



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CAMPUS ARARANGUÁ  
TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

Jonas da Silva Elias

**Recuperação e Análise de Jurisprudências como Suporte a Elaboração de  
Defesas em Processos Jurídicos por meio de Modelos de Linguagem**

Araranguá

2024

Jonas da Silva Elias

**Recuperação e Análise de Jurisprudências como Suporte a Elaboração de Defesas em Processos Jurídicos por meio de Modelos de Linguagem**

Trabalho de Conclusão de Curso submetido ao Curso de Graduação em Tecnologias da Informação e Comunicação do Centro de Ciências, Tecnologia e Saúde do Campus de Araranguá da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Tecnologias da Informação e Comunicação.

Orientador(a): Prof. Alexandre Leopoldo Gonçalves, Dr.

Coorientador(a): Rafael Cândido, Me.

Araranguá

2024

Elias, Jonas da Silva

Recuperação e Análise de Jurisprudências como Suporte a  
Elaboração de Defesas em Processos Jurídicos por meio de  
Modelos de Linguagem / Jonas da Silva Elias ; orientador,  
Alexandre Leopoldo Gonçalves, coorientador, Rafael  
Candido, 2024.

65 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Campus Araranguá,  
Graduação em Tecnologias da Informação e Comunicação,  
Araranguá, 2024.

Inclui referências.

1. Tecnologias da Informação e Comunicação. 2.  
Jurisprudência. 3. Elaboração de defesa. 4. Busca  
semântica. 5. Vetores densos. I. Gonçalves, Alexandre  
Leopoldo. II. Candido, Rafael. III. Universidade Federal  
de Santa Catarina. Graduação em Tecnologias da Informação  
e Comunicação. IV. Título.

Jonas da Silva Elias

**Recuperação e Análise de Jurisprudências como Suporte a Elaboração de Defesas em  
Processos Jurídicos por meio de Modelos de Linguagem**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel e aprovado em sua forma final pelo Curso Tecnologias da Informação e Comunicação.

Araranguá, 27 de Junho de 2024.

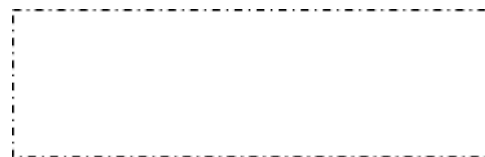


Prof. Fernando José Spanhol, Dr.  
Coordenação do Curso

**Banca examinadora**



Prof. Alexandre Leopoldo Gonçalves, Dr.  
Orientador



Prof.(a) Marina Carradore Sérgio, Dr.(a)  
Universidade Federal de Santa Catarina



Prof. Fabrício Herpich, Dr.  
Universidade Federal de Santa Catarina

Araranguá, 2024.

## RESUMO

Jurisprudências são essenciais para a prática jurídica, sendo fundamentadas por uma linguagem técnica que ultrapassa o vocabulário comum. Os desafios técnicos da abordagem jurídica requerem atenção para superar desafios relacionados a sua linguagem peculiar. Neste contexto, este trabalho objetiva a recuperação de jurisprudências e elaboração de defesas jurídicas através de busca semântica com o auxílio de Modelos de Linguagem de Grande Porte (do inglês *Large Language Model* - LLM). As jurisprudências utilizadas foram adquiridas a partir da técnica de *scrapping* e armazenadas em um banco de dados relacional. Em seguida, ocorre a sumarização das jurisprudências, transformação em vetores densos (*embeddings*) e indexação em banco de dados vetorial. Isto permite que o profissional do direito realize buscas nas jurisprudências indexadas e, considerando documentos de sua escolha, efetue a geração de defesa jurídica. Para análise do método, foram estabelecidos cinco cenários específicos. Os quatro primeiros cenários avaliam a ordenação efetuada por profissionais jurídicos em relação a ordenação provida pelo protótipo que instancia o método proposto através do coeficiente de Spearman. Apesar da maioria das correlações serem positivas fracas ( $< 0.3$ ) ou negativa, com discrepâncias consideráveis entre os profissionais, existem correlações positivas com certa relevância, demonstrando potencial de entrega de documentos aderentes à demanda dos profissionais. O quinto cenário específico trata da avaliação de um dos advogados quanto à elaboração de sugestão de defesa, sendo percebida como positiva. Diante dos resultados, o método proposto de recuperação de jurisprudências e elaboração de defesas utilizando LLMs demonstra potencial em servir de suporte a atividades consideradas complexas por profissionais na área do direito.

