SARA: Leveraging LLM Agents and Jurisprudential Ontologies for Automated Legal Reasoning

Francisco C J Bonfim[1], Sara Pessoa Silva[1], Alicia S Neves[1], Henrique Santos[2], João A Monteiro Neto[1], Jorge L B Araujo[1], Rilder S Pires[1], Ricardo Alexandre Da Silva Costa[1,3], Vasco Furtado[1,4], Vladia Pinheiro[1]

[1]University of Fortaleza (UNIFOR)
[2]Rensselaer Polytechnic Institute (RPI)
[3]Court of Justice of the State of Ceará (TJCE)
[4]IT Company of the State of Ceará (ETICE)
fjuca@unifor.br, sarapessoa@edu.unifor.br, aliciasouzan@gmail.com, oliveh@rpi.edu,
joaoneto@unifor.br, jorgearaujo@unifor.br, rilder@unifor.br, ricardoascpai@gmail.com,
vasco@unifor.br, vladiacelia@unifor.br

Abstract

Delivering judicial decisions requires interpreting complex legal texts, analyzing evidence, and reasoning over jurisprudence and legal principles. Recent advances in Generative Artificial Intelligence, particularly Large Language Models (LLMs), have shown potential to automate parts of this process, yet practical, measurable benefits in real-world judicial settings remain limited. This paper introduces SARA, an LLM-powered legal reasoning platform deployed in a regional Brazilian court, which demonstrates significant efficiency and quality gains through the integration of LLM agents with a Jurisprudential Knowledge Graph (Jur-KG). SARA automatically extracts and structures key elements from legal documents—including claims, requests, and evidence—and generates reasoning grounded in retrieved jurisprudential precedents. The Jur-KG, modeled through an ontology encompassing concepts such as *LegalRelation*, *LegalGrounds*, and *LegalClaims*, enables semantic matching and retrieval of relevant case law. By representing cases according to the Legal Case Ontology for the Brazilian Judicial System, SARA supports traceable reasoning and addresses competence questions to assess coverage, coherence, and justification of AI-generated outputs. Deployment results indicate measurable improvements in processing time, consistency, and explainability, while ensuring compliance with ethical and legal guidelines established by Brazil's National Council of Justice. This work demonstrates that combining LLM-based agents with domain-specific knowledge graphs can yield both innovative capabilities and proven impact in judicial decision-making.

Introduction

Formulating judicial decisions is a demanding task that requires coordinating extensive case files, interpreting legal norms, and balancing jurisprudence, principles, and doctrinal interpretations. Judges, clerks, and judicial assistants must navigate large volumes of information, including testimonies and expert reports, making it difficult to produce well-reasoned decisions efficiently. Recent studies have shown that Artificial Intelligence (AI) can improve the management of judicial data by enhancing accuracy and efficiency (Rajasekar and Vezhaventhan 2024). However, its adoption in courts also raises concerns regarding algorithmic bias, data quality, hallucinations, privacy, and the verifiability of AI-generated outputs. A recent case in Brazil illustrates these risks: a judge relied on an AI application that produced fictitious precedents attributed to the Superior Court of Justice, prompting an investigation by the National Council of Justice (CNJ) (G1 2023; Conselho Nacional de Justiça 2025). In response, CNJ Resolution No. 332/2020 established ethical guidelines for AI in the judiciary, underscoring the need for transparency, grounded reasoning, and protection of fundamental rights.

To address these challenges, we present **SARA** (Bomfim et al. 2025b) (Bomfim et al. 2025a) , a Generative AI system powered by Large Language Models (LLMs) and deployed in the Court of Justice of Ceará, Brazil. Developed since 2022 and officially launched in 2024, SARA supports the drafting of judicial decisions by automatically identifying claims, requests, and evidence across case documents, linking evidence to specific claims, and retrieving relevant jurisprudence. Through integration with the Jurisprudential Knowledge Graph (JUR-KG), SARA grounds its reasoning in structured legal knowledge. The JUR-KG encodes jurisprudential data via an ontology that models concepts such as *LegalRelation*, *LegalGrounds*, and *Claims*. Together with the Legal Case Ontology for the Brazilian Judicial System (LCO-BR), this infrastructure allows SARA to build comprehensive knowledge graphs of each case, ensuring semantic consistency, interpretability, and traceability of AI-generated reasoning.

Beyond reasoning generation, these knowledge graphs act as an interpretability layer, enabling the verification of whether claims, evidence, and jurisprudence have been properly addressed in the decision. This strengthens the fairness, accountability, and transparency of AI-assisted judgments.

We evaluated SARA quantitatively and qualitatively. Using a gold-standard dataset of five cases (16 textual documents and 321 images, totaling ~528,000 tokens), semantic similarity analysis showed a 2.97% improvement in alignment with human-authored reasoning when leveraging the JUR-KG, as measured by the *ALIGNSCORE* metric. In qualitative assessment, 22 judges formally appointed to evaluate the system reported a 94.4% positive rate, highlighting SARA's ability to produce well-structured reasoning, correctly analyze preliminary matters, and integrate relevant jurisprudence; negative feedback is incorporated into iterative system refinement.

