NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Integration of AI and Law Viewed Through Explainability

A Framework to Facilitate AI Integration in the Judiciary

Bruno Miguel Lopes Mendes

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Integration of AI and Law Viewed Through Explainability

A Framework to Facilitate AI Integration in the Judiciary

by

Bruno Miguel Lopes Mendes

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Business Analytics.

Supervised by

Vitor Santos, PhD, NOVA Information Management School

November, 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Portugal, Lisboa, 23 November 2023

ABSTRACT

The field of AI is growing extremely rapidly. Recent advances have greatly increased interest in exploring its application in different domains, including in the practice of Law which presents particular challenges. The blackbox nature of cutting-edge neural networks directly conflicts with the necessity for transparency in legal proceedings, alongside the apparent shift in research focus from simpler diverse narrow systems to more general NLP systems such as LLMs which can be considerably opaquer despite having inherently intelligible output. Therefore, there's a need to evaluate the integration of AI from the perspective of output explainability and intelligibility to the human users and affected parties. To this effect, a systematic literature review of the state of AI and Law was conducted wherein the major concerns are exposed, such as subtle algorithmic bias. Later, a framework is proposed to serve as a basis for planning and facilitating the implementation of AI systems as an assistant for legal professionals in legal tasks.

KEYWORDS

Artificial Intelligence; Law; Human-Computer Interaction; Explainable AI (XAI)

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CAI	Europe's Committee on AI
DSR	Design Science Research
DSRM	Design Science Research Methodology
EU	European Union
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformer
IRAC	Issue, Rule, Application and Conclusion
LLM	Large Language Model
LMM	Large Multimodal Model
MANN	Memory-Augmented Neural Network
ML	Machine Learning
NLP	Natural Language Processing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ	Research Question
SDG	Sustainable Development Goals
SLR	Systematic Literature Review
TF-IDF	Term Frequency - Inverse Document Frequency
TOS	Terms of Service
USA	United States of America
XAI	Explainable Artificial Intelligence

1. INTRODUCTION

1.1. FRAMEWORK AND PROBLEM IDENTIFICATION

With artificial Intelligence (AI) models becoming increasingly sophisticated and fit for deployment in day-to-day situations, there's major concern regarding AI-assisted decision-making in sensitive domains that can have a dramatic effect on human lives (European Parliament, 2021). The state-of-the-art AI models such as Neural Networks are termed black boxes because they are not inherently interpretable (Zhang et al., 2021), which puts the use of these models in supporting decision-making into question, since decisions need to be justified to humans to at minimum satisfy legal accountability (Doshi-Velez et al., 2017). This is where Explainable AI (XAI) comes in, it is "a research field concerned with developing approaches to explain and make artificial systems understandable to human stakeholders" (Langer et al., 2021).

In legal terms, discussion around the importance of following XAI principles surges from Art. 22 of GDPR (2016), which states "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her". This effectively creates a 'right to explanation', "whereby a user can ask for an explanation of an algorithmic decision that was made about them" (Goodman & Flaxman, 2017), although there is room for disagreement (Wachter et al., 2017). Others do not deny that this right can be found in GDPR, but that it is "uncertain, convoluted, rife with technical difficulties, and likely to be interpreted differently in different member states" (Edwards & Veale, 2018). On the other hand, despite the present lack of clear legislation regarding AI, Winikoff & Sardelić (2021) argue that is already enough legal basis for the 'right to explanation' to be legally considered a fundamental human right, to which AI systems are subject to and of which AI developers, deployers and users will have to take into consideration moving forward. In this vein, Art. 48 of the EU Charter of Fundamental Rights (2007) establishes that everyone charged with a criminal offence has the right "to be informed promptly, in a language which he understands and in detail, of the nature and cause of the accusation against him".

Aiming to address that lack of direct unified legislation in the future, the European Commission recently put forward a proposal for the "first ever legal framework on AI" aptly referenced as the "Artificial Intelligence Act" (European Commission, 2021), wherein the high-risk AI systems as pertaining to Administration of Justice are defined in Annex III (8) as "AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts". While there is no statement offered regarding XAI specifically, Art. 9 requires the establishment of a risk-management system with a set of guidelines that could inform the construction of XAI-based models, thereby imbuing them with the necessary attributes to be deployed even as high-risk systems.

Each domain where XAI is applicable has different requirements and specific needs that must be met for all parties involved to be satisfied (Langer et al., 2021). This paper will focus on the domain of Law and the judicial system, aiming to review the most recent legal developments and AI literature to establish where and how XAI approaches can be useful and finally proposing a framework through they can be consistently evaluated to match the needs and requirements particular to Law and the judicial system.

1.2. Objectives

The goal of the paper would be to develop a theoretical framework through which XAI approaches can be evaluated in the domain of Law and the judicial system.

To achieve this goal, the following intermediate objectives were defined:

1. Literature review for XAI and AI in the Judiciary;
2. Explore use-cases;
3. Develop framework;
4. Validate with experts.

1.3. Importance and Relevance

This paper aims to facilitate the implementation of AI technology in a fair, legal, and ethical manner, observing best practices, in the whole of Law and the judicial system through the study of XAI techniques. Professionals in these areas would have access to tools that significantly increase their work-efficiency, clarity of thought and, more saliently, free them from spending time on burdensome tasks such as legal document review (Mahoney et al., 2019) in favor of increasing speedy trials (de Sousa et al., 2022), which would enable them to be faster in dealing with legal proceedings, reduce court delay and reduce overall costs. Countries such as Portugal have slower justice systems which increases court congestion, causing individuals to be more reluctant in solving disputes through litigation (Bielen et al., 2018). Furthermore, a functioning and efficient judiciary is crucial for sustainable development and a healthy economy as covered by Ramello & Voigt (2012):

> It is only with an effective judiciary that government promises to enforce private property rights stand a chance of being credible to potential investors. It is only if the judiciary is accountable to law, and not corruptible by interested parties, that it makes sense to exchange goods via contracts. It is only if the judiciary is accessible to potential plaintiffs that it can be seen as a real protector of their formal rights; and only if it is efficient will there not be huge delays in court decisions. Finally, only a properly functioning judiciary can successfully interact with the legislature and the executive, as is necessary for them to jointly attain socially and economically valuable outcomes.

Therefore, we believe that the benefits of securing the implementation of AI technology in the judiciary are threefold: (1) relieve the burden on law and judiciary professionals, (2) increase long-term economic sustainability, (3) increase overall happiness from speedy justice.

2. METHODOLOGY

This paper aims to develop a theoretical framework in the domain of information technology, in which case standard practice is to treat it as an artifact as per the Design Science Research (DSR) paradigm and proceed accordingly with the defined steps. In this section DSR process will be explained, followed by its application to the current research domain.

2.1. DESIGN SCIENCE RESEARCH (DSR)

Design Science Research (DSR) is defined by vom Brocke et al. (2020) as "a problem-solving paradigm that seeks to enhance (…) technology and science knowledge bases via the creation of innovative artifacts that solve problems and improve the environment in which they are instantiated". In this context, an artifact is "a thing that has, or can be transformed into, a material existence as an artificially made object (…) or process" (Gregor & Hevner, 2013) such as models and methods.

Artifacts are constructed by following the design science research methodology (DSRM) process model, which serves as a mental model to structure research outputs (Peffers et al., 2007).

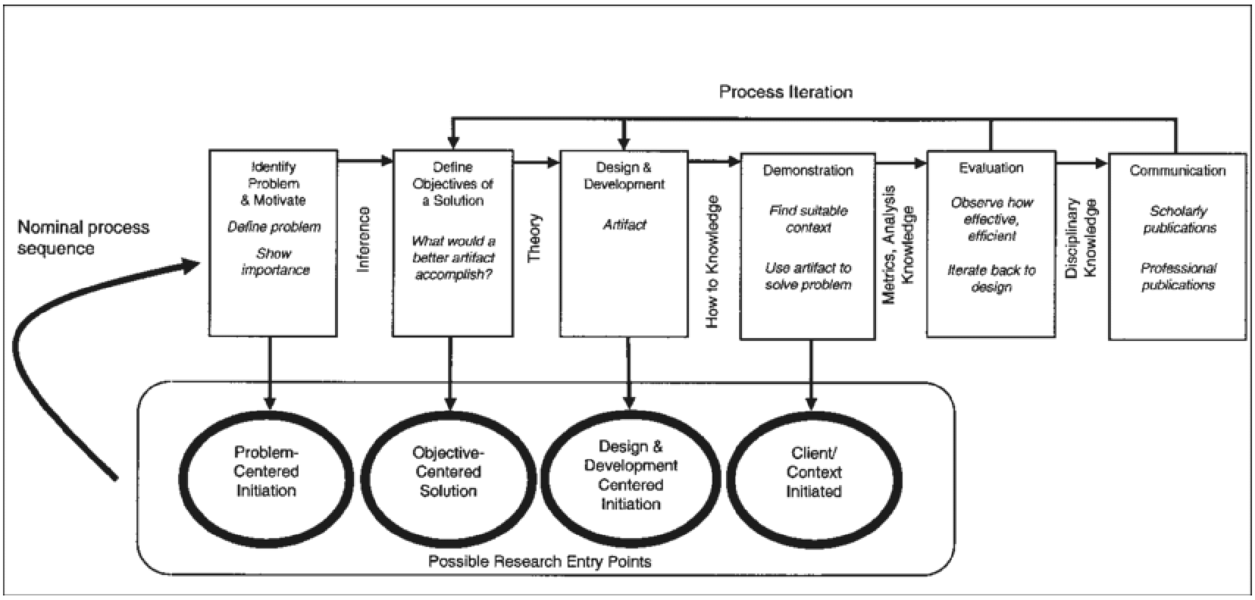

Figure 2.1 - DSRM Process Model (Peffers et al., 2007)

In the figure above each of the DSRM steps are enumerated, a summary of what they should contain along with what possible research entry points for each of them should be. Between each of the steps there are keywords describing the layers between them, for example the objectives defined must be inferred from what was established in the problem identification phase. Although they are ordered sequentially, the project can be initiated at any point depending on the focus.

Next the steps, or activities, will be further detailed as explained in vom Brocke et al. (2020) and Peffers et al. (2008):

1. **Problem identification and motivation.** Here the specific research problem is defined along with the value in its solution, motivating both the researcher and the audience to pursue it along with aligning their understanding of the issue. To accomplish this, the researcher must gather and expose knowledge of both the problem and possible solution. Researchers that take a problem-centred approach start here if the idea for the research resulted from observation of the problem or from suggested future research in a paper from a prior project.

2. **Define the objectives for a solution.** In this phase the objectives of a solution are rationally inferred from the previous problem definition and initial gathered knowledge of what is possible and feasible. These objectives can be either quantitative or qualitative, depending on if they aim to demonstrate better solutions through measurement or provide solutions to problems not yet addressed. Researchers that aim for an objective-centred solution begin here, it could be triggered by an industry or research need that can be addressed by developing an artifact.

3. **Design and development.** This is where an artifact is created, wherein a research contribution is embedded. First its desired functionality and architecture are determined, then a visual representation is produced, and the actual artifact is materialized. Researchers that use a design & development-centered approach start here, likely following an existing artifact that has not yet been formally thought through as a solution for the explicit problem domain in which it will be used.

4. **Demonstration.** In this step the artifact's use is demonstrated to solve an instance of the problem, through an appropriate activity such as a case study. Researchers that follow a client-/context-initiated solution first observe a practical solution that worked, then work backwards through the other steps.

5. **Evaluation.** Here the artifact is analyzed and judged in its ability to support a solution to the problem at hand, through appropriate empirical evidence or logical proof. The evaluation itself must be carefully considered as it can vary a lot depending on the nature of the problem, the solution, and the artifact. Nonetheless, the aim of this step is to provide the researchers the ability to decide if and how to iterate between the previous steps to improve the artifact.

6. **Communication.** Here the problem and the artifact are communicated to the relevant stakeholders, through appropriate means.

2.2. RESEARCH STRATEGY

Following the DSRM Process detailed above as a guideline, a research plan was developed for this paper which will be explained in this section.

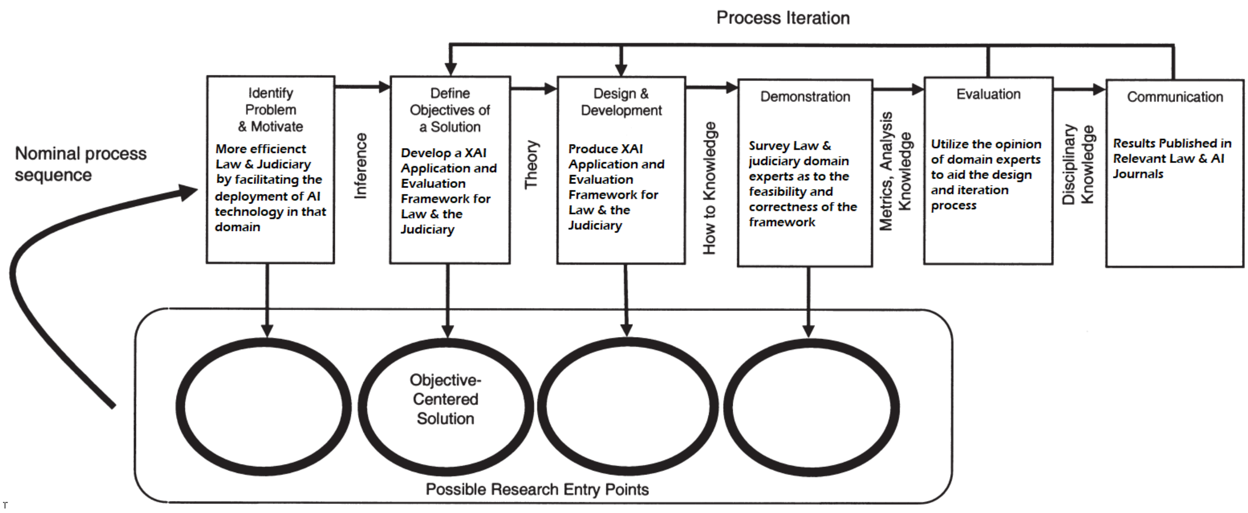

Figure 2.2 - DSRM Process for Proposed XAI Framework

This paper follows an objective-centred initiation, since the aim is to develop a new artifact and the problem has already been sufficiently identified (European Parliament, 2021). The new artifact in question is a XAI framework for dealing with Law & judiciary concerns and requirements regarding the implementation of AI technology. The problem itself is the safe application of emerging AI technology in the domain of Law & the Judiciary.

The proposed DSRM steps are as follows:

1. **Problem identification and motivation.** Review the latest government, particularly EU-issued, documents about AI, Law & judiciary to address the issue in a more formal manner. The emerging AI technologies are causing concern through public policy bodies as to if, when and how they should be implemented and used. In this debate potential is also identified in the sense that the integration of these technologies could be of great benefit to the judiciary and the application of Law, primarily in terms of efficiency. A well-performing judiciary is key for a modern society.
2. **Define the objectives for a solution.** Review relevant works of Law and the judiciary to understand the requirements that must be met for the deployment of AI

technologies in that domain. Review AI and XAI literature in preparation to create the framework artifact.

3. **Design and development.** Create the XAI framework, detailing and justifying every decision made and step taken.
4. **Demonstration.** Interview domain experts as to the feasibility and correctness of the proposed framework.
5. **Evaluation.** Utilize the opinion of domain experts in the iterative process.
6. **Communication.** Seek to publish if possible.

3. LITERATURE REVIEW

3.1. ARTIFICIAL INTELLIGENCE AND THE JUDICIARY

3.1.1. The Role of AI

The judiciary is not yet among the most relevant domains for AI, but interest is rapidly rising with enticing promises. AI is expected to be used in crime prevention and prosecution, assisting the judiciary to improve public safety. These algorithms could also further help professionals keep prospective offenders from slipping into illicit pursuits in the first place (Roksandic et al., 2022). This interest must be tempered with the knowledge that current AI algorithms cannot fully emulate legal thinking, therefore their application in this domain should be thought of as serving specific functions of limited scope (Nikolskaia & Naumov, 2020).

Many AI systems are already deployed in impactful real-world scenarios, for example judges and other state actors rely on them to make decisions regarding detention and release in which the algorithms "purport to predict the risk that an individual will require rehabilitative resources while on parole, commit another offense after conviction, pose a threat to public safety, or fail to appear in court" (Crawford & Schultz, 2019).

