Chapitre 1
Enhancing the Accessibility of Case Law

Leveraging Artificial Intelligence for the Retrieval of Court Decisions

Kalliopi Terzidou
Doctoral Researcher
University of Luxembourg

LARCIER INTERSENTIA

Abstract: Access to court decisions is achieved through their publication in online or analogue formats. However, the publication of case law does not alone ensure its accessibility by lawyers and self-represented litigants seeking to formulate their representation strategy. Information retrieval systems (IRS) integrated in court websites are called to render court decisions more accessible to the public. Various methods underline the functioning of IRS; however, Artificial Intelligence (AI) techniques can retrieve more relevant court decisions, according to users' information needs, in an understandable manner. This chapter seeks to address the question of how Artificial Intelligence can contribute to information retrieval systems to enhance the accessibility of court decisions. The question is answered through literature review on the legal background on access to courts through the accessibility of court decisions and on the methods underlying systems that support the retrieval of court decisions. The analysed AI techniques for IR are then evaluated under the notion of the accessibility of court decisions. The paper concludes that AI systems for the retrieval of court decisions can provide context-specific case law that responds to users' information needs, without prejudice to further technical and regulatory conditions that must be implemented for the successful integration of AI in courts' IRS.

Keywords: Artificial Intelligence, Generative AI, Information Retrieval Systems, Court Decisions, Access to Courts

INTRODUCTION

Citizens all over Europe have the right to a fair trial before independent and impartial courts, including the right to witness the public pronouncement of (their) case.[1] Access to court decisions, through their publication in analogue or digital format,[2] is further important to inform lawyers and self-represented litigants on the legal precedent for the management of current or future cases before the courts. Court decisions can be diffused not only manually but also automatically in online platforms, through the assistance of digital technologies that can help organize and index the data contained within the decision's text to render them easily retrievable. The European Commission considers the digitalisation of the justice sector as a priority, since it can facilitate access to justice by offering better court services to all citizens in the context of a more effective and efficient justice system.[3]

* Doctoral Researcher, Department of Law, University of Luxembourg. The paper has been concluded in the framework of the author's doctoral research, which is supported by the Luxembourg National Research Fund (PRIDE19/14268506).

1 Council of Europe, European Convention on Human Rights – Official Texts, Convention and Protocols, Art. 6. <https://www.echr.coe.int/Pages/home.aspx?p=basictexts&c=> [accessed 22 October 2023]; European Commission, Charter of Fundamental Rights of the European Union, [2012] OJ C 326, Art. 47. <http://data.europa.eu/eli/treaty/char_2012/oj/eng> [accessed 18 October 2023].

2 See, e.g. App. no. 7984/77, *Pretto and Others v. Italy*, Judgment (Merits), Court (Plenary), 8 December 1983, para. 26. <https://hudoc.echr.coe.int/?i=001-57561> [accessed 20 October 2023].

3 General Secretariat of the Council of the European Union, European e-Justice Strategy 2024-2028, p. 4. <https://data.consilium.europa.eu/doc/document/ST-15509-2023-INIT/en/pdf> [accessed 10 April 2024].

'Disruptive technologies', including Artificial Intelligence (AI), are not only automating but transforming the working practices of lawyers in an innovative way, reducing research time to minutes and offering answers to legal questions.[4] Large law firms started employing advanced technological means for various tasks related to the daily legal practice, including legal research. Generative AI chatbots, similar to ChatGPT, have been reported to respond to lawyers' queries on regulatory matters.[5] LegalTech companies also offer AI solutions not only for the retrieval of legislation and case law, but also for additional analysis of the search results, including the identification of the average duration of a litigation procedure based on the retrieved case law.[6] Self-represented litigants also have access to online AI tools that provide them with information on case law for the purpose of supporting their legal claims before courts.[7]

The aim of this chapter is to determine how AI techniques can be integrated in the IRS of courts to enhance the accessibility of court decisions. The chapter, thus, analyses the access to and accessibility of court decisions according to the principles of the rule of law, access to justice, and fair trial. It then explores the technical features of IRS, underlined among other by AI techniques. Finally, it reviews further technical and regulatory conditions for the integration of AI techniques in courts' IRS.

This paper, therefore, seeks to answer the following question:

HOW CAN ARTIFICIAL INTELLIGENCE CONTRIBUTE TO INFORMATION RETRIEVAL SYSTEMS TO ENHANCE THE ACCESSIBILITY OF COURT DECISIONS?

The chapter focuses on court decisions published by judicial institutions, the access to which is necessary for the consultation and (self-)representation of litigants before courts. Court decisions may derive from both higher and lower courts from all European Union (EU) Member States, since the publication of judgments is a matter of transparency and accountability common to all types and levels of courts, under the principles of the rule of law and access to justice.[8] The accessibility of court decisions is distinguished from their access, since the former denotes not merely how to obtain but how easy it is to obtain and understand a court decision.[9] In other words, the accessibility of court decisions denotes their availability in an open data format and their readability by all audiences.

4 R. SUSSKIND, *Tomorrow's Lawyers: An Introduction to Your Future*, Oxford, OUP, 2017, pp. 13-15 and 184-188. <https://global.oup.com/academic/product/tomorrows-lawyers-9780192864727?cc=be&lang=en&> [accessed 25 March 2024].

5 See, e.g. T. HEINDRICHS, 'Chatbot "Harvey Has the Potential to Revolutionise the World of Legal Advice"', *Delano*, 15 February 2023, <https://delano.lu/article/harvey-has-the-potential-to-re> [accessed 29 January 2024].

6 See, e.g. Predictice, Analyse de la jurisprudence. <https://predictice.com/analyse> [accessed 25 March 2024].

7 See, e.g. DoNotPay. <https://donotpay.com/> [accessed 30 January 2024].

8 European Commission, Treaty on European Union, [2012] OJ C 326, Art. 2. <http://data.europa.eu/eli/treaty/teu_2012/oj/eng> [accessed 21 July 2023]; See Council of Europe, above n. 1; See European Commission, above n. 1.

9 Cambridge Dictionary, Accessibility. <https://dictionary.cambridge.org/dictionary/english/accessibility> [accessed 3 April 2024].