Currently, SARA is integrated with Brazil's nationwide Electronic Judicial Process System (PJe) and actively used in the Court of Justice of Ceará. With 240 users, SARA generated 510 reports for 476 cases in the last month alone, demonstrating its practical impact.

Jurisprudential Knowledge Graph of the Brazilian Judicial System (JUR-KG) and Ontologies

The Jurisprudential Knowledge Graph (Jur-KG) is a database of court decisions from Brazil. It currently includes 22,106 judgments from the private law chambers of the Court of Justice of Ceará, issued between 2023 and 2024 and sourced from one of the court's internal systems. Jur-KG is being constantly expanded to include decisions from public law chambers, appeal panels, and other Brazilian courts. The data is organized by "Justice Body" and "Class," as shown in Table 1. To enable a structured semantic representation of jurisprudential precedents, the Jur-KG ontology was defined (see Figure 1). We use the European Legislation Identifier (ELI)[1], which provides a set of essential legal concepts and has good potential for reuse and connection with various legal sources, similarly to how it is used in *Brazilian Legal Knowledge Graph* (Pires et al. 2022). In this ontology, the jurisprudential precedents are represented as the class **jus-br:Jurisprudence** that is a subclass of **eli:LegalResource**, and inherits the following properties from ELI ontology: the case number (**eli:id_local**); the justice body (**eli:agent _document**); the judging body (**eli:service**); the judgment date (**eli:date _publication**), the class of the document (**eli:type_document**); and the full text of the judgment (**eli:description**). In addition, the **jus-br: Jurisprudence** class is associated with other classes such as **jus-br:LegalRelation**, **jus-br:LegalRequest**, **jus-br:LegalGround**, **jus-br:CaseSummary**, and **jus-br:ReportingJudge** that provide other information about the jurisprudence. In addition, we use a Legal Case Ontology for the Brazilian Judicial System (LCO-BR), which provides the semantic foundation for representing legal cases integrated with a knowledge graph of jurisprudential precedents. LCO-BR builds upon existing ontologies, particularly PROV-O (Lebo et al. 2013) (W3C 2013), which is a W3C standard for representing provenance information using RDF and OWL, and the JUR-KG ontology (see Figure 1). LCO-BR integrates them with semantics describing the internal structure of legal cases, including documents, claims, requests, pieces of evidence, agents, jurisprudence, and judicial activities, all of which contribute to legal reasoning. By supporting the creation of provenance-aware knowledge graphs, LCO-BR facilitates the verification of completeness and provenance not only for evidence but also for other critical elements of a case, such as jurisprudential precedents related to the case. This verification contributes to a more robust, well-founded, and explainable legal reasoning.

Judging Body	Class	Count	%
1st Chamber	Civ Appeal	5343	24.2
1st Chamber	Civ Internal Appeal	137	0.6
1st Chamber	Civ Jurisdiction Conflict & Others	92	0.4
2nd Chamber	Civ Appeal	7357	33.3
2nd Chamber	Civ Jurisdic Conflict	91	0.4
2nd Chamber	Civ Internal Appeal & Others	75	0.3
3rd Chamber	Civ Appeal	3380	15.3
3rd Chamber	Civ Internal Appeal	126	0.6
3rd Chamber	Civ Jurisdiction Conflict & Others	95	0.4
4th Chamber	Civ Appeal	5188	23.5
4th Chamber	Civ Internal Appeal	128	0.6
4th Chamber	Civ Jurisdiction Conflict & Others	94	0.4
Total		**22106**	**100.0**

Table 1: Jur-KG jurisprudential entries count by Judging Body and Class

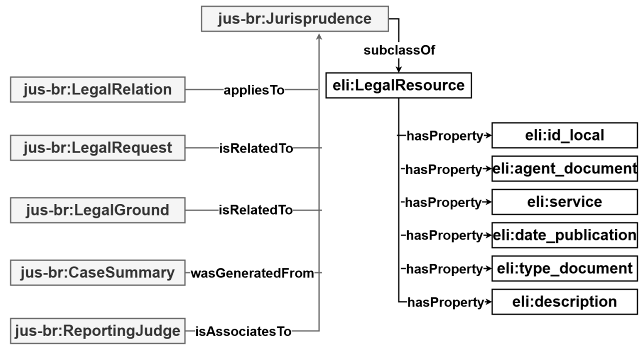

Figure 1: Jur-KG Ontology

SARA - System for Analysis, Summarization and Jurisprudence-aware AI Legal Reasoning

This section presents the SARA - System for Analysis, Summarization, and Jurisprudence-aware AI Legal Reasoning

[1]https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01)