**Palavras-chave:** Jurisprudência. Elaboração de defesa. Busca semântica. Vetores densos.

## ABSTRACT

Jurisprudences are essential to legal practice, often grounded in technical language that surpasses common vocabulary. The technical challenges of legal approaches require attention to overcome issues related to their peculiar language. In this context, this work aims to retrieve jurisprudences and draft legal defenses through semantic search using Large Language Models (LLMs). The jurisprudences were acquired through scraping techniques and stored in a relational database. Next, jurisprudences are summarized, transformed into dense vectors (embeddings), and indexed in a vector database. This allows legal professionals to search indexed jurisprudence and generate legal defenses based on selected documents. To analyze the method, five specific scenarios were established. The first four scenarios evaluate the ranking performed by legal professionals compared to the ranking provided by the prototype that instantiates the proposed method using the Spearman coefficient. Despite most correlations being weakly positive ( $< 0.3$ ) or negative, with considerable discrepancies among professionals, there are positive correlations of some relevance, demonstrating potential to retrieve documents tailored to professionals' demands. The fifth specific scenario evaluates one lawyer's assessment of defense suggestion drafting, perceived positively. Given the results, the proposed method of retrieving jurisprudences and drafting defenses using LLMs shows potential to support activities considered complex by legal professionals.

**Keywords:** Jurisprudence. Defense preparation. Semantic search. Embeddings.

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>9</b>
1.1 OBJETIVOS	11
<b>1.1.1 Objetivo geral</b>	<b>11</b>
1.1.2 Objetivos Específicos	11
1.2 JUSTIFICATIVA E RELEVÂNCIA DO TEMA	11
1.3 ESTRUTURA DO TRABALHO	12
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>14</b>
2.1 JURISPRUDÊNCIAS	14
2.2 PROCESSAMENTO DE LINGUAGEM NATURAL	15
2.2.1 Embeddings	16
2.2.2 Large Language Models (LLM)	17
2.3 RECUPERAÇÃO DE INFORMAÇÃO	19
2.3.1 Banco de Dados Vetorial	21
2.4 TRABALHOS CORRELATOS	22
<b>3 METODOLOGIA</b>	<b>29</b>
3.1 ENQUADRAMENTO METODOLÓGICO	29
3.2 DESIGN SCIENCE RESEARCH METHODOLOGY	30
3.3 DESENVOLVIMENTO DA PESQUISA	31
<b>3.3.1 Identificação do problema e motivação</b>	<b>32</b>
<b>3.3.2 Definição dos objetivos</b>	<b>33</b>
<b>3.3.3 Projeto e desenvolvimento</b>	<b>33</b>
3.3.3.1 Coleta de dados	33
3.3.3.2 Transformação dos dados	34
3.3.3.3 Indexação e busca semântica	35
<b>3.3.4 Demonstração</b>	<b>36</b>
<b>3.3.5 Avaliação</b>	<b>37</b>
<b>3.3.6 Comunicação</b>	<b>37</b>
<b>4 MÉTODO PROPOSTO</b>	<b>39</b>
4.1 ETAPA 1: COLETA DE DADOS	40
4.2 ETAPA 2: LIMPEZA E SUMARIZAÇÃO	41
4.3 ETAPA 3: GERAÇÃO DOS EMBEDDINGS	43
4.4 ETAPA 4: BANCO DE DADOS VETORIAL	43
4.5 ETAPA 5: SOLICITAÇÃO DA JURISPRUDÊNCIA	43
4.6 ETAPA 6: RECUPERAÇÃO DE DOCUMENTOS	44
4.7 ETAPA 7: MODELO DE LINGUAGEM	45
4.8 ETAPA 8: DEFESA JURÍDICA	45
<b>5 ANÁLISE E DISCUSSÃO DOS RESULTADOS</b>	<b>49</b>
5.1 CENÁRIO DE ESTUDO	49
5.2 IMPLEMENTAÇÃO DO MÉTODO PROPOSTO	49

5.3 APRESENTAÇÃO DOS RESULTADOS	51
<b>6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS</b>	<b>60</b>



## 1 INTRODUÇÃO

O cenário jurídico enfrenta desafios significativos ao lidar com um alto volume de informações, especialmente no que se refere às jurisprudências. A transformação digital, promovendo crescente digitalização de documentos legais, resulta em uma elevada quantidade de dados não estruturados (Fernandes *et al.*, 2022). Todavia, este cenário representa oportunidades, bem como uma complexidade adicional ao domínio jurídico, que é responsável por gerenciar um vasto conjunto de jurisprudências.