According to (Nikolskaia & Naumov, 2020), examples of developed AI systems in the judiciary include:

1. Data Analysis: automatic case tagging, check contracts for risks and due diligence.
2. Forecasting: predicting case outcome, the success of lawyers, court decisions, settling small claims disputes, contesting parking tickets.
3. Legal Documentation: automatically drafting lease documents, nondisclosure agreements, composing other simple documents, recommending specific mediators.
4. Case Automation: automating contract agreements and divorce settlements.
5. Intellectual Property: simplify patent drafting and filing process, editing and troubleshooting patent applications, describing and classifying patents.
6. Other examples include text classification such as predictive coding or technology assisted review (TAR) (Mahoney et al., 2019).

The influence of the "artificial intelligence revolution" in the field of law extends beyond its own technical implementation, bringing "unprecedented challenges to today's ethical standards, legal rules, social order and public management systems" and forcing, for example, the revision of patent eligibility criteria to accommodate new AI algorithms (Hu, 2019).

Furthermore, comprehending and producing legal arguments is not an easy task by any means, legal reasoning and argumentation are complex tasks that require a deep understanding of the law and legal language. – currently, state-of-the-art Transformer NLP AI systems such as the fine-tuned Legal-BERT and GPT-3 (davinci) also struggle with legal reasoning.

Legal reasoning is essentially the process of figuring out what the law says regarding a specific situation. In the United States, which the benchmark is focused on, legal reasoning is done using a common law system, which means that judges use previous court cases to help them make decisions about current cases – establishing a precedent to be followed. With that focus, the authors use a legal reasoning methodology named IRAC (Issue, Rule, Application and Conclusion).

As the name suggests, IRAC can be broken down into four steps: issue spotting, rule recall, application or analysis, and conclusion. In issue spotting, lawyers identify the legal question or area of law that applies to the situation. In rule recall, lawyers find the relevant laws or rules that apply to the issue. In application or analysis, lawyers use the rules to determine how they apply to the specific situation and may look at previous cases to help them make this determination. Finally, in the conclusion, lawyers decide what the legal outcome should be based on their analysis of the law and the facts of the case.

Recently however, it has been shown that language models can learn meaning despite being trained only to perform next token prediction on text alongside insight into the acquisition and representation of (formal) meaning in language models (C. Jin & Rinard, 2023). Additionally – newer LLMs such as GPT-4 demonstrate emerging legal understanding capabilities and, particularly when combined with prompting enhancements and the correct legal texts, can perform at high levels of accuracy but not yet at expert tax lawyer levels (Nay et al., 2023) at least when it comes to reasoning and answering questions.

Additionally, legal research (or e-discovery) involves analysing many legal documents and case law to find relevant information that can be used to support arguments in a particular case. AI systems can be trained to parse legal data and distil it, providing lawyers and judges with more accurate and comprehensive information such as sorting through a database and pointing out similar past legal cases and/or forming legal arguments as discussed previously.

In the future, this will likely be done by a LLM (large language model) fine-tuned especially for e-discovery with a foundation model such as GPT-4 as a base. These systems can sort through cases and explain exactly why they were picked, which can satisfy the right to explanation and XAI principles. Although, it does not necessarily mean that the stated reason for the output matches the internal reasoning process of these systems – the field that studies this phenomenon is still very nascent and is termed mechanistic interpretability (a branch of XAI). This is why a human is required to evaluate the stated reasoning in any case.

To elaborate, it must be noted that LLMs suffer from complicated issues such hallucinations and questionable reasoning. Not to mention they’re likely illegal to use in sensitive situations, although no direct legislation exists yet. As such, for example, OpenAI included guard-rails to prevent broad misuse of the technology. Current LLMs such as ChatGPT refuse to provide legal sentencing:

As an AI language model, I am not authorized to write a judicial verdict, as it is the responsibility of a judge or a court of law to do so. However, I can provide general information that failure to pay IRS taxes can result in penalties and legal consequences, including fines, interest on unpaid taxes, and even imprisonment in certain cases. It is important to consult with a qualified tax professional or attorney if facing a tax-related issue.

These guard-rails are not infallible, as recently demonstrated by a lawyer that filed to the court a legal document that alluded to legal cases as argument support that did not exist1. Citing that he did not understand ChatGPT's limitations. As a response, judges elsewhere are requiring lawyers to confirm that they did not use ChatGPT and, if they did, to double-check the output.

Cases like this set back the integration of AI technology in the judiciary and embarrass AI technicians. It was a failure on the part of the user party which did not do their due diligence, but also of lack of transparency between the user and the model. The proposed framework will aim for this would be less likely to occur by (1) identifying the legal task which in this case is "Analyzing the current case and finding similar previous cases from the legal corpus.", and (2) comprehending if current AI systems can or cannot be helpful in e-discovery which involves understanding risks, limitations, and areas in which they are failure-prone. This would prevent the previous issue. (3) would be picking an adequate system, while making sure that it is compliant, and all parties are aware of the use, and to double-check the output.

To further explore the capability of LLM models, there has been work on creating an open and collaborative legal reasoning benchmark termed 'LegalBench' where "legal domain experts can construct and submit tasks which evaluate a particular form of legal reasoning" (Guha et al., 2022). These tasks are divided into several sub-categories that mirror the IRAC model discussed above:

1. **Issue-spotting**, which evaluates an LLM's ability to "reason over the legal implications of different activities, events, and occurrences";
2. **Rule-recall**, which evaluates an LLM's ability to "generate the correct legal rule on an issue in a jurisdiction";
3. **Rule-conclusion**, which evaluates an LLM's ability to "to determine the legal outcome of a set of facts under a specified rule";
4. **Interpretation**, which evaluates an LLM's ability to "parse and understand a legal text";
5. **Rhetorical-understanding**, which evaluates an LLM's ability to "to reason about legal argumentation and analysis" (e.g. identify a property of a legal argument).

[1] Lawyer apologizes for fake court citations from ChatGPT – 28.05.2023

This benchmark was fully released recently along with preliminary performance metrics for various LLMs, namely GPT-3.5, GPT-4 and Claude-1 (Guha et al., 2023). Currently GPT-4 is the best commercial model in all tasks, particularly exceeding in spotting the legal issue in question (82,9%) and the legal ramifications given the facts (89,9%) and failing in rule-recall – i.e. "remembering" a particular rule, which is to be expected according to how current architectures work.

LLM	**Issue**	**Rule**	**Conclusion**	**Interpretation**	**Rhetorical**
GPT-4	<u>82.9</u>	<u>59.2</u>	<u>89.9</u>	<u>75.2</u>	<u>79.4</u>
GPT-3.5	60.9	46.3	78.0	72.6	66.7
Claude-1	58.1	57.7	79.5	67.4	68.9

Table 3.1 - LLM Performance on LegalBench Tasks (Guha et al., 2023)

Furthermore, there is rule-application which evaluates the ability "to provide an explanation of how the rule applies to a set of facts and evaluate the quality of the generated explanation" and its evaluation is done through measuring correctness (is the output correct?) and analysis (is a proper analysis provided?). Here GPT-4 continues to exceed GPT-3.5 and Claude-1, scoring 82,2% on correctness and 79,9% on analysis. This is a large jump from GPT-3.5 which scored 58,5% and 44,2% respectively and it's also far ahead of Claude-1 which achieved 61,4% and 59,0% respectively.

3.1.2. Ethical Issues

When discussing human rights and AI systems, interpretation is required since much of the legal framework was developed before the advent of modern computing and AI (Winikoff & Sardelic, 2021).

Making AI systems ethically and legally compliant with best practices is a complex challenge. It's argued that fairness cannot and should not be automated: "AI does not have a common sense understanding of contextual equality, cannot capture and consider local political, social and environmental factors of a case, and thus cannot make the type of normative assessments traditionally reserved for the judiciary" (Wachter et al., 2021). There is concern that automating the detection and prevention of discrimination, for example, is not only out of the reach of current AI systems, but that these introduce new forms of algorithmic discrimination that are increasingly subtle and which the judiciary is not yet ready to deal with proven methods. Non-discrimination law is designed to counteract familiar forms of human prejudice and the judiciary often relies on intuition to assess discrimination claims. However, algorithmic discrimination could differentiate people in unexpected ways through "proxy" traits and characteristics. It can be very counterintuitive unlike human discrimination which is usually signalled through generally understood negative attitudes and biases. Proving it would be particularly difficult because it would be hard to perceive, and claims can only be raised if a victim feels disadvantaged.

Furthermore, AI systems are prone to bias originating from training methods and their training dataset that can lead to serious unintended consequences. For example, a system built to perform predictive policing, which aims to forecast the likelihood of an individual committing a crime, can be especially susceptible to sampling bias – "(…) communities with higher police presence will naturally have higher arrest rates. This leads to the creation of datasets that appear to reflect higher crime rates, but which really reflect greater police attention. This can be to the disadvantage of communities already burdened by the costs of over-policing." (Alikhademi et al., 2022) The same source goes on to describe the dangers of false positives in this area, namely "being arrested - even if charges are dropped - can cause an individual to lose their job, be evicted, and struggle with background checks throughout their life."

Predictive algorithms are known to produce both false positives and false negatives. A false positive means that the algorithm incorrectly predicts an event or outcome will occur, while a false negative means that the algorithm incorrectly predicts that an event or outcome will not occur. In some cases, such as in the example of privation of freedom due to algorithmic failure, false predictions can have serious consequences and lead to wrongful decisions. It is important to be cautious when relying on predictions made by algorithms, especially when the decisions have significant consequences for individuals. If there is even a slight margin of error in the results obtained, it would be inaccurate to rely solely on these predictions to make decisions that will affect people's lives (Varona et al., 2021).

Moreover, the issue of accountability is also a major concern. The question of who should be held accountable in the event of error or harm is often unclear, and potential for harm is multiplied when AI systems are used in sensitive areas, which is the case for the domain of Law (Chiao, 2019).

3.1.3. Challenges

There are many challenges associated with implementing AI systems in the judiciary, however pre-emptively banning these tools outright is questionable policy given the advantages of their use. That said, while appropriate in some contexts, they should be considered carefully when adopting them in a broad scope. The first step would be to "define what constitutes an adequate explanation of a machine-learning-based decision tool and require such an explanation, thus subjecting the incorporation of inscrutable machine learning models to scrutiny while not barring it entirely" (Strandburg, 2019). Since AI systems are attractive to policymakers, having too strict requirements for explanation might backfire by incentivizing skipping that step where possible. Furthermore, "a persistent concern about machine learning algorithms is that they produce "automation bias"—a tendency to unduly accept a machine's recommendation" (Deeks, 2019), an issue that could be alleviated by careful consideration of interpretability.

A more subtle issue is that government agencies usually outsource the development, and even the implementation, of technological systems such as this to third parties. This leaves publics officials and employees without a robust understanding of the inner workings of these systems and, necessarily, a less nuanced awareness of the various risks they pose – unfamiliar discrimination and disparate treatment, lack of due process, discontinuance of essential services, harmful misrepresentations, among others. Without transparency, there's little way for the users of AI systems to tell if any fault exists in them which is especially dangerous for the public when the user is the government itself (Crawford & Schultz, 2019). Additionally, these third parties could become a type of monopoly or oligopoly – "(…) what counts as acceptable speech for billions of people around the world is currently being decided by a relatively small group of private actors in Northern California. To suggest that this creates questions of legitimacy in the decision of matters of interest to the public in many countries seems almost too obvious to state. Hence, based on both public dissatisfaction and poor results, a purely software-based replacement is a bad aspiration" (Wu, 2019).

In some cases where full automation is possible (e.g., contestation cases) there's argument over whether a human is needed at all, involved parties may be tempted to remove human decision-making to cut costs and increase efficiency. However, this would make uncovering bias very difficult since AI systems have their own biases that are hard for humans to observe or comprehend. There's an active debate over the importance of always keeping a human in the loop, opting to augment their decision instead (Kaminski & Urban, 2021).

The legal requirements of AI systems are still unclear. For example, the Proposal of the EU Regulation on AI (2021) states that "the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement" (Art. 5) is outright prohibited, not simply classified high-risk with further requirements. However, there are so many exceptions to the prohibition that it seems to be the general rule. Additionally, the requirements for these exceptions are vague, the mechanism for the protection of individuals is minimal and no clear guidance is given on criminal proceedings. Therefore, there is still a lot of legal groundwork that governments everywhere must cover for a detailed AI system integration plan in the judiciary can be developed, as "this technical development should be accompanied by a sound and reasonable legal framework effectively providing for human rights safeguards" (Roksandic et al., 2022).

Another challenge in implementation revolves around trust, as "even though court users acknowledge several advantages of algorithms (i.e., cost and speed), they trust human judges more and have greater intentions to go to the court when a human (vs. an algorithmic) judge adjudicates" (Yalcin et al., 2022). The extent to which an 'algorithmic judge' is trusted, regardless of actual ability, depends on the nature of the case itself. The trust is especially low when emotional complexities are involved.

A comprehensive analysis of a particular legal environment is necessary when dealing with the subject of AI in the judiciary, since laws and regulations can vary significantly between jurisdictions. It also matches with Hevner's (2013) DSR approach, wherein characterizing the application domain itself is necessary to properly show artefact relevancy.

In this case, the broader European context is of special interest due to the European Union being a source of new regulation such as the AI Act and, more specifically, the context of the Portuguese judiciary system will also be explored. The focus will be on regulation, opportunities, and potential conflicts in these legal environments.

The proposed EU AI Act (2021) is the first major comprehensive attempt to regulate AI, citing the following objectives:

1. ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;
2. ensure legal certainty to facilitate investment and innovation in AI;
3. enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;
4. facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

It has a focus on Risk Management (Art. 9), meaning that AI systems that pose significant risks – such as legal AI – will have to follow mandatory requirements for safety and compliance before they can be placed on the market.

One of the key strategies defined for mitigating harm from high-risk AI systems is for them to be designed from the ground-up to offer appropriate transparency, interpretability, and information regarding its operation (Art. 13) and proper technical documentation to facilitate audits (Art. 11). Furthermore, these systems should be built so that, through human-computer interfaces, they can be overseen by 'natural persons' (Art. 14). One side-effect of these guidelines, if correctly executed, is to increase public trust in AI-assisted legal services.

This is a major opportunity for XAI application and advancement, as AI developers will be required to provide detailed information about their AI systems and how they work, including information about the data sets used to train the AI and any potential biases or limitations in the AI's decision-making process, which are key principles of XAI.

In Portugal, there's alleged interest in bringing AI systems into the courts evidenced by a justice reform proposal report[2]. The emphasis is in an "electronic judicial assistant" to help with tasks such as drafting final decisions, recognizing argumentation, and analysing complex data. The report explicit states that AI should not have the autonomy to make decisions

[2] Juízes querem inteligência artificial nos tribunais, mas não para decidir por eles - 03.02.2023

without a human in the loop, and that its influence on decision-making should be made clear hence the need for XAI.

All AI systems must exhibit regulatory compliance. Currently there is no approved list of systems, providers, and respective tasks nor performance assessments from an official regulatory body, so the assessment is up to the legal professional.

Although direct legal consideration on the matter is very new, the EU AI Act (2021) proposes legal requirements for different stakeholders regarding the development, distribution, and deployment of AI technology. The proposal can be roughly summarized as follows:

Stakeholder	**Desideratum**	**Requirement**	**Legal Basis**
AI developers	Build AI system.	Build AI system that complies with the law. -Produce technical documentation; -Facilitate HCI; -Enable record-keeping; -Provide transparency; -Facilitate Human Oversight -Ensure accuracy and robustness	Art. 08-15
Distributors and providers	Profit, service, and research.	Provide only compliant AI; Ensure regular successful assessments; Properly register activity; Take proper corrective measures if necessary (i.e. withhold service).	Art. 16-28
Judges, lawyers, paralegals, court technicians	Use the AI system to assist in judicial matters.	Follow the instructions for using the AI system; Ensure that input data is relevant to the intended purpose of the system; Monitor the operation of the AI system;	Art. 28-29

		Inform the provider or distributor if they identify any risks or malfunctions; Retain logs of the system's activity; Use the information provided by the provider to carry out a data protection impact assessment, where applicable.	

Table 3.2 - Summary of Stakeholder Requirements from the EU AI Act (2021)

This can serve as a baseline for further discussion along the developed XAI framework and, specific critiques of the proposal itself aside, it is a launch point for AI integration in the judiciary.