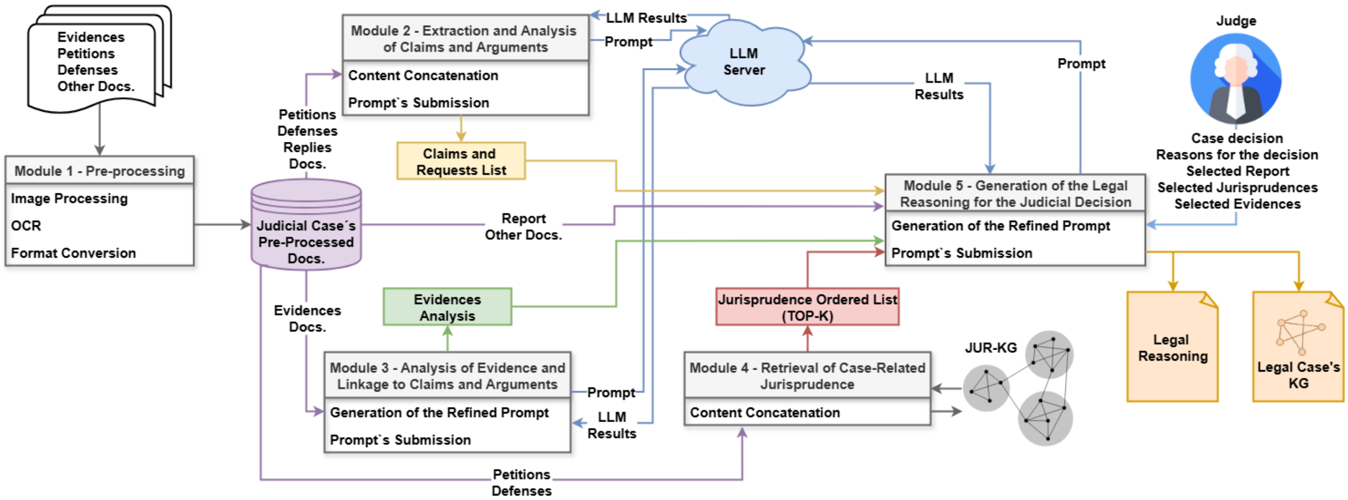

Figure 2: SARA's Pipeline for generating the legal reasoning of judicial decisions and the knowledge graph of legal cases.

of Legal Cases (Bomfim et al. 2025b) (Bomfim et al. 2025a), which leverages Generative Large Language Models (LLMs) such as GPT-4 (OpenAI 2023) and LLaMA 3 (Meta AI 2024). The SARA's pipeline (see Figure 2) system consists of sequential modules that employ robust prompt engineering, which incorporates evidentiary elements, the parties' claims and requests, and related jurisprudence, ensuring that all relevant arguments are analyzed and that the reasoning is clear, transparent, and legally substantiated.

In general, the workflow begins with the collection of relevant legal documents, including petitions, defenses, expert reports, and other materials related to the case, along with the evidence submitted by the parties. In Module 1 (Pre-Processing), these documents and evidence are analyzed using image processing models and optical character recognition (OCR) tools. This step aims to extract content by converting images and scanned documents into machine-readable text. In Module 2, the textual content of documents such as the initial petition, defenses, and replies is analyzed by the LLM to extract and structure the claims and requests of the parties. Module 3 focuses on the analysis of evidence using refined prompts submitted to the LLM, which results in the organization of evidence, verification of its provenance, and linkage of evidence to the claims and arguments. Module 4 handles the integration with the Jurisprudential Knowledge Graph (Jur-KG), enabling the retrieval of legally and contextually relevant precedents. This supports the identification of applicable jurisprudence that aligns with the legal issues, facts, and claims presented in the current case. Finally, in Module 5, the outputs of the previous modules, combined with the case report and other case-related documents, are submitted via prompt to the LLM to generate the legal reasoning grounding the judicial decision.

In the following subsections, we detail the process by which SARA generates legal reasoning, starting from the operations performed in Module 2.

Module 2 - Extraction and Analysis of Claims and Requests The goal of this module is to analyse and extract, from the set of documents (e.g., initial petitions, defenses, replies, etc.), the claims and requests presented by both the plaintiff and the defendant, as well as any additional requests made by other entities (such as judges), if applicable. A specialized prompt (see (Bomfim et al. 2025b)) is submitted to the LLM, and the output is structured into lists of claims and requests separated by plaintiff, defendant, and other entities.

Figure 3 presents an excerpt from SARA's analytical output, illustrating the identification of claims and associated requests submitted by the plaintiff. As an example, the following claim was extracted: <ap1> = "The plaintiff did not request the service contract and was a victim of fraud."

In addition, the legal claims and requests are modeled as instances of the LCO-BR ontology classes **lco-br:LegalClaim** and **lco-br:LegalRequest**, respectively. For instance, the claim <ap1> = "The plaintiff did not request the service contract and was a victim of fraud." was represented as :PlaintiffClaim001 and related with :LegalDocument001 and :Plaintiff001, as presented in Turtle syntax in Figure 5. This structured format for claims and requests facilitates subsequent integration and retrieval of each party's submissions, enabling their linkage to the evidence presented by each party and their relationship to the reasoning in the legal decision.