Sansone e Sperlí (2022) indicam que a transição para o digital proporcionou um amplo acesso às informações jurídicas, no entanto, tem promovido um desafio significativo em termos de eficiência na busca e recuperação desses dados. Métodos tradicionais de pesquisa mostram-se insuficientes diante da quantidade informacional presente no ambiente jurídico contemporâneo. A necessidade de ferramentas inovadoras que estejam capacitadas para lidar eficazmente com a complexidade linguística e vastidão de dados torna-se cada vez mais premente (Ma *et al.*, 2024).

O domínio jurídico, de forma geral, é um ambiente que se destaca por características linguísticas distintas, envolvidas por terminologias complexas e sutilezas semânticas específicas. Diante deste cenário, as jurisprudências, fundamentais para a prática jurídica, são frequentemente expressas por uma linguagem técnica e jurídica que vai além do vocabulário comum (Sansone; Sperlí, 2022). Esta complexidade linguística, com diversos termos especializados, expressões específicas e construções legais com particularidades, demanda uma compreensão profunda e uma metodologia avançada para a representação eficaz dos documentos jurídicos. Em um contexto computacional recente, a Inteligência Artificial (do inglês *Artificial Intelligence* - AI) e seus subcampos, Aprendizado de Máquina (do inglês *Machine Learning* - ML) e de Aprendizado Profundo (do inglês *Deep Learning* - DL), tem promovido suporte para tal.

A abordagem técnica das jurisprudências requer uma maneira cuidadosa para superar os desafios impostos por sua linguagem peculiar. A compreensão completa do contexto legal e a captura precisa das sutilezas semânticas são determinantes para o desenvolvimento de modelos eficazes de DL. Isto se amplia ao lidar com a terminologia específica e também considerar a estrutura única das

sentenças legais, a lógica argumentativa e os detalhes interpretativos que são inerentes ao campo jurídico (Jain; Borah; Biswas, 2021).

Quanto ao estado da arte em recuperação de jurisprudências, técnicas e tecnologias têm sido adotadas para enfrentar os desafios complexos apresentados pelo cenário jurídico. Predominantemente, a abordagem destaca-se no uso de métodos de AI baseados em DL para lidar com o alto volume de dados não estruturados voltado ao contexto das jurisprudências.

Modelos de linguagem pré-treinados, como BERT (*Bidirectional Encoder Representations from Transformers*) e, outros mais recentes, como o GPT (*Generative Pre-trained Transformer*), são ferramentas úteis na representação semântica avançada de documentos. Tais modelos, treinados em grandes conjuntos de dados textuais, são capazes de capturar relações específicas e variações semânticas, proporcionando uma base sólida para a compreensão profunda do conteúdo legal (Bhattacharya *et al.*, 2022).

Ademais, técnicas específicas de Processamento de Linguagem Natural (do inglês *Natural Language Processing* - NLP) possibilitam representações textuais adequadas para os mais diversos domínios na forma de vetores densos (do inglês *embedding*) visando lidar com as características linguísticas distintas do domínio jurídico. O reconhecimento de entidades jurídicas, a extração de informações relevantes e a análise de sentimentos em textos legais contribuem para uma interpretação mais refinada das jurisprudências. A utilização de estruturas conceituais e modelos semânticos também destaca-se como uma estratégia eficaz para a representação conceitual e organização estruturada dos documentos jurídicos (Mumcuoğlu *et al.*, 2021).

Considerando o contexto apresentado, a seguinte pergunta de pesquisa foi elaborada: “Como armazenar e recuperar jurisprudências considerando suas características semânticas visando promover suporte à elaboração de defesas em processos jurídicos?”.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

Este trabalho tem como objetivo geral propor e desenvolver um método voltado ao armazenamento e recuperação de jurisprudências como suporte à elaboração de defesas em processos jurídicos.