AI technology is examined to the extent that it can enhance the accessibility of court decisions by being integrated in the IRS of courts. AI is a field of computer science dedicated to the development of intelligent machines, especially of advanced computer programs, that can achieve goals in the environment where they operate.[10]

AI systems can operate based on different techniques, which are restricted in this chapter in the performance of language processing tasks, since the input of AI-based IRS is legal text. These techniques include Natural Language Processing (NLP) for the understanding of human language embedded in court decisions and Machine Learning (ML) for the classification or prediction of IR tasks.[11] Generative AI systems for question answering are also examined through Large Language Models (LLMs) that are trained on vast amounts of data to perform IR and generative tasks.[12]

The methodology followed in this chapter is literature review and analysis to understand how the reviewed AI techniques can enhance the accessibility of court decisions through their integration in courts' IRS. The paper is accordingly divided into four main sections. First, the legal background of the access to courts and the accessibility of court decisions is examined, based on the fundamental principles of the rule of law, access to justice, and fair trial. Second, the characteristics and methods traditionally underlying IRS are explained, including Boolean methods and citations metrics. Third, the AI techniques for IR that can enhance the accessibility of court decisions are analysed, including NLP, ML, and Generative AI. Finally, the interoperability of judicial systems, the existence of open-source software, and the regulation of the development and release of AI systems in the market are reviewed as further conditions for the integration of AI in the IRS of courts. The chapter concludes with an overview of the research findings.

I. ACCESS TO COURT DECISIONS

Access of citizens to court decisions is primarily founded on the principle of the rule of law. The principle indicates, among other, that there must be legal certainty through the accessibility of law and access to justice through the right to a fair trial before independent and impartial courts.[13] The law must be definite and understandable, so citizens can regulate their conduct according to foreseeable rules. State institutions are thus hindered from abusing their power by diverging from the prescribed rule. The decisions of courts are no exception from the principle of legal certainty, since they constitute the authoritative

10 J. MCCARTHY, 'What Is Artificial Intelligence?', *John McCarthy's Original Website*, p. 2. <http://jmc.stanford.edu/artificial-intelligence/index.html> [accessed 19 August 2023].

11 IBM, 'What Is Natural Language Processing?' <https://www.ibm.com/cloud/learn/natural-language-processing> [accessed 14 September 2023]; IBM, 'What Is Machine Learning?' <https://www.ibm.com/cloud/learn/machine-learning> [accessed 14 September 2023].

12 IBM, 'What Are Large Language Models (LLMs)?' <https://www.ibm.com/topics/large-language-models> [accessed 23 March 2024].

13 Council of Europe, Venice Commission, Rule of Law. <https://www.venice.coe.int/WebForms/pages/?p=02_Rule_of_law&lang=EN> [accessed 3 April 2024].

output of one of the branches of the state, i.e. the judiciary. Therefore, case law must be made accessible to citizens in a binding and unambiguous manner.

The courts themselves must be independent and impartial, meaning that judges must not be unduly influenced by the executive and legislative powers or show signs of prejudice or bias in a given case.[14] The principles of independence and impartiality of the judiciary contribute to the fairness of the trial and, more broadly, access to justice, since judges possess the institutional safeguards to award binding and foreseeable decisions. Citizens' trust in the judicial system is further enhanced and they are more likely to accept and respect the judgments of the courts.

A. Access to Court Decisions through their Publication

Citizens can access court decisions when these are published by the courts. The obligation of courts to publish their decisions has been reviewed by the European Court of Human Rights in cases concerning violations of Article 6 of the European Convention of Human Rights prescribing the right to a fair trial.[15] Paragraph 1 of Article 6 states that 'Judgment shall be pronounced publicly […]', an obligation which has been interpreted by the Court to also require the publication of court decisions in several cases. According to the Court's case law, the Court does not have to literally interpret the term "judgment shall be pronounced publicly" but leaves a margin of appreciation to the national courts to assess the meaning of the phrase according to 'the special features of the proceedings in question' and the scope of Article 6(1).[16] Many signatories to the Convention view 'public pronouncement' not only as the reading aloud of the decision, but also as the making public of court decisions through their depositing in a registry accessible to the public, among other means of publication.[17]

Courts in EU Member States select which decisions to publish based on negative and positive criteria.[18] A negative selection means that all decisions are published, unless the grounds on which the decisions are based are stated according to a standard formula or when the decisions concern questions of evidence in agreement with existing case law. A positive selection means that decisions are in principle not published unless certain grounds apply. Among other, decisions must explain a concept or formulate or amend a rule of law, apply a line of arguments in a way that defers from previous decisions, or treat a rule of law and/or a body of facts which is of general interest.

14 European Court of Human Rights, Guide on Article 6 of the European Convention on Human Rights – Right to a fair trial (civil limb), pp. 60–72. <https://rm.coe.int/1680700aaf> [accessed 31 August 2023].
15 See Council of Europe, above n. 1.
16 See *Pretto and Others v. Italy*, above n. 2.
17 *Sutter v. Switzerland*, no. 8209/78, 22 February 1984, para. 33, <https://hudoc.echr.coe.int/?i=001-57585> [accessed 20 October 2023].
18 Committee of Ministers of the Council of Europe, The Selection, Processing, Presentation, and Archiving of Court Decisions in Legal Information Retrieval Systems (Recommendation No R(95)11), Appendix II, Chapter II, para. 5. <https://vkksu.gov.ua/sites/default/files/field/file/eng/17.11.22/8%20Recommendation%20%2895%29%2011%20of%20the%20Committee%20of%20Ministers.pdf> [accessed 3 April 2024].

According to a 2017 report by van Opijnen *et al.*, a negative selection is usually performed by higher courts while a positive selection is applied in lower courts.[19] The authors further point out that more than half of first instance courts and courts of appeal in EU Member States do not publish their decisions (at all or to a significant extent), while higher courts have a better record. In any case, many EU Member States' courts have more than one online database, either public or non-public (for internal purposes), while commercial publishers also have access to court decisions which they re-publish in their own websites. The version of the decision published online is considered authoritative, although it might not be authentic since it could be altered according, for example, to pseudonymisation requirements prior to its publication.

B. Pseudonymisation of Court Decisions

The pseudonymisation of personal data is a method for obscuring personal data to protect the privacy of data subjects. Contrary to anonynimisation, the results of pseudonymisation are reversible, so that additional information can lead to the re-identification of the data subject.[20]

The pseudonymisation of the personal information of the participants to court proceedings is in compliance with the requirements of the General Data Protection Regulation (GDPR), including the principle of purpose limitation.[21] In particular, the processing of the personal data of litigants could be limited to a time before the publication of the court decision, since these personal information are primarily processed for case management purposes regarding the conduct of court proceedings.[22] Certain courts are also implementing algorithmic solutions for the partial automation of the pseudonymisation process, including national courts in Austria and Luxembourg.[23]

19 M VAN OPIJNEN *et al.*, 'On-Line Publication of Court Decisions in the EU: Report of the Policy Group of the Project "Building on the European Case Law Identifier"'. *SSRN Scholarly Paper*, pp. 11–15. <https://papers.ssrn.com/abstract=3088495> [accessed 5 April 2024].