Claims and Requests of the Plaintiff

List 1: Claims of the Plaintiff

1. Allegation that the plaintiff did not request the service contract and was a victim of fraud <ap1>.
2. Declaration that there was no consent or signature for the contract, with indications of forgery <ap2>.
3. Statement that the received amount was transferred without authorization or prior knowledge <ap3>. (...)
6. Declaration that there was an attempt at extrajudicial resolution, but without success <ap6>.

List 2: Requests of the Plaintiff

1. Request for the granting of free legal aid <pa1>.

2. Request for procedural priority, based on specific legal requirements <pa2>.
3. Request for preliminary injunction to immediately suspend charges and negative credit reporting <pa3>. (...)
7. Request for the defendant to be ordered to pay costs and attorney's fees <pa7>.

Figure 3: An example from SARA's analytical output with the claims and requests submitted by the plaintiff in a judicial case.

Evidence Analysis Document

Title: Institutional Contact Document – Service Protocol
Subtitle: Institution Contact Information
Date: Not specified in the image
Page: Page(2)
Sheet: fls(18)
Produced by: Financial institution
Requested by: Plaintiff

Objective: To comply with the request to prove the attempt of administrative resolution made by the plaintiff.
Evidence Outcome: The image contains contact information for the institution, including email and phone numbers, as well as details about previous service interactions.
Content Summary: The document presents an email address for contacting the institution and phone numbers for support services, along with a protocol number associated with a specific interaction (000000XXXX), where the attendants' names are generically mentioned.

Legal Basis: The plaintiff alleges that they attempted to resolve the issue administratively directly with the institution but were unsuccessful. The presented evidence corroborates this narrative.
(...)

Figure 4: An example of SARA's analytical output of a specific evidence.

Module 3 - Analysis of Evidence and Connection to Claims and Requests The objective of this module is to analyze the content of the evidence documents submitted in the case, alongside the outputs from the previous phase (a structured list of claims and requests), utilizing a refined prompt (see (Bomfim et al. 2025b)). At this stage, the LLM is responsible not only for describing the evidence but also for establishing explicit links between the evidence and specific claims or requests made by the parties. Figure 4 illustrates the result of the analysis of the evidence from the plaintiff, including the mandatory elements such as title, subtitle, date, produced by, requested by, among others. In this example, the **Objective** of the evidence presented by the plaintiff in the judicial case was "to prove the attempt to resolve the issue administratively by the petitioner.". By explicitly including claims and requests in the evidence analysis prompt, the LLM receives more targeted context, improving the relevance and precision of its reasoning. This approach strengthens the alignment between evidence and claims, supporting more coherent and reliable legal reasoning.

Module 4 – Retrieval of Case-Related Jurisprudence Jurisprudential precedents represent the largest volume of legal information and pose the greatest challenges for verification and retrieval by judges and their clerks. This complexity arises both from the large volume of available decisions and the diversity of topics, formats, and terminologies involved. As a result, manually searching for relevant precedents is a time-consuming and resource-intensive task, requiring detailed analysis to ensure consistency and legal grounding in judicial decisions.

The jurisprudence retrieval module of the SARA is responsible for submitting selected legal case documents to the Jurisprudential Knowledge Graph (JUR-KG) via an API call. Specifically, the module sends the initial petition and the replies, which are procedural documents filed by the plaintiff, the defendant, or both parties. Before submission, all relevant documents from the current legal case undergo a pre-processing stage, where they are cleaned, normalized, and concatenated into a single textual representation ("query"). This unified text is then transmitted to the JUR-KG through an HTTP POST request, accompanied by the following parameters:

- "query": textual content extracted from documents of the legal case;
- "is_semantic": related to the type of syntactic/semantic search that will be performed;
- "search_type": related to the use of "active search" as a search method. This method adds an extra layer of information extraction to the search, allowing the selection of the most appropriate jurisprudence for the new legal case.

In response, the JUR-KG API returns an ordered (by similarity score) list of relevant and curated precedent decisions (jurisprudence), each accompanied by detailed metadata, including: case number, adjudicating court or tribunal, judicial branch (e.g., civil, criminal, administrative), date of the ruling, case summary (ementa), full text of the judgment, thematic or subject-matter classification, and additional information such as the name of the reporting judge (relator) and a computed similarity score indicating the degree of relevance to the current legal case.