### 1.1.2 Objetivos Específicos

Os objetivos específicos definidos para o presente trabalho são apresentados a seguir:

- Definir a fonte de dados jurídicos a fim de suportar a extração de jurisprudências;
- Definir a abordagem de sumarização textual mais adequada ao contexto jurídico;
- Elencar modelos de representação e armazenamento de jurisprudências sumarizadas de modo que as características semânticas sejam preservadas;
- Elaborar cenário de estudo que possibilite avaliar o método proposto neste trabalho.

## 1.2 JUSTIFICATIVA E RELEVÂNCIA DO TEMA

O cenário jurídico contemporâneo enfrenta desafios frequentes ao lidar com o alto volume de informações, especialmente no contexto das jurisprudências. A transformação digital resultou em uma crescente digitalização de documentos legais, gerando uma vasta quantidade de dados não estruturados. Nesse contexto, a pesquisa em recuperação de jurisprudências por meio de bancos de dados vetoriais com busca semântica possui uma relevância social significativa. A eficiência na gestão e compreensão dessas informações é crucial para a comunidade jurídica, permitindo um acesso mais ágil e preciso a precedentes e decisões judiciais (Sansone; Sperlí, 2022).

Para enfatizar a importância de estudos voltados ao contexto deste trabalho, vale mencionar que o volume de processos no Brasil demanda eficiência na compreensão e gestão de informações jurídicas. Em 2021, o Poder Judiciário concluiu 26,9 milhões de processos e recebeu 27,7 milhões de novas ações, sendo que 97,2% destes já estavam em formato eletrônico. Com 62 milhões de ações judiciais em andamento no final do ano de 2021, a busca por ferramentas mais eficazes de gestão e recuperação de informações jurídicas se torna relevante para assegurar a agilidade e a qualidade na prestação de serviços jurisdicionais (CNJ, 2022).

Desta forma, este trabalho visa contribuir para a área com um método de recuperação de jurisprudências e suporte à produção de textos jurídicos, com auxílio de bancos de dados vetoriais, busca semântica e modelos de linguagem. Ao superar desafios linguísticos e semânticos específicos do campo jurídico, busca-se uma abordagem para representar e acessar informações legais através de métodos baseados em NLP, com o intuito de beneficiar a comunidade jurídica em suas práticas cotidianas. Considerando a evolução contínua do conhecimento jurídico, esta pesquisa pretende contribuir para avanços no campo, fornecendo ferramentas de auxílio para profissionais do direito.

### 1.3 ESTRUTURA DO TRABALHO

Para além deste capítulo que apresenta a temática, e os objetivos juntamente com a justificativa deste trabalho, os demais capítulos são estruturados como segue:

- O segundo capítulo é constituído pela fundamentação teórica, abordando os principais assuntos relacionados à pesquisa, como jurisprudências, processamento da linguagem natural e recuperação de informação, além de apresentar os trabalhos correlatos;
- O terceiro capítulo descreve a metodologia de pesquisa adotada no desenvolvimento deste trabalho;
- O quarto capítulo é composto pelo método proposto com o seu respectivo detalhamento;
- O quinto capítulo apresenta os resultados e discussões deste trabalho;

- Por fim, o sexto e último capítulo, apresenta as considerações finais do trabalho, assim como enfatiza possíveis evoluções futuras.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 JURISPRUDÊNCIAS

Concede-se à jurisprudência o compromisso de garantir a convicção e previsibilidade em processos judiciais, assegurando a isonomia entre os jurisdicionados, destacando o respeito, a sabedoria acumulada e a experiência dos magistrados (Bidinotto; Martins, 2022). Segundo Leung (2019), distintas escolas de filosofia definem os tipos de jurisprudência de forma particular, todavia, chega-se à conclusão entre os envolvidos de que a jurisprudência geral é voltada às especulações ligadas à definição comum da lei, porém diferente de especulações sobre uma lei específica.