20 European Parliament and Council of the EU, 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, and Repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119, Recital 26 and Art. 4(1). <https://eur-lex.europa.eu/eli/reg/2016/679/oj> [accessed 20 October 2023].

21 *Ibid.*, Art. 5(1)(c).

22 K. TERZIDOU, 'Automated Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems', in F. ANDRADE (Conference Chair), *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law, ICAIL '23*, New York, Association for Computing Machinery, 2023, pp. 3–4. <https://doi.org/10.1145/3594536.3595151> [accessed 9 September 2023].

23 Unknown, 'Einsatz von künstlicher Intelligenz bei der Anonymisierung von Gerichtsentscheidungen: Justiz gewinnt eAward 2022' *Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung*, <https://www.brz.gv.at/presse/eaward2022.html> [accessed 30 January 2023]; Unknown, 'Lancement d'une application pour faciliter la pseudonymisation de décisions judiciaires', *Communiqué du Parquet général*, <https://justice.public.lu/fr/actualites/2022/07/communique-parquet-general-juano.html> [accessed 11 July 2022].

II. ACCESSIBILITY OF COURT DECISIONS

'Accessibility' is defined as 'the fact of being able to be reached or obtained easily' or 'the quality of being easy to understand or enjoy'.[24] In the context of access to published court decisions, 'accessibility' can be viewed as the easiness to obtain and understand court decisions. This can be achieved through their open data format and their understandability, not only be legal experts but also by individuals without legal expertise, including self-represented litigants.

A. Open Data Policies Regarding Court Decisions

The European e-Justice Strategy 2024-2028 stresses the importance of a data-driven justice in the EU through access to open justice data, which enhances transparency of justice and citizens' trust in the judiciary.[25] According to the Strategy, an open data format allows for the collection and analysis of data, two processes that can guide targeted actions that address specific needs, as well as mitigate risks associated with cybersecurity, data protection, and technological biases. The Directive on open data and the re-use of public sector data establishes rules for the re-use of public sector information, so that certain public sector bodies, including judicial authorities, are under the obligation to make their data available for re-use.[26]

According to Article 2(11) of the Directive, re-use means 'the use by persons or legal entities of documents held by… public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the documents were produced…'. The re-use of information shall be granted after request and shall be free of charge, while the format of the document must be, where possible, electronic, open, machine-readable, accessible, findable, and re-usable, together with the respective metadata (Articles 4-6).

The movement of 'open data' of, among other, court decisions in the European Union (EU) underlines the importance of access to and accessibility of court decisions.[27] The transformation of court decisions into an open data format contributes to the transparency of the justice system, that can ultimately lead to the accountability of justice officials and to an increased trust in the judicial mechanisms by citizens seeking to commence court proceedings.[28] Citizens can re-use the decisions by downloading them, so that they can get

24 See Cambridge Dictionary, above n. 9.

25 See General Secretariat of the Council of the European Union, above n. 3, pp. 14–15.

26 European Parliament and Council of the EU, Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information (Recast), [2019] OJ L 172, <http://data.europa.eu/eli/dir/2019/1024/oj/eng> [accessed 5 April 2024].

27 See, e.g. European Commission, Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA Relevance), [2022] OJ L 152/1, <https://eur-lex.europa.eu/eli/reg/2022/868/oj> [accessed 25 March 2024].

28 Directorate-General for Communication of the European Commission, The Social Justice System: Leveraging Open Data for Transparent Judicial Activities, <https://data.europa.eu/en/news-events/news/social-justice-system-leveraging-open-data-transparent-judicial-activities> [accessed 25 March 2024].

informed on case law updates that are important for the defence of their case or their legal education, in general. Court decisions can also be re-used when their data and metadata are 'fed' in an algorithmic system for training purposes. Consequently, new (AI) systems can be developed to achieve goals, including the pseudonymisation and translation of court decisions.

On an EU level, the EU Open Data Portal constitutes a single point of free access to datasets produced by EU and national bodies, including datasets on justice matters.[29] Court decisions by national judicial authorities can also be found, to an extent, at the websites of EUR-Lex and the European e-Justice Portal.[30] Occasionally, available court decisions in EUR-Lex will be accompanied by legal analysis and links to related information, while the e-Justice Portal also provides information on the organisation of justice systems across EU Member States.

B. Readability and Comprehensibility of Court Decisions

Accessibility of court decisions also means that the public, and not only legal experts, should understand the legal terms present in a given (set of) law(s) that are later applied by courts on the facts of the case to reach a judgment. This is especially important given the principle of 'ignorance of the law excuses not' ('*ignorantia juris non excusat*') or ignorance of law excuses no one ('*ignorantia legis neminem excusat*') that is prescribed in legislative documents of many European jurisdictions.[31] This doctrine of criminal law states that an individual cannot escape liability or punishment for violating a law with the excuse that they were not aware of the law.[32] However, it could be argued that awareness of the law does not necessarily mean 'readability' of the law, which must exist so citizens can make conscious decisions on how to perform their rights and obligations.

'Readability' of the law can denote 'the ease of understanding or comprehension due to the style of writing', as well as 'the degree to which a given class of people find certain reading matter compelling and comprehensible'.[33] Readability is especially important for individuals with no legal expertise who

29 Publications Office of the European Union, The Official Portal for European Data, <https://data.europa.eu/en> [accessed 5 April 2024].

30 Directorate-General for Communication of the European Commission, EUR-Lex – Access to European Union Law, <https://eur-lex.europa.eu/collection/n-law/n-case-law.html> [accessed 7 February 2023]; European Commission, European E-Justice Portal – European Case Law Identifier (ECLI), <https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-en.do> [accessed 7 February 2023].

31 See, e.g. Ministry of Justice of France, French Criminal Code, Art. 122-3, <https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006070719/LEGISCTA000006136037/#LEGISCTA000006136037> [accessed 31 March 2024]; Ministry of Justice of Italy, Italian Criminal Code, Article 5, <https://www.gazzettaufficiale.it/dettaglio/codici/codicePenale/5_1_1> [accessed 31 March 2024].

32 P. MATTHEWS, 'Ignorance of the Law Is No Excuse' (2018) *Legal Studies* 3(2), pp. 174–175, <https://www.cambridge.org/core/journals/legal-studies/article/abs/ignorance-of-the-law-is-no-excuse/31F800ED44C5CF1FAFA1562889D8ED0D> [accessed 4 April 2024].