Module 5 - Generation of the Legal Reasoning for the Judicial Decision Upon completing the execution of the modules outlined above, a comprehensive dataset and KG of the legal case are prepared for this phase of drafting the judicial decision's reasoning:

- The systematic organization of all claims and arguments submitted by the parties;
- The processing of evidence images and documents, ensuring their conversion into a format interpretable by Large Language Models (LLMs);

```
@prefix lco-br: <http://example.org/lco-br#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix : <http://example.org/instances#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-
schema#> .

# Legal Claim
:PlaintiffClaim001 a lco-br:LegalClaim ;
    rdfs:label "The claimant did not request the
service contract and was a victim of fraud." ;
    prov:wasDerivedFrom :LegalDocument001 .

# Related Legal Document
:LegalDocument001 a lco-br:LegalDocument ;
    rdfs:label "Initial Petition by Plaintiff" ;
    lco-br:documentDate "2025-01-05"^^xsd:date ;
    lco-br:documentType "Initial Petition" ;
    prov:wasGeneratedBy :DocumentSubmission001 .
```

```
# Document Submission
:DocumentSubmission001 a lco-br:DocumentSubmission
;
    rdfs:label "Submission of initial petition" ;
    prov:qualifiedAssociation [
        prov:hadAgent :Plaintiff001 ;
        prov:hadRole :RolePlaintiff
    ] ;
    prov:wasGeneratedBy :ActivityFiling001 .

# Legal Agent (Plaintiff)
:Plaintiff001 a lco-br:LegalAgent ;
    rdfs:label "Vladia Pinheiro (Plaintiff)" .

# Role of the Agent
:RolePlaintiff a prov:Role ;
    rdfs:label "Plaintiff" .
```

Figure 5: An example of one specific plaintiff's claim represented using the LCO-BR classes in Turtle syntax.

- The description of each evidence presented by the parties, integrating the contextual nuances of the claims and arguments;
- The set of jurisprudential precedents related to the case, retrieved from the Jur-KG.

Based on these datasets and the Knowledge Graph (jurisprudential precedents), judges and their clerks are responsible for evaluating and selecting the jurisprudences and evidences to be considered in the legal reasoning, according to the case outcome: granted, denied, or partially granted. In addition, the judge must describe the reasons for the decision and may optionally provide a template for SARA to follow (if the default template is not to be used). Figure 6 presents the SARA User Interface prior to the generation of the legal reasoning

Finally, SARA generates the legal reasoning text that supports the judicial decision using a prompt with the following parameters: ***claims_requests_content*** (the list of claims and requests), ***docs_content*** (all case documents including petition, defenses, and case report), ***evidence_content*** (Module 3 evidence analysis selected by the user), ***jurisprudence_content*** (selected jurisprudential entries retrieved from the JUR-KG ranked list), and ***verdict*** (the judgment outcome: granted (G), denied (D), or partially granted (PG)).

Figure 7 presents an example of the legal reasoning for a legal case, generated by SARA. In the example, the tags **<begin evN> / <end evN>** delimit the citation to the evidence; the tags **<begin apN> / <end apN>** delimit the citations to plaintiff's claims; the tags **<begin arN> / <end arN>** delimit the citations to defendant's claims; and the tags **<begin juN> / <end juN>** delimit the citations to jurisprudential precedents. In summary, the presence of tags is not limited to mere formal labeling: they play a practical role by enabling easy review of the coverage of requests, claims, evidence, and jurisprudence in the case by legal advisors and judges. Figure 8 presents the SARA user interface with a partial view of the Legal Reasoning generated by the legal case 0255892-30.2021. On the right side, the selected jurisprudential entries are grouped based on their similarity with the argumentative origin: plaintiff, defendant, or both parties. The arrow indicates the portion of the legal reasoning where one of the jurisprudential precedents recommended by the JUR-KG is cited—namely, precedent No. 0174992-02.2022.0178.060001. The interface also provides features designed to facilitate the practical use of the retrieved information, such as the ability to directly copy the case summary or the full ruling to the clipboard. This functionality is intended to support the legal reasoning process by allowing users to select and reuse relevant excerpts as needed, thereby streamlining the identification of legal grounds aligned with the case under review. Finally, SARA generates the case-specific knowledge graph based on the LCO-BR ontology. Figure 9 illustrates the KG of the legal Case 0255892-30.2021 with 12 requests (purple nodes), 10 claims (red nodes), and 5 jurisprudential precedents (green nodes). From the Knowledge Graph (KG), it is possible to retrieve that several requests, claims, and jurisprudential precedents were not cited in the legal reasoning **lco-br:legalReasoning_0255892**. This insight is particularly valuable as it enables the identification of potential gaps in legal argumentation. By detecting uncited claims, the system can help ensure that all relevant case elements are considered, improving the completeness and transparency of

Figure 6: SARA User Interface displaying the decision parameters prior to the generation of the legal reasoning.

the legal reasoning. Additionally, this feature enables legal professionals to check whether any essential evidence has been missed, strengthening the soundness and interpretability of judicial decisions.

Experimental Evaluation

SARA is currently in operation with over 240 active users at the Ceará State Court of Justice (TJCE) and has played a key role in helping the court achieve consecutive productivity records in its judicial workflows. In 2024, the TJCE surpassed 737,000 rulings, marking a 15.3% increase in productivity compared to 2023 (Tribunal de Justiça do Estado do Ceará 2025).