Conforme Mendes *et al.* (2021), tradicionalmente o termo de jurisprudência estava relacionado com a interpretação pacífica e reiterada dos tribunais acerca de um tema específico. O entendimento dos pilares pacífico e reiterado, podem ser questionados de certa maneira nos dias atuais, tendo em vista que a interpretação pacífica não pode caracterizar somente o consenso, já que na verdade, caracteriza-se como um entendimento no tribunal, podendo ser uma maioria larga ou menor. A jurisprudência provém de julgamentos proferidos reiteradamente, ou seja, advém normalmente de órgãos fracionários, não oportunizando um precedente qualificado, porém, pode prover de julgamentos concentrados, fixados em teses e precedentes qualificados. Existe portanto uma intersecção entre jurisprudência e precedentes, mas não uma união de identidade absoluta.

Segundo Scheleder e Noschang (2018), o precedente fornece uma abordagem “universalizável”, podendo dispor a aplicação como regra de decisão no caso sucessivo em função da identidade ou até em razão da analogia entre os acontecimentos de um segundo caso. Os precedentes representam os princípios que visam orientar o entendimento de normas dialéticas de argumentação legal complexas, oferecendo suporte à compreensão entendida como legítima no contexto da argumentação justificativa da decisão. De acordo com Avelino (2020), os precedentes em si, como decisão jurídica, não possuem como atribuição principal o efeito vinculante. O efeito vinculante implica que as decisões tomadas em tal modalidade de controle devem ser obrigatoriamente seguidas (Ferraz; Barbosa, 2017). Pelo fato de recuperar consigo uma tese ou princípio jurídico construído a

partir da análise do caso concreto, não caracterizando obrigatoriamente um caso análogo futuro. É relevante impedir a permissão do reconhecimento referente ao efeito vinculante como precedente, que apesar de muitas vezes utilizado, em verdade refere-se ao seu *ratio decidendi* específico, que é o elemento identificador da decisão judicial, mas não se refere ao precedente efetivamente.

Nesse sentido, Scheleder e Noschang (2018), levam em consideração que a jurisprudência diferencia-se dos precedentes em variados aspectos, incluindo a falta de análise comparativa dos fatos e a forma de sua respectiva constituição, sendo em grande parte dos casos, representadas por sumulados construídos por serventuários dos tribunais em questão, por meio de breves textos e propósitos que por vezes são desvinculados dos casos concretos originários, no qual criam determinadas regras jurídicas. Por outro lado, o precedente está relacionado com um caso particular, diferente da jurisprudência, que foca em indicar uma quantidade detalhada de decisões referente aos diversos casos classificados como concretos, ou seja, o precedente, em regra, é constituído a partir de uma decisão exclusiva ao contrário da jurisprudência, no qual tem origem baseada em distintas decisões relacionadas ao assunto, não sendo possível identificar qual das decisões deu origem para a formação da jurisprudência.

Bidinotto e Martins (2022) entendem que, no Brasil, somente o Supremo Tribunal Federal (STF) e o Superior Tribunal de Justiça (STJ) formulam precedentes, responsabilizando os Tribunais Regionais Federais e os Tribunais de Justiça de decidirem quanto à categorização da jurisprudência. Nesse aspecto, os precedentes não correspondem às decisões judiciais, porém existe a possibilidade de serem elaborados a partir desses. Apesar da utilização da jurisprudência como precedente por muitos, nem toda decisão judicial pode ser considerada um precedente. A afirmação anterior gera instabilidade aos litigantes, pois a jurisprudência em si não proporciona a segurança, mas sim as razões significativas quanto ao fundamento da decisão judicial.

## 2.2 PROCESSAMENTO DE LINGUAGEM NATURAL

A linguagem é uma competência eminente nos seres humanos, amplamente utilizada para expressar e comunicar com outros indivíduos, desenvolvendo-se nos primeiros momentos da infância e evoluindo ao longo da vida. As máquinas, porém,

por sua natureza não podem compreender as técnicas de comunicação de forma equivalente ao ser humano, a menos que sua infraestrutura esteja configurada com algoritmos de Inteligência Artificial (do inglês *Artificial Intelligence* - AI) (Zhao *et al.*, 2023). Desta forma, tem sido um desafio conquistar esse objetivo, sendo a autonomia relacionada à leitura, escrita e comunicação, um obstáculo para a computação (Castelfranchi, 2013).