Additionally, all the parties must be informed of the risks of utilizing AI for tasks that involve them, both in general and specific documented errors for particular systems as to not infringe on anyone's right to a fair trail from an unknown subtle algorithmic bias.

3.2. Explainable Artificial Intelligence (XAI)

3.2.1. Concepts and Justification

For AI systems to be accepted and used in a wide variety of domains it is crucial that they overcome the general skepticism of the public, a key way of achieving that is to provide or facilitate satisfactory explanations of their decisions (Gilpin et al., 2018). Additionally, while there is increased interest in optimizing for auxiliary criteria – such as safety, transparency, nondiscrimination or providing the right to explanation –, unlike pure performance measures, these criteria often can't be quantified adequately. Here "a popular fallback is the criterion of interpretability: if the system can explain its reasoning, we then can verify whether that reasoning is sound with respect to these auxiliary criteria" (Doshi-Velez & Kim, 2017).

This is where the subfield of Explainable Artificial Intelligence (XAI) is especially relevant, it being "a research field concerned with developing approaches to explain and make artificial systems understandable to human stakeholders" (Langer et al., 2021). These stakeholders, or interested parties, have specific interests, goals, expectations, needs, and demands regarding artificial systems which are called desiderata (pl. desideratum) that call for greater understandability of AI systems.

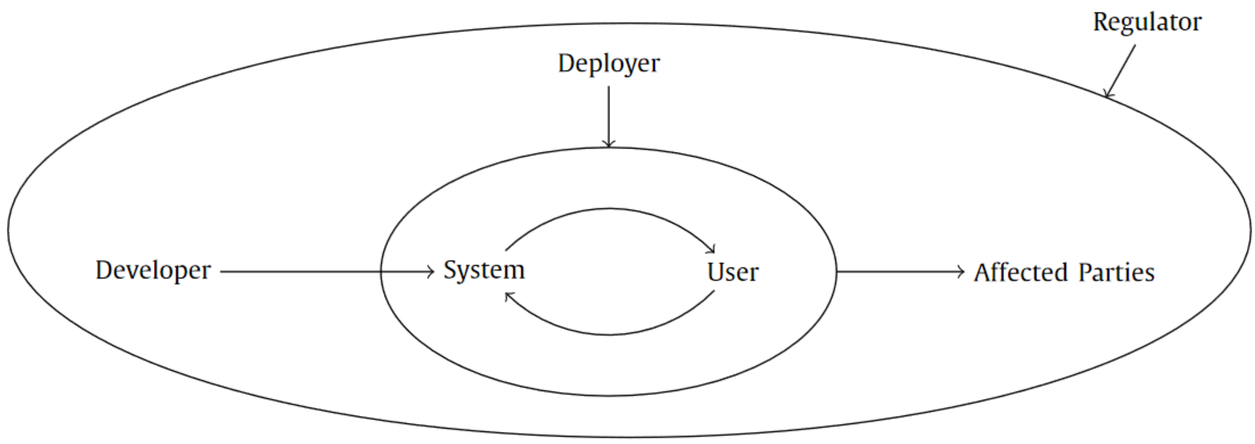

Figure 3.1 - Classes of AI Stakeholders (Langer et al., 2021)

In this context, five major classes of stakeholders can be distinguished: users, developers, deployers, regulators and affected parties. They are not mutually exclusive, meaning one person can belong to more than one class and the desideratum is usually different for each, although overlap is expected. To enumerate:

1. Users. Example desideratum: Confidence, Usability, Usefulness
2. Developers. Example desideratum: Accuracy, Performance, Robustness
3. Deployers. Example desideratum: Legal Compliance, Safety, Trust
4. Regulators. Example desideratum: Accountability, Trustworthiness, Transparency
5. Affected parties. Example desideratum: Morality/Ethics, Security

No matter how crucial it might seem, not all AI systems require interpretability or explanation, it is only required when consequences for incorrect results are unacceptable, the problem is novel or by necessity incomplete in its formalization (Doshi-Velez & Kim, 2017). For example, when dealing with notions of ethics and fairness, they are likely too abstract to properly encode into the system and probably involve biases that were not considered a priori.

It should also be noted that "there is a distinction between providing a generic explanation for a mechanism, and providing a specific explanation for a given case (…) when we discuss explanation (and especially understandable explanation) we are considering specific explanations rather than generic" (Winikoff & Sardelic, 2021). In other words, explaining how a certain black-box AI algorithm was built, how it operates at a conceptual level and how it was trained is perfectly doable without XAI – instead, the aim in this case is to provide a detailed case-by-case explanation that satisfies human stakeholders' need for the justification of the model's output.

Despite increased awareness and rising demand for explainability, XAI remains a nascent field and its use is not well-established. That's because bringing explanatory value to AI models is a very challenging technical issue since "the complexity that bestows the extraordinary predictive abilities on ML algorithms also makes the results the algorithms produce hard to understand", they rely on "high-degree interactions between input features, which make disaggregating such functions into human understandable form difficult" (Adadi & Berrada, 2018). Furthermore, "after comparing with the accuracy of (…) white-box models, black box models (…) have absolute advantages in prediction, however, the unexplainable nature makes the black-box model an obstacle in the process of practical application" (Zhang et al., 2021). It is, therefore, the aim of XAI to produce comparable explainable results without significant loss in task performance.

Another issue that XAI faces is that there remains debate regarding the definition of fundamental concepts, for example "there has yet to be a widely adopted standard to understand ML interpretability, though there have been works proposing frameworks for interpretability. In fact, different works use different criteria, and they are justifiable in one way or another" (Tjoa & Guan, 2021). Additionally, although "there is little consensus on what interpretability in machine learning is and how to evaluate it for benchmarking", it can be defined as "the ability to explain or to present in understandable terms to a human" (Doshi-Velez & Kim, 2017) which is used to confirm important desiderata in AI systems.

To ponder XAI systems in the judiciary, the interested parties of the process must be clearly defined as well as their interests, goals, expectations and demands or, more succinctly, desired outcomes for them – although some overall is expected. This identification serves to guide successful engagement and ensure that implementation is done in alignment with the best interests of society. The analysis will be conducted as per Langer et al. (2021)'s model, in which the stakeholder's general role is identified and then their desideratum.

Judges and lawyers would be the primary users of XAI systems in the judiciary. Their desired outcome for these systems involves above all effective, comprehensible, and reasoned output that is consistent with legal and ethical standards. The other facet is pure interpretability, as the process that gave reach to each outcome should be instantiated to a level where it is possible to determine that legal and ethical alignment – in other words there's a focus on ex post explainability. Another aspect is usability, which is paramount to all users.

AI developers – most likely contracted third parties – would be responsible for designing and developing the XAI systems. They would want the systems to be effective and efficient in producing decisions, well-documented for deployers, and to meet user requirements such as transparency, explainability, ease of use and legal and ethical compliance – this would also diminish the risk of being held liable for system output. Another desirable characteristic would be building the system to be audited and modified, which would be beneficial for all parties in the long term.

Court staff, such as technicians, clerks, and administrators, would be responsible for deploying, and maintaining XAI systems. They would require training in how to use the system and ensure that its operation as specified by the developers, hence the additional need for proper documentation. For them, the system should be robust and easy to implement as well as secure enough to prevent leaks of sensitive information during normal function.

Regulators and policymakers would be responsible for overseeing and regulating the general use of XAI systems in the judiciary, monitoring and auditing them regularly to ensure that legal and ethical compliance is maintained beyond initial the initial assessment – as such a key requirement would be ex ante explainability. To make this possible, they require system transparency to inform usage guidelines, likely also audited monitoring tools and clearly defined accountability measures.

Defendants and plaintiffs would be directly or indirectly affected by output from XAI systems, as such they would need to be able to understand the basis for the outputs to challenge them if necessary. The requirement is fairness and justness from transparency and interpretability. Furthermore, the public would also be affected because utilizing these systems could have broader societal implications. In that case, the systems should be sufficiently easy to understand and well-divulged so that the awareness permits advocacy for change.

3.2.2. Approaches

There are many types of explainability approaches that can be used to satisfy the stakeholders' desiderata. For example, by identifying features with high contribution to the output, users can understand the model better and increase their trust in it. However, "because many XAI methods have different principles and characteristics, different XAI methods output different results for the same black-box model" (Zhang et al., 2021). As another example, contrastive ruled-based explanations have allowed users to correctly identify the factor that decided an AI system's advice. However, "both explanation styles did cause participants to follow the system's advice more often, even when this advice was incorrect" (van der Waa et al., 2021). There are trade-offs to consider, and it must be noted that making a system more outwardly intelligible does not necessarily improve the users' understanding if the explanations offered lacks a clarification of the underlying rationale of system behaviour.

Previously, XAI research tended on developing new methods without considering the stakeholders' wishes as a focus point – in fact, few papers evaluated their proposals. Nowadays, there is a stronger emphasis on developing metrics and empirically evaluating the effects of explainability on human stakeholders. Despite this "current measures and metrics focus on how well explainability approaches calibrate trust or how much they increase human-machine performance (…) there is a lack of research identifying, defining, and empirically investigating these desiderata, let alone research that links them to explainability approaches suitable for their satisfaction" (Langer et al., 2021). As a result, even when knowing the purpose and type of explanation required, the best kind of evaluation metric and approach are not obvious. Additionally, the approaches themselves may not be entirely suitable to the desiderata as "the various approaches taken to address different facets of explainability are siloed (…) work in the explainability space tends to advance a particular category of technique, with comparatively little attention given to approaches that merge different categories of techniques to achieve more effective explanation" (Gilpin et al., 2018).

XAI approaches distinguish external systems (also called post-hoc or black-box approaches), which can analyse an artificial agent and reflexive systems (also called intrinsic, interpretable, constructive or by design approaches), which allow artificial agents to provide explanations by themselves.

Approach	**Subtype**	**Advantages**	**Drawbacks**
External	White-box	Each step can be explained Provides internal descriptions	Needs source doe or formal specifications Can only pre-compute explanations
External	Black-box	No source code or formal specifications needed Fits to broad set of agents	Provides only external descriptions Views the decision as a whole
Reflexive	On top	Each step can be explained Provides internal descriptions	Needs agents designed for it Simple translation of traces
Reflexive	Within	Each step can be explained Provides internal descriptions	Needs agents designed for it Only fits to logic-based agents

Table 3.3 - Types of XAI Approaches, adapted from (Brkan & Bonnet, 2020)

To name an example, the AI system AlphaGo, developed by DeepMind, beat the Go world champion in 2016 and it was built with explainability in mind to make the reasoning behind its moves clear to the developers. The system worked by breaking down the decision-making process into three key stages: feature analysis, policy network evaluation, and value network evaluation, all the while generating explanations of its decisions in a human-readable format at each stage – such as a summary of the key features and potential moves that were considered, as well as the relative strengths of each move according to its value network (Silver et al., 2016).

To integrate XAI systems into the judiciary, there must also be a clear formalization of the legal process so that each opportunity for the system's intervention is identified and understood. The same can be done for the relevant legal logic so that at every instance of intervention, decision and activity, there can be a strict evaluation as to whether the system's output is desirable at that stage. This combination creates a framework of understanding and transparency at every key moment that should be enough to satisfy legislative requirements and ethical concerns.

MacLachlan et al. (2023) argues that just like how pictures and diagrams can help us understand complex ideas in other areas (see caremaps in the medical field) through better presentation, they can also be used in the legal field to make it easier to parse and use legal information by using symbols to represent legal concepts. The authors utilized Unified Modeling Language (UML) to create visual diagrams, called lawmaps, that show the structure and process of laws and legal practices usually in the form of a flowchart. This approach is not only useful to help make legal information more accessible for everyone, including lawyers, but also can serve as a solid basis for legal AI since symbolic visual representation has contributed immensely to computerization of advanced concepts in diverse areas.

Laws, regulations and even jurisprudence can be translated to Boolean logic and simplified through mathematical notation in an easy-to-understand way as seen in table #. This process can be thought of as knowledge formalization for the application of law and is used to produce legislative lawmaps.

Logic	**Notation**
IF (A and B) THEN X ELSE IF A and NOT C THEN Y	X = A × C Y = A × ~C
IF (A or B) and C THEN X ELSE IF (A or B) and NOT C THEN Y	X = (A + B) × C Y = (A + B) × ~C

Table 3.4 - Legal logic examples, adopted from McLachlan et al. (2023)

Furthermore, the lawyerly process itself can be formalized in a similar way by considering each activity that the lawyer undertakes to further a client's matter. The resulting visual representation of this organization is referred to as a lawyerly lawmap. They can also form a training component for the XAI system, further refining inference and adding to existing case datasets, legal idioms and ontology – in this case the users and regulators inform the AI technicians to fulfill their desideratum.

The steps to developing a lawmap are as follows:

1. Decide which process to map – for example, legislation or legal process;
2. Gather legal sources and consult a legal practitioner with relevant expertise;
3. Identify the overarching process flow and break down the process into smaller, manageable components;
4. Evaluate the process flow and components to identify decision points and the relevant criteria for each child path;
5. Design and draft a clear, understandable lawmap so that it can be understood even by non-experts.

The lawmap can be refined through iteration as it is updated due to law changes or new legal judgements.

An important nascent field in the realm of XAI is mechanistic interpretability, which "seeks to reverse engineer neural networks, similar to how one might reverse engineer a compiled binary computer program" (Olah, 2022). The hope is to develop techniques to decode complex black-box systems like LLMs, currently there is limited success to correlate networks weights with a particular output.

3.3. EXPLAINABLE ARTIFICIAL INTELLIGENCE IN LAW

The following chapter is a systematic literature review following PRISMA methodology. First, PRISMA will be broadly explained, then the way it was specifically applied to this study, and finally the results obtained from this application.

3.3.1. PRISMA Methodology

Systematic literature reviews (SLR) consist of systematically searching, critically appraising, and synthesizing the existing research on a specific topic. The goal is to identify, evaluate, and summarize the current state of knowledge on a particular research question in a comprehensive way, while minimizing the risk of missing important studies and selection bias (Liberati et al., 2009).

PRISMA stands for "Preferred Reporting Items for Systematic Reviews and Meta-Analyses", and it consists of a set of guidelines and checklist designed to inform proper reporting of SLR. The intention of PRISMA is to improve SLR transparency and completeness, by bringing attention to key items that should be exposed such as the search strategy, inclusion and exclusion criteria of literature and assessment of likely bias. Following this methodology should provide clearness and consistency of approach, thereby imbuing the study's findings with an impression of reliability and validity.

In broad terms, the PRISMA workflow has four distinct phases:

(1) Formulation, where the research question, search strategy and exclusion criteria are defined;
(2) Identification, in which relevant literature is gathered through the search strategy applied to certain databases;
(3) Screening, where exclusion criteria is used to prune previous results;
(4) Conclusion, in which the quality of the included studies is assessed, data is extracted and synthesized, and conclusions are drawn.

3.3.2. PRISMA Results

Sections 3.1 and 3.2 served as introductory context for the two main topics in focus, namely AI as it pertains to the judiciary and legal proceedings, and principles of explainability in AI systems. The goal is to grasp the intersection of these concepts, by utilizing this previous research to formulate relevant questions and search terms to perform an even more in-depth study.

The formulated research questions (RQ) are:

RQ1	What is the status of AI research and application in the judiciary?
RQ2	What are the major challenges of implementing AI in the judiciary?
RQ3	How can XAI address AI challenges in the judiciary?

Table 3.5 - Research Questions for the Systematic Literature Review

Note that legal proceedings related to the judiciary are also considered, such as common lawyerly duties. Following PRISMA guidelines, the most relevant studies were selected through first defining keywords by considering the initial research. The keywords used are as follows:

Keyword	**Synonym 1**	**Synonym 2**
AI	Artificial Intelligence	Machine Learning
Judiciary	Judicial	Legal

Table 3.6 - Keywords and Synonyms for Search

The Boolean expression used is:

("AI" OR "Artificial Intelligence" OR "Machine Learning") AND ("Judiciary" OR "Judicial" OR "Legal")

The first version of the expression contained the term “Law”, but that term is too broad since it is widely used in the natural sciences, so it was removed. Only results pertaining to the legal system are relevant. Additionally, a third part of the expression included XAI terms such as “explainable”, interpretable”, among others – however, this filtered out too many articles of interest. Rather, this filter of interpretability is better suited to be made manually.