For all experiments reported in this section, SARA used GPT-4o as the default LLM for reasoning generation. However, SARA is model-agnostic: the LLM is fully parametrizable, allowing the selection among multiple homologated models (e.g., GPT-4, GPT-5-mini, GPT-5). The TJCE continuously evaluates and approves the set of available models integrated into SARA.

In a quantitative experimental evaluation, we aimed to assess the semantic similarity between the legal reasoning generated by SARA, with and without the inclusion of jurisprudential precedents retrieved from the JUR-KG, and the reference legal reasoning authored by judges and their clerks. For this purpose, we employed the ALIGNSCORE_embeddings metric (Zha et al. 2023) and BERTScore (F1) (Zhang et al. 2020), both well-established, reference-based metrics for semantic alignment. For BERTScore, we used the RoBERTa-large pretrained model, with the default configuration of the bert-score package v0.3.13. A curated dataset comprising five legal cases, selected by the judges themselves, was assembled to support this evaluation. The results of this analysis are summarized in Table **??**.

According to the results, the legal reasoning generated with the support of the JUR-KG showed, on average, a 2.97% improvement in the ALIGN-SCORE metric, indicating a measurable gain in semantic alignment with the reference reasonings authored by judges. However, when evaluated using the BERTScore metric, two out of three AI-generated reasonings with JUR-KG support exhibited a decrease in semantic similarity compared to the human-authored versions. These results suggest that while the integration of structured jurisprudential knowledge may enhance the contextual relevance and traceability of legal justifications, its impact on surface-level semantic similarity remains inconclusive. This reinforces the need for a larger and more diverse set of annotated legal cases to support statistically robust evaluations.

The knowledge graphs generated for the legal cases enabled the system to answer a set of key competence questions (CQs), providing structured insights into the legal reasoning process. As summarized in Table 3, the knowledge graphs allowed for the extraction and organization of a wide

Legal Reasoning

The central point of the controversy is to decide whether ¡ **Plaintiff** ¿ claims regarding the declaration of non-existence of debt, (..). Initially, I will analyze the (...), the evidence presented by the plaintiff **<begin ev1>** notably the graphotechnical expert report (pages 179–193) **<end ev1>**, indicates that the signature on the contested contract is not authentic, demonstrating the possibility of fraud. Consequently, I reject the defendant's claim that there was no fraud.

On the merits, analyzing the (...), according to article 6, VIII, of the CDC. In addition, articles 186 and 927 of the Civil Code, which determine liability for an unlawful act, are also applicable.
(...)
In the case, ¡ **Plaintiff** ¿ proved a fact constituting the alleged right **<begin ev2>** through the expert evidence that attests to the forgery of the signature (pages 179–193) **<end ev2>**, **<begin ev3>** as well as the police report that reports the occurrence of fraud (page 17) **<end ev3>**. It also alleges that it did not authorize the payroll loan, did not touch the amount deposited and tried, unsuccessfully, to

resolve the situation administratively (p. 18). **<begin ap1>** This evidence proves its allegation of fraud and makes it impossible for the bank to defend itself **<end ap1>**.
In turn, ¡ **Defendant** ¿ claimed that the plaintiff would have enjoyed the amount and that the contract was entered into in a regular manner, **<begin ev4>** presenting documentation of a simplified proposal (fls. 50) **<end ev4>**, **<begin ev5>** and electronic transfer of the amount (fls. 48) **<end ev5>**. However, **<begin ar1>** it did not prove the existence of a fact modifying or extinguishing the plaintiff's right, since the evidence presented is based on documents with signatures already proven to be altered or false. **<end ar1>**

(...)
<begin ju1> The case law corroborates this conclusion, and it is relevant to mention the summary of case No. (XXX) of the TJCE **<end ju1>**, which recognizes inconsistencies in bank contracts arising from fraud and consequent strict liability of the bank in such occurrences. **<begin ju2>** As decided by the Court of Justice of the State of Ceará in Civil Appeal No. 02031563620228060151, judged on 10/21/2020, "the bank contract presented does not meet the minimum validity requirements. . . " **<end ju2>**.(...).

Figure 7: Legal reasoning generated by SARA for a legal case, with citations explicitly tagged to indicate claims, arguments, evidence, and jurisprudential precedents.

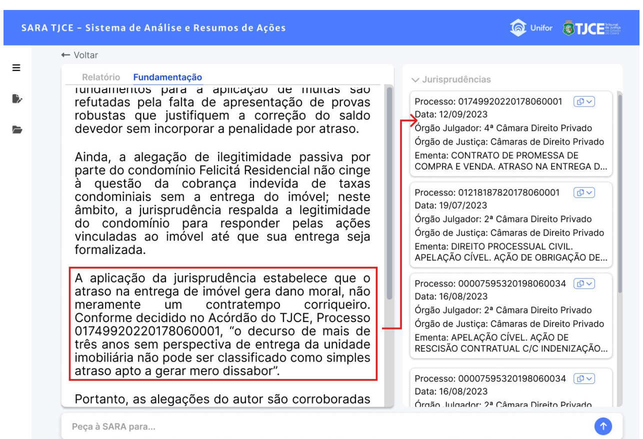

Figure 8: SARA User Interface displaying the legal reasoning and retrieved jurisprudence for Legal Case No. 0255892-30.2021