Com o propósito de alcançar a compreensão da comunicação através das máquinas, nos últimos anos, o desenvolvimento de Modelos de Linguagem de Grande Porte (do inglês *Large Language Model* - LLM) tem revolucionado o campo do Processamento de Linguagem Natural (NLP). Alguns LLMs lidam com uma variedade de tarefas de NLP, desde a compreensão da linguagem natural, até tarefas de geração de texto, englobando o contexto da AI de forma geral. A utilização destes modelos de forma proficiente requer uma compreensão prática de suas capacidades e limitações, bem como das informações e atividades relacionadas à NLP (Yang *et al.*, 2024).

No contexto jurídico, a quantidade de documentos digitais aplicados ao setor legal tornou-se importante para o desenvolvimento de metodologias específicas baseadas em NLP. A representação concisa de questões legais em um documento é essencial para profissionais do direito, visto que os documentos legais possuem uma estrutura de difícil compreensão e semanticamente diferentes em relação ao domínio jurídico. Documentos legais utilizam com frequência fórmulas rituais ou retóricas, sendo as rituais, fáceis de identificar em relação às retóricas. Textos jurídicos realizam também a citação de outras normas, como referências legislativas ou citações de outras decisões judiciais, em consequência, algumas técnicas de análise de texto são altamente limitadas em sua circunstância caso as partes citadas não sejam consideradas (Sansone; Sperlí, 2022).

### **2.2.1 *Embeddings***

Os *embeddings* constituem-se em representações de palavras em espaços semânticos. São obtidos a partir do treinamento em conjuntos de dados textuais, como os LLMs, o que permite posicionar vetores de palavras em um espaço semântico. O significado de uma palavra pode ser inferido a partir de sua ocorrência estatística com outras palavras, conforme a associação com a linguística. De modo



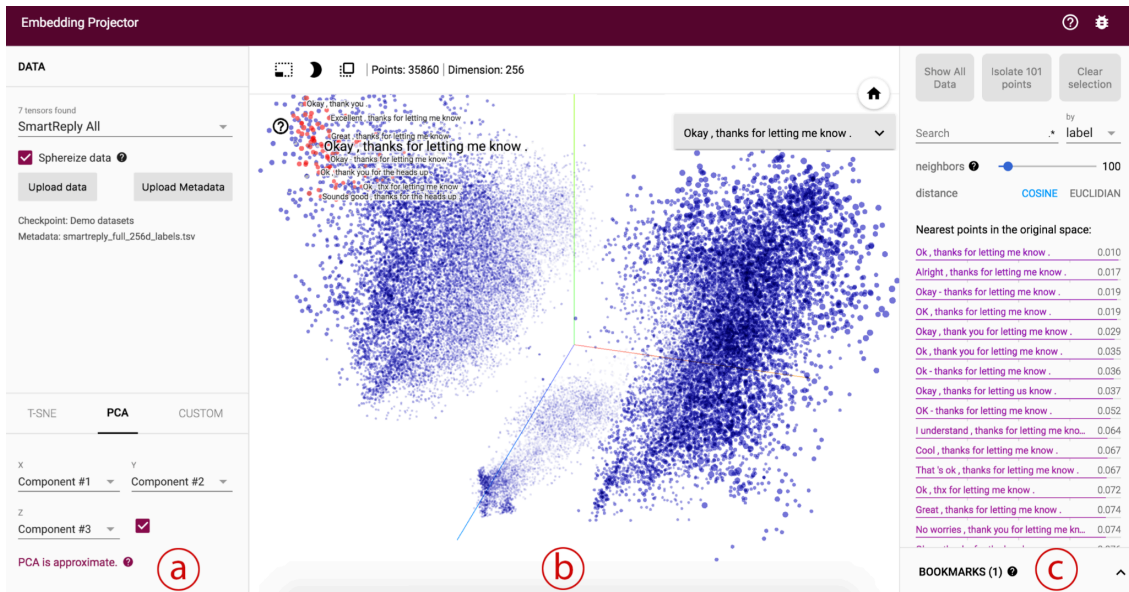
geral, os *embeddings* têm demonstrado bom desempenho em tarefas como a similaridade semântica (Mumcuoğlu *et al.*, 2021).

O uso de *embeddings* em documentos legais pode ser motivado pela necessidade de lidar com uma grande quantidade de texto e pela complexidade inerente à linguagem jurídica. Ao treinar modelos voltados à documentos legais, é possível capturar informações semânticas e contextuais importantes sobre os textos. Esses *embeddings* representam não apenas palavras isoladas, mas também o contexto em que essas são usadas, o que pode ser crucial para entender o significado e a intenção por trás de cada frase ou parágrafo (Bhattacharya *et al.*, 2022).