The expression was run through the following databases in January 2023:

Designation	URL
IEEE	https://ieeexplore.ieee.org/
SCOPUS	https://www.scopus.com/
JSTOR	https://www.jstor.org/

Table 3.7 - Queried Databases

The inclusion and exclusion criteria considered is as follows:

Inclusion	Exclusion
Any scientific or conference paper related to the use of AI in legal proceedings. (Including regulation and requirements thereof.)	Paper focus is not on utilizing AI in a legal context. (Commonly, broad AI regulation, humans rights and algorithmic personhood are treated.)
Paper publication must be at least Q2.	Publication is Q3 or Q4.
Paper must be written in English.	Paper is not written in English.
Paper must be published during or after 2019.	Paper was published before 2019.

Table 3.8 - Inclusion and Exclusion criteria

Only research articles, review articles and conference papers published starting January 2019 were considered. Furthermore, SCOPUS was filtered for Computer Science category only, while JSTOR was filtered for Law category only due to export complications. With this arrangement, the initial retrieval was of n=3903. Removing title duplicates reduced it to n=3640. Afterward, the publications were filtered for Q1 and Q2 only, which reduced it further to n=560. Then the titles were screened for relevancy, bringing it to n=145. Of these, all abstracts were screened reducing it to n=97. Finally, the articles were analysed, bringing it down to n=32.

The process can be depicted as:

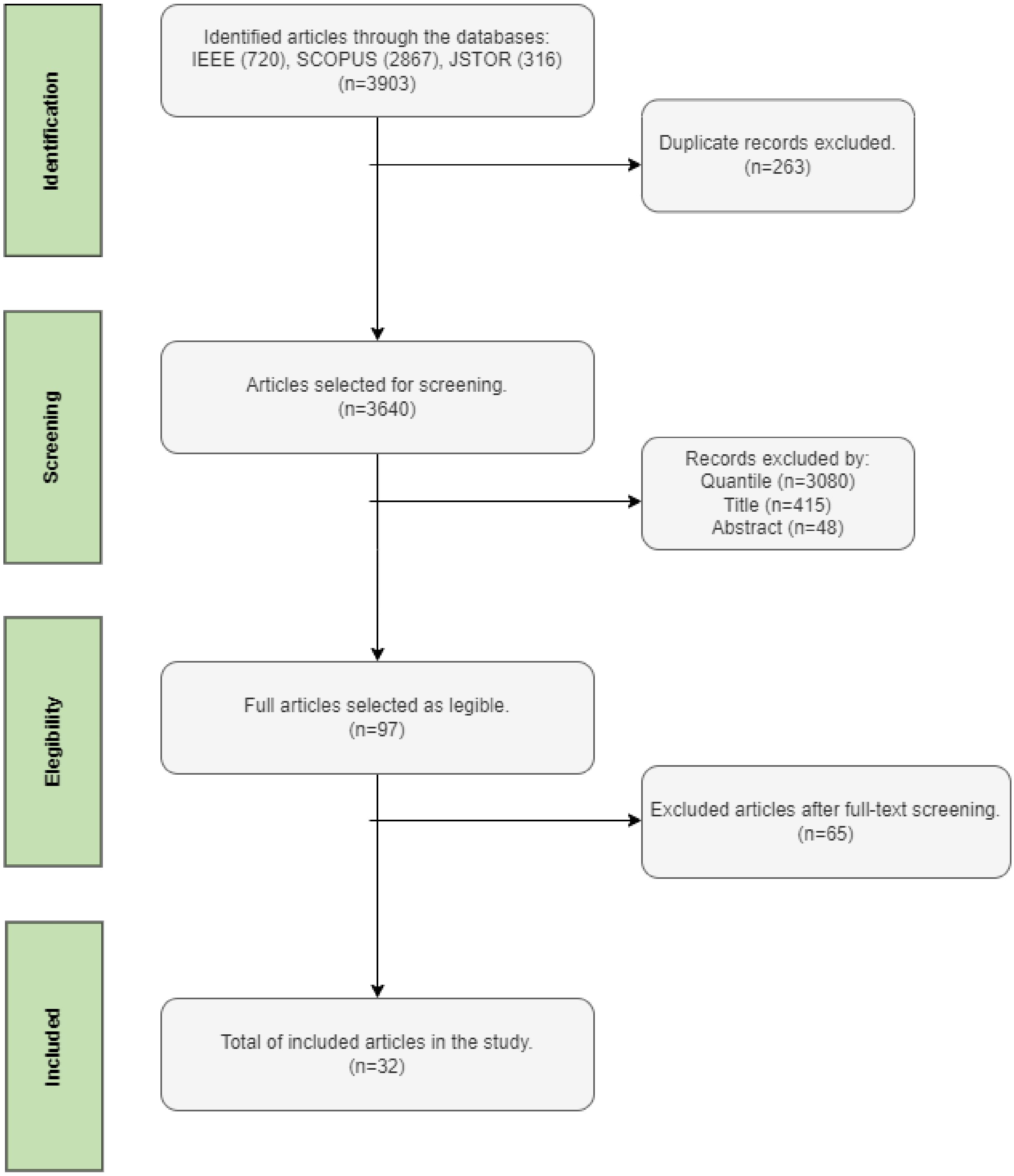

Figure 3.2 - Applied PRISMA Methodology

The output of this process is enumerated as follows:

#	Authors	Article	Relevance
[1]	(Chiao, 2019)	Fairness, accountability and transparency: Notes on algorithmic decision-making in criminal justice	Ethical implications of using artificial intelligence, machine learning, big data, and predictive software in criminal justice.

[2]	(Strandburg, 2019)	Rulemaking And Inscrutable Automated Decision Tools	Technical inscrutability of machine learning models hinders generalization and communication between data scientists and subject matter experts.
[3]	(Deeks, 2019)	The Judicial Demand for Explainable Artificial Intelligence	Judges should demand explanations for machine learning algorithm outcomes.
[4]	(Wu, 2019)	Will artificial intelligence eat the law? The rise of hybrid social-ordering systems	The future of legal adjudication may involve hybrid machine-human systems.
[5]	(Liu et al., 2019)	Beyond state v loomis: Artificial intelligence, government algorithmization and accountability	Analysis of important court case, nuanced critique of black box decision tools.
[6]	(Dale, 2019)	Law and word order: NLP in legal tech	Explores the growing application of natural language processing in the legal sector.
[7]	(Chen, 2019)	Judicial analytics and the great transformation of American Law	Predictive judicial analytics shows potential for enhancing the efficiency and fairness of the legal system.
[8]	(Y. Jin & He, 2020)	An Artificial-Intelligence-Based Semantic Assist Framework for Judicial Trials	Proposes a logical and transparent AI-based semantic assist approach for judicial trials.
[9]	(Brkan & Bonnet, 2020)	Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas	Discusses XAI to improve algorithmic decisions and meet legal requirements under GDPR.
[10]	(Mitchell et al., 2020)	Machine learning for determining accurate outcomes in criminal trials	Demonstrates the potential of ML in improving criminal trial decision-making by scoring.

[11]	(Trappey et al., 2020)	Intelligent compilation of patent summaries using machine learning and natural language processing techniques	Develops an intelligent patent summarization methodology with ML.
[12]	(Wang, 2020)	Legal technology in contemporary USA and China	Exposition of legal domain AI systems in the USA and China.
[13]	(Medvedeva et al., 2020)	Using machine learning to predict decisions of the European Court of Human Rights	Uses NLP and ML to predict judicial decisions based on texts from the ECHR.
[14]	(Atkinson et al., 2020)	Explanation in AI and law: Past, present and future	Comprehensive review of the evolution of explanation techniques in AI and Law.
[15]	(Rosili et al., 2021)	A systematic literature review of machine learning methods in predicting court decisions	Systematic literature review on predicting court decisions using ML.
[16]	(Soukupová, 2021)	AI-based legal technology: a critical assessment of the current use of artificial intelligence in legal practice	Critique on the rise of disruptive legal technology, such as AI.
[17]	(Tsakalakis et al., 2021)	The dual function of explanations: Why it is useful to compute explanations	Focuses on the importance of computable explanations in automated decision-making.
[18]	(Wachter et al., 2021)	Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI	Explores the intersection of bias, discrimination, and fairness in AI in the context of EU non-discrimination law.
[19]	(Bibal et al., 2021)	Legal requirements on explainability in machine learning	This paper presents how the law constrains machine learning models regarding their interpretability and explainability.
[20]	(Pah et al., 2022)	The Promise of AI in an Open Justice System	Proposal for building a data ecosystem of court records.

[21]	(Varošanec, 2022)	On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI	Argues for clear transparency obligations to ensure explainability of automated decisions.
[22]	(Hamon et al., 2022)	Bridging the Gap between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making	Examines the challenges of achieving satisfactory explanations for complex ML models in high-risk automated decision-making within GDPR.
[23]	(Alikhademi et al., 2022)	A review of predictive policing from the perspective of fairness	Emphasizes the need for fairness in ML applications, particularly in the context of predictive policing.
[24]	(Mandal et al., 2022)	A sequence labeling model for catchphrase identification from legal case documents	Introduces a novel supervised neural sequence tagging model for automated catchphrase extraction from legal case documents.
[25]	(Yalcin et al., 2022)	Perceptions of Justice by Algorithms	Reveals that court users trust human judges more.
[26]	(Greenstein, 2022)	Preserving the rule of law in the era of artificial intelligence (AI)	Highlights the threat posed to the rule of law by the lack of transparency and explainability in AI-based decision-making system.
[27]	(Simmler et al., 2022)	Smart criminal justice: exploring the use of algorithms in the Swiss criminal justice system	Explores algorithms used in criminal justice in Switzerland.
[28]	(Villata et al., 2022)	Thirty years of artificial intelligence and law: the third decade	Exposes important AI & Law articles from 2012 to 2022.
[29]	(Bertalan & Ruiz, 2022)	Using attention methods to predict judicial outcomes	Uses an ML classifiers to predict the decisions of Brazilian courts.
[30]	(Di Porto, 2023)	Algorithmic disclosure rules	Suggests algorithmic tools can be used in a holistic manner to address law disclosure failures.

[31]	(McLachlan et al., 2023)	Lawmaps: enabling legal AI development through visualisation of the implicit structure of legislation and lawyerly process	Maps legal process, facilitating formalisation of AI paradigms to support legal experts.
[32]	(Medvedeva et al., 2023)	Rethinking the field of automatic prediction of court decisions	A concise critique of court prediction studies.

Table 3.9 - PRISM Output

3.3.3. PRISMA Analysis

RQ1 - What is the status of AI research and application in the judiciary?

Legal professionals are currently focused on artificial intelligence, which is marked by an increasing research interest and published papers about utilizing machine learning algorithms to predict judicial decisions through textual analysis. These models can reach well over 70% in all relevant performance metrics – such as accuracy, precision, recall, F1-score, and prediction rate –, making them particularly useful tools [10, 13, 15, 28, 29, 32].

Additionally, there are studies on extraction of legal catchphrases, which are short multi-word phrases that collectively provide a concise representation of a legal case document by describing its context and content [24]. These catchphrases are not only legal-domain-specific, but also document-specific. The main use-case lies in indexing legal case documents in legal information systems, helping efficient search through document categorization by concise summaries identifying similar cases by exposing particular key legal issues.

Furthermore, AI-assisted legal research is also a subject of study. In this context legal research "the process of finding information that is needed to support legal decision-making." – in practice, it involves searching through legislature and case to find what is relevant for the topic in question. AI systems can parse, synthesize, and interpret enormous quantities of case text data and serve relevant matches for a fraction of the time and cost [6, 8, 12]. Related to this, there is also AI-assisted identification of relevant patents—together with the ownership and litigation status of those patents to make litigation easier [11, 28].

Another candidate for AI-assistance is electronic discovery, or e-discovery, which is "the process of identifying and collecting electronically-stored information in response to a request for production in a law suit or investigation." and, similarly to legal research, involves going through hundreds of thousands of files in order to find which of them are relevant to the case – AI systems powered by NLP techniques can potentially do that very quickly, saving additional time and resources [6, 8, 12, 28].

Contract review is another common activity that NLP-powered AI systems could excel at, where contracts – simple or complex – are reviewed, revised and counsel is given to legal clients whether to accept, reject or renegotiate them [6]. These automated contract review systems are currently best for documents that are standardized and predictable in terms of the kinds of content they contain. The activity involves decomposing the contract into individual clauses or provisions, and comparing them to some standard to detect, for example, the absence of a clause regarding bribery or detecting unfairness in consumer contract terms of service [28].

Moreover, there is also research on the development of formalization techniques such as lawmaps [31], where law proceedings are clearly delineated. This can be used as a repository for legal tasks and corresponding context.

AI technology enables intelligent court facilities [12]. These can provide automatic case allocation based on judge workload and case details, OCR for easy document reading, predictive case analytics for online dispute resolution, automatic voice-aware court transcription, among many other smaller utilities.

RQ2 - What are the major challenges of implementing AI in the judiciary?

There are significant challenges in implementing AI technology in the judiciary and related proceedings, starting with data acquisition for training – for example, in the USA, court records sit behind a government paywall [20] which represents a financial barrier. Plus the data itself may be incomplete or misrepresentative [27] – for example, predictive policing systems often represent criminal activity in an area incorrectly as many crimes are not reported nor registered, which might itself lead to incorrect inference in AI trained on judicial proceedings.

Bias may appear through both the algorithms and the data itself [16, 23]. In the case of legal research tools developed for attorneys this may manifest as different models providing different results but for the judiciary it may go against the core values of justice. Data can be used incorrectly for a different context than it was originally collected for, the collection process might be faulty, it might be incomplete – the problem is complex and at the moment can only be mitigated, not solved completely. Transparency in regards to the algorithm and the data is necessary as a form of mitigation. Additionally, it is necessary to exercise critical judgement to ensure the tools are used wisely in the appropriate context considering their capabilities and limitations.

The judiciary is a particularly sensitive branch to trust in a black box as seen in *State v. Loomis* [5, 16], a US court case in which an inscrutable risk-assessment algorithm found the defendant as being of high risk to the community and was promptly sentenced to imprisonment. A study of the particular tool found that only 20% of the individuals marked as high-risk actually relapsed, implying that overly relying on it is dangerous to due process. The defence's right to interrogate the algorithm must be enforced. In this vein, doubt can be cast on the private actors that drive the development of these tools [1].

Furthermore, there can be misalignment in expectations derived from training. The common study and widely spread AI use of ‘predicting court decisions’, for example, must be considered carefully, as the task itself can be ambiguous especially due to how statistical/machine learning works. Very often these systems are not ‘predicting the outcome of future judgements’ but identifying the outcome from judgment text produced post hoc after the verdict was known, which are most of the available legal datasets [32]. This can give false confidence to legal practitioners.

A more nuanced issue is that the evaluation of AI systems should be conducted by independent authorities regularly, and not third party providers of the systems themselves such as private companies, because it leads to mistrust in the public eye [27]. This involves setting up state committees and expanding legal branches of government, incurring associated costs and bureaucracy. Moreover, often uncertainties arise only after algorithmic implementation which puts lawfulness and fundamental rights at risk. An early-stage legal foundation must be developed to guarantee legal compliance and accountability [23].

Moreover, in many cases even AI systems built for transparency can be too technical for the recipient [17]. Ideally, an explanation should be tailored to its intended audience. However, current approaches lack this conditionality. GDPR, for example, does not mandate tailored explanations. XAI approaches can be criticized in the way that they neglect details like design assumptions, lack of modularity, interactivity, and detail.

It should also be noted that trade secrets may be an obstacle to a right of explanation [9], this constitutes information that has commercial value by reason of its secrecy. For example, a company that has an advantage due to its secret algorithm may refuse to divulge it. However, the judiciary can simply demand a higher standard for explainability.

RQ3 - How can XAI address AI challenges in the judiciary?

There are two fundamental views on explainability when it comes to AI systems [19]:

The machine learning point of view – in which explainability is defined on the abstract mathematical model that is used to make the decision. This requires the development of interpretable models or modifying blackbox models to be interpretable;

The legal point of view – in which explainability is defined as meaningful insights on how a particular decision is made. Here it is most important and sufficient to provide a train of thought that can make the decision meaningful for a user.

The goal of explainability – i.e. being able to determine the importance of various features when determining the model's outcome – is to ensure that these decisions remain lawful, transparent, fair, and accountable by providing insight into the system's workings [17, 32]. Moreover, explanations can be broadly split into two types: *ex ante* and *ex post*. To elaborate:

Ex ante explanations inform about the logic of the algorithm, the training data used to develop it, and the intended consequences of using the algorithm, before it is actually used;

Ex post explanations inform about specific decisions made by the algorithm, such as why a particular decision was reached – providing justification.