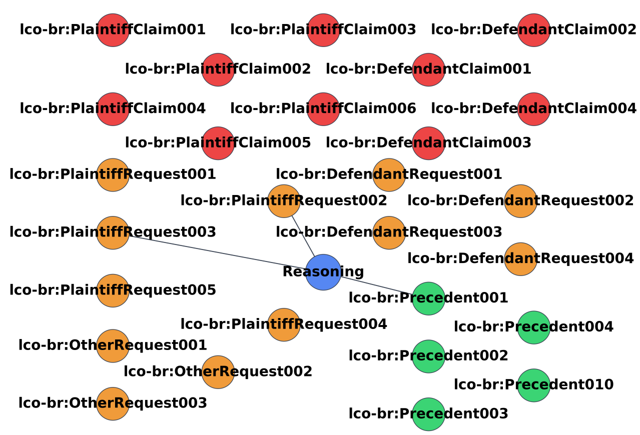

Figure 9: The Knowledge Graph of the Legal Case 0255892-30.2021.

range of elements, such as claims, requests, evidence, and jurisprudence, for each legal case, supporting a more detailed and traceable analysis.

- **CQ1 -** Which elements of a legal case (claims, requests, evidence, and jurisprudence) were not quoted in the legal reasoning generated by SARA?
  - 87% of the extracted claims;
  - 92.8% of the extracted requests;
  - 0,09% of the extracted evidence;
  - 95.9% of the extracted jurisprudential precedents.
- **CQ2 -** Which legal claims or requests should be dismissed because they rely on invalid evidence?
  - no claim or request was dismissed due to invalid evidence.
- **CQ3 -** Which claims and requests quoted in the legal rea-

Case	ALIGN	BERT	ALIGN	BERT
	w/ JUR-KG		w/o JUR-KG	
0255892-30.2021	0.5161	0.6875	0.4966	0.7163
0207225-76.2022	0.3791	0.6808	0.3598	0.6726
0181397-83.2019	0.3004	0.6601	0.2519	0.6574
0154724-53.2019	0.3524	0.6905	0.3257	0.7230
0131659-97.2017	0.3497	0.6737	0.4091	0.6692

Table 2: Comparative evaluation of semantic similarity metrics for reasoning with and without Jur-KG support.

Caso	Claims (E / C)	Requests (E / C)	Evidence (E / C)	Juris (E / C)
0006309-46.2010	10 / 5	14 / 3	1 / 1	5 / 0
0130337-08.2018	13 / 5	13 / 0	3 / 2	5 / 0
0131659-97.2017	12 / 0	12 / 0	0 / 0	5 / 0
0154724-53.2019	8 / 0	12 / 0	2 / 2	4 / 0
0181397-83.2019	9 / 0	9 / 0	0 / 0	5 / 0
0207225-76.2022	10 / 0	11 / 0	0 / 0	5 / 0
0238593-06.2022	12 / 0	16 / 0	0 / 0	5 / 0
0243585-78.2020	11 / 0	13 / 0	2 / 2	5 / 1
0253487-55.2020	13 / 4	13 / 4	3 / 3	5 / 0
0255892-30.2021	10 / 0	12 / 2	0 / 0	5 / 1
Total	108 / 14	125 / 9	11 / 10	49 / 2

Table 3: Resumo estatístico da extração e organização dos elementos legais (claims, requests, evidence e jurisprudência) por processo. Os pares "E / C" indicam valores "Extraídos / Citados".

soning of the case have no supporting evidence?
  - 69.5% of the claims and requests have no supporting evidence.
- **CQ4 -** Which jurisprudences are relevant to a given legal claim or request?
  - 0,04% of the recovered jurisprudence was relevant.
- **CQ5 -** Which jurisprudential precedents referenced in the case documents were not matched or retrieved from the JUR-KG?
  - In these legal cases, all jurisprudential precedents cited in the legal reasoning were retrieved from the JUR-KG.

To enable qualitative expert assessment, SARA incorporates a continuous evaluation module where judges rate generated legal reasoning and provide feedback. A group of 22 judges and appellate judges at the Court of Justice of Ceará was formally appointed to review system outputs. Between June 14, 2024, and August 13, 2025, they evaluated 390 legal reasonings, with 94.4% rated positive. In SARA's integrated evaluation module, judges provide binary feedback—'positive' or 'negative'—together with an optional written justification. Therefore, the 94.4% score reflects the proportion of reasonings explicitly marked as positive by the evaluators. A 'positive' rating indicates that the judge considered the generated reasoning to be structurally adequate, legally coherent, and suitable for use as a basis for drafting a final judicial decision. Conversely, 'negative' evaluations required judges to provide a textual justification, identifying omissions, inconsistencies, hallucinated arguments, or structural deficiencies. This binary feedback mechanism is part of SARA's internal quality monitoring system and is used by the TJCE technical team for continuous iterative refinement of prompts and system behavior. Judges emphasized that SARA produced well-structured reasoning, properly addressed preliminary matters and evidence, and effectively integrated jurisprudence, often aligning with decision models. Negative evaluations (5.6%) cited superficial or generic outputs, omissions of key elements, misapplied jurisprudence, and occasional structural flaws or hallucinations, such as irrelevant arguments and formatting inconsistencies. This feedback is continuously incorporated into system improvements and underscores the value of knowledge graphs as an interpretability layer, helping assess reasoning coverage, track case elements, and strengthen the fairness, accountability, and transparency of AI-assisted decision-making.

