and, most importantly, how this furthers access to justice. Thereafter, we explain the technical processes used in CREA2 in detail (section 4), and the implementation and evaluation (section 5), before concluding (section 6).

2. Access to Justice and the 'Three Waves'

This paper contributes to the broader query of how the digitalisation of dispute resolution bodies – inside or outside the judiciary – can enhance access to justice. In this section, we examine access to justice in the context of human rights, specifically, as considered in academic scholarship and in legal practice. We focus on the civil

¹ Conflict Resolution with Equitable Algorithms (CREA2), Grant Agreement n. 101046629, in the Call: JUST-2021-Ejustice.

² Conflict Resolution with Equitable Algorithms (CREA3), Grant Agreement no. 101160564, in the Call: JUST-2023-JACC-EJUSTICE.

justice system (as opposed to the criminal justice system) as that is the domain in which our project operates.

2.1 What is 'Access to Justice'?

The notion of access to justice is inextricably linked to the legal system, i.e. a 'system by which people may vindicate their rights and/or resolve their disputes under the general auspices of the state'.³ It represents a practical avenue for mitigating the unjust legal repercussions of societal inequalities.⁴ According to Dworkin, the notion of access to justice is based on the idea that each person should have an effective means of protecting his/her rights or entitlements under the law. This principle is derived from the belief that the foundations of justice rest on the state's recognition of the values of human dignity and political equality, which are substantiated via access to justice.⁵

Access to justice can be defined in a precise sense, as suggested by Lord Woolf, or in a broader sense, as suggested by Galanter. Following Lord Woolf, who was solely concerned with the civil justice system, access to justice entails: i) just results; ii) fair treatment of litigants; iii) the capacity to deal with cases at a reasonable speed and reasonable cost; as well as iv) a justice system that is understandable to those who use it.⁶ According to Galanter, access to justice is best explained through a spatial metaphor, which can be crudely translated as 'getting people and their grievances into court'.⁷ This is an oversimplification, for modern legal systems encompass much more than just the courts as a forum for dispute resolution. Recognising this, Galanter expands his definition: 'access [to justice] is refreshed when justice in the courtroom gives way to justice in many rooms'.⁸ Thus, Galanter's notion of access to justice can be conceived as two-fold: firstly, that grievances may include two or more disputing parties; and secondly, that the forum for dispute resolution should be expanded from courtrooms to include the multiple and varied 'rooms' of alternate dispute resolution.⁹ As we will illustrate in the following sections, LLM-based legal tools can contribute to access to justice in both the precise and broader sense.

³ Bryant G Garth and Mauro Cappelletti, 'Access to Justice: The Newest Wave in the Worldwide Movement to Make Rights Effective' (1978) *Buffalo Law Review* 181, 182.

⁴ Ronald Sackville, 'Some Thoughts on Access to Justice' (2004) 2 *New Zealand Journal of Public and International Law* 85, 86.

⁵ Ronald Dworkin, *Taking Rights Seriously* (Harvard University Press, 1977) 199.

⁶ Lord Woolf, *Access to Justice: Final Report* (HMSO, 1996) 108–115.

⁷ Marc Galanter, 'Justice in Many Rooms: Courts, Private Ordering, and Indigenous Law' (1981) 13(19) *The Journal of Legal Pluralism and Unofficial Law* 1.

⁸ Quoted in: Christine Parker, *Just Lawyers: Regulation and Access to Justice* (Oxford University Press, 1999) 207, and deriving from: Marc Galanter 'Justice in Many Rooms' in Mauro Cappelletti (ed) *Access to Justice and the Welfare State* (Sijthoff, Alphen aan den Rijn, 1981) 147.

⁹ On the involvement of multiple parties see: Arno Lodder, 'The Third Party and Beyond: An Analysis of the Different Parties, in Particular the Fifth, Involved in Online Dispute Resolution' (2008) 15 *Information & Communications Technology Law* 143.

2.2 Access to Justice in the European Legal Order

Since our project, CREA2, is a European project, it must be viewed against the backdrop of access to justice within the European legal order; a brief exposition of relevant laws is therefore necessary. According to several European legal sources, effective access to justice must be considered a core fundamental right and a general principle of EU law.¹⁰ The notion of access to justice, in all its multi-faceted and multi-layered complexity, is not found in a single place; rather, it can be pieced together across several instruments and in case law. The right of access to justice is explicitly articulated in Articles 67(4) and 81(2)(e) of the Treaty on the Functioning of the European Union (TFEU). Article 67(4) provides that 'the Union shall facilitate access to justice, particularly through the principle of mutual recognition of judicial and extrajudicial decisions in civil matters'. Article 81(2)(e) provides that the EU shall adopt measures, particularly when necessary for the proper functioning of the internal market, to ensure effective access to justice in civil justice.

Further references to access to justice, albeit not as explicit as in the TFEU, can be found in other European treaties, namely, the EU Charter of Fundamental Rights (EU Charter), the European Convention on Human Rights (ECHR), and the Agreement on the European Economic Area (EEA Agreement). First, Article 47 of the EU Charter provides the right to an effective remedy and fair trial: 'Everyone whose rights and freedoms guaranteed by the law of the Union are violated has the right to an effective remedy before a tribunal [...]' In addition, everyone must have access to a fair and public hearing within a reasonable timeframe by an independent and impartial court previously established by law. Article 47 foresees the right to legal advice and representation, and states that legal aid must be made available to those lacking sufficient resources. Although Article 47 is relatively extensive in what it sets out, it does not fully define or account for the concept of access to justice.

Articles 6 and 13 of the ECHR outline the right to a fair trial and the right to an effective remedy, respectively. CJEU case law (Cases C-715/20¹¹, C-216/22¹², C-118/22¹³ and C-281/22¹⁴) has recognised the right to effective legal protection before national courts based on the requirements of Articles 6 and 13 ECHR, and the right to obtain efficient remedies before a competent court based on the principles of equivalence and

¹⁰ Gilles Cuniberti, 'The Recognition of Foreign Judgments Lacking Reasons in Europe: Access to Justice, Foreign Court Avoidance, and Efficiency' (2008) 57(1) *International and Comparative Law Quarterly* 25; Nathy Rass-Masson and Virginie Rouas, *Effective Access to Justice in the EU: A Study Based on Petitions to the European Parliament* (European Asylum Support Office, Malta, December 2017) 1-165 <<https://policycommons.net/artifacts/2175746/effective-access-to-justice/2931437/>> accessed 11 April 2025.

¹¹ *K. L. v X sp. z o.o.* (Case C-715/20, EU:C:2024:139)

¹² *A. A. v Bundesrepublik Deutschland* (Case C-216/22, EU:C:2024:122)

¹³ *NG v Direktor na Glavna Direktsia 'Natsionalna Politsia' pri MVR – Sofia* (Case C-118/22, EU:C:2024:97)

¹⁴ *G.K. & Others (European Public Prosecutor's Office)* (Case C-281/22, EU:C:2023:1018)

effectiveness of enshrined within EU law.¹⁵ It is also worth mentioning the Magna Carta of Judges¹⁶ – made by the Consultative Council of European Judges – which stresses the fundamental criteria of the rule of law, the independence of the judiciary, access to justice, and the principles of ethics and responsibility in a national and international context.¹⁷ Finally, it must also be borne in mind that effective access to justice played a noteworthy role inside the EU Courts: according to the CJEU, access to justice is not only one of the constitutive elements of a Union based on the rule of law (Case C-78/98¹⁸), but also a general principle of Union law (Case C-14/08¹⁹).

Finally, within the European region, both the European Union's (EU) AI Act and the Council of Europe's AI Framework Convention will impact access to justice via AI mechanisms.²⁰ Domestic regimes may also impact access to justice within particular national settings, and international policies and standards might also influence access to justice.²¹ It is outside the remit of this paper to assess the specific impact of these frameworks and instruments on access to justice; rather, we approach the matter from the opposite direction: how can AI-based tools (here, an asset divider) enhance access to justice?

2.3 Three Waves of Access to Justice

A seminal contribution to the notion of access to justice is Cappelletti and Garth's paper, 'Access to Justice'. According to Cappelletti and Garth, 'effective access to justice can be seen as the most basic requirement – the most basic human right – of a modern, egalitarian legal system which purports to guarantee, and not merely proclaim, the legal rights of all'.²² This aligns with the approach to access to justice in supranational European legal frameworks. In 'Access to Justice', Cappelletti and Garth

¹⁵ M Elvira Méndez Pinedo, 'Access to Justice as Hope in the Dark in Search for a New Concept in European Law' (2011) 19(1) *International Journal of Humanities and Social Science* 9, 12.

¹⁶ Consultative Council of European Judges, *Magna Carta of Judges (Fundamental Principles)* (Consultative Council of European Judges 17 November 2010) <<https://rm.coe.int/16807482c6>> accessed 11 April 2025.

¹⁷ Graham Greenleaf and Ginervra Peruginelli, 'A Comprehensive Free Access Legal Information System for Europe' (UNSW Law Research Paper No 2012-9, 29 February 2012) <<https://ssrn.com/abstract=2012956>> accessed 11 April 2025.