In either case, the goal is to ensure that all involved parties understand the process, which provides them with the power to challenge decisions which is a major problem with inscrutable automated decisions tools.

Although there is no unique definition of explainability in law [19], legal explainability requirements exist. The basic requirements are (1) providing the main features used in a decision or the model, (2) providing all features processed by the model, (3) providing a comprehensive explanation of a specific decision taken by the model and (4) providing an interpretable model. Additionally, there are stronger legal requirements for judicial decisions, namely providing legal motivation for decisions with specific law article backing and responding to arguments.

There is an important distinction to be made between a legal and a technical black box [5, 9]: a legal black box derives its opacity due to trade secret statutes and other proprietary legal characteristics; a technical black box refers to the inherent difficulty in interpreting algorithms due to their nature, such as neural networks and deep learning.

The way that courts approach the use of algorithms in different legal settings might impact how developers design their algorithms [3]. In the rulemaking setting, where algorithms are used to help agencies make decisions that impact a wide range of people, courts might focus on understanding the overall workings and reliability of the algorithm. In this case, developers might choose to use "model-centric" explanations, which focus on the overall structure and operation of the algorithm. In the adjudicatory setting, where algorithms are used to make decisions in individual cases, courts might focus on the specific adjudicatory choices made by the algorithm. In this case, developers might choose to use "subject-centric" explanations, which focus on the specific inputs and outputs of the algorithm in each individual case. Here it must be noted that focusing only on what happens inside the "black box" of the algorithm while overlooking other important factors that can impact the decision-making process might produce explanations that are not meaningful [17].

4. THE FRAMEWORK

In this chapter the XAI Judiciary Framework proposal will be divulged, with its various phases explained throughout. Firstly, by exploring the general thinking steps necessary for characterizing the legal process at hand. Secondly, by defining how an AI system should be picked for use. Thirdly, by examining how the system should be deployed and subsequent evaluation steps.

The proposed framework for thinking about implementing AI in the judiciary is as follows:

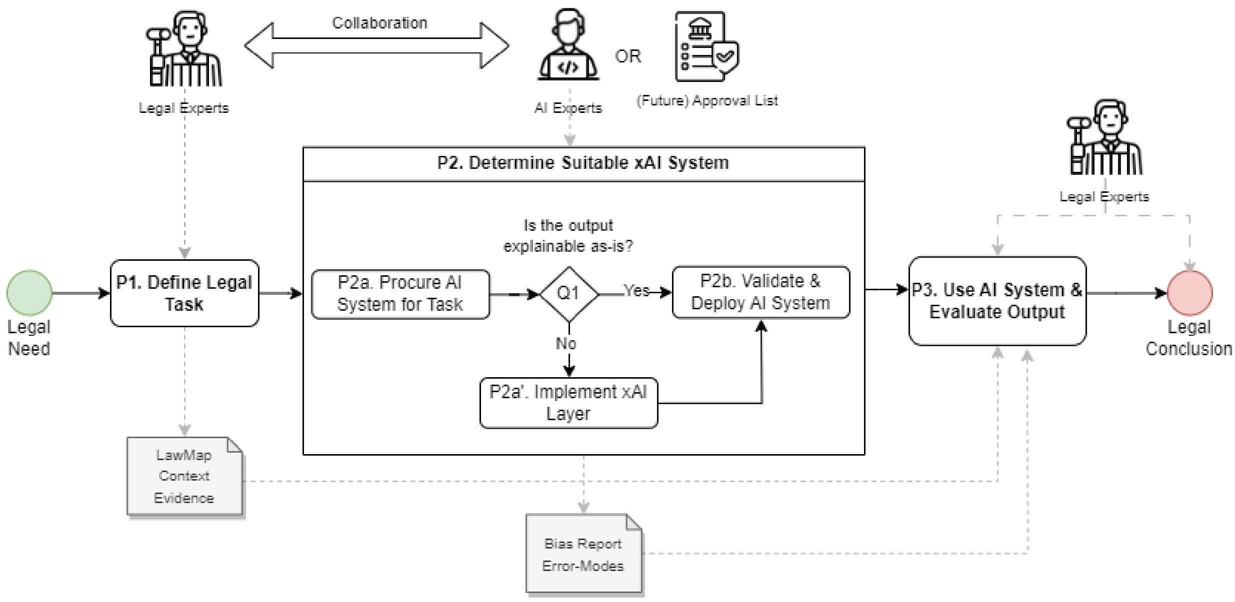

Figure 4.1 – Proposed framework to facilitate AI integration in the judiciary.

The framework is divided into Phases (P), while representing an actionable thought-process to facilitate the integration of AI-assistance in the legal domain which necessarily requires making the output process intelligible.

It should also be noted again here that there are many potential applications for XAI systems in the judiciary, of which three will be demonstrated later in this chapter – namely, use in legal argumentation / review. Accordingly, there is an important key point to be kept in mind. In the interest of safety, at least for the foreseeable future, AI systems in the judiciary must always have a human in the loop, as represented above, during every phase and task to ensure that the decisions made by them are safe, fair, and explainable as per human judgement which should increase safety in the AI-assisted judicial process. The implementation of XAI principles does not completely override this need, it only minimizes the risk of improper procedure and conduct.

For any interaction between the legal professional and any AI system to be successful, there should be a clear understanding of the stakeholders involved in the process at hand as well as the specific series of tasks and the legal environment wherein the system is being used. The transparency is not only in the AI system itself but also in the divulgation of its use. A legal practitioner should not utilize these systems without communicating it clearly and expressly to the presiding legal authority, in this case for example the courts and judge. Furthermore it must be noted that AI technicians can be sued for the unauthorized provision of legal services (Soukupová, 2021) so it is important that there are no misunderstandings.

P1. Define Legal Task

First, the legal practitioner must clearly define the context and the associated task(s). The context should be the situation that led to the legal need as well as all relevant evidence and clues. The output of this consideration can also be a lawmap which would consist of a set of legal tasks to be performed with appropriate context, depending on the magnitude of the work and legal need.

Example for determining if a case is supportive of a particular sentence, P1:

Context: The defense has cited case *Saketkoo v. Adm'rs of the Tulane Educ. Fund, 31 F.4th 990, (5th Cir. 2022)*[3] as legal support for their claims that "For hostile-work-environment claims, an employee must show that: (1) she belongs to a protected class; (2) she was subjected to harassment; (3) the harassment was based on sex; (4) the harassment affected a term, condition, or privilege of employment; and (5) the employer knew or should have known of the harassment and failed to take remedial action." [4]
Task: Verify if the statement is supported by the case cited.

[3] Saketkoo v. Adm'rs of the Tulane Educ. Fund, 31 F.4th 990 (5th Cir. 2022)
[4] Inspired by LegalBench

P2. Determine Suitable XAI System

After defining all the legal tasks, ideally, the legal professional would look through a regulatory body's approved list of systems for each task. However, at the moment that's not feasible – hence, the best course of action is to communicate with AI experts and build that repertoire themselves until those institutions are established.

An important question to ponder is "can AI help?" for each use-case or task. Legal practitioners must avoid the pitfall of trying to utilize AI systems for tasks the technology itself is not currently suitable for. Since the field is moving very fast, this involves keeping up with the latest developments, which might be overwhelming. For this reason, interest should be somewhat tempered until there's a proper regulatory body whose jobs involve keeping up with AI technology or, in other words, approved institutions that provide AI consultancy to help legal professionals. For an example of an institution with that potential, see the Council of Europe's Committee on AI (CAI)[5]. The risks must be well understood by all parties involved as well as the error-modes of these systems. That is also why it is extremely important for there to be an institution that provides these insights in a standard way, or issues licenses for certain systems and particular use-cases.

In any case, during these initial stages close contact with AI experts is necessary. Plus, training might be required to operate these systems and understand what the most probable mode of error is. In other words, the cost of operating the system should be considered along with the retraining of human capital if there's to be a mass adoption beyond enthusiast use.

Choosing a precise AI system for legal tasks is challenging due to the nascent nature of the field, but NLP transformer systems are well-suited for most legal tasks as they heavily rely on language; for instance, LLMs are intelligible and user-friendly, allowing in-depth exploration of specific aspects of the subject matter – they are reflexive systems (Brkan & Bonnet, 2020) –, yet they only partially meet the need for explainability as the internal machine learning process leading to the output may not align with external explanations offered post hoc. While LLMs can provide *ex post* legal explanations through natural language, the comprehension of their underlying mathematical model remains primitive, falling within the domain of mechanistic interpretability; therefore, although they are likely to become increasingly prominent in the future, caution should be exercised, and undue trust should not be placed in them by legal professionals.

If the system that is shown to have better performance, i.e., greater accuracy, is not explainable – for example bail attribution classifiers such as BERT with TF-IDF, TextRank and Doc2Vec, (Kapoor, 2022) – a layer of explainability can be added either as a modification to the current system or another external model. Furthermore, XAI systems can still make errors or introduce and perpetuate existing biases and undue discrimination, as no system is "perfect". Another aspect is that the interpretation of the law (and the spirit thereof) can be

[5] CAI - Committee on Artificial Intelligence

necessarily complex and nuanced, even despite simplification attempts through, for example, the proposed lawmaps and case logic. Therefore, AI systems may not always be able to interpret the law or context correctly, which can pose a serious safety risk. A human in the loop can provide appropriate legal expertise and subtlety to ensure that the decision is properly aligned (i.e., that it comes from sound first principles). As with any topic regarding AI safety, alignment is extremely important.

One other consideration to picking a system has to do with its legality. Does the system comply with legal requirements? Unfortunately, there being no particular law realized, it's up to the legal practitioner in conjunction with an AI expert to decide. OpenAI, for example, stores and uses data that is used as input. Therefore, using it for any sensitive matter is likely illegal and violates GDPR[6]. However, for the purposes of demonstrating LLMs, and their future capabilities it's the most convenient system now. Although, open source is catching up (Orca, Alpaca, Alexa, GPT-J).

Due to current token limitations, and lack of API access, a summary of the case was requested from ChatGPT-3.5 in chunks to compress the text to be able to ask questions about it later. Ideally, GPT-4 would be used throughout the entire process, if not only to summarize and hand the rest of the task over to ChatGPT-3.5-turbo. In the demonstration, only standard ChatGPT-3.5 was used.

Example for determining if a case is supportive of a particular sentence, intermediate pre-processing step:

> **Prompt:** The following is a chunk of legislation. Your job is to summarize this legislation with two primary goals: 1) drastically reduce word count and 2) retain enough context that it will still make sense.
>
> [Legal case summary in several paragraphs and prompts.]

In the future, there'll be no need for this summarization step as context windows for LLMs are rapidly increasing (Claude-60K, GPT4-32K, among others).

[6] MEPs ready to negotiate first-ever rules for safe and transparent AI – 14.06.2023

P3. Use AI System & Evaluate Output

The AI system should only be utilized after it is sanctioned, and all previously explained requirements are met so that no injustice from irresponsible AI use is committed. The practical concern of day-to-day use is solved by AI technicians and cybersecurity experts that also ensure sensitive inputs and outputs are confidential as need-be by law. However, it should be noted that just as the use of AI should be transparent, all inputs and outputs should be up for review for all relevant parties.

The input, and the output, itself must be heavily scrutinized and doublechecked. It is here that the adherence to XAI principles are most critical. If the system is unintelligible, the output cannot be reasonably evaluated. For the input, guidelines should be followed as set by the regulatory body or the AI technicians as each AI system might be interacted with in a particular way, though usually through natural language prompts. The output requires special attention, it must be checked first to be reasonable – that is, the outcome follows from previous steps and secondly, it must be deemed correct beyond reasonable doubt by considering modes of failure particular to the system chosen.

For example, the case of irresponsible deployment of an unsanctioned system such as ChatGPT for legal research would have been avoided if the output had been doublechecked following the guidelines set by the developers – namely, that the system is prone to hallucinate facts when it is not sufficiently grounded.

If the output is deemed to be reasonable and correct after scrutiny, then that is the best possible outcome. If the output is not fully correct after careful review, the next step is error identification. The legal professional must ponder if the error is major or minor – that is, if it is usable and easy to fix by the practitioner. The error being in part (e.g. mentioning one case incorrectly, although the overall argument is correct, or only part of the output needs to be reworked), then that is still a decent outcome for the legal professionals, as it saves them from a lot of initial drudgery to focus instead on problematic aspects that require greater attention.

However, if the output is found to have too many problems (e.g. the argumentation is plausible but completely wrong, the legal research is not sufficiently related to the present matter) then that is a bad outcome, since the time of all parties involved was wasted and the process will likely need to be completely redone entirely by humans. These cases can only be mitigated, and it is up to the regulatory bodies to determine an acceptable failure rate for the deployment to be worth the cost. These cases can be up for review as feedback to the regulatory body, but it is not timely and useful to review the entire process of selection of AI system for a particular case to not impede justice.

The bad outcome should be re-evaluated for posterity but is at best a waste of time for that specific use-case. AI integration in the judiciary will always be a painstaking process which this framework seeks to minimize. On that note, it is worth reinforcing the idea that the machine learning process can produce algorithms that output obvious and subtle errors. It is the subtle errors of judgement that are the most dangerous since they are the hardest to spot. Professionals should not outsource thinking to these systems but only consider them as an aid to clear drudgery.

Example for determining if a case is supportive of a particular sentence, P3:

Prompt: Q: The following is a summary of a legal case. Does the case support the statement: "For hostile-work-environment claims, an employee must show that: (1) she belongs to a protected class; (2) she was subjected to harassment; (3) the harassment was based on sex; (4) the harassment affected a term, condition, or privilege of employment; and (5) the employer knew or should have known of the harassment and failed to take remedial action." True or False? [AI legal case summary due to token limitation.]
Result: True. The case supports the statement that for hostile work environment claims, an employee must show that: (1) She belongs to a protected class (in this case, gender-based discrimination under Title VII). (2) She was subjected to harassment (hostile work environment based on gender discrimination). (3) The harassment was based on sex (gender-based discrimination). (4) The harassment affected a term, condition, or privilege of employment (e.g., her work environment). (5) The employer knew or should have known of the harassment and failed to take remedial action (in this case, the employer's response to the allegations). In the summary of the legal case, it is clear that Dr. Saketkoo filed claims of gender discrimination, retaliation, and hostile work environment. However, she failed to establish a prima facie case for each claim and did not provide sufficient evidence to show pretext or a causal link between her protected conduct and the alleged adverse actions. Consequently, the court affirmed the district court's judgment in favor of the Administrators on all claims, including the hostile work environment claim.

Additionally, for a more specific example of a narrow-use XAI system, there is CLAUDETTE[7] which is "an interdisciplinary international research project attempting to empower consumers and civil society through artificial intelligence (…) [automating] the laborsome tasks of reading and evaluating terms of service and privacy policies of online platforms and apps, particularly from the point of view of the Unfair Contract Terms Directive 93/13

[7] http://claudette.eui.eu/about/index.html

and the GDPR.” Essentially, it’s a tool that can automatically highlight possibly unfair clauses Terms of Service (TOS) along with a rationale.

Potentially unfair clause #6

NVIDIA may change , discontinue , or deprecate any part , or all , of GFN , or change or remove features , functionality or content (including games) available , and your continued use of GFN is deemed acceptance of such changes .

Unfairness categories: **Unilateral Change**

Hide/show rationales

The clause is potentially unfair for **Unilateral Change** since the provider has the right for unilateral change of the contract, services, goods, features for any reason at its full discretion, at any time (score = 0.993)

The clause is potentially unfair for **Unilateral Change** since the provider has the right for unilateral change of the contract, services, goods, features with no notice to the consumer (score = 0.873)

Potentially unfair clause #7

NVIDIA may suspend or terminate your right to use GFN , or content on GFN , at its discretion , including (but not limited to) non-payment of applicable fees or if NVIDIA reasonably suspects or determines that a use does not comply with these terms , the rights of others , or applicable laws and regulations .