Conclusion

This work presented an integrated approach for generating AI-assisted legal reasoning through the SARA system, enhanced by the incorporation of structured jurisprudential knowledge from the JUR-KG. Our experimental results demonstrate that leveraging semantic precedents improves the alignment between AI-generated and human-authored legal justifications, as reflected in a measurable gain in semantic similarity metrics. Furthermore, the use of knowledge graphs enables the identification and traceability of legal elements, such as claims, evidence, and jurisprudence, supporting more transparent, auditable, and comprehensive judicial decision-making.

However, the study also presents limitations. The evaluation was conducted on a relatively small dataset of five legal cases, which may not fully capture the diversity and complexity of broader judicial contexts. Additionally, while semantic similarity metrics such as ALIGNSCORE provide quantitative insights, they do not fully account for legal correctness or argumentative sufficiency. Another important limitation is the dependency on OCR quality when processing image-based evidence, which may affect the accuracy of element extraction.

Future research should expand the dataset to include a wider range of cases, jurisdictions, and document types to enhance generalizability. Further efforts are also needed to incorporate formal argumentation structures and legal ontologies into the reasoning process, enabling more robust justification mechanisms. Lastly, user studies with judges and legal professionals could help assess the perceived reliability and usefulness of AI-generated reasoning in real-world judicial workflows.

References

Bomfim, F. C. J.; Monteiro Neto, J. A.; Bezerra Filho, G.; Furtado, V.; and Pinheiro, V. 2025a. SARA – A Generative AI for Legal Process Summarization Based on Chain of Density Prompt Engineering. In *BRACIS 2024*, volume 15413 of *LNAI*, 370–383. Cham: Springer.

Bomfim, F. C. J.; Neves, A. S.; Silva, S. P.; Monteiro Neto, J. A. M.; Furtado, V.; and Pinheiro, V. 2025b. Automating Legal Reasoning with Generative AI: A Framework for Judicial Decision Support. In *Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL)*. Forthcoming.

Conselho Nacional de Justiça. 2025. Metas Nacionais aprovadas pela Justiça Estadual para 2025. https://www.cnj.jus.br/gestao-estrategica-e-planejamento/metas/justica-estadual/. Accessed: 2025-01-26.

G1. 2023. Juiz usa IA para fazer decisão e cita jurisprudência falsa; CNJ investiga. https://g1.globo.com/politica/blog/daniela-lima/post/2023/11/13/juiz-usa-inteligencia-artificial-para-fazer-decisao-e-cita-jurisprudencia-falsa-cnj-investiga-caso.ghtml. Accessed: 2025-05-13.

Lebo, T.; Sahoo, S.; Mcguinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D.; Garijo, D.; Soiland-Reyes, S.; Zednik, S.; and Zhao, J. 2013. *PROV-O: The PROV Ontology*.

Meta AI. 2024. The LLaMA 3 Herd of Models. https://ai.meta.com/research/publications/the-llama-3-herd-of-models/. Accessed: 2025-05-13.

OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774.

Pires, R. S.; Santos, H.; Guedes, R.; Monteiro Neto, J.; Caminha, C.; and Furtado, V. 2022. Building and Analyzing the Brazilian Legal Knowledge Graph. In *AI4Legal Workshop at ISWC*.

Rajasekar, K. P.; and Vezhaventhan, D. 2024. Artificial intelligence revolutionizing legal and forensic practices: A comprehensive analysis. In *Proceedings of the 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–12.

Tribunal de Justiça do Estado do Ceará. 2025. TJCE supera 737 mil julgamentos em 2024... https://www.tjce.jus.br/noticias/recorde-historico-tjce-supera-737-mil-julgamentos-em-2024-e-aumenta-em-153-a-produtividade/. Accessed: 2025-05-13.

W3C. 2013. PROV-O: The PROV Ontology. https://www.w3.org/TR/prov-o/. Accessed: 2025-05-13.

Zha, Y.; Xu, Y.; Xu, J.; Liu, Y.; Neubig, G.; and Liu, P. 2023. ALIGNSCORE: Evaluating Factual Consistency with a Unified Alignment Function. https://github.com/yuh-zha/AlignScore. Accessed: 2025-05-13.

Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K. Q.; and Artzi, Y. 2020. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations (ICLR)*.