33 W. DUBAY, 'The Principles of Readability', *Impact Information*, p. 3. <https://www.researchgate.net/publication/228965813> [accessed 4 April 2024].

reportedly think that online legislative materials are difficult to access, while they find legislative language hard [to understand], a view that is shared also by legal experts.[34] In the absence of plain language that can be understood, especially by individuals with no legal expertise, or of widely available glossaries explaining legal terms, citizens will remain unable to represent themselves before courts and thus dependent on often unaffordable legal services. The adverse effects of the lack of readable law can extend to court decisions, where legal terms are also present due to the application of legislation or the use of specialised terms characterising court proceedings.

Readability in the sense of the understandability or comprehensibility of the law can also be viewed in terms of the user-friendliness of the medium through which the law is displayed. According to Newman and Doherty, individuals with no legal expertise often need significant time to carry out information retrieval tasks on online legal sources, compared to lawyers.[35] The authors found that non-lawyers perceive databases as 'confusing', because they face difficulties, among other, in navigating from the home page of the website to the appropriate field dedicated to browsing or searching, in filling out the search form correctly, and in locating the required information in the retrieved document. Therefore, law-related websites need to present an information architecture, including a navigation system, and provide the appropriate tools to help users search for information and interpret the search results to determine what is relevant to them.[36] Courts offering access to case law should also consider increasing the user-friendliness of their websites to enhance the understandability of court decisions.

III. INFORMATION RETRIEVAL SYSTEMS FOR COURT DECISIONS

Court decisions that are published in a pseudonymised and open data format are then available to be uploaded in the databases of courts' IRS, where legal experts and individuals with no legal expertise can search for case law, ideally in a user-friendly environment. Information retrieval is defined by Britannica as '[…][the] recovery of information, especially in a database stored in a computer' by 'matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links'.[37] The Encyclopaedia mentions the underlying techniques of natural language, hyperlinks, and keyword searching, highlighting AI as a technique employed for more precision in the retrieval of information. An IRS collects documents with

34 M. CURTOTTI *et al.*, 'Citizen Science for Citizen Access to Law' (2015) 3(1) *Journal of Open Access to Law*, p. 53, <https://ojs.law.cornell.edu/index.php/joal/article/view/34> [accessed 4 April 2024].

35 D.R. NEWMAN and U. DOHERTY, 'Making the Law Accessible to Non-Lawyers: Effects of Different Kinds of Expertise on Perceived Usability of Online Legal Information Services: Behaviour & Information Technology' (2008) 27(5) *Behaviour & Information Technology*, pp. 429–431, <https://doi.org/10.1080/01449290601111002> [accessed 4 April 2024].

36 M. HAGAN, 'The User Experience of the Internet as a Legal Help Service: Defining Standards for the Next Generation of User-Friendly Online Legal Services' *SSRN Scholarly Paper*, p. 419. <https://papers.ssrn.com/abstract=2942478> [accessed 4 April 2024].

37 Britannica, 'Information Retrieval | Definition, Methods, & Facts', <https://www.britannica.com/technology/information-retrieval> [accessed 26 March 2024].

the help of a 'crawler', creates an index based on the collected documents using an 'indexer' for their analysis, extracts relevant terms, and maps them according to their corresponding documents.[38] The system can then process the queries to match them to the indexed documents, through a 'query processor', while the 'ranking algorithm' can decide on the relevance of the documents to the query to return them to users.

Relevance is of particular importance for the designing of IRS. The Cambridge Dictionary defines relevance as 'the degree to which something is related or useful to what is happening or being talked about'.[39] According to Saracevic, there is an interdependent system of relevances that dynamically interact with each other.[40] The author states that an effective IRS must return documents that contain the terms used in the query (algorithmic relevance) and documents with a topic related to that of the query (topical relevance). He further mentions that retrieved results must correspond to users' knowledge state (cognitive relevance), be useful in addressing their situation or problem (situational relevance), and respond to their goals and intents (motivational relevance). A relevant case for lawyers can be considered to be the one that satisfies their information needs, for example by helping them to form a legal strategy for their assigned case.[41]

The retrieval of legal documents can be challenging for digital systems that have to adapt to their length, specific structure, legal terminology, temporal relativity, and personal data.[42] According to van Opijnen and Santos, legal documents are more extensive in length in comparison to other scientific domains, which implicates their full-text search, while they are characterised by intricate internal structures that may not always be machine-readable. Moreover, a single legal document may have different versions. For example, a case may be followed by multiple court decisions depending on the stage of court proceedings. Therefore, digital systems may struggle to determine the most relevant version of a document to retrieve based on users' needs. The legal jargon is another challenge due to the specialised nature of the legal terminology, the terms of which can have multiple synonyms and definitions [depending on the context and the domain]. Lastly, legal documents can encompass personal data. This is especially true for court decisions containing personal information of

38 P. RAGHURAMAN, 'Advancing AI: From Information Retrieval to Large Language Models and Personalized Question…', *Medium*, 18 August 2023, <https://medium.com/@pr0031/advancing-ai-from-information-retrieval-to-large-language-models-and-personalized-question-add28cd35bec> [accessed 28 March 2024].

39 Cambridge Dictionary, Definition of Relevance. <https://dictionary.cambridge.org/dictionary/english/relevance> [accessed 12 July 2023].

40 T. SARACEVIC, 'Relevance reconsidered', in P. INGWERSEN and N.O. PORS (eds.), *Information science: integration in perspectives, Proceedings of the Second Conference on Conceptions of Library and Information Science*, Copenhagen, 1996, pp. 210–213. <https://www.bibsonomy.org/bibtex/2c3f30a1af84eb099284f266679ff41f2/aschmidt> [accessed 12 October 2023].

41 S.A. SUTTON, 'The Role of Attorney Mental Models of Law in Case Relevance Determinations: An Exploratory Analysis' (1994) 45(3) *Journal of the American Society for Information Science*, p. 187, <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%28199404%2945%3A3%3C186%3A%3AAID-ASI8%3E3.0.CO%3B2-F> [accessed 13 July 2023].

42 M. VAN OPIJNEN and C. SANTOS, 'On the Concept of Relevance in Legal Information Retrieval' (2017) 25(1) *Artificial Intelligence and Law*, pp. 67–69. <https://doi.org/10.1007/s10506-017-9195-8> [accessed 11 July 2023].

the parties to a case and justice officials, such as their names. As van Opijnen and Santos highlight, legal memory frequently relies on the identification of individuals and locations. However, national courts in the EU are gradually pseudonymising court decisions before their publication to comply with the requirements of the GDPR.[43] The pseudonymisation process can render the retrieval of court decisions more challenging when it is based on the names of the parties and related information capable of (indirectly) identifying them.

A. Information Retrieval through Boolean Methods

IRS employ different techniques, including the Boolean search. Boolean models allow for full-text search, search focusing on specific sections of the text, and search on its metadata. Boolean models additionally make use of operators, such as 'OR', 'AND', or 'NOT', that enable the inclusion or exclusion of certain terms in the search query[44]. The effectiveness of these models in terms of relevance is measured under the metrics of recall, i.e. the system's ability to retrieve all relevant documents from the entire collection of relevant documents available, and precision, i.e. the system's ability to retrieve only the relevant documents among all the retrieved documents.[45] Ideally, there is a balance between the two metrics, so the search query retrieves a large number of relevant documents while minimising the number of irrelevant ones. Low values of recall can occur when users struggle to choose the right (and correctly spelled) search terms that will return all relevant documents, without retrieving many irrelevant ones.[46] To improve the precision-recall ratio and address the user's information need, the query can be adjusted by combining different search terms and operators, adding more specific terms, and using synonyms. Consequently, an effective IRS for court decisions would ideally return by default most or all of the relevant court decisions included in the database of a given court, with a minimum number of irrelevant decisions.

In the context of Boolean searches, IRS likely prioritise algorithmic relevance as Saracevic determines it, i.e. matching the search terms used in the query with the available documents. For a high precision and recall value in the retrieved decisions, the lawyer or self-represented litigant would need to insert the appropriate terms (such as the name of the parties, the date of the decision, or the competent court) combined with the appropriate operators. Courts might also employ search forms that assist users in searching for a court case, by guiding them on which terms to enter for retrieving relevant search

43 See TERZIDOU, above n. 22, pp. 3–4.

44 University of Illinois Urbana-Champaign, University Library, Lib Guides, 'What is Boolean Searching?', <https://guides.library.illinois.edu/c.php?g=563215&p=3877584#:~:text=Boolean%20Searching%3A%20A%20search%20technique,by%20linking%20two%20terms%20together> [accessed 1 March 2025].

45 C. D. MANNING, P. RAGHAVAN and H. SCHÜTZE, *Introduction to Information Retrieval*, Cambridge CUP, 2008, pp. 154–157. <https://nlp.stanford.edu/IR-book/information-retrieval-book.html> [accessed 26 March 2024].

46 D.C. BLAIR and M.E. MARON, 'An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System' (1985) 28(3) *Communications of the ACM*, pp. 290–297. <https://doi.org/10.1145/3166.3197> [accessed 17 July 2023].

results according to their information need. For instance, the Court of Justice of the EU hosts a search form within its website that contains multiple cells which the user can fill out to retrieve court decisions, including cells on the case number, the name of the parties and the judges, and the subject matter of the decision.[47]

B. Information Retrieval through Citation Metrics

Apart from Boolean methods, case law websites may apply citation metrics as an additional method of IR. Citations of legal documents are essentially used for their ranking in IRS and function based on the number of times a document has been referenced in another document as an indication of its impact on other scholars.[48] Citation metrics present several limitations for the retrieval of relevant documents. First, in the absence of a common citation guide, each Member State's justice sector or individual courts apply their own citation style, notwithstanding some similarities among judicial institutions, including the reference of the case number.[49] This lack of uniformity of citation styles risks the accessibility of court decisions, as it might be challenging for IRS to uniformly apply their rules on different types of citations of court decisions.

The European Case Law Identifier (ECLI) addresses the challenge of having to access multiple databases across jurisdictions to find national case law by establishing a single identifier and citation style.[50] The ECLI consists of five mandatory elements: 'ECLI' to identify it as a European Case Law Identifier, the country code, court code, year of judgment, and a unique ordinal number. Metadata standards are also established to enhance understanding and searchability of case law documents.

Second, even if citations offer valuable information about the relevance of legal publications for their ranking in IRS, they cannot be the sole criterion of relevance.[51] Citations within legal documents reflect a broader spectrum of relevance within the legal domain, since scholars and practitioners consider similar criteria in assessing the relevance of retrieved documents. For IRS of courts, this means that the ranking of cases within the database cannot be solely based on their citations, because these might not reflect the same level of impact for users other than legal professionals and scholars. Further methods of information retrieval must be alternatively used or combined.

Third, multiple citations of a court decision do not always reflect its legal importance or authority, i.e. the importance of a decision for legal theory and

47 Court of Justice of the European Union, CURIA – Search Form. <https://curia.europa.eu/juris/recherche.jsf?language=en> [accessed 27 March 2024].

48 G. WIGGERS and S. VERBERNE, 'Citation Metrics for Legal Information Retrieval Systems', in *BIR 2019 Workshop on Bibliometric-Enhanced Information Retrieval*, Cologne, 2019, pp. 40–41. <https://ceur-ws.org/Vol-2345/paper4.pdf> [accessed 5 April 2024].

49 See VAN OPIJNEN *et al.*, above n. 19, pp. 36–37.

50 See European Commission, above n. 30.

51 See WIGGERS and VERBERNE, above n. 47, pp. 48–49.

practice according to the opinion of the legal community.[52] Instead, there are different ways to measure the legal relevance of a court decision, including the 'wisdom of the crowd' that is based on the variables of the publication in jurisprudence magazines, the publication of annotations (i.e. scholarly notes), the incoming citations from case law, and the incoming citations from literature that might be updated overtime.[53]

IV. INFORMATION RETRIEVAL OF COURT DECISIONS THROUGH AI SYSTEMS

AI technology is another method for IRS, that could render the retrieval process context-specific, capable of addressing users' information needs, and characterised by user-friendly interfaces that make the contents of a judgment more understandable.

AI technology is manifested through different techniques, including Natural Language Processing (NLP), Machine Learning (ML), and Generative AI.

A. The Techniques of Natural Language Processing and Machine Learning

NLP is a technique that makes computers able to understand and interpret human language through a combination of computational linguistics, statistical modeling, and ML techniques.[54] ML algorithms are trained to perform classifications or predictions that reveal patterns in the processed data, including those in IRS.[55]

These two techniques, separately or in combination, can be distinguished from Boolean and citation metrics methods for IR for several reasons. AI systems can understand not only the semantics but also the context of a given query.[56] This can be a dynamic process as AI systems, especially ML-based ones, are continuously trained with new input data,[57] so that they can adapt to the users' preferences or search behaviour. Search results can then be ranked based not simply on their matching with the query's term(s), but on their relevance (according to the underlying information need) and/or the users' behavioural data.[58]

52 M. VAN OPIJNEN, 'A Model for Automated Rating of Case Law', in: E. FRANCESCONI (Conference Chair), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law ICAIL '13,* New York, Association for Computing Machinery, 2013, p. 140, <https://doi.org/10.1145/2514601.2514617> [accessed 5 April 2024].

53 *Ibid.*, p. 141.

54 See IBM, 'What Is Natural Language Processing?', above n. 11.

55 See IBM, 'What Is Machine Learning?', above n. 11.

56 See, e.g. M. THOMAS. 'Artificial Intelligence for Information Retrieval', in J.R. RABUÑAL DOPICO *et al.* (eds.), *Encyclopedia of Artificial Intelligence*, Hershey, IGI Global, 2009, p. 152. <https://doi.org/10.4018/978-1-59904-849-9.ch023> [accessed 3 April 2024].

57 See IBM, above n. 54.

58 M. BHANGALE, 'Relevance, Ranking and Search' *Medium*, 7 September 2019, <https://medium.com/@mayurbhangale/relevance-ranking-and-search-a98b35ebc7b3> [accessed 28 March 2024].

In other words, the understanding of the users' information need and the adaptation of the IRS according to their search patterns can help them receive more relevant search results, depending on their situation, level of legal knowledge, and intents. This can be particularly useful when users search for case law based on the subject matter of the court proceedings, so it is hard to insert specific search terms to formulate a Boolean-like query. The model can also learn the search patterns of users that frequently use a court's database (most likely lawyers), so it can adapt to their preferences.

B. The Technique of Generative AI through Large Language Models

Generative AI systems are being used to retrieve and present information in a more user-friendly and understandable manner, by prompting the system to generate answers in a conversational tone.[59] Large Language Models (LLMs) offer new opportunities for information retrieval due to their ability to understand natural language prompts and generate answers, by being trained on vast amounts of input data.[60] A popular demonstration of the generative powers of these models is ChatGPT, an application that is pre-trained on trillions of natural language examples serving as input data to react to users' prompts in a dialogue format.[61]

Generative AI systems can provide some additional efficiencies for information retrieval in comparison to the AI techniques of NLP and ML. In particular, Generative AI systems can answer (open) questions and generate content,[62] so that these systems are not only able to retrieve documents but also to directly address the user's query by providing an answer based on the relevant documents. In this way, users can 'digest' the retrieved information with greater easiness. Another feature of Generative AI systems is that they can generate content that supplements the search results through, for example, summaries and translations. Summaries and translations can help users comprehend the retrieved information more easily and access sources that they could not normally understand due to language restrictions, respectively.[63] The user's queries can be further expanded through the generation of additional terms by the system.[64] These terms could consist in synonyms or related or contextually relevant search terms and phrases.

The above efficiencies can be particularly useful for the retrieval of court decisions, especially when the intended audience are individuals with no legal expertise. More specifically, IRS for court decisions could be complemented by chatbot-like functions that allow users to pose their query in the form of a

59 See, e.g. OpenAI, 'Introducing ChatGPT', <https://openai.com/blog/chatgpt> [accessed 7 March 2023].
60 See IBM, above n. 12.
61 See OpenAI, above n. 58.
62 See RAGHURAMAN, above n. 38.
63 F. HASAN, 'The Future of Document Analysis with Generative AI', *Medium*, 8 November 2023, <https://medium.com/@fahimhasan007/the-future-of-document-analysis-with-generative-ai-7687bd7a015f> [accessed 28 March 2024].
64 See, e.g. R.JAGERMAN *et al.*, 'Query Expansion by Prompting Large Language Models', *arXiv - CS - Information Retrieval*, pp. 2–4. <http://arxiv.org/abs/2305.03653> [accessed 28 March 2024].

question, thus not requiring strict Boolean structures including exact terms and operators. The system does not only address the query but also the information need of the user in a direct way. The IRS can additionally provide summaries, translations, and documents that could be related to the decisions, such as orders. During the research phase, the system could also suggest terms and their synonyms that are related to the subject matter, court type, or other particularities of court proceedings.

C. The Risks of AI Systems for Information Retrieval

AI systems, in general, are characterised by a number of restrictions, that may render their outputs unreliable. Algorithmic models can perpetuate and amplify biases present in the training data, so that certain perspectives or information are over-represented or under-represented in the search results or autocompleted search predictions. Search engines have been shown to promote racial discrimination when responding to users' queries, due to the commercial interests of the search engine promoting webpages which it has been paid to advertise.[65] Autocompletion of search predictions in Generative AI models has been found to entail gender and ethnic biases, which may in turn reinforce disinformation when they redirect users to sources that contain biases.[66] Courts cannot be characterised as search engines that pursue an economic interest; however, their IRS can contain biases depending on the data that the model is trained on. Such biases can underrepresent case law that concerns minorities or other marginalised groups in society, so that the respective search results are ranked lower in the retrieved list of case law.

The tracking of users' preferences through their search patterns can also lead to 'filter bubbles', so that users are exposed to information that primarily aligns with their preferences and are likely to ignore more diverse types of information that are ranked lower in the search results.[67] Frequent users of courts' websites, such as lawyers, could be leaving a digital trace when posing their queries, which reveals the subject matters of case law in which they are usually interested. Consequently, cases which might be relevant by responding to their information needs but do not correspond to their search pattern could be ranked lower and easily ignored.

Moreover, the processing of court decisions that contain personal data of the participants in court proceedings, including the litigants, their lawyers, and court officials, may not always comply with the requirements of data protection. According to the principles of purpose limitation and data minimisation,

65 S.U. NOBLE, *Algorithms of Oppression: How Search Engines Reinforce Racism,* New York, New York University Press, 2018, pp. 34–40. <https://doi.org/10.18574/nyu/9781479833641.001.0001> [accessed 2 April 2024].

66 P. THOMAS *et al., Evaluating Identity Bias in GPT-3 and Google Search Autocompletion,* Medauras Global and Diplomatic Courier, 2022, pp. 26–27, <https://issuu.com/medauras/docs/disinfolab_report_jan22> [accessed 2 April 2024].

67 Unknown, 'Researchers Devise Algorithm to Break Through "Search Bubbles"', *New York University*, 20 July 2023, <http://www.nyu.edu/content/nyu/en/about/news-publications/news/2023/july/researchers-devise-algorithm-to-break-through--search-bubbles-> [accessed 29 March 2024].

the processing of personal data should not exceed the original purpose or the necessary extent, respectively.[68] The IRS could be breaching the above requirements if it re-purposes the processing of personal information from case management to the training of the IRS. A privacy breach can also consist in the failure to apply pseudonymisation requirements of personal data within the court decisions, which is regulated differently by each EU Member State or individual court.[69] Failing to update the publicly available documents within the court's database or storing them for more time than necessary, depending on the applicable regulations, could also violate data protection rights of the participants in the court proceedings.[70]

Generative AI systems, in particular, are known to present further limitations, including 'hallucinations'. LLMs merely replicate statements based on word patterns identified within a given context, so that the generated output may seem sophisticated but not be accurate.[71] This limitation can be especially severe for the retrieval of court decisions, since popular LLMs, notably ChatGPT, have reportedly been returning fictitious cases that cannot be found in the precedent of the referenced court.[72] It has to be noted that ChatGPT is a general purpose system, as it can be used in multiple contexts for multiple purposes. Courts' IRS are likely fine-tuned models, 'specialised' in the retrieval of court decisions or related documents, such as orders. Therefore, the risk of the model misinforming users on fictitious cases is minimised, since it is trained on a more controlled set of data rather than data crawled from the web.

The management of the above-mentioned risks is implicated due to the lack of transparency of AI systems, which can be viewed both in terms of the visibility of the algorithmic processes leading to certain outputs, as well as the way these outputs contribute to human decisions.[73] This drawback can undermine trust in search results and impede the verification of the accuracy of the retrieved information. Courts' IRS returning fictitious cases, as a result of the Generative AI model's hallucinations, could impact the decisions of lawyers or self-represented litigants who rely on these fictitious results, thus undermining the successful representation of their case. Failing to verify the accuracy of the results and lacking the ability to address any technical issues due to the opacity of the model, the results of malfunctioning AI-based IRS may be severe for applications in the judicial domain.

Lastly, it must be noted that users can be affected by 'automation bias', whereby they trust the machine's outputs and over-rely on them, ignoring

68 See European Commission, above n. 20, Art. 5(1)(b) and (c).
69 See TERZIDOU, above n. 22, p. 3.
70 See European Commission, above n. 20, Art. 5(1)(d) and (e).
71 C. SHAH, 'AI Information Retrieval: A Search Engine Researcher Explains the Promise and Peril of Letting ChatGPT and Its Cousins Search the Web for You', *The Conversation*, 1 March 2023, <http://theconversation.com/ai-information-retrieval-a-search-engine-researcher-explains-the-promise-and-peril-of-letting-chatgpt-and-its-cousins-search-the-web-for-you-200875> [accessed 28 March 2024].
72 See, e.g. S. MERKEN, 'New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief' *Reuters*, 26 June 2023, <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/> [accessed 25 January 2024].
73 M. ALMADA, 'Governing the Black Box of Artificial Intelligence', *SSRN Scholarly Paper*, 7 November 2023, pp. 9–10, <https://doi.org/10.2139/ssrn.4587609> [accessed 3 April 2024].

information that might be contradictory to them.[74] In this way, users may lose their critical thinking when evaluating the accuracy of the search results and their relevance, according to their information needs. Consequently, they may stop conducting further queries that may retrieve more accurate or relevant documents, depending on their situation, problem, or intent. In the case of question-answering systems, the consequences may be more severe when users rely on answers that may contain misinformation due to hallucinations or biases in the models. Failing to verify false information provided by the system can lead to serious consequences, particularly for lawyers' professional conduct obligation to stay competent when advising or representing their clients by possessing the necessary legal knowledge.[75]

V. THE INTEGRATION OF AI-BASED INFORMATION RETRIEVAL SYSTEMS IN NATIONAL COURTS

AI techniques have been shown to facilitate the retrieval of context-specific results that address users' information needs, in an understandable and user-friendly manner, thus enhancing the accessibility of court decisions.

The efficiencies of AI systems for the accessibility of court decisions can drive courts into the transformation of their existing databases or the creation of new websites incorporating these advanced methods for the retrieval of court decisions. Apart from the mitigation of the above-mentioned risks that primarily require technical measures, the successful transformation of courts' IRS depends on interoperable, open-source, and regulated AI systems.

A. Interoperability of Justice Systems

The interoperability of justice systems is deemed as one of the operational principles of the e-Justice Strategy 2024-2028. The Strategy explains that interoperability contributes to the communication and data exchange between systems.[76] Better communication allows for the coordination and cooperation between judicial institutions at the national level and across different jurisdictions, which in turn enables the mitigation of cybersecurity risks. The European Commission has set out the new European Interoperability Framework that aims at encouraging public administrations to design and deliver their public services to other national or international public administrations, citizens, and businesses in a digital and open data format by default.[77] The Commission has also proposed a regulation

74 B. HOFFMAN, 'Automation Bias: What It Is And How To Overcome It' *Forbes*, 10 March 2020, <https://www.forbes.com/sites/brycehoffman/2024/03/10/automation-bias-what-it-is-and-how-to-overcome-it/> [accessed 29 March 2024].

75 See, e.g. International Bar Association, 'IBA International Principles on Conduct for the Legal Profession', p. 30, <https://www.icj.org/wp-content/uploads/2014/10/IBA_International_Principles_on_Conduct_for_the_legal_prof.pdf> [accessed 13 February 2024].

76 See General Secretariat of the Council of the European Union, above n. 3, pp. 13–14.

77 Directorate-General for Digital Services (European Commission), 'New European Interoperability Framework: Promoting Seamless Services and Data Flows for European Public Administrations', Publications Office of the European Union, pp. 5–6. <https://data.europa.eu/doi/10.2799/78681> [accessed 9 April 2024].

establishing obligations for public authorities to render new or existing databases interoperable through proactive legislative measures, in respect of the principles of good governance and free movement of data in the EU.[78]

The e-CODEX project is another European initiative that facilitates the connection between judicial systems through building blocks, such gateways, that ensure secure connections with counterparts in other Member States, and connectors, that handle adaptations necessary for receiving encrypted data from service providers in other Member States.[79] The project is based upon an interoperability layer that verifies electronic signatures using certificates issued in the signer's home country, thus enhancing the reliability of electronic legal proceedings.