Smilkov *et al.* (2016) definem *embedding* como sendo uma incorporação de palavras em um mapa de dados de entrada para pontos no espaço euclidiano  $n$ -dimensional. Vale mencionar que, de maneira genérica, cada ponto pode representar uma instância de determinada unidade de informação, como um texto, uma imagem, um vídeo, um áudio, ou mesmo, uma tupla em um banco de dados relacional. Sob uma perspectiva gráfica, *embeddings* podem ser projetados em espaços bi ou tridimensionais. Um exemplo para representar o conceito de *embedding* foi proposto por Smilkov *et al.* (2016) (Figura 1). A aplicação proposta por Smilkov *et al.* (2016) conta com as seguintes características: a) um painel de dados à esquerda, onde usuários podem selecionar colunas de dados para colorir e rotular os pontos; b) uma área de visualização ao centro; e c) um painel de inspeção dos pontos projetados à direita onde usuários podem buscar pontos específicos e visualizar uma lista dos vizinhos mais próximos.

### **2.2.2 Large Language Models (LLM)**

Normalmente, os Modelos de Linguagem de Grande Porte (LLMs) referem-se a modelos de linguagem baseados em arquiteturas de Redes Neurais (do inglês *Neural Networks* - NN) do tipo *transformers*, que possuem bilhões ou mais parâmetros e são treinados a partir de grandes conjuntos de dados baseados em textos. Os LLMs demonstram significativa capacidade de compreensão da linguagem natural e resolução de tarefas complexas, por meio da geração de texto (Zhao *et al.*, 2023).