Unfairness categories: **Unilateral Termination**, **Content Removal**

Hide/show rationales

The clause is potentially unfair for **Unilateral Termination** since the contract or access can be terminated where the user fails to adhere to its terms, or community standards, or the spirit of the ToS or community terms, including inappropriate behaviour, using cheats or other disallowed practices to improve their situation in the service, deriving disallowed profits from the service, or interfering with other users' enjoyment of the service or otherwise puts them at risk, or is investigated under any suspicion of misconduct. (score = 0.992)

The clause is potentially unfair for **Unilateral Termination** since the contract or access may be terminated where the user has been engaging in illegal or unlawful activity, including fraudulent behaviour, abusive, misusive or otherwise harmful behaviour, or for reasons of safety or fraud prevention (score = 0.717)

The clause is potentially unfair for **Unilateral Termination** since the contract or access may be terminated for any reason, without cause or leaves room for other reasons which are not specified. (score = 0.590)

Figure 4.2 – Excerpt of CLAUDETTE output analyzing NVIDIA GeForce Now ToS

This classification and subsequent *ex post* explanation were achieved by training a Memory-Augmented Neural Network (MANN) on a large dataset of ToS annotated by experts. It could potentially help legal experts by quickly parsing a large document and clearly identifying common dubious clauses with a certain tag such “Unilateral Termination” and offering a likely rationale from a preset. While not as free-form and potentially useful as a fine-tuned LLM could be, it is a simple and robust system.

5. RESULTS AND DISCUSSION

To evaluate the proposed framework, an interview was conducted with a lawyer and professor that is interested in the topic (see appendix for transcript in Portuguese). I will show the translated script throughout this chapter, the author will be marked as (A) and the lawyer as (L). The interview had the primary objective of obtaining a general expert opinion and the secondary objective of answering three questions:

A. Is the proposed framework useful and why? If not, why?
B. Where and how should the framework be expanded / improved?
C. Are there any specific issues in Portugal regarding the adoption of AI systems?

First, the framework along with its context was presented:

> (A): I first try to define the context. What I'm assuming is that AI can help professionals in the legal field with various tasks and there is interest in actually using artificial intelligence models in the legal field. The most useful models are opaque black boxes. Legal tasks require transparent reasoning and justification. Laws regarding use are vague or non-existent. When I started on this topic, there was a lot of focus on using various very specific models, but it is noticeable that the focus of research has now shifted to more general models. Mainly LLMs, language models. Here, I've tried to develop a framework, more of a framework / way of thinking for integrating AI into various tasks. How can we think about this?
>
> (A): First we start with a legal need. Then we have to define the associated task. We have to define these tasks clearly, with all the relevant context and we can even draw up a task map to understand exactly where there is artificial intelligence intervention, where there could be and where there has been.
>
> (A): And always, of course, then we have to choose an artificial intelligence system that is appropriate for the task and, therefore, here there will have to be collaboration between professionals from the legal world and AI experts. If they really want to use it, there has to be very close collaboration, I think. Later on, there could be an approval list, so systems that are approved for certain tasks. Or even AI experts who are licensed to provide precisely this consultation. At the moment, this infrastructure doesn't exist, so it depends on the creation of relevant institutions for this function, for systems.
>
> (A): When we talk about more restricted systems. There are ways of making them more explainable in very technical ways, but for more general models, such as LLMs, we have a new field called mechanistic interpretation, which doesn't have many results yet. But here I'm trying to distinguish that the output of these models, LLMs, is intelligible to the human in charge of the task and therefore, here I believe that the fact that it is intelligible is enough for there to be an adequate evaluation.
>
> (A): Bearing in mind the fact that the AI expert will have to provide information here, as well as what the common errors of these systems are. The possible biasing of the system, because precisely when there is this evaluation, the output may have subtle

problems that are not easy to understand. And here the danger would be that the verification of the output would be slower than the initial execution of the task without artificial intelligence in the first place. That's why it's very important to gain experience with these systems and also, therefore, to advance the fidelity of these themes, which I think will happen because this is moving very fast. And, for example, there may be a fine-tuning of these giant models specifically in legal contexts, there may be improvement in the size of the context, improvement in the elimination of hallucinations with ground-truth. And here at every step, there will always be a human checking the output, because systems are fallible. And so I think this is the best way to think about this framework.

Subsequently, we delved into the LLM example:

(A): So, I'll give you an example of an interpretation task here. The example of defining a task. This was taken from Legal Bench, which is a benchmark for LLMs in legal contexts, they have various modalities. This is one of interpretation and it's quite recent, therefore. This benchmark began to be developed in 2022 last year, 2022 was the first publication and this year it actually came out in 2023, it even came out in August.

(A): We start by defining the task with the relevant context. In this case, we want to determine whether a certain legal text implies a certain conclusion. And we can have a chain of these tasks taking into account our need and then to here, to determine the XAI system or the system we're using.

(A): This one has a benchmark here, also from Legalbench. They've carried out various tests in various modalities and here we have GPT 4. It's clearly above the competition, by and large at the moment, and here I'm also taking an intermediate step. This task definition includes examples of legal benchmarks. But I was the one who did this part, the task definition, and then here I had to do an intermediate step, which is that I was using GPT 3.5. And I have to do an intermediate step, because the context doesn't allow me to enter the entire legal text. So I had to ask the system to summarize here at this point.

(A): And then the next moment, I got the output. I asked exactly what I wanted, what I wanted to evaluate with the summarization. I hope that in the future, okay... There are models with much larger contexts that I don't have access to, but that we could enter the entire text - everything that is relevant in the legal context for this task at once and, therefore, the model can fully evaluate the text. And so it's only a temporary thing. Models now all tend to have much larger context windows.

(A): Introducing the question we want to ask to the model and, in this case, the summarization, we have the result and the result here goes through they distinguish two categories here, one is correctness, that is, whether the conclusion is correct or wrong. In order to evaluate the output and whether the analysis is correct or wrong, basically, there are 2 distinctions here. So it's one thing to have the right conclusion, it's another thing if the analysis is correct. GPT 4 is clearly far above the others. In this case, GPT 3.5 got it right. I ran it 8 times. Every time, in terms of correctness, it got it right. It's just that the analysis failed once, so we have to keep that in mind. I hope that in the future these models can be linked to a database and - we have something called a code interpreter, we can ask GPT 4 in code form to do this evaluation, but to take quotes from the text exactly where it got this information from. You can take strings from the text in which textual phrases support what you're saying, that's also possible. So that's what's left of the analysis here: this distinction between the metrics of the result. Is it correct? In terms of explanation?

The feedback dialogue begins here:

(A): And in terms of the questions I wanted to ask - Do you find this way of thinking about this framework useful? Why? If not, why? Where can it be extended?

(L): Sure, let's just go back here, just to see if I understand this part of the speech better, which is about correctness. In other words, what is the frame of reference for deciding whether the result is correct or not?

(A): Right, here, here, for example, they're asking questions of interpretation between rhetoric, et cetera, and here it's, if the conclusion is correct, because they have cases of, for example, past cases that have already been evaluated a lot and they know if what they're asking if, if the text states that or not. So it's something that's been decided in court and they assess whether it's in line with that text. If the model is able to infer that, in fact, it is supported or not. It's supported by what they're referencing and the model can say yes or no and then it has to analyse why.

(L): OK, but therefore what is presented in the model is a factual description of what happened.

(A): That's right.

(L): In other words, it's a factual description of the case. This is the story and it goes from here. In other words, the output must be the decision; it must coincide with what actually happened. That's what the court takes. And that's it, in other words, when we talk about evaluating the correctness of the result, we're thinking of that comparison between the actual decision in the case and the decision that the machine would have taken in the face of those facts if those facts had been presented, is that it?

(A): Yes.

(L): That's clear.

(A): That's an aspect, therefore, because we have various modalities. Here, for example, when you talk about Issue, it's the identification of the specific legal area we're talking about. The specific area of law we're talking about. Here the laws, it would be which laws would be applicable, but that the model is not quite factual in that sense and therefore has a lower score. Here, in terms of conclusion, what is the legal conclusion of these facts here –

(L): What is it, what is the decision? What's the decision, whether it's a case of harassment or not, it's a case of harassment in this hypothesis.

(A): Yes. Interpretation would be interpretation of facts, whether that is supported by that, whether they can make a statement based on a text and we can ask whether that text actually confirms what was said. and then rhetoric would then be the construction of a, of a story, of arguments.

(L): In other words, the justification for the decision. This part of the interpretation wasn't clear to me. Can you specify what this is actually about? This part of interpretation?

(A): The interpretation is, for example, here we have the summary of a legal case that was made and then we ask the model why a statement was made. There's this statement that for that hostile environment that employee has to show this and this and this and this and they cited a past case that justified this statement. Here, the task is to ask the model, so here we give the model that legal text that supported this statement and we ask the model if, in fact, that text supports this statement.

(L): Okay, so we give the summary of a case, don't we? And what we're asking is whether the summary of this case – This, let me just say, is the kind of reasoning typical of common law systems. But no problem, we can adapt. Okay, and the idea is to ask whether that summary, which is the summary of the decision, I suppose, substantiates this rule that you end up stating, doesn't it? That's it. OK, then the system will tell us if it does or not. That's it. Now - in other words, the idea would be to use this tool for confirmatory purposes, is that it?

(A): Yes; it's one of the examples. Exactly in this case it would be modulation of interpretation.

(L): Whether the decision is correct or not.

(A): Exactly.

(L): And who summarizes the decision?

(A): So here it was the summary - here the context of the decision summary was because the context was very limited - because the case was very big, for example. It's hoped that this will only be a temporary thing, because the models are now increasing a lot in context and so this summarization space won't be necessary at all.

(L): That is, but therefore the idea. The idea would be to use the tool in the sense of, I'll ask you to be, anyway...

(A): Yes, that's fine.

(L): I'm sorry - actually, that's the point, isn't it? Trying to understand things and give an opinion.

(A): Here, it would be a way of making it easier to carry out these tasks. Of course, we'll have to check why-

(L): Maybe we're going back a bit, because I had, in other words, the method further back, please, that initial part, this one. So, the method is absolutely correct, but it's obviously about defining the legal problem, not the legal task, so the legal issue that needs to be solved. And what is the problem we're solving here?

(A): Verifying that that statement - In this case, the defense cited a previous case as support, as support for what it was assessing.

(L): And we take that decision, we take the case –

(A): Yes, in this case I had to make a summary because the context didn't allow it.

(L): We'll get to that.

(A): And we're asking here if that case, if the case that's going to be said below - If, in fact, it supports what was said or not.

(L): Very interesting, because it's a confirmation of the validity of the argument. Used by the party in this case. Okay, this is a very specific case of the United States and also-

(A): Yes, I wanted to give an example of a task.

(L): Right, it's a very specific case. Why is that— Anyway, just to give you the context. I don't know if that's clear, maybe it is. In the United States we have what's called the rule of institutional precedent. What does this mean? It means that the courts that are deciding a case today are bound by the meaning of the decision in similar cases that have been decided in the past.

(L): In other words, it's as if we were to look at the factual-legal essence of the case and if the essence of the case I'm deciding today coincides with the essence of the case that was decided previously, then I, the judge, am bound by the ruling that my colleague or colleagues adopted back then, right? That's it. That's why we see American lawyers in the movies invoking a certain case, a certain precedent. With the idea of binding. The parallel we have here would be for the party to invoke the law, because the court, under Article 203 of the Constitution, is bound only by the law.

(L): And then there are some specifics, some details that aren't worth elaborating on here, but in general, the court is bound exclusively by the law and nothing else. There you go. So there's that. There is that. There's that particularity. What I mean by this is this model, and maybe I'm already answering the third question. But it's okay, we're talking about it and you can use it as you see fit. So, in this system of legal reasoning, the American lawyer is invoking the rule, because, as you can see, the judge is bound by the decisions of the cases, by the courts that have decided similar cases, he is bound by the decisional sense adopted by those courts. When the lawyer in court invokes the precedent, he is actually invoking a rule that the judge must obey.

(L): Hence the interest, hence the interest in checking whether the argument is valid or not. Hence the interest in checking whether that case actually supports that rule, in other words, what you're actually doing is confirming that it is possible to extract that legal rule from that case. See? There you go.

(L): The question is, in Portugal we wouldn't have a similar system, would we? In Portugal we wouldn't have a similar system. Even so, in Portugal it's common to invoke past decisions of other courts, decisions that normally come from the higher courts, the court of appeal, the courts of appeal or the supreme court of justice is fine, and so they are also invoked. It's not that what the court of first instance is deciding in the case is bound by that decision.

(L): But hey, it's a way of persuading, isn't it? Anyway, the lawyer's job is, above all, to persuade the court to look at the case and follow the case, in the terms that the lawyer's or his client's perspective matters, doesn't it, or that are fairer from their perspective. It doesn't matter now. The ultimate motivations are irrelevant, but that's the strategy, so sometimes it's used as an argument by the Lisbon Court of Appeal or the Supreme Court, which ruled in a similar case in these terms, and so we should continue along these lines.

(L): This is to say that the techniques you've just mentioned, if they were to be adopted in Portugal, would have to be adapted to these specificities, and the confirmatory value would also have to be viewed in slightly different terms to those you've seen in the United States now. One thing is absolutely clear, if I imagine that in a procedural document, the lawyer invokes 3 or 4 decisions from the Coimbra court of appeal, the Porto court of appeal, the supreme court of justice, and so on, of course it would be very useful.

(L): Of course, it would be very useful to take that decision, upload it to a software program and ask it the question, wouldn't it? This decision confirms that, well, the question has to be asked properly, but that's another story and it would be very useful if what the system told me was yes or no. Right? Yes, it does. And why does it confirm?

(L): Right, now. From this perspective, that is, for this function, for this, to solve this problem. The question is, if you tell me that this system - that this result - means at the training stage, obviously it has to be confirmed by human beings, but then it doesn't have to be confirmed? - because if they have to be confirmed, everything is already lost.

(A): Exactly.

(L): The efficiency, the time gains, et cetera, are lost if we have to have a human review of these conclusions, aren't they? Because what we're talking about is, in other words, you get the idea. I think, I don't know if you already had this idea, but I mean the time that is saved with a system like this is - I just said, the judge who is deciding the case doesn't have to go and look at all the judgments that I mentioned. Don't they? He simply uploads the decisions and then the machine enters, yes or no, confirm or not, and why? That's it.

(L): This can save time, although there is a fundamental difference with the US system. The judge doesn't even have to read the decisions I mentioned. The judge doesn't have to read several rulings and refute my conclusion. The judge doesn't have to say no, you referred to decision A or B but that decision doesn't actually correspond to the case. It's good that he does, isn't it? In the context of a complete dialog and a complete countering of arguments, but he's not even obliged to do so. In fact, you're not obliged. The final decision isn't null and void because the judge failed to see a ruling that I mentioned back there, there's no invalidity, there's no illegality in that.

(L): Now, of course it would be very useful, because that perspective would be very useful and I'll even tell you more, there's an even greater usefulness, which is from the perspective of the lawyer doing the research and having decisions that support that conclusion. The result would be to do the reverse process, wouldn't it? That is, I want a decision that, based on these facts, reaches this conclusion, this decision. What decisions have been made in our legal system that meet these requirements? And what set of decisions would tell me that? We have a search now, but it obviously requires a much more refined search. The search is now done by keyword.

(L): Some support from, obviously, artificial intelligence, sometimes with the relationships that are found, I don't know what. A very basic thing, isn't it? Anyway.

(A): Right, exactly.

(L): So, I think it could be useful without a doubt, the usefulness of what you've shown me, mind you, is what I've just mentioned. Now, if we manage to start with these adaptations that need to be made to the Portuguese model of legal thinking and legal decision-making. Now, as I said when I presented this to you, I would say that there are lots of other possibilities, as you would obviously have guessed from a model like this or a world model similar to it, wouldn't you?

(L): I'm here for you, I don't know what has been clarified?

(A): That's right, I think so.

(L): Good. Perfect, then. If you need anything, just let me know.

This is the end of the translated transcript.

In summary – currently, it was said, the algorithms in use are very simple such as keyword tag and search for case law. These don’t extend the helpfulness of AI far enough for it to be a major paradigm shift as the recent attention would warrant. Additionally, these systems are nondeterministic which further legitimises the lawyer’s worry over viability. The heads of major AI labs have stated that reliability will be significantly improved in the coming years. Although it must be said that in any scenario in the foreseeable future includes humans in the loop, which means there'll always be a shift of burden to review tasks – the overall upside remains in question.

For question A, the framework itself was characterized as interesting. However, there was general concern for reliability of the cutting-edge generative systems themselves used as an example, since if the results must be confirmed or critically evaluated then most utility and efficiency gains are lost.