Interoperable systems are in line with the EU's once-only principle, which allows natural and legal persons to submit their data only once to public administrations within their Member State or beyond.[80] According to the European Commission, once-only policies facilitate the delivery of digital public services, foster mobility of citizens and businesses, and lower the cost of administrative procedures. The implementation of such policies in the context of courts' IRS would allow lawyers and self-represented litigants to search for court decisions only through a single website, provided that the systems of all courts within and even outside a Member State communicate and can safely exchange information among them. The existence of pseudonymised court decisions is crucial in that regard, to ensure that the protection of the personal data of participants to the trial will not be compromised during their exchange in the various national or international judicial databases.

B. Open-Source Software

Another operational principle of the e-Justice Strategy 2024-2028 is the open-source nature of software used in the judicial context.[81] According to the Strategy, there is an increasing recognition of the significance of developing and employing software through open-source licenses to take advantage of the reduced total cost of ownership, promote innovation, and improve the transparency and interoperability of justice systems. The EU's Open Source Strategy 2020-2023 encourages the sharing and reuse of software, applications, data, information, and knowledge, for the delivery of better services across the EU.[82]

78 European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act), Brussels, 18 November 2022, COM(2022), 720 final, 2022/0379 (COD), Preambles 1–8. <https://commission.europa.eu/system/files/2022-11/com2022720_0.pdf> [accessed 9 April 2024].

79 Eu-LISA, Technical Solutions | E-CODEX Website. <https://www.e-codex.eu/technical-solutions> [accessed 9 April 2024].

80 European Commission, 'The Once Only Principle System: A Breakthrough for the EU's Digital Single Market', <https://commission.europa.eu/news/once-only-principle-system-breakthrough-eus-digital-single-market-2020-11-05_en> [accessed 9 April 2024].

81 See General Secretariat of the Council of the European Union, above, n. 3, p. 15.

82 European Commission, 'Open Source Software Strategy', <https://commission.europa.eu/about/departments-and-executive-agencies/digital-services/open-source-software-strategy_en> [accessed 9 April 2024].

The Strategy focuses on a set of concrete actions, including enhancing the software repository, revising software distribution practices, enabling innovation with open source labs, and integrating open source in internal IT governance.

Open-source software can adhere to open data standards, thus promoting compatibility and integration with other systems and supporting interoperability efforts across EU national justice systems for safe data exchanges. Accessible case law, characterised among other by standardised open data formats, can enable the competent authorities to render IRS open-source software. Examples of open data standards in the legal sector include the ECLI mentioned in Section III.B.

C. Regulation of (Generative) AI Models

The EU's AI Act imposes requirements for the development and use of AI systems, including 'high-risk' AI systems. High-risk AI systems pose a risk for democracy, the rule of law, and individual rights and should abide by certain requirements, including risk management, data governance, and technical documentation.[83] IRS for case law could be characterised as high-risk under the Act's text, when these systems are used by judicial authorities for the research, interpretation, and application of the law to a set of facts.[84] However, there is no indication that AI systems assisting lawyers and self-represented litigants with their research in preparation of their defence strategy can constitute high-risk AI systems.

Specific obligations are prescribed for providers of General-Purpose AI systems that have a generative power, namely the technical documentation of the model, its compliance with Union copyright law and related rights, and a sufficiently detailed summary about the content used for the training of the model.[85] Consequently, providers of IRS based on Generative AI have to be transparent on the development of these systems and compliant with conditions related to the robustness of the models, especially considering the hallucination risks described in Section IV.C. It could be argued that AI-based IRS could pose a risk to individuals' rights when they retrieve, for example, fictitious court decisions that users might rely on, thus endangering the successful representation of their case.

83 European Parliament and of the Council of the EU, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>, Art. 8-15.

84 *Ibid.*, Preamble 61 and Annex III, point 8.

85 European Parliament and of the Council of the EU, Artificial Intelligence Act, Art. 51.

CONCLUSION

This paper sought to address the question of how Artificial Intelligence can contribute to information retrieval systems to enhance the accessibility of court decisions. Access to court decisions by lawyers and self-represented litigants, among others, is mandated by the principles of the rule of law, access to justice, and the right to a fair trial (in the sense of the public announcement of the judgment) and is made possible through their publication, in analogue or online format. Nevertheless, court decisions are not characterised as accessible (i.e. easy to be understood and obtained) until they are transformed into an open data format and made readable for legal experts and individuals with no legal expertise alike.

Courts possess their own IRS designed to retrieve (accessible) case law following users' queries. These systems have been viewed under the aspects of the relevance and complexity of court decisions. An IRS must retrieve relevant court decisions, which can be understood to mean, among other, decisions that address the specific situation, problem, or intent of the user, i.e. the representation of a case by a lawyer or self-represented litigant before the court. IR is challenged by the complexity of court decisions due to their length and legal terminology that can be hard to understand for individuals with no legal expertise.

IR is achieved through various methods. The Boolean method performs full-text or specific text search through the use of operators, such as 'AND', 'OR', and 'NOT'. Citation metrics is another method for the retrieval of case law, whereby the decisions are ranked based on the number of times they have been referenced in another decision. However, these methods demonstrate limitations in retrieving relevant case law, which can be summarised in the failure to capture the context of users' queries and respond to their underlying information need.

AI systems based on the techniques of NLP, ML, and Generative AI can address these limitations and render the retrieved search results more relevant, according to users' information needs. NLP and ML methods can retrieve more context-specific information that responds to the user's situation, problem, or intent, considering for example their preferences and search patterns. Generative AI, through LLMs, has the added benefit of generating content that can help users better understand the court decision, for instance summaries or translations of the judgments.

In any case, AI systems have several inherent risks that must be taken into consideration when developing IRS for court decisions. AI models can produce biases when under-representing case law that is related to marginalised groups and can create filter bubbles when returning non-diverse search results that primarily align with users' preferences. Additionally, Generative AI models are prone to return fictitious information due to incidents of hallucinations. These technical defects cannot always be identified and addressed easily because the model may lack transparency in the way it functions. Users are also prone to over-reliance on the output of the AI-based IRS, without verifying the accuracy of the information.

Apart from the mitigation of these risks that are mostly of a technical nature, further conditions must be present to reap the benefits of AI models for the retrieval of court decisions. Interoperability is important for the safe communication and exchange of information and data among judicial systems on a national European level. In this way, coordination among judicial authorities is facilitated and citizens and businesses can access case law from at least national courts by visiting only one website, due to the safe connection among judicial systems. Open-source software that adheres to open data standards can further support interoperability by facilitating the compatibility and interaction of courts' IRS. Finally, regulations on AI systems, including the EU's AI Act, must be respected when establishing obligations for providers of (high-risk) AI systems regarding their transparency and robustness, among other.