¹⁸ *PreussenElektra AG v Schleswag AG* (Case C-78/98, EU:C:2001:160)

¹⁹ *Roda Golf & Beach Resort SL* (Case C-14/08, EU:C:2009:395)

²⁰ Committee of Ministers, 'Council of Europe adopts first international treaty on artificial intelligence' (Council of Europe, 17 May 2024) <<https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence>> accessed 11 April 2025; 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) [2024] OJ L230/1.

²¹ For example, ISO standards: <<https://www.iso.org/standard/81230.html>>; <<https://www.iso.org/standard/84834.html>> accessed 11 April 2025.

²² Garth and Cappelletti (n 3) 185.

identify and classify the most common barriers and obstacles preventing ordinary citizens from accessing the legal system, namely: the lack of information and awareness; the excessive cost of litigation; the undue length of procedures; the asymmetry between private parties and organisations; and the lack of standing for collective, general or diffuse interests. Significantly, they analyse the legal history of access to justice, categorising the history of reforms that promote access to justice in three waves. These three waves progress chronologically starting from the seventeenth century, but of course, they 'can be continuing in one form or another, while others may be in retreat'.²³ We view the current generation of digital and AI-enhanced legal reforms in law as the 'fourth wave', and thus situate CREA2 within the narrative arc of access to justice.

The first wave – the charitable model – involves reforming institutions to provide legal aid to everyone by creating more efficient legal assistance or advice systems. This model mainly targets enhancing access to justice for those from poorer socio-economic backgrounds. The second wave extends legal representation to special interests, particularly consumer and environmental protection. This was achieved through class actions, the rise of public interest lawyers, and the granting of standing to sue consumer and environmental groups. The third wave is described as an alternative to traditional courts. This wave focuses on the full panoply of procedures, rules and institutions comprising our dispute-processing machinery. It involves simplifying law and legal procedures, and significantly, the mainstream integration of ADR and ODR within the legal framework.

Integrating ODR into existing legal architectures is an ongoing project.²⁴ The Legal Education Foundation published a blueprint for digital justice, considering recommendations from two workshops in October and November 2018. During these two workshops, a minimum definition of access to justice that can operate as an empirical standard was developed (England and Wales – still a part of the EU at the time – served as the test jurisdiction). Four interrelated, mutually supportive and non-divisible components were identified. Stakeholders agreed that any reform evaluation must examine the impact of reformed services against each of these four components to determine the outcomes vis-à-vis access to justice. The four components of this minimum standard are as follows:²⁵

1. Access to the formal legal system: Any evaluation of the reform programme's impact on access to justice must examine both the operation of assisted digital and paper channels and the experience of individuals who use them.

²³ Sackville (n 4) 90.

²⁴ Ayelet Sela, 'The Effect of Online Technologies on Dispute Resolution System Design: Antecedents, Current Trends and Future Directions' (2017) 21 *Lewis and Clark Law Review* 633, 653.

²⁵ Natalie Byrom, *Digital Justice: HMCTS Data Strategy and Delivering Access to Justice* (The Legal Education Foundation, 2019) 17–22.

2. Access to an effective hearing: Any evaluation of new processes intended to replace the function of physical hearings, such as continuous online resolution or virtual hearings, should look at the impact of changes in mode on judicial attitudes, behaviour and decision-making to ensure that changing the mode of hearing does not impact the way in which evidence is heard and understood.
3. Access to a decision by substantive law: Any evaluation of the impact of reform programmes on access to justice must capture the impact of reform on the types of cases that are being decided before the Courts and the individuals who bring them to understand whether the impact of the reform programme is to replicate existing trends or create new ones.
4. Access to remedy: Data should be captured on enforcement rates, in addition to the time from decision to enforcement, as part of any evaluation of the reform programme's impact on access to justice.

These elements are worth bearing in mind, as technology continues to evolve and impact legal services.

2.4 Digital Journey: 'the Fourth Wave' beyond Procedural Justice and towards Social Justice?

Cappelletti and Garths's 'three waves' fundamentally shape how we, as contemporary legal scholars, conceive of access to justice. A fourth wave or even just a rip current, is arguably occurring now. However, legal scholars differ vastly in how they characterise the fourth wave.

In the 1990s, Parker identified a fourth wave in common law jurisdictions on competition policies designed to allocate resources to access to justice. For example, she recognised how the introduction of less restrictive advertising policies by law societies resulted in lower fees for some legal services; the problem is that, as Parker acknowledged, 'those without discretionary resources will be excluded from the market and therefore from participation in the legal system'.²⁶ In June 2014, at the UCL International Conference on Access to Justice and Legal Services in London,²⁷ Economides identified a fourth wave investigating access to justice in remote communities, a wave able to resist the colonial expansion of Western rights – of the kind described by Cappelletti and Garth – destroying unique forms of social organisations. In December 2018, at the International Forum on Online Courts in

²⁶ Parker (n 8) 41, as cited in: Mary Jane Mossman, Karen Schucher and Claudia Schmeing, 'Comparing and Understanding Legal Aid Priorities: A Paper Prepared for Legal Aid Ontario' (2010) 29 *Windsor Review of Legal and Social Issues* 149, 157.

²⁷ UCL Centre for Empirical Legal Studies, UCL Centre for Ethics and Law & UCL Centre for Access to Justice, 'International Conference on Access to Justice and Legal Services' (UCL, London, UK, 19–20 June 2014).

London,²⁸ Cortes acknowledged a fourth wave in online courts, which will allow judges and lawyers – working online – to be allocated by speciality. Harvey, meanwhile, points to the rise in voluntary negotiations (and similar methods) as a dispute resolution method, with the public court system operating as a necessary fallback when such methods fail or are impossible.²⁹ We contend that the fourth wave is characterised by the themes identified by both Cortes and Harvey: a simultaneous shift towards technology-enabled legal processes. This is a much broader phenomenon than online courts or even ODR. It is a shift away from traditional court-based procedures in favour of alternate methods of dispute resolution, mainly in web-based settings.

So, we arrive at the question: How are advances in web technology being integrated into dispute resolution? Two clear trends have emerged: firstly, applying technology in a purely imitative fashion, that is to say, mirroring the current legal system and reflecting its prevailing practices and procedures. Secondly, using technology in a more transformative way allows the technology itself to lead the development of the system. The first practice, the so-called ‘imitative’, focuses on increasing efficiency and reducing costs; here, the application of technology to the current court process could be seen mainly as a substitute for paper-based and physical systems. The second practice, the so-called ‘innovative’, emphasises new scenarios for widespread reforms, keeping the common values of efficiency, equality, fairness, justice and legitimacy.³⁰ It is within the latter, the ‘innovative ground’, that artificial intelligence (AI) and, particularly, generative AI is starting to show potential. In the next section, we critically reflect on generative AI in legal practice generally, before proceeding to outline what we see as a strong use case for generative AI in legal practice via the example of our project, CREA2.

3. Digitalisation and Access to Justice

3.1 Generative AI in Legal Practice

Generative AI, in its short life span, has been greeted enthusiastically by legal practitioners, primarily commercial law firms and consultancies.³¹ Rhetoric on how generative AI will enhance efficiency and output in legal practice is widespread. Yet, it is challenging to find concrete examples of how precisely generative AI is being or will be used to augment legal practice, whilst maintaining professional ethics and standards. To date, instances of using generative AI in legal practice that have reached

²⁸ HM Courts & Tribunals Service and Society for Computers & Law, ‘First International Forum on Online Courts’ (London, UK, 3–4 December 2018).

²⁹ David Harvey, ‘From Susskind to Briggs: Online Court Approaches’ (2016) 5(2) *Journal of Civil Litigation and Practice* 84.

³⁰ Ethan Katsh and Leah Wing, ‘Ten Years of Online Dispute Resolution (ODR): Looking at the Past and Constructing the Future’ (2006) 38 *University of Toledo Law Review* 19.

³¹ Reena SenGupta, ‘Generative AI: A Legal Revolution Is Coming — Eventually’ (*Financial Times*, 22 September 2023) <<https://www.ft.com/content/0f36eb4e-b90f-4ffe-befc-daf01829c182>> accessed 11 April 2025.

the public sphere are largely negative. For example, a New York lawyer who used generative AI to assist in legal research and drafting found himself caught up in generative AI's biggest flaw – hallucination.³² Generative AI can offer incorrect or even manufactured responses, such as inventing plausible-sounding cases and, what is more, defending them as real when probed about their authenticity.³³

How else might generative AI be deployed in law? The legal field is often identified as ideally placed to benefit from so-called 'predictive analytics'.³⁴ Here, the power of generative AI is harnessed to analyse data and extrapolate legal predictions, the most obvious being the outcome of litigation.³⁵ This is problematic for two reasons. Firstly, as Reiling mentions, the output data should be considered a forecast rather than a definitive prediction. It is similar to a weather forecast, which naturally involves a degree of variability.³⁶ The very nature of legal proceedings and judicial decision-making means that the outcomes are inherently unpredictable. This understanding curtails the utility of AI-based forecasting for users considering litigation and for lawyers as part of a litigation strategy.