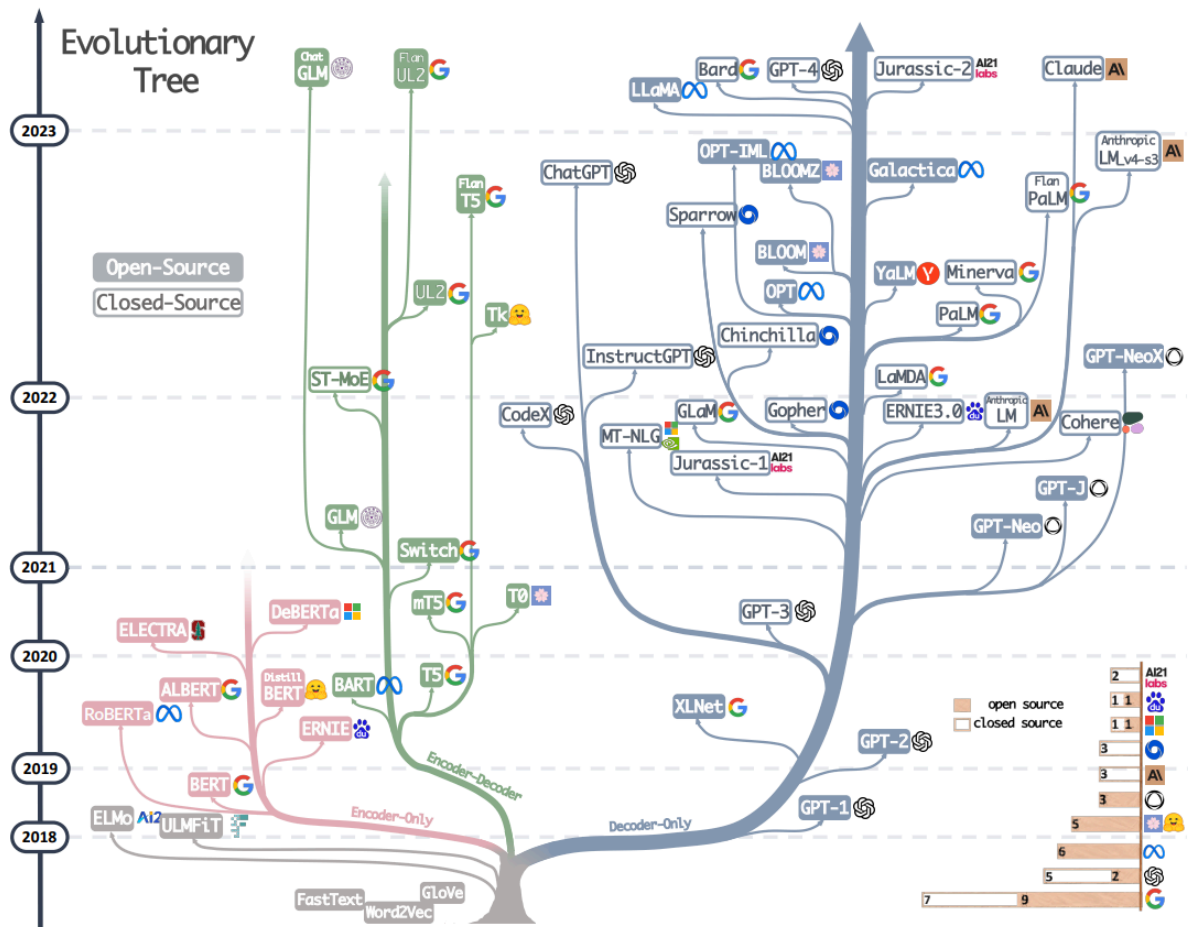
Figura 1 - Projeção de *embeddings*

Fonte: Smilkov *et al.* (2023).

A árvore de evolução dos LLMs modernos, conforme Figura 2, acompanha a construção de modelos de linguagem nos últimos anos e destaca os modelos geralmente mais conhecidos. Esta representação, proposta por (Yang *et al.*, 2024), destaca os modelos baseados em *transformers* (aqueles com cores não cinza), dividindo em modelos decodificadores na ramificação azul, modelos codificadores na ramificação rosa e modelos codificadores-decodificadores na ramificação verde.

A posição vertical dos modelos na linha temporal representa as respectivas datas de disponibilização. Modelos *open source* são representados por quadrados compactos enquanto modelos proprietários são exibidos por quadrados vazios. O gráfico de barras contido no canto inferior direito mostra o número de modelos de algumas empresas e instituições (Yang *et al.*, 2024).

Figura 2 - Árvore de modelos de linguagem



Fonte: Yang *et al.* (2024).

### 2.3 RECUPERAÇÃO DE INFORMAÇÃO

A Recuperação de Informação (do inglês *Information Retrieval* - IR), é uma área de pesquisa relacionada ao contexto de NLP, concentrando-se em capturar e associar informações relevantes a partir de dados não estruturados, geralmente na forma de textos. As técnicas de IR estimularam o desenvolvimento de aplicações no mundo real, incluindo motores de busca, plataformas de recomendação e sistemas de perguntas e respostas. A área preocupa-se com problemas relacionados ao armazenamento eficiente, disponibilidade e controle de informações textuais. A informação está se tornando cada vez mais disponível e acessível nos computadores, visto que as redes de computadores e a *internet*, bem como a *World Wide Web*, vem oferecendo a comunicação de informações de forma mais prática, enquanto novas arquiteturas de computadores vem diminuindo o custo. Novas tecnologias e a constante evolução das técnicas de AI, tornam viável a introdução de

algoritmos com capacidade de armazenar, recuperar e apresentar volumes extensos de informações em diversas mídias de forma contemporânea e destacada (Hou; Cosma; Finke, 2023), além de permitir uma representação que leve em consideração a semântica de determinado texto.

Neste sentido, Kraft e Colvin (2017) afirmam que a busca semântica em documentos tem o objetivo de encontrar informações que não se baseiam somente na existência de termos, mas também em seu significado. Essa tarefa constitui-se em uma abordagem diferenciada da IR, sendo os documentos recuperados com base na relevância, além das palavras. A premissa básica é bastante semelhante, sendo que um documento é caracterizado pelo conjunto de *tokens*<sup>1</sup> que compreendem o seu teor, ignorando sua estrutura. Já a abordagem básica de IR considera os radicais das palavras como *tokens*. No caso da busca semântica, a indexação é tipicamente uma combinação de palavras, conceitos que transmitem o significado de algumas dessas palavras e relação entre as categorias específicas. Essa abordagem, permite que caso a busca no documento esteja sendo realizada sobre o Reino Unido, também encontre documentos que mencionam Cambridge, por exemplo.

Lashkari, Bagheri e