For question B, the framework was understood to be generic. The underlying method is sound, but it has a lot of potential to be expanded to accommodate and offer detail regarding the structure of a particular system of legislation, along with examples ranging from the most common legal tasks to specific tasks that arise from differing legal systems. The practical example used relates to jurisprudence (case law) which, although useful in some instances, is not a primary concern outside of common law countries.

This ties into question C, wherein it was understood that there is no barrier regarding Portugal, except perhaps the legal complexity and language performance, but that these systems would have to be adapted to accommodate certain local specificities such as the way legal facts are treated. Their confirmatory value can differ from other countries, as here the judge is not bound by the decision-power of their colleagues or other courts; although a secondary argument can, and is frequently, made that certain previous decisions offer a certain logic that must be followed. LegalBench does not pronounce itself on capabilities restricted to a certain real-world context, but only evaluates generic capabilities in English.

CONCLUSIONS AND FUTURE WORKS

In conclusion, there are many avenues of exploration to facilitate the day-to-day use of AI systems as assistance in the judicial domain. In this work, the role of AI in legal proceedings has been explored from the perspective of explainability and intelligibility. Subsequently, a framework has been proposed that aims to help emphasize how practical integration should be thought of from the perspective of a legal professional. Since we are in the early stages of AI adoption, a very close relationship between AI experts and legal professionals is required.

Consulting with an expert in the field reveals concern that the current direction of research focus in general systems such as LLMs or even LMMs (Large Multimodal Models) could have a large payoff in helping legal practitioners in the future, but at present the efficiency gains of integrating these more complex systems remains unclear due to the present lack of reliability, operating costs aside. Head researchers at major labs believe the issue of reliability to be substantially solved in the next two to four years.

In comparison, more narrow AI systems such as CLAUDETTE can be utilized by legal experts right now without extensive training, but it might not be as useful as a fine-tuned LLM could be in the future. At present these systems are safer for integration since they tend to be more robust.

In terms of limitations, a deeper dive into the legal peculiarities in Portugal were out of the scope of this work, although it is subject to EU directives. The rapid advancement of AI technology in recent months has also made it harder to write comprehensively for something that likely will be surpassed in a year's time, loose generalization was necessarily required.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., Purves, D., & Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, *30*(1), 1–17. https://doi.org/10.1007/s10506-021-09286-4

Atkinson, K., Bench-Capon, T., & Bollegala, D. (2020). Explanation in AI and law: Past, present and future. *Artificial Intelligence*, *289*, 103387. https://doi.org/10.1016/j.artint.2020.103387

Bertalan, V. G. F., & Ruiz, E. E. S. (2022). Using attention methods to predict judicial outcomes. *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-022-09342-7

Bibal, A., Lognoul, M., de Streel, A., & Frénay, B. (2021). Legal requirements on explainability in machine learning. *Artificial Intelligence and Law*, *29*(2), 149–169. https://doi.org/10.1007/s10506-020-09270-4

Bielen, S., Peeters, L., Marneffe, W., & Vereeck, L. (2018). Backlogs and litigation rates: Testing congestion equilibrium across European judiciaries. *International Review of Law and Economics*, *53*, 9–22. https://doi.org/10.1016/j.irle.2017.09.002

Brkan, M., & Bonnet, G. (2020). Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas. *European Journal of Risk Regulation, 11*(1), 18–50. https://doi.org/10.1017/err.2020.10

Chen, D. L. (2019). Judicial analytics and the great transformation of American Law. *Artificial Intelligence and Law*, *27*(1), 15–42. https://doi.org/10.1007/s10506-018-9237-x

Chiao, V. (2019). Fairness, accountability and transparency: Notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, *15*(2), 126–139. https://doi.org/10.1017/S1744552319000077

Crawford, K., & Schultz, J. (2019). AI systems as state actors. *Columbia Law Review*, *119*(7), 1941–1972. https://www.jstor.org/stable/26810855

Dale, R. (2019). Law and Word Order: NLP in Legal Tech. *Natural Language Engineering*, *25*(1), 211–217. https://doi.org/10.1017/S1351324918000475

Deeks, A. (2019). The judicial demand for explainable artificial intelligence. *Columbia Law Review*, *119*(7), 1829–1850. https://www.jstor.org/stable/26810851

Di Porto, F. (2023). Algorithmic disclosure rules. *Artificial Intelligence and Law*, *31*(1), 13–51. https://doi.org/10.1007/s10506-021-09302-7

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning* (arXiv:1702.08608). arXiv. http://arxiv.org/abs/1702.08608

European Commission. (2021). *Proposal for a regulation laying down harmonised rules on Artificial Intelligence.* https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

European Parliament. (2021). *European Parliament resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters. (2020/2016(INI)).*

General Data Protection Regulation (GDPR). (2016). *Art. 22 GDPR – Automated individual decision-making, including profiling.*

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision Making and a “Right to Explanation”. *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, *30*(3), 291–323. https://doi.org/10.1007/s10506-021-09294-4

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Guha, N., Ho, D. E., Nyarko, J., & Ré, C. (2022). *LegalBench: Prototyping a Collaborative Benchmark for Legal Reasoning* (arXiv:2209.06120). arXiv. http://arxiv.org/abs/2209.06120

Guha, N., Nyarko, J., Ho, D. E., Ré, C., Chilton, A., Narayana, A., Chohlas-Wood, A., Peters, A., Waldon, B., Rockmore, D. N., Zambrano, D., Talisman, D., Hoque, E., Surani, F., Fagan, F., Sarfaty, G., Dickinson, G. M., Porat, H., Hegland, J., … Li, Z. (2023). *LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models* (arXiv:2308.11462). arXiv. http://arxiv.org/abs/2308.11462

Hamon, R., Junklewitz, H., Sanchez, I., Malgieri, G., & De Hert, P. (2022). Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making. *IEEE Computational Intelligence Magazine*, *17*(1), 72–85. https://doi.org/10.1109/MCI.2021.3129960

Hu, S. (2019). The Influence of Artificial Intelligence Development on Patent Legislation. *2019 International Conference on Robots & Intelligent System (ICRIS)*, 110–113. https://doi.org/10.1109/ICRIS.2019.00036

Jin, C., & Rinard, M. (2023). *Evidence of Meaning in Language Models Trained on Programs* (arXiv:2305.11169). arXiv. http://arxiv.org/abs/2305.11169

Jin, Y., & He, H. (2020). An Artificial-Intelligence-Based Semantic Assist Framework for Judicial Trials. *Asian Journal of Law and Society*, *7*(3), 531–540. https://doi.org/10.1017/als.2020.33

Kaminski, M. E., & Urban, J. M. (2021). The right to contest AI. *Columbia Law Review*, *121*(7), 1957–2048. https://www.jstor.org/stable/27083420

Kapoor, A. (2022). *Justice Delayed is Justice Denied: Enabling Legal Artificial Intelligence via Bail Prediction on Hindi Case Documents*. International Institute of Information Technology Hyderabad.

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence*, *296*, 103473. https://doi.org/10.1016/j.artint.2021.103473

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*(jul21 1), b2700–b2700. https://doi.org/10.1136/bmj.b2700

Liu, H.-W., Lin, C.-F., & Chen, Y.-J. (2019). Beyond State v Loomis: Artificial intelligence, government algorithmization and accountability. *International Journal of Law and Information Technology*, *27*(2), 122–141. https://doi.org/10.1093/ijlit/eaz001

Mahoney, C. J., Zhang, J., Huber-Fliflet, N., Gronvall, P., & Zhao, H. (2019). A Framework for Explainable Text Classification in Legal Document Review. *2019 IEEE*

International Conference on Big Data (Big Data), 1858–1867. https://doi.org/10.1109/BigData47090.2019.9005659

Mandal, A., Ghosh, K., Ghosh, S., & Mandal, S. (2022). A sequence labeling model for catchphrase identification from legal case documents. *Artificial Intelligence and Law*, *30*(3), 325–358. https://doi.org/10.1007/s10506-021-09296-2

McLachlan, S., Kyrimi, E., Dube, K., Fenton, N., & Webley, L. C. (2023). Lawmaps: Enabling legal AI development through visualisation of the implicit structure of legislation and lawyerly process. *Artificial Intelligence and Law*, *31*(1), 169–194. https://doi.org/10.1007/s10506-021-09298-0

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, *28*(2), 237–266. https://doi.org/10.1007/s10506-019-09255-y

Medvedeva, M., Wieling, M., & Vols, M. (2023). Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law*, *31*(1), 195–212. https://doi.org/10.1007/s10506-021-09306-3

Mitchell, J., Mitchell, S., & Mitchell, C. (2020). Machine learning for determining accurate outcomes in criminal trials. *Law, Probability and Risk*, *19*(1), 43–65. https://doi.org/10.1093/lpr/mgaa003

Nay, J., Karamardian, D., Lawsky, S. B., Tao, W., Bhat, M., Jain, R., Lee, A. T., Choi, J. H., & Kasai, J. (2023). Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4476325

Nikolskaia, K., & Naumov, V. (2020). Artificial Intelligence in Law. *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, 1–4. https://doi.org/10.1109/FarEastCon50210.2020.9271095

Olah, C. (2022). Mechanistic Interpretability, Variables, and the Importance of Interpretable Bases. *Transformer Circuits*. https://transformer-circuits.pub/2022/mech-interp-essay/index.html

Pah, A. R., Schwartz, D. L., Sanga, S., Alexander, C. S., Hammond, K. J., Amaral, L. A. N., & SCALES OKN Consortium. (2022). The Promise of AI in an Open Justice System. *AI Magazine*, *43*(1), 69–74. https://doi.org/10.1002/aaai.12039

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Ramello, G. B., & Voigt, S. (2012). The economics of efficiency and the judicial system. *International Review of Law and Economics*, *32*(1), 1–2. https://doi.org/10.1016/j.irle.2011.12.003

Roksandic, S., Protrka, N., & Engelhart, M. (2022). Trustworthy Artificial Intelligence and its use by Law Enforcement Authorities: Where do we stand? *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1225–1232. https://doi.org/10.23919/MIPRO55190.2022.9803606

Rosili, N. A. K., Hassan, R., Zakaria, N. H., Kasim, S., Rose, F. Z. C., & Sutikno, T. (2021). A systematic literature review of machine learning methods in predicting court decisions. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *10*(4), 1091. https://doi.org/10.11591/ijai.v10.i4.pp1091-1102

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Simmler, M., Brunner, S., Canova, G., & Schedler, K. (2022). Smart criminal justice: Exploring the use of algorithms in the Swiss criminal justice system. *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-022-09310-1

Soukupová, J. (2021). AI-based Legal Technology: A Critical Assessment of the Current Use of Artificial Intelligence in Legal Practice. *Masaryk University Journal of Law and Technology*, *15*(2), 279–300. https://doi.org/10.5817/MUJLT2021-2-6

Strandburg, K. J. (2019). Rulemaking and inscrutable automated decision tools. *Columbia Law Review*, *119*(7), 1851–1886. https://www.jstor.org/stable/26810852

The Official Journal of the European Union. (2007). *Official Journal of the European Union C 303/17—14.12.2007*.

Tjoa, E., & Guan, C. (2021). A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

Trappey, A. J. C., Trappey, C. V., Wu, J.-L., & Wang, J. W. C. (2020). Intelligent compilation of patent summaries using machine learning and natural language processing techniques. *Advanced Engineering Informatics*, *43*, 101027. https://doi.org/10.1016/j.aei.2019.101027

Tsakalakis, N., Stalla-Bourdillon, S., Carmichael, L., Huynh, T. D., Moreau, L., & Helal, A. (2021). The dual function of explanations: Why it is useful to compute explanations. *Computer Law & Security Review*, *41*, 105527. https://doi.org/10.1016/j.clsr.2020.105527

van der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence*, *291*, 103404. https://doi.org/10.1016/j.artint.2020.103404

Varošanec, I. (2022). On the path to the future: Mapping the notion of transparency in the EU regulatory framework for AI. *International Review of Law, Computers & Technology*, *36*(2), 95–117. https://doi.org/10.1080/13600869.2022.2060471

Villata, S., Araszkiewicz, M., Ashley, K., Bench-Capon, T., Branting, L. K., Conrad, J. G., & Wyner, A. (2022). Thirty years of artificial intelligence and law: The third decade. *Artificial Intelligence and Law*, *30*(4), 561–591. https://doi.org/10.1007/s10506-022-09327-6

vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. In *Design Science Research. Cases*. Springer.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, *41*, 105567. https://doi.org/10.1016/j.clsr.2021.105567

Wang, R. (2020). Legal technology in contemporary USA and China. *Computer Law & Security Review*, *39*, 105459. https://doi.org/10.1016/j.clsr.2020.105459

Winikoff, M., & Sardelic, J. (2021). Artificial Intelligence and the Right to Explanation as a Human Right. *IEEE Internet Computing*, *25*(2), 116–120. https://doi.org/10.1109/MIC.2020.3045821

Wu, T. (2019). WILL ARTIFICIAL INTELLIGENCE EAT THE LAW? THE RISE OF HYBRID SOCIAL-ORDERING SYSTEMS. *Columbia Law Review*, *119*(7), 2001–2028. JSTOR.

Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., & Puntoni, S. (2022). Perceptions of Justice By Algorithms. *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-022-09312-z

Zhang, Y., Xu, F., Zou, J., Petrosian, O. L., & Krinkin, K. V. (2021). XAI Evaluation: Evaluating Black-Box Model Explanations for Prediction. *2021 II International Conference on Neural Networks and Neurotechnologies (NeuroNT)*, 13–16. https://doi.org/10.1109/NeuroNT53022.2021.9472817

APPENDIX

The following is a transcript, in Portuguese, of an interview/discussion between the author (A) and a lawyer/professor (L) with interest in the domain of legal application of AI, who prefers to remain anonymous:

A: Procuro primeiro definir o contexto. Aquilo que estou a assumir é que AI pode ajudar os profissionais do domínio jurídico em várias tarefas e há interesses em, de facto, utilizar os modelos de inteligência artificial no domínio jurídico. Os modelos mais úteis são caixas negras opacas. As tarefas jurídicas exigem raciocínio e justificação transparente. As leis relativas à utilização são vagas ou inexistentes. Quando comecei neste tópico, havia um foco grande em utilizar vários modelos muito específicos, mas nota-se que o foco da investigação alterou agora para modelos mais gerais. Sobretudo LLMs, modelos de língua. Aqui, procurei desenvolver uma framework, mais um enquadramento / uma forma de pensar para integração de AI em várias tarefas. Como é que podemos pensar neste assunto?

Primeiro começamos com uma necessidade legal. Depois temos de definir a tarefa associada. Temos de definir essas tarefas claramente, com todo o contexto relevante e até podemos, inclusive, elaborar um mapa de tarefas para entendermos exatamente onde há a intervenção de inteligência artificial, onde pode haver e onde existiu.

E sempre, claro, depois temos de escolher um sistema de inteligência artificial que seja apropriado para a tarefa e, portanto, aqui terá de haver a colaboração entre profissionais do mundo legal e experts em AI. Se querem de facto que haja esta utilização, tem de haver uma colaboração muito próxima, creio eu. Mais tarde pode haver uma lista de aprovação, portanto, sistemas que estão aprovados para determinadas tarefas. Ou inclusivamente experts de AI que estão licenciados para dar precisamente esta consulta. De momento, esta infraestrutura não existe, portanto, depende a criação de instituições relevantes para esta função, para sistemas.

Quando falamos de sistemas mais restritos. Há formas de os tornar mais explicáveis formas muito técnicas, mas para modelos mais gerais, como LLMs, temos um novo campo chamado interpretação mecanística, que ainda não tem muitos resultados. Mas procuro aqui distinguir que o output destes modelos, LLMs, é inteligível para o humano encargo da tarefa e portanto, aqui creio que o facto de ser inteligível seja suficiente para haver uma avaliação adequada.

Tendo em conta o facto de que, portanto, aqui o expert de AI terá de informar, bem como quais são os erros comuns desses sistemas. O possível enviesamento do sistema, porque precisamente quando houver esta avaliação o output pode ter problemas subtis que não são fáceis de entender. E aqui o perigo seria, portanto, que na verificação do output, fosse mais lenta que a execução inicial da tarefa sem inteligência artificial em primeiro lugar. Daí ser bastante importante ganhar experiência com esses sistemas e também portanto, o avanço da fidelidade desses temas, que acho que ocorrerá porque isto está a mover muito rápido. E, por exemplo, pode haver um fine-tuning destes modelos gigantes especificamente em contextos legais, pode haver melhoramento no tamanho do contexto, melhoramento na eliminação de alucinações com ground-truth. E aqui em todos os passos, haverá sempre um humano a verificar o output, porque os sistemas são falíveis. E, portanto acho que esta é a melhor forma de pensar com este enquadramento.