Furthermore, as Medvedeva and McBride point out, research claiming to engage in legal predictive analytics and boasting high accuracy rates is most often undertaken after the fact, with data taken from cases that have already been decided.³⁷ Thus, the final output data is not an accurate 'prediction', as it is based on data that already contains the actual outcome(s).³⁸ This, too, compromises the utility of AI-based forecasting in legal practice. Medvedeva and McBride's study reveals that forecasting, undertaken with data that would be available to users in real time is both less

³² Benjamin Weiser, 'Here's What Happens When Your Lawyer Uses ChatGPT' (*The New York Times*, 27 May 2023) <<https://www.nytimes.com/2023/05/27/nyregion/aviair-airline-lawsuit-chatgpt.html>> accessed 11 April 2025; Lei Huang and others, 'A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions' (2023) 43(2) *ACM Transactions on Information Systems* 1 <<http://arxiv.org/abs/2311.05232>> accessed 11 April 2025.

³³ Frank Pasquale and Glyn Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' (2018) *University of Toronto Law Journal*, 68(supplement 1), 63.

³⁴ Nikolaos Aletras and others, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 *PeerJ Computer Science* 93; Daniel Martin Katz, Michael J Bommarito and Josh Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' (2017) 12(4) *PLOS ONE* e0174698.

³⁵ Nicola Lettieri, 'Law, Rights, and the Fallacy of Computation On the Hidden Pitfalls of Predictive Analytics' (2020) 2 *Jura Gentium – Journal of Philosophy of International Law and Global Politics* 72, 77.

³⁶ A Dory Reiling, 'Courts and Artificial Intelligence' (2020) 11 *International Journal for Court Administration* 1, 4.

³⁷ Masha Medvedeva and Pauline McBride, 'Legal Judgment Prediction: If You Are Going to Do It, Do It Right' in Daniel Preoțiuc-Pietro and others (eds), *Proceedings of the Natural Legal Language Processing Workshop 2023* (Association for Computational Linguistics 2023) 75 <<https://aclanthology.org/2023.nlpl-1.9>> accessed 11 April 2025.

³⁸ Ninareh Mehrabi and others, 'A Survey on Bias and Fairness in Machine Learning' (2021) 54(6) *ACM Computing Surveys* 1, 3.

impressive and limited in its utility as the input data is incomplete (i.e. the would-be claimant can only input data from her perspective, not the defendant's, so the output data is based on one side of the legal arguments only).

Bearing in mind these limitations, we assert that there is nevertheless scope to use AI-based forecasting to good effect. The process of analysing data and extrapolating evidence-based insights based on that data (i.e. text mining) can be useful in law to both lay users and practitioners. It is important to underline that the focus here is on extrapolating information in the form of insights – this is very different to attempting to extrapolate legal outcomes. Here, what is being harnessed is the processing power of generative AI. Legal outcomes, on the other hand, are the result of legal reasoning; the processes of determining the relevant facts, identifying the legal issue(s), identifying which legal rules to apply, interpreting those rules and applying them to a given set of facts, are all inherently variable processes. In particular, interpreting and applying rules towards a particular end is an exercise in legal reasoning, a task that generative AI is not able to perform properly, at least for now.³⁹ Text mining, however, can provide useful insights into quantifiable matters, and this is how we leverage the technology in the CREA2 platform.

For now, it is important to underline the point that, at present, generative AI cannot engage in critical thinking, an essential part of the task of legal reasoning. Thus, we do not foresee generative AI substituting the counsel of a qualified lawyer. For the same reason, nor do we foresee generative AI stepping into the role of a judge. Rather, we see the role and function of generative AI – for now, at least – as limited to providing information and insights, based on the provided legal data. This is best demonstrated by example, hence the next section sets out our European co-funded project, CREA2, which integrates generative AI into its platform.

3.2 CREA, CREA2 and CREA3

Algorithms, and more recently, machine learning algorithms, have been developed to assist in asset division in the context of divorce or separation, in the form of negotiation support systems (NSS). These asset division developments draw on game theory principles and have primarily emerged within common law jurisdictions.⁴⁰

³⁹ Francesco Contini, Alessandra Minissale and Stina Bergman Blix, 'Artificial Intelligence and Real Decisions: Predictive Systems and Generative AI vs. Emotive-Cognitive Legal Deliberations' (2024) 9 *Frontiers in Sociology* 1417766.

⁴⁰ Emilia Bellucci, Deborah Macfarlane and John Zeleznikow, 'How Information Technology Can Support Family Law and Mediation' in Witold Abramowicz, Robert Tolksdorf and Krzysztof Węcel (eds), *Business Information Systems Workshops* (Springer, 2010) 243–255; Emilia Bellucci and John Zeleznikow, 'Developing Negotiation Decision Support Systems That Support Mediators: A Case Study of the Family Winner System' (2005) 13 *Artificial Intelligence and Law* 233; Elisabeth Wilson-Evered and John Zeleznikow, *Online Family Dispute Resolution: Evidence for Creating the Ideal People and Technology Interface*, Vol 45 (Springer International Publishing, 2021); Steven J Brams and Alan D Taylor, *Fair Division: From Cake-Cutting to Dispute Resolution* (Cambridge University Press, 1996).

Elsewhere, the history and practical development of algorithms and AI for asset division and other aspects of family law have been traced in detail.⁴¹ Our project contributes both theoretically and practically to this ongoing discourse in the European context. The Conflict Resolution with Equitative Algorithms project (CREA, followed by CREA2 and now CREA3) was established as an interdisciplinary European initiative to develop an algorithm able to assist parties in asset division. The project received funding from the European Commission during the period 2017–2019 (CREA), then for the period 2022–2024 (CREA2) due to its innovative contributions to civil dispute resolution within the EU, and currently for the period 2024–2026 (CREA3). Each iteration of the project builds on its predecessor.

3.2.1 CREA (2017–2019)

CREA was conceptualised as a dispute resolution model based on game-theoretical algorithms and was proposed as a decision support system for parties involved in settlement negotiations concerning asset division after divorce, separation or death.⁴² Notably, the CREA approach to civil dispute resolution addresses the disparities among the national legal systems of the EU Member States participating in the project by establishing a European Common Ground of Available Rights (ECGAR). This involves setting aside the mandatory rules of each EU Member State and focusing on the remaining ‘rights available’. This novel theoretical approach allows for a broader application of the CREA algorithmic decision support tool across civil law jurisdictions within the EU, encompassing both national and cross-border cases. The source code for the game-theoretical algorithm is publicly available.⁴³

3.2.2 CREA2 (2022–2024)

Building upon the outcomes of its predecessor, CREA2 enhanced the existing CREA platform by introducing innovative AI-based features specifically designed to facilitate online dispute resolution in asset division cases. In the CREA2 platform, an LLM-powered chatbot was developed to guide the parties through the asset division process. Generative AI was not integrated into the game-theoretical algorithm itself; instead, the chatbot was developed to accompany the user through each step, explaining the process and how the algorithm operates. The game-theoretical algorithm then provides a potential solution to the parties, who remain free to accept or reject it. The incorporation of generative AI allows the CREA2 platform to extend its remit: as well as guiding the user through the process of using our asset division tool, the LLM-powered chatbot will also be able to guide the user through their digital legal journey holistically (asset division is just one part of the legal process of divorce/separation). Incorporating an LLM also enabled the solution explorer, a tool

⁴¹ Nishat Hyder-Rahman, Elisabeth Alofs and Marco Giacalone, ‘Artificial Intelligence and Algorithms in Family (Patrimonial) Law’ in Robin Fretwell Wilson and Carbone June (eds), *International Survey of Family Law 2023* (Intersentia, 2023) 355.

⁴² John Zelezniak, ‘Using Web-Based Legal Decision Support Systems to Improve Access to Justice’ (2002) 11 *Information & Communications Technology Law* 15, 29.

⁴³ Source Code <<https://github.com/giabit/crea>> accessed 11 April 2025.

that shows the user the different solutions available at every stage (including, where appropriate, the CREA2 game-theoretical algorithm). Additionally, CREA2 will implement blockchain technology, which has extensive use in ODR procedures as it facilitates transactions of digital assets and can, therefore, enable property settlement. Thus, if the parties accept the algorithm's solution or reach an agreement themselves, they can record the agreement on the electronic medium (the blockchain system), which can improve long-term sustainability and ensure the trust and reliability of the system.

The platform developed in the CREA2 project falls within the category of ODR platforms. The target user group is varied. The platform is operable by lay citizens, affording them the autonomy to resolve asset division disputes without resorting to judicial or extra-judicial bodies and to make informed decisions about their legal matters. At the same time, the platform is valuable to a diverse array of legal professionals, including lawyers, public notaries and facilitators of alternative dispute resolution mechanisms, such as mediators, whose practice encompasses alternatives to traditional court-focused legal procedures. This project aimed to establish a robust platform that serves the dual purpose of providing a tech-oriented avenue for access to justice and alleviating the judicial system's caseload in both national and cross-border contexts.

3.2.3 CREA3 (2024–2026)

Building upon the advancements of CREA and CREA2, the CREA3 project aims to further enhance access to justice and interaction with the rule of law through digitalisation, particularly in the context of family disputes in both national and European settings. Recognising the essential role of an effective European area of justice in civil matters for societal growth, CREA3 focuses on strengthening connections with legal professionals to make the CREA algorithms and platform more accessible to citizens.

In response to the growing digitalisation of society and the need to build trust in cross-border situations, CREA3 integrates features that allow parties and relevant individuals in family hearings to participate via videoconferencing, thereby enhancing accessibility and reducing reliance on physical court appearances. Importantly, the project is committed to guaranteeing access to justice for vulnerable individuals – such as those with dyslexia and visual impairments – who may lack the necessary technical skills or resources. By providing technical assistance and user-friendly interfaces, CREA3 ensures that personal capacity does not hinder citizens from accessing legal support.

The algorithms and chatbot developed in CREA and CREA2 are incorporated into a comprehensive digital pathway that accompanies parties from the moment they recognise the need to divide assets (without there being an existing agreement), to the actual division. This 'digital journey' tool serves as a trustworthy digital assistant, offering precise and reliable information that surpasses conventional search engines.

It not only informs users, but also connects them with a curated list of legal professionals—such as notaries, mediators and lawyers—who specialise in asset division and can provide personalised assistance.

Suppose the parties choose not to accept the suggestions provided by these professionals? In that case, the platform can indicate the competent judicial authorities where they can file their cases, including average processing times and associated costs. Moreover, by utilising natural language processing and AI, the digital journey tool allows parties to detail their specific situations and receive possible outcomes along with the statistical likelihood of each, based on similar past cases. This approach empowers parties by making them more aware of possible scenarios and guiding them toward informed decisions, ultimately making justice more attainable for all citizens.

3.3 Integrating Generative AI in CREA2 – a Legal Overview

In this section, we offer an overview of how generative AI was integrated into CREA2 from a legal perspective and how we envision the platform as a mechanism to enhance access to justice. In section 4 below, the technical integration is explained in detail.

3.3.1 LLM-Powered Chatbot

Generative AI was integrated into CREA2 via an LLM and a database of legal rules, past legal case resolutions, and how legal rules have been applied in those cases. Firstly, a chatbot can guide non-expert users through the platform and asset division process using the original CREA game-theoretical tool. The CREA2 chatbot can also further assist users by explaining and guiding them digitally through their legal journey. The digital journey starts when a user queries, for example, what their rights are in a given situation, or whether they have a claim. The journey continues, guiding the user through the processes involved in determining whether a right has been violated, how best to enforce that right, and reaching a suitable resolution. Along the way, the chatbot can respond to user queries. Currently, the CREA2 chatbot can offer guidance on the law relating to asset division in inheritance and divorce/separation, specific to the user's indicated jurisdiction. The next iteration (CREA3) will offer a more personalised experience, drawing on past cases and settlements with similarities to the user's situation, to comparatively analyse and offer evidence-based insights to the user's queries.

CREA2 uses a pre-trained LLM; by providing legal context via documents and prompt engineering, the LLM is better equipped to guide users in their digital legal journey. Thus, from a user perspective, the information imparted via CREA2 is much more precise and reliable than the traditional methods of searching for information via web-based search engines. Moreover, the information imparted will be specific to the users' position and delivered in an appropriate language and tone. Integrating an LLM results in a smoother, more personalised experience for the user, compared with the older algorithmic rule-based chatbots. This is important, as users will be accessing the

CREA2 platform at a challenging time in their lives, and a clear, easy-to-use user interface is therefore essential.

3.3.2 Solution Explorer

This component was inspired by a tool of the same name, developed in British Columbia, Canada, for the Civil Resolution Tribunal. The Canadian tool aims to classify the user's issue and lead them towards the appropriate claim form on the basis of the legal information provided by lawyers and opinions by legal experts, forming a legal knowledge base. In contrast, our Solution Explorer takes a distinct, data-driven approach. We utilise data from previous cases to suggest possible legal journeys tailored to individual circumstances. By analysing precedents and outcomes, we can provide users with information-based guidance, empowering them to navigate their legal challenges confidently and clearly.

The Solution Explorer is aimed at citizens who wish to explore their options for divorce or separation. It provides a personalised guide through the options available to the user throughout the divorce/separation process. For example, a person considering divorce will be presented with possible options for reaching a divorce agreement within their jurisdiction (e.g. privately, via a mediator, lawyer-led or before a judge) and information on how to pursue each option. We analyse previous cases to draw out evidence-based insights to support such options. These insights can also be used to offer further practical information into each option, for example, to indicate the estimated cost of mediation versus lawyer-led negotiations versus litigation before a judge, as well as the estimated timeframe for each of those procedures. Thus, to draw on the ideas of Lassègue and Garapon, we seek to gain an overview of the past to augment our understanding of the present.⁴⁴ As with the chatbot, currently, the Solution Explorer can only offer guidance on the legal framework and the options within the framework specific to the user's indicated jurisdiction. The next iteration will offer a more personalised experience, drawing on past cases and settlements with similarities to the user's situation, to comparatively analyse and offer evidence-based insights into the user's query solution pathways.

Finally, within the pathways and circumstances that allow it, the option of using the integrated CREA2 game-theoretical algorithm will be presented to the user. In time, we expect to integrate other services within the platform. Thus, if mediation is an option, users will not only be directed to a list of qualified mediators within the area, but the option of arranging an appointment from within the platform will also be available. Likewise, if a lawyer is needed, the user will be directed to a list of qualified lawyers within the area, whom they can contact within a few clicks. Eventually, we intend that the platform will be able to host online meetings.

⁴⁴ Jean Lassègue and Antoine Garapon, *Justice Digitale: Révolution Graphique et Rupture Anthropologique* (PUF, 2018).

3.3.3 CREA2 and Access to Justice

Significantly, for us, both the game-theoretical algorithm and Solution Explorer are intended to increase access to justice in concrete ways. The opacity and complexity of the legal process for divorce/separation may be daunting or overwhelming to citizens; the Digital Journey and Solution Explorer aim to address that by presenting tailored information to citizens in an easy-to-follow, interactive platform. In doing so, these tools will address the inherent asymmetry of information between citizens and the profession within the current framework of legal practice, thus empowering citizens to make informed decisions about their own lives. We contend that the game-theoretical algorithm offers users a fast, cost-effective and just alternative to mediation, lawyer-led negotiation or court. As it is inherently based on the users' own preferences and is easily and flexibly accessible via the platform, it hands control over the division of property back to the parties.

CREA2 is developed for use in six Member States within the EU. It is developed in English to ease the technical development process. However, it will eventually be available to users in the national language(s) of the country in which they live. Indeed, the availability of the platform in local languages is an essential element in ensuring access to justice. The prototype with the implemented core functionalities was completed in June 2024. It can be found at <https://crea2-dev.azurewebsites.net>. As of June 2024, the project has moved into the testing phase to test both the technology and empirically test the claims about access to justice. Testing, along with the continued development of the platform, is currently being undertaken as part of the ongoing CREA3 project.

3.4 Mitigating Risks

Undoubtedly, several risks within the CREA2 project had to be considered throughout the development and implementation stages. At a practical level, because the tools within the platform require users' sensitive information, compliance with the GDPR is essential. Furthermore, compliance with the EU's legislative regime on AI must also be assured, notably the AI Act. Finally, the platform had to comply with pre-existing national and EU legal frameworks. Beyond the practical issues of ensuring security within the platform and the legality of the platform, we identified two key areas that demanded particular attention.

3.4.1 Technology-User Alignment

First, the CREA2 tools had to be developed with a proper understanding of the capabilities of generative AI in order to create appropriately aligned applications. Above, we considered the limitations of so-called predictive justice. We argued that one of the traps that users of generative AI might fall into is failing to align the machine's capabilities with the users' desired outputs. By carefully limiting our demands to evidence-based insights drawn from data, we reached an appropriate and feasible alignment between the capabilities of generative AI (processing case

data) and our requirements (calculating quantifiable matters of costs, timeframes, etc.). This will be a key theme during the testing phase.

3.4.2 Trust

Second is the issue of trust. Trust is key. How can we build trust between end users (citizens and legal professionals) in the platform and tools? This matter has previously been considered in the context of assessing the success or failure of online divorce platforms.⁴⁵ In that context, two factors were identified as essential to building trust, which are also relevant here in relation to generative AI-based tools: first, consulting stakeholders and experts throughout the technology design process; and secondly, ensuring that the technology's purpose, capabilities and processes are delineated and communicated. We took steps to safeguard both trust-building elements in the CREA2 project. The latter is significant when the technology is used in extremely sensitive and complex personal life events, such as relationship breakdown. Significantly, the CREA2 platform employs an LLM that indicates the sources of legal information used to formulate an answer. This keeps the process as transparent as possible for the user and will help build trust in the platform. Furthermore, as noted above, we have limited our technological demands to align with our understanding of its capabilities. Open communication about the platform's functionalities also required rigorous and ongoing testing and reporting, which was built into the CREA2 project and will continue in its follow-up project, CREA3. Regarding the first point, CREA2 is a multijurisdictional, interdisciplinary project that has involved civil procedure lawyers, family lawyers, engineers and technology specialists from the outset. We have and will continue to adapt the project as our understanding of the technology being used deepens, as it advances, and in response to peer and user feedback.

We anticipate that the CREA2 platform will offer tools to augment access to justice in the first instance via the Solution Explorer and the Digital Journey. However, we are aware that this will depend on building trust in the platform, and this is, therefore, a critical overarching objective through the project.

4. Technical Solution Overview

The focus of this section is the technical integration of generative AI into the CREA2 project, namely via the mechanism of a chatbot. This chatbot is built upon an LLM that handles the actual text generation, a reasoning framework that handles complex queries by generating reasoning traces and retrieval actions, and a repository of relevant legal rules and past legal cases used as an external trusted knowledge base to be queried for retrieval to support the generation.

The assistant comprises two main parts: the Solution Explorer and the Digital Journey. The first part leverages text embedding models to find relevant documents, such as

⁴⁵ Hyder-Rahman, Alofs and Giacalone (n 41).

legal rules and past legal case resolutions, stored within a vector database to provide the right context to generate factual help to user queries. The second part employs an LLM and a reasoning framework to guide individuals through the necessary steps to solve a legal problem.

As for the current state of the project, we provide guidance on the relevant legal framework to non-expert and expert users alike employing natural language. The generated responses are based on a trusted legal database that the chatbot can use to provide context behind any answer. The sources of information needed to formulate such answers are disclosed to the user to increase trust in the tool.

This approach translates into two aspects:

1. To formulate the answer, the chatbot looks at civil codes and past legal cases, exploring previous solutions and uses them as basis to formulate the response.
2. To get help with conflict resolution, the user has a conversation with the chatbot and receives an answer by means of reasoning traces and trusted legal content, i.e. the user is guided through their digital legal journey.

These components are illustrated in more depth in Figure 1:

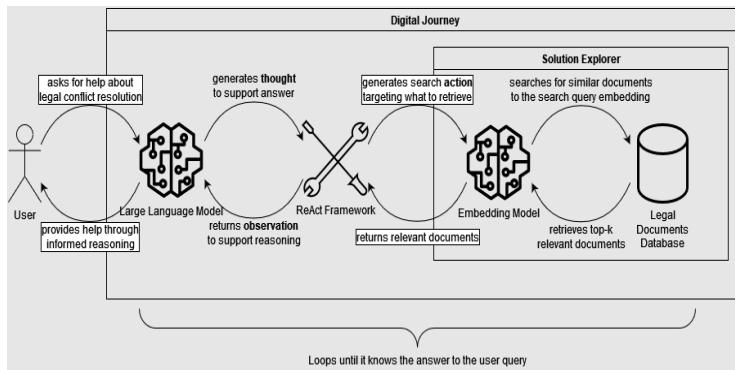


Figure 1. Diagram depicting the proposed methodology

4.1 Solution Explorer

This component plays a crucial role in the process of retrieving and analysing relevant documents for legal conflict resolution. Its methodology unfolds into the following steps:

- query analysis; and
- document retrieval.

4.1.1 Query Analysis

LLMs are proficient at generating text. However, a crucial aspect to support such generation is the ability to obtain text representation: a set of numerical vectors that translate the meaning of the text and effectively capture its semantics. These numeric vectors are commonly called text embeddings, as shown Figure 2.

In response to a user request relating to conflict resolution for divorce or inheritance, the reasoning framework generates a search action to retrieve a legal document that is able to address the query. A search action is simply text that represents the content to be searched for within a specific knowledge base.

The Solution Explorer uses a text embedding model to capture the semantic meaning of the search action and look for pieces (chunks) of legal documents to address that align with this meaning. A text embedding model converts textual data into numerical representations in a higher-dimensional space with semantic relationships.⁴⁶

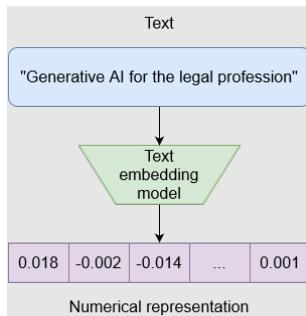


Figure 2. Text embeddings

This methodology is advantageous because, once the text is converted into this form, it allows for comparison with other texts to evaluate semantic similarity, perform groupings, classifications and much more.⁴⁷ Using a similarity function, like cosine similarity, it is possible to calculate a semantic similarity index between two text embeddings to determine whether the two texts deal with the same topic. Exploiting such a property, it is possible to find relevant legal content within the knowledge base, based on the semantic similarity between the two texts.

⁴⁶ Bhaskar Mitra and Nick Craswell, 'Neural Text Embeddings for Information Retrieval' in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Association for Computing Machinery, New York 2017) 813.

⁴⁷ Bohan Li and others, 'On the Sentence Embeddings from Pre-Trained Language Models' (2020) *arXiv preprint arXiv:2011.05864* <<https://arxiv.org/abs/2011.05864>> accessed 11 April 2025.

4.1.2 Document Retrieval

After the search action embedding computation, the Solution Explorer compares its semantic vector representation with the pre-computed semantic vector representations of each piece of the legal documents of the legal knowledge base stored within a vector database.⁴⁸ Splitting the knowledge base into smaller pieces is called 'chunking', and it is done by essentially preserving sentences together so that each chunk has a concrete meaning. Regarding the different legal rules encoded within the vector database, a chunk provides content from just one legal rule (no overlap with different rules or countries) and preserves full sentences together.

Vector databases are an alternative to traditional databases that best suit the efficient storage and fast retrieval of embeddings and their corresponding chunks. Since the legal document chunks are always the same, pre-computing their embeddings represents an optimisation in terms of both speed and efficiency. The retrieval process involves lots of comparisons by means of a similarity metric like cosine similarity or dot product. A higher similarity score indicates greater semantic similarity between the two originating texts in the embedding vector space.⁴⁹ The vector database calculates a similarity score for each tuple (search action embedding, pre-computed legal rule chunk embedding) and returns the top-ranked chunks based on that calculation. An early implementation of the CREA2 chatbot focused mainly on similar document retrieval.⁵⁰

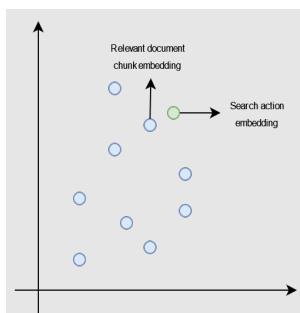


Figure 3. 2D example of relevant document chunk retrieval in the text embedding space

⁴⁸ Jan Rygl and others, 'Semantic Vector Encoding and Similarity Search Using Fulltext Search Engines' *arXiv preprint arXiv:1706.00957* <<https://arxiv.org/abs/1706.00957>> accessed 11 April 2025.

⁴⁹ Jiapeng Wang and Yihong Dong, 'Measurement of Text Similarity: A Survey' (2020) 11(9) *Information* 421.

⁵⁰ Flora Amato and others, 'An Intelligent Conversational Agent for the Legal Domain' (2023) 14 *Information* 307.

This process of retrieving relevant legal information forms the basis for generating insightful responses by the LLM using the Reason + Act (ReAct) framework.

4.2 Digital Journey

This component is responsible for guiding the user's decision-making process through legal conflict resolution. Its methodology involves the following:

- large language model; and
- reasoning framework.

4.2.1 Large Language Model

One of the main components of the generative chatbot is the LLM, which generates the text. A language model (LM) is a probability distribution over sequences of tokens (smaller parts of text) that assigns a probability to each sequence within a vocabulary (collection of tokens). The probability indicates how 'good' a sequence is and can be used for generation purposes. Although one can generate sequences of tokens by sampling from this probability distribution, we are interested in exploiting conditional generation by specifying a prefix sequence (prompt) to generate the rest (completion).⁵¹ An effective way to perform such generation is using autoregressive (AR) LMs, which can efficiently compute the conditional probability of the next token given the previous ones. An important step forward for such language models has come from neural language models and the Transformer architecture which is able to compute this conditional probability.⁵²

An LLM, by contrast, is mainly a huge LM in terms of the number of model parameters and the size of the training data, able to perform tasks which were previously considered impossible. Open-source LLMs like Llama 2, Llama 3, Mistral or Gemma are sophisticated models that come pre-trained over a huge corpus of textual data, such that they correctly handle most topics, at least in the English language.⁵³ It is worth noting that most open-source models are not multilingual, which represents a limitation for multi-language generative applications. A curated collection of AI applications for legal analytics is also available.⁵⁴

⁵¹ Hermann Ney, Ute Essen and Reinhard Kneser, 'On Structuring Probabilistic Dependences in Stochastic Language Modelling' (1994) 8 *Computer Speech & Language* 1.

⁵² Ashish Vaswani and others, 'Attention Is All You Need' (2017) 30 *Advances in Neural Information Processing Systems* 1; Yoshua Bengio, Réjean Ducharme and Pascal Vincent, 'A Neural Probabilistic Language Model' (2000) 3 *Journal of Machine Learning Research* 1137.

⁵³ Aaron Grattafiori and others, 'The Llama 3 Herd of Models' (2024) *arXiv preprint arXiv:2407.21783* <<https://arxiv.org/abs/2407.21783>> accessed 11 April 2025; Albert Q Jiang and others, 'Mistral 7B' (2023) *arXiv preprint arXiv:2310.06825* <<https://arxiv.org/abs/2310.06825>> accessed 11 April 2025; Gemma Team and Google DeepMind, 'Gemma: Open Models Based on Gemini Research and Technology' (2024) *arXiv preprint arXiv:2403.08295* <<https://arxiv.org/abs/2403.08295>> accessed 11 April 2025.

⁵⁴ Kevin D Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge University Press, 2017).

4.2.2 Reasoning Framework

The next component is the reasoning prompting framework, one of this solution's novelties. To augment the reasoning capabilities of the Digital Journey to address more complex queries, the LLM utilises the ReAct prompting framework, which offers a structured methodology for generating responses by guiding the LLM's reasoning process (NB: this is different to the process of legal reasoning, discussed above).⁵⁵ It concurrently produces both reasoning traces and search actions, making the chatbot agentic by design. Through reasoning, the model can create search actions using tools that enable access to external knowledge sources, such as the vector database containing trusted legal content. Each tool refers to just one country and one main topic covered by CREA2 (divorce or inheritance), e.g. `search_inheritance_italy`, and it is identified by a description telling the LLM what it does and what it expects as input, e.g. 'This tool provides articles regarding the division of assets for inheritance purposes in Italy. Input to this tool should be what to look for within the Italian civil code articles.'

This approach, inspired by the Chain-of-Thought prompting technique, results in an overall better performance for complex queries, along with improved interpretability and trustworthiness since both the reasoning and the sources upon which the responses are generated are given to the user. Essentially, the framework guides the LLM in generating the answer step-by-step while supporting the reasoning process at each step with the proper trusted legal information.⁵⁶

From a broader perspective, the reasoning framework can be seen as a complex prompt engineering technique, which is how users interact with an LLM. Through prompt engineering, one can guide an LLM to perform tasks for which the model has never been explicitly instructed, as well as format the output in a structured way, usually by just providing a few examples of input–output pairs.⁵⁷

The ReAct framework advances the concept of guided prompting by empowering the LLM with reasoning and the possibility to interact with external tools by means of actions. It forces the LLM to produce text following a thought–action–observation structure. Here, the thought serves as an intermediary question that must be addressed to reach the final solution; the action entails retrieving information needed to answer the intermediary question, and the observation represents the output resulting from the action. This agentic approach has demonstrated notable efficacy

⁵⁵ Shunyu Yao and others, 'React: Synergizing Reasoning and Acting in Language Models' (2022) *arXiv preprint arXiv:2210.03629* <<https://arxiv.org/abs/2210.03629>> accessed 11 April 2025.

⁵⁶ Jason Wei and others, 'Chain-of-Thought Prompting Elicits Reasoning in Large Language Models' (2022) 35 *Advances in Neural Information Processing Systems* 24824 <<https://arxiv.org/abs/2201.11903>> accessed 11 April 2025.

⁵⁷ Tom Brown and others, 'Language Models Are Few-Shot Learners' (2020) 33 *Advances in Neural Information Processing Systems* 1877 <<https://arxiv.org/abs/2005.14165>> accessed 11 April 2025.

when compared to more conventional methods, such as classic retrieval-augmented generation (RAG).⁵⁸

Table 1 below depicts an example taken from the development-stage implementation of the proposed solution that illustrates how the framework operates when queried about legal rules. It shows that when a question is more complex than usual and often requires additional knowledge to be addressed, the ReAct framework succeeds in providing the correct answer.

User question	Are the properties considered personal and not part of the community in a marriage the same both in Italy and Lithuania?
Thought	I need to know which properties are considered personal and not part of the community in a marriage in Italy
Action	search_inheritance_italy
Action input	personal and community properties in marriage
Observation	[ITALIAN] 'CIVIL CODE\nARTICLE 210.\nCONVENTIONAL CHANGES TO LEGAL COMMUNITY OF PROPERTY\nThe spouses may, by means of an agreement concluded in accordance with Article 162, i.e. by public instrument or by declaration in the marriage certificate, modify the rules governing community of property, provided that the agreements are not contrary to the provisions of Article 161, i.e. the spouses can not stipulate in a general manner that their property relationships shall be governed in whole or in part by laws to which they are not subject or by custom.\nThe property referred to in points (c), (d) and (e) of Article 179, i.e. property for the strictly personal use of each spouse and its accessories, property used in the exercise of the spouse's profession, except for property used to run a business forming part of the community, and property obtained by way of compensation for damage as well as the pension relating to the partial or total loss of the capacity to work cannot be included in the contractual community.\nThe rules of legal community of property regarding the administration of community property and the equality of shares limited to the property that would be the subject of legal community of property cannot be derogated from.'

⁵⁸ Patrick Lewis and others, 'Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks' (2021) 33 *Advances in Neural Information Processing Systems* 9459 <<https://arxiv.org/abs/2005.11401>> accessed 11 April 2025.

Thought	I need to know which properties are considered personal and not part of the community in a marriage in Lithuania
Action	search_inheritance_lithuania
Action input	personal and community properties in marriage
Observation	[LITHUANIAN] 'CIVIL CODE\n3.89.\nPersonal property of the spouses\n. Property shall be recognised as the personal property of the spouses:\n1) property acquired separately by both spouses before the marriage;\n2) property given to a spouse as a gift or inherited by a spouse after the marriage, unless the gift agreement or the will specifies that the property is to be transferred to the community property of the spouses;\n3) property for the personal use of the spouses (footwear, clothing, tools for professional activity);\n4) intellectual and industrial property rights, with the exception of income derived from intellectual activities;\n5) funds and goods for the spouse's private business, with the exception of funds and goods for a business carried on jointly by both spouses;\n6) funds received by one spouse as compensation for damages or other forms of compensation for personal injury and non-pecuniary damage, targeted material support and other benefits which are exclusively attributable to the person of the spouse who received them and which are not transferable to other persons;\n7) property acquired by a spouse with his/her own funds or with the proceeds of the realisation of property in his/her personal possession, provided that, at the time of the acquisition of that property, the spouse's intention to acquire the property in his/her personal possession was expressly expressed.'
Final answer	Yes, the properties considered personal and not part of the community in a marriage are the same in both Italy and Lithuania. They include property owned before the marriage, property acquired by gift or inheritance during the marriage, property for personal use, property used for professional purposes, and property obtained through compensation for damages.

Table 1. ReAct framework example

5. Solution Implementation and Evaluation

The proposed solution was implemented in Python using the LangChain open-source library in the development stage. We employed the open-source Mixtral 8x7B as the

LLM for the text generation, Mistral Embed as the text embedding model for semantic comparing of texts, the ReAct prompting framework as the reasoning framework for guidance in the answer generation and information retrieval, and the open-source FAISS as the vector database and retrieval strategy for storing the knowledge base embeddings and related chunks.

Mixtral 8x7B is an open-source Sparse Mixture of Experts language model that outperforms other LLMs like Llama 2 70B or GPT-3.5 across various benchmarks.⁵⁹ It has the same architecture as Mistral 7B, but it employs eight feed-forward blocks per layer, namely the experts, and a gating network that selects the best two experts per token to process the current state and combine their outputs. This selection of experts varies, making optimal usage of parameters during inference while retaining optimal performance. FAISS is an open-source library for efficient similarity search of high-dimensional vectors. It pushes forward the state-of-the-art prior to its release for GPU similarity search in sets of vectors of any size, taking strong advantage of GPU parallelisation potential.⁶⁰

In production, however, we transitioned away from a self-hosted open-source model with the ReAct reasoning framework for two key reasons:

- hosting an LLM locally in production proved unreliable due to inconsistent uptimes and slow output generation times; and
- while reasoning with tracing (as in ReAct) is powerful, it is too slow for a fast-evolving environment like a web application.

In this context, tracing refers to the explicit step-by-step reasoning process where the model outputs intermediate thoughts and decision-making steps before reaching the final answer. This method, while insightful and effective for complex reasoning tasks, adds significant latency, making it unsuitable for scenarios requiring rapid response times. To address these challenges, we adopted GPT-3.5 Turbo with a function calling for the production environment. This approach retains the reasoning component by enabling tool calls for legal document retrieval actions, but without tracing the reasoning process step-by-step. This ensures faster and more reliable performance suited to the web application's needs, while preserving the reasoning process under the hood. (To clarify: the legal sources used to generate the answer are still shared with the user, but the tool's reasoning process remains hidden.) The source code for the production-stage solution is distributed open-source and available at: https://github.com/Uzare1/crea2_api.

⁵⁹ Albert Q Jiang and others, ‘Mixtral of Experts’ (2024) *arXiv preprint arXiv:2401.04088* <<https://arxiv.org/abs/2401.04088>> accessed 11 April 2025.

⁶⁰ Jeff Johnson, Matthijs Douze and Hervé Jégou, ‘Billion-Scale Similarity Search with GPUs’ (2019) 7(3) *IEEE Transactions on Big Data* 535 <<https://arxiv.org/abs/1702.08734>> accessed 11 April 2025.

To assess the performance of the developed agentic chatbot, a comprehensive evaluation has been conducted. The assessment focused on the two main topics covered by CREA, i.e. the division of assets for divorce and inheritance over the six different countries involved in the project: Belgium, Croatia, Estonia, Italy, Lithuania and Slovenia. The assessment employed several metrics to gauge the chatbot's effectiveness in providing accurate and relevant responses to user queries. The evaluation pipeline can be summarised into two main steps:

- Synthetic test data generation, which involves synthetically generating a dataset of possible user queries and related correct answers.
- Evaluation by metrics, which involves using the queries from the test set to generate answers by adopting the developed agentic chatbot and finally comparing these answers to the correct ones provided in the test set to compute the evaluation metrics.

The following evaluation refers to the solution developed during the development stage.

5.1 Synthetic Test Data Generation

Performance evaluation of RAG-like pipelines is both difficult and essential. It can be time-consuming to manually create hundreds of triplet question-context-answer (QCA) samples from a knowledge base of legal documents. Furthermore, human-generated questions could not be as complex as those needed for a comprehensive assessment, ultimately lowering the assessment's quality. Having a systematic approach to QCA generation, supervised by human experts, is almost mandatory. Development time can also be reduced by using synthetic data generation, as it can lead to an overall better codebase by continuous performance evaluation.

To perform such generation, we took inspiration from this intuition: a perfect evaluation dataset would include questions of different levels of difficulty and other types of questions encountered in real-case scenarios. Using LLMs for generation tends to be repetitive, resulting in poor samples by default. Our approach uses an evolutionary generation algorithm inspired by Evol-Instruct, implemented in Python by means of the Ragas open-source library.⁶¹ From the different sets of legal documents that are provided to the generation algorithm, different kinds of harder questions – such as reasoning, conditioning and multi-context questions – are crafted. LLMs can efficiently generate easy questions and convert them into harder ones. We used this idea to create medium-to-hard samples from the given legal documents.

Going into a little more depth, it proceeds as follows:

⁶¹ Can Xu and others, 'Wizardlm: Empowering Large Language Models to Follow Complex Instructions' (2023) *arXiv preprint arXiv:2304.12244* <<https://arxiv.org/abs/2304.12244>> accessed 11 April 2025.

1. It randomly takes one or more document chunks (context) from the legal knowledge base and generates a clear yet simple question from it.
2. The newly generated question (seed) can be ‘evolved’ into a harder version of it according to a prior distribution over the possible evolutions.
3. For such a question, an answer according to the context provided is formulated and stored as ground truth.

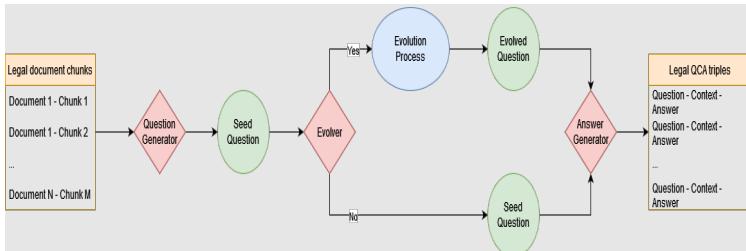


Figure 4. Synthetic evaluation dataset generation

The possible evolutions are defined prior to the generation, and for these an evolution distribution is given, to have a certain proportion of simple/medium-to-hard questions. The probability distribution used for our generation is as follows:

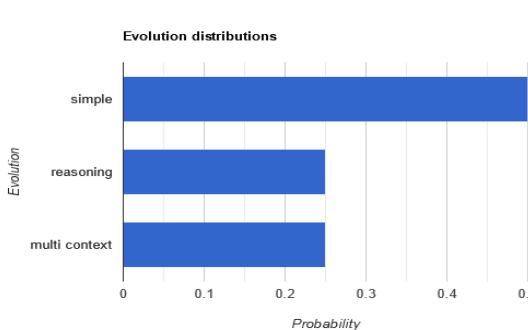


Figure 5. Evolution probability distribution

The evolutions involved in our test set generation are the following:

- Simple: The seed question as it is, with no evolution involved.
- Reasoning: Rephrased question such that answering it successfully requires an additional reasoning step.
- Multi-context: Augmented question such that answering it requires knowledge from two connected legal documents.

The prompt given to GPT-4 to get these evolutions are as follows:

- Simple: *'Generate a question that can be fully answered from a given context.'*
- Reasoning: *'Complicate the given question by rewriting it into a multi-hop reasoning question based on the provided context. Answering the question should require the reader to make multiple logical connections or inferences using the information available in a given context.'*
- Multi-context: *'Complicate the given question in a way that answering it requires information derived from both context1 and context2.'*, where the additional context is taken from the same knowledge base.

The synthetic dataset generation has been conducted by exploiting the generation capabilities of GPT-4 by OpenAI, which is a trusted model in literature usually taken as instructor for less-accurate LLMs.⁶² We created twelve knowledge bases corresponding to the six countries and two legal areas (divorce and inheritance) dealt with by CREA2. These knowledge bases are grounded in the Civil Code for each country. For each of the twelve legal knowledge bases, we generated ten questions following the evolution distributions above, gathering 120 QCA triplets that have been finally supervised by the CREA2 team of legal experts for possible errors. The knowledge bases used for such generation are as follows:

- divorce_belgium
- divorce_croatia
- divorce_estonia
- divorce_italy
- divorce_lithuania
- divorce_slovenia
- inheritance_belgium
- inheritance_croatia
- inheritance_estonia
- inheritance_italy
- inheritance_lithuania
- inheritance_slovenia

Some examples from the synthetic test dataset generated are reported in Table 2 below just for illustrative purposes:

⁶² Baolin Peng and others, 'Instruction Tuning with GPT-4' (2023) *arXiv preprint arXiv:2304.03277* <<https://arxiv.org/abs/2304.03277>> accessed 11 April 2025.

Better models to date (at the time of writing) are GPT o1, Gemini 1.5 Pro, Claude 3.5 Sonnet.

Question	Contexts	Answer
<p>Under what circumstances is the community property not liable for obligations contracted by either spouse before the marriage in Italy?</p>	<p>“CIVIL CODE\nARTICLE 189\nOBLIGATIONS CONTRACTED SEPARATELY BY THE SPOUSES\nThe community property, up to the value corresponding to the obligated spouse’s share, are liable, when creditors cannot be satisfied on personal property, for obligations incurred, after the marriage, by one of the spouses for the performance of acts exceeding ordinary administration without the necessary consent of the other.\nThe particular creditors of one of the spouses, even if the claim arose prior to the marriage, may be satisfied on a subsidiary basis on the community property, up to the value corresponding to the share of the obligated spouse. If unsecured, the community creditors have preference over them.\n\nCIVIL CODE\nARTICLE 190 \nSUBSIDIARY LIABILITY OF PERSONAL PROPERTY\nCreditors can take subsidiary action on the personal property of each of the spouses, up to half of the debt, when the community property is insufficient to satisfy the debts of the community.”, “CIVIL CODE\nARTICLE 211\nOBLIGATIONS OF SPOUSES CONTRACTED BEFORE MARRIAGE\nThe community property are liable for obligations contracted by one of the spouses before the marriage limited to the value of the property owned by that spouse before the marriage which, by agreement entered into, became part of the community of property.\n\nCIVIL CODE\nARTICLE</p>	<p>The community property is not liable for obligations contracted by either spouse before the marriage, except in certain circumstances such as when creditors cannot be satisfied on personal property or when the particular creditors of one spouse can be satisfied on a subsidiary basis. However, the community property can be liable for obligations incurred by one spouse for acts exceeding ordinary administration without the necessary consent of the other spouse. Additionally, the community property can be liable for obligations contracted by one spouse before the marriage if the property was owned by that spouse before the marriage and became part of the community property through an agreement.</p>

Question	Contexts	Answer
	<p>215 \nSEPARATION OF PROPERTY\nThe spouses may agree that each of them shall retain sole ownership of the property acquired during the marriage.\n\nCIVIL CODE\nARTICLE 217\nADMINISTRATION AND USE OF PROPERTY\nEach spouse has the enjoyment and administration of the property of which he or she is the sole owner. [...]</p>	
Who succeeds as third order intestate successors if there are no first or second order successors in Estonia?	<p>ESTONIAN SUCCESSION ACT\n15.\nThird order intestate successors\nThird order intestate successors are the grandparents of the bequeather and their descendants.\nIf at the time of opening of a succession all the grandparents are alive, they succeed to the entire estate in equal shares.\nIf by the time of opening of a succession a paternal or maternal grandparent is deceased, his or her descendants replace him or her. If he or she has no descendants, the other grandparent on the same side succeeds to his or her share. If the other grandparent is also deceased, his or her descendants succeed.\nIf by the time of opening of a succession both paternal or maternal grandparents are deceased and they have no descendants, the grandparents on the other side or their descendants succeed.\nUpon replacement of parents by their descendants, the provisions concerning first order successors apply. [...]</p>	<p>If there are no first or second order successors, the third order intestate successors are the grandparents of the bequeather and their descendants. If all the grandparents are alive, they succeed to the entire estate in equal shares. If one grandparent is deceased, their descendants replace them. If there are no descendants, the other grandparent on the same side succeeds to their share. If both grandparents are deceased, the grandparents on the other side or their descendants succeed.</p>

Table 2. Some QCA triplets from the synthetically generated test dataset

5.2 Evaluation by Metrics

As with any machine learning system, the overall experience of a chatbot is greatly influenced by the performance of its individual parts. These metrics can be roughly divided into two main categories:

- generation metrics, such as faithfulness and answer relevancy; and
- retrieval metrics, such as context recall.

The retrieval-based agentic chatbot evaluation was conducted by exploiting, just as for the synthetic dataset generation, the capabilities of GPT-4 by OpenAI. For each metric, a more detailed explanation can be provided:

- Faithfulness: it evaluates the generated response's factual consistency in relation to the provided context. First, a set of claims from the generated answer are identified with GPT-4. Next, to establish whether each of these claims can be deduced from the provided context, they are cross-checked with the context by GPT-4. The prompt for that is the following: '*Perform natural language inference. Use only 'Yes' (1), 'No' (0) and 'Null' (-1) as verdict.*', which is then augmented with few examples to guide GPT-4's generation in a few-shot prompting fashion. If every assertion made in the generated response can be deduced from the provided context, the answer is considered faithful. The faithfulness score is given as the proportion between the number of relevant claims over the total number of claims, which is in the range (0,1).
- Answer relevancy: it measures how relevant the generated answer is to the provided query. Responses with redundant or incomplete information receive a lower score when compared to relevant responses. When a response appropriately and directly answers the original question, it is considered relevant. The mean cosine similarity, that is to say, a semantic similarity score within the range (0,1), between a set of generated questions (which are derived by repeatedly asking GPT-4 to generate a question that corresponds to the generated answer) and the original question is computed to get this score. The intuition here is that GPT-4 should be able to produce questions from the generated answer that corresponds with the original question if it appropriately answers the original question.
- Context recall: it measures the extent to which the retrieved contexts align with the answer. It is computed based on the ground truth answers and the retrieved contexts, and the values range between 0 and 1, with higher values indicating better performance. To estimate context recall from the ground truth answer, each sentence in the ground truth answer is analysed by GPT-4 to determine whether it can be attributed to the retrieved context or not. The prompt for that is the following: '*Given a context, and an answer, analyse each sentence in the answer and classify if the sentence can be attributed to the given context or not. Use only 'Yes'*

(1) or 'No' (0) as a binary classification.', which is then augmented with a few examples. In an ideal scenario, all sentences in the ground truth answer should be attributable to the retrieved context. The metric is given by the ratio between ground truth answers that are compatible with the provided contexts over the total number of ground truth answers.

To assess the answers' quality and the relevance of the contexts for generating such answers, these three metrics have been employed over the twelve synthetically generated test sets defined above. For each of them, the original questions and answers are taken and appended to the reasoning chatbot's generated answers and contexts retrieved for answering those. The obtained results are the reported in the Table 3.

Topic per country	faithfulness	answer_relevancy	context_recall
divorce_belgium	0.667	0.967	0.889
divorce_croatia	0.889	0.970	1.000
divorce_estonia	0.889	0.957	0.500
divorce_italy	0.875	0.948	0.889
divorce_lithuania	1.000	0.965	0.978
divorce_slovenia	0.571	0.944	0.639
inheritance_belgium	0.714	0.959	0.778
inheritance_croatia	0.778	0.956	0.667
inheritance_estonia	0.875	0.955	0.556
inheritance_italy	0.625	0.836	0.778
inheritance_lithuania	1.000	0.950	0.889
inheritance_slovenia	0.857	0.972	1.000

Table 3. Evaluation metrics over the different synthetic datasets

6. Conclusion

Looking at the results indicated in the preceding section, two comments can be made on the technical progress in CREA2. Firstly, the reasoning chatbot demonstrates strong performance in terms of faithfulness and answer relevancy across most countries and topics. However, the performance results were not as strong in Belgium and Slovenia for divorce, or in Italy for inheritance. Secondly, the context recall shows variability, suggesting that the chatbot's ability to retrieve all necessary information differs across topics and countries. We are confident these matters can be improved

via further refinements to retrieval techniques, e.g. query rewriting⁶³ and hypothetical document embeddings (HyDE), and this will be done in the course of CREA3.⁶⁴ While the production-stage solution does not implement reasoning tracing, it still achieves strong results, balancing performance and response efficiency. The next challenge is to integrate past legal cases and settlement agreements for data mining, which will enable the chatbot to give the user personalised evidence-based insights drawn from comparative data analysis to provide options regarding conflict resolution.

From a legal perspective, CREA2 responds to the ever-pressing need to address and enhance access to justice via cutting-edge technology. Often, in legal processes, there are decisions to be made. Traditionally, lawyers and judges have held a privileged position within those processes, possessing the specialist knowledge and experience to make informed decisions. While the expertise that legal professionals can provide will continue to be demanded in complex situations, many other situations can be addressed via alternate channels. Historically, ADR initially filled the gap during the 'third wave' of access to justice. We are now in the 'fourth wave' of access to justice, characterised by digitalisation in the form of ODR and AI-enhanced digital justice. These new technologies, including CREA2, aim to address the continuing burdens on the court – and indeed the ADR system – as well as the opacity of those systems by offering a digital alternative.

Falling squarely within the fourth wave of access to justice and aligning with the EU E-Justice Strategy 2024–2028 on improving access to digital justice and enhancing digital judicial cooperation, CREA2 enables citizens to access justice in concrete ways via the game-theoretical algorithms and the Solution Explorer. The game-theoretical algorithms allow citizens to directly engage in decision-making to arrange their own affairs (e.g. reaching a divorce agreement), side-stepping lengthy, time-consuming and expensive traditional legal/ADR processes. The Solution Explorer, upon full completion, will allow the user to explore and understand their legal journey through a parallel digital journey and thus make informed legal decisions. As demonstrated above, both of these features responsibly use generative AI to deliver an accessible and evidence-based service.

The advances achieved through CREA2 have laid a robust foundation for transforming digital legal services in Europe. Building upon this groundwork, the upcoming CREA3 project aims to further integrate cutting-edge technologies into the legal framework,

⁶³ Xinbei Ma and others, 'Query Rewriting for Retrieval-Augmented Large Language Models' in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (2023) 5303 <<https://arxiv.org/abs/2305.14283>> accessed 11 April 2025.

⁶⁴ Luyu Gao and others, 'Precise Zero-Shot Dense Retrieval without Relevance Labels' in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (2023) 1762 <<https://arxiv.org/abs/2212.10496>> accessed 11 April 2025.

with a particular focus on enhancing accessibility and efficiency for EU citizens. By strengthening connections with legal professionals and making the CREA algorithms and platform more accessible, CREA3 addresses critical needs such as secure electronic communication via blockchain with e-signatures, facilitation of videoconferencing in family hearings and ensuring access to justice for vulnerable individuals through technical assistance.

The integration of advanced algorithms and chatbots into a comprehensive digital pathway empowers parties from the moment they recognise the need to divide assets to the actual resolution. This pathway provides a trustworthy digital assistant that offers precise and reliable information, connects users with specialised legal professionals, and, if necessary, illustrates judicial options with transparent data on costs and processing times. Utilising natural language processing and AI, the platform predicts possible outcomes based on similar cases, guiding parties toward informed decisions.

By making parties more aware of potential scenarios and encouraging amicable agreements through dialogue – supported by both technology and legal expertise – CREA3 aspires to modernise the delivery of legal services. This initiative represents a significant advancement in integrating digital technologies into legal frameworks, ultimately enhancing accessibility, efficiency and trust in legal processes across Europe. Through these innovative tools and approaches, we are confident that CREA3 will foster a more equitable and efficient legal system, making justice more attainable for all citizens.

Acknowledgments

This paper was made possible by funding from the European Union (under CREA2 Project, Grant Agreement no. 101046629). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.