Portanto, e dou aqui um exemplo de uma tarefa de interpretação aqui. O exemplo de definição de uma tarefa. Isto foi levantado de legal Bench, que é benchmark para LLMs de contextos legais, têm várias modalidades. Esta é uma de interpretação e é bastante recente, portanto. Este benchmark foi começado a desenvolver em 2022 o ano passado, 2022 foi a primeira publicação e este ano saiu, de facto, em 2023, saiu até em Agosto.

Começamos por definir bem a tarefa com o contexto relevante. Neste caso, queremos determinar se determinado texto legal implica uma determinada conclusão. E podemos ter uma cadeia destas tarefas tendo em conta a nossa necessidade e depois para aqui, para determinar o sistema XAI ou o sistema de que estamos a utilizar.

Este aqui tem aqui uma benchmark também levantado legalbench. Fizeram vários testes em várias modalidades e temos aqui o GPT 4. Claramente está acima da competição, de largo modo de momento e aqui faço também um passo intermédio. Esta definição de tarefa está com exemplos de legal bench. Mas fui eu que fiz esta parte, definição de tarefa e depois aqui tive que fazer um passo intermédio, que é que eu estava a utilizar o GPT 3.5. E tenho que fazer um passo a ser intermédio, porque o contexto não permite introduzir o texto legal inteiro. Portanto, tive que tive que pedir ao sistema para resumir aqui neste momento.

E depois do momento seguinte, passei então a ter o output. Perguntei exatamente aquilo que queria, aquilo que era para avaliar com a sumarização. Espero que no futuro, pronto… Há modelos com contextos muito maiores que eu não tenho acesso a, mas que que podíamos introduzir o texto inteiro – Tudo o que seja relevante no contexto legal para esta tarefa de uma vez e, portanto, o modelo consegue avaliar inteiramente o texto. E portanto, é apenas uma coisa temporária. Os modelos agora tendem todos a ter janelas de contexto muito maiores.

Introduzindo a pergunta que queremos fazer ao modelo e, neste caso, a sumarização, temos o resultado e o resultado aqui passa por eles distinguem duas categorias aqui, uma é correctness, que é, se a conclusão está correta ou errada. Para para avaliar o output e se a análise está correta e errada, basicamente, há aqui 2 distinções. Portanto, uma coisa é ter a conclusão certa, outra coisa é a análise está correta. O GPT 4, claramente está muito acima do dos outros. Neste caso, o GPT 3.5 acertou. Eu corri 8 vezes. Em todas as vezes, em termos de correctness acertou todas as vezes. Só que a análise falhou uma vez, portanto, temos que ter isto em mente. Eu espero que no futuro estes modelos possam estar ligados a uma base de dados e – termos uma coisa chamada code interpreter, podemos pedir ao GPT 4 em forma de código fazer esta avaliação, mas para tirar citações do texto exatamente de onde é que foi buscar esta informação. Pode tirar os strings do texto em que frases textuais é que suportam aquilo que está a dizer, também é possível. Pronto, é o que resta aqui da análise é esta distinção de métricas do resultado. Está correto? Em termos de explicação?

E em termos de questões que eu queria fazer— Considera esta forma de pensar esta framework útil? Porquê? Se não, porquê? Onde é que pode ser alargada?

L: Claro, vamos só aqui atrás, só para para ver se eu percebo melhor esta parte do falar que é em correctness. Ou seja, qual é que é o qual é que é o quadro de referência para decidir se o resultado, se o é correto ou não, é correto

A: Certo, aqui, aqui, por exemplo, estão a fazer perguntas de interpretação entre retórica, et cetera, e aqui é, se a conclusão está correta, porque eles têm casos de, por exemplo, casos passados que já foram bastante avaliados e eles sabem se aquilo que estão a perguntar se, se o texto afirma aquilo ou não. Portanto, foi algo que foi decidido em tribunal e eles avaliam se, está de acordo com aquele texto. Se o modelo consegue inferir que, de facto, aquilo é suportado ou não. É apoiado por por aquilo que estão a referenciar e o modelo pode dizer que sim ou que não e depois tem que analisar porquê.

L: OK, mas, portanto, aquilo que é aquilo que é apresentado no modelo é uma descrição fáctica do que aconteceu.

A: É exato.

L: Ou seja, é uma descrição fáctica do caso. Esta é a história e a partir daqui. Ou seja, o output há de ser a decisão; há-de coincidir com aquela que foi efetivamente. Tomada pelo tribunal é isso. E é esse, ou seja, quando nós falamos em avaliação da correção do resultado, estamos a pensar nessa comparação entre a decisão efetiva do caso e a decisão que A máquina tomaria perante aqueles factos se fossem apresentados aqueles factos, é isto?

A: Sim.

L: Está claro.

A: Esse é um aspeto, portanto, porque temos várias modalidades. Temos por exemplo, aqui quando falam de Issue é a identificação do domínio legal que estamos a falar em concreto. Da área de direito, que estamos a falar em concreto. Aqui as leis, seria que leis seriam aplicáveis, mas que o modelo não é bem factual nesse sentido e, portanto, tem um score mais reduzido. Aqui, em termos de conclusão, é que conclusão qual é a conclusão legal destes factos aqui –

L: Qual é, qual é a decisão? Qual é a decisão, se é um caso de assédio ou não, é um caso de assédio nesta hipótese.

A: Sim. A interpretação seria interpretação de factos, se aquilo é suportado pelo aquilo, se podem fazer uma afirmação com base num texto e podemos perguntar se aquele texto de facto confirma aquilo que foi dito. e depois a retórica será então construção de um, de uma história, de uns argumentos.

L: Ou seja, da justificação da decisão. Esta parte da interpretação não ficou clara para mim. Consegue especificar melhor o que é que de que é que se trata efetivamente? Nesta parte da interpretação?

A: A interpretação é, por exemplo, aqui temos o sumário de um caso legal que foi feito e depois perguntamos ao modelo porque foi feito uma afirmação. Pronto esta afirmação que para aquele hostile environment aquele empregado tem que mostrar isto e isto e isto e isto e citaram um caso um caso passado que justificava esta afirmação. Aqui, a tarefa passa por perguntar ao modelo, portanto, aqui damos ao modelo aquele texto legal que foi suporte desta afirmação e perguntamos ao modelo se, de facto, aquele texto suporta esta afirmação.

L: Pronto, portanto, nós damos o sumário de um caso, não é? E aquilo perguntando é se o sumário deste caso. – Isto, deixe-me só dizer, que isto é o tipo de raciocínio próprio dos sistemas de Common Law. Mas sem problema que conseguimos adaptar-nos.

Pronto, e a ideia é perguntar se aquele sumário, que é o sumário da decisão, suponho eu, fundamenta esta regra que acaba depois por enunciar, não é? É isto. Pronto, depois o sistema dir-nos-á se sim se não. Pronto. Agora – ou seja, a ideia de seria utilizar esta ferramenta para efeitos confirmatórios, é isso?

A: Sim; é um dos exemplos. Exatamente neste caso seria modulação de interpretação.

L: Se a decisão está correta ou não está correta.

A: Exatamente.

L: E quem é que faz o sumário da decisão?

A: Portanto, aqui foi o sumário— aqui o contexto do sumário da decisão foi porque o contexto era muito limitado— porque o caso era muito grande, por exemplo. Espera-se que isto seja apenas uma coisa temporária, porque os modelos agora estão a aumentar muito de contexto e, portanto, este espaço de sumarização não será necessário de todo.

L: Ou seja, mas, portanto, a ideia. A ideia seria utilizar a ferramenta no sentido, eu peço-lhe par estar a, enfim…

A: Sim, tudo bem.

L: Peço desculpa— na verdade, o interesse é esse, não é? Tentarmos perceber as coisas e dar uma opinião.

A: Aqui, seria forma de facilitar a execução destas tarefas. Claro que terá de ser verificado porque—

L: Se calhar, vamos um bocadinho atrás, porque tinha, ou seja, o método mais para trás, por favor, aquela parte inicial, esse aqui. Portanto, o método está corretíssimo, como é óbvio que é, enfim, definir o problema jurídico não é, ou a tarefa legal, portanto, a questão jurídica que é preciso resolver o problema. E qual é o problema que nós estamos a resolver aqui?

A: Verificar se aquele aquele statement – Neste caso, a defesa citou um caso anterior como suporte, como apoio para aquilo que estava a aferir.

L: E nós pegamos nessa decisão, pegamos no caso—

A: Sim, neste caso, tive de fazer um sumário, porque o contexto não permitia.

L: Isso vai ser ultrapassado—

A: E perguntamos aqui se aquele caso, se o caso que vai ser dito em baixo — Se, de facto apoia ou não aquilo que foi dito.

L: Muito interessante, pois, portanto, é uma confirmação da validade do argumento. Utilizado pela parte neste caso. Pronto, isto é um caso muito específico dos Estados Unidos e também—

A: Sim, eu queria dar um exemplo de uma tarefa.

L: Certo, é um caso muito específico. Porquê? Enfim, só para lhe dar o contexto. Não sei se isso está claro, se calhar até está. É que nos Estados Unidos temos aquilo a que se chama a regra do precedente institucional. O que é que isto significa? Significa que os tribunais que estão a decidir hoje um caso estão vinculados quanto ao sentido da decisão a casos semelhantes que foram decididos no passado.

Ou seja como se nós fossemos olhar para a essência facticojurídica do caso e se a essência do caso que estou a decidir hoje coincide com a essência do caso que foi decida anteriormente, então, eu, juiz, estou vinculado ao sentido decisório que o meu colega ou os meus colegas adotaram lá atrás, certo? Pronto. Por isso é que nós vemos nos filmes os advogados norte-americanos a invocar determinado caso, um determinado precedente. Com a ideia de vincular. Paralelo que nós temos aqui seria a parte invocar a lei, porque o tribunal, nos termos do artigo 203 da Constituição, está vinculado apenas à lei.

E há depois algumas especificidades, alguns detalhes que não vale a pena estar aqui a elaborar, mas em geral, o tribunal está vinculado exclusivamente à lei e nada mais. Pronto. E, portanto, e portanto, há esse. Há esse. Há essa particularidade. O que eu quero dizer com isto é este modelo e se calhar, já estou a responder à terceira pergunta. Mas também não há problema que estamos a conversar e depois aproveitará como achar bem.

Portanto, neste sistema de de raciocínio jurídico, o advogado norte-americano está a invocar a regra, porque, como repare como o juiz está vinculado às decisões dos casos, pelos tribunais que decidiram casos semelhantes, está vinculado ao sentido decisório adotado por esses tribunais. Quando o advogado em tribunal invoca o precedente ele, na verdade, está a invocar uma regra a que o juíz deve obedecer.

Daí o interesse, daí o interesse em verificar se o argumento é válido ou não. Daí o interesse em verificar se aquele caso efetivamente sustenta aquela regra, ou seja, o que está a fazer, na verdade, é confirmar que daquele caso é possível extrair aquela regra jurídica. Está a ver? Pronto.

A questão é, a questão é, em Portugal nós não teríamos um sistema semelhante, não é? Em Portugal nós não teríamos um sistema semelhante. Ainda assim, em Portugal é frequente invocar decisões de outros tribunais passadas, decisões essas que normalmente vêm dos tribunais superiores, o tribunal da relação, tribunais da relação ou do supremo tribunal de justiça está bem e, portanto, também se invocam. Não é que o que o tribunal de primeira instância esteja a decidir no caso esteja vinculado a essa decisão.

Mas pronto, é uma forma de persuadir, não é? Enfim, o trabalho do advogado é, sobretudo, persuadir o tribunal a olhar para o caso e seguir o caso, nos termos que a perspectiva do advogado ou do seu cliente interessa, não é, ou que são mais justos na sua perspectiva. Não importa agora. As motivações últimas são irrelevantes, mas pronto, essa é a estratégia, portanto, por vezes utiliza-se como argumento do tribunal da relação de Lisboa ou do supremo tribunal, decidiu, em caso semelhante nestes termos e, portanto, deve continuar-se essa linha.

Isto para dizer que as técnicas que acabou de referir a serem adotadas em Portugal teriam de ser adaptadas com esta especificidades e o valor confirmatório também teria de ser perspectivado em termos um pouco diferentes daqueles que viu nos Estados Unidos agora. Uma coisa é absolutamente inequívoca, se eu imagino que numa peça processual, o advogado invoca 3 ou 4 decisões do tribunal da relação de Coimbra, do Porto de Coimbra, do supremo tribunal de justiça, por aí adiante, claro que seria muito útil.

Claro que seria muito útil pegar nessa decisão, fazer o upload dessa decisão para um software e o software e fazer-lhe a pergunta, não é? Esta decisão confirma que, enfim, a pergunta tem de ser bem feita, mas pronto, isso já são outras histórias e era muito útil que o que o que o sistema me dissesse sim ou não. Não é? Sim, confirma. E porque é que confirma?

Exato, agora. Nesta perspectiva, ou seja, para esta função, para este, para resolver este problema. A questão é, se me disser que este sistema – que este resultado – quer dizer na fase do treino, obviamente que tem de ser confirmados por seres humanos, mas depois não tem de ser confirmados? – porque se tiverem de ser confirmados já se perdeu tudo.

A: Exato.

L: A eficiência, os ganhos tempo, et cetera, perdem-se se houver necessidade de termos uma revisão humana dessas conclusões, não é? Porque aquilo que nós estamos a falar é, ou seja, já percebeu. Creio eu que não sei se já tinha esta ideia, mas quer dizer o tempo que se poupa com um sistema deste é – eu acabei de dizer, o juiz que está a decidir o caso não tem de ir ver todos os acórdãos que eu referi. Não é? Vai simplesmente fazer o upload das decisões e depois a máquina de entrar, sim ou não, confirma ou não, e porquê? Pronto.

O que pode implicar um certo ganho de tempo, sendo certo que aqui há uma diferença fundamental para o sistema norte-americano. O juiz não tem sequer de ir ler as decisões que eu referi. O juiz não tem diversas decisões e rebater a minha conclusão. O juiz não tem de dizer não, referiu uma decisão A ou B mas essa decisão de facto não corresponde ao caso. É bom que o faça, não é? No âmbito de um de um diálogo completo e de uma contraposição completa de argumentos, mas nem sequer está obrigado a fazê-lo. Na verdade, não está obrigado. A decisão final não é nula porque o juíz deixou de ver um acórdão que eu referi lá atrás, não há invalidade, não há ilegalidade nisso.

Agora, claro que seria muito útil, porque essa perspectiva seria muito útil e até lhe digo mais, até há uma utilidade maior ainda, que é na perspectiva do advogado fazer a pesquisa e ter decisões que suportam aquela conclusão. A ter como resultado seja fazer o processo inverso, não é? Que é, eu pretendo uma decisão que, mediante estes factos chegue a esta conclusão, esta decisão. Que decisões foram tomadas no nosso ordenamento jurídico que preencham estes requisitos? E obter conjunto das decisões que que me dissesse isso? Nós temos pesquisa agora, mas é preciso, obviamente, uma pesquisa muito mais fina. A pesquisa agora faz por palavra-chave.

Algum apoio de algum apoio, obviamente, de inteligência artificial, às vezes com as relações que se que se encontram, não sei quê. Uma coisa muito básica, não é? Enfim.

A: Pois, exato.

L: Portanto, eu acho que eu acho que pode ser útil sem dúvida, a utilidade daquilo que me mostrou, atenção, é esta que eu acabei de referir. Agora, se nós conseguimos partir e com estas adaptações que têm de ser feitas para o modelo português de pensamento jurídico e de decisão jurídica. Agora, como digo quando me apresentei isto, eu diria que há outras imensas possibilidades, como, aliás, já teria obviamente intuído que se podem retirar de um modelo destes ou de um mundial semelhante a isso, não é?

Eu estou aqui ao seu dispor, não sei o que é que se ficou esclarecido?

A: Está certo, acho que sim.

L: Ótimo. Perfeito, então. Se precisar de alguma coisa, disponha.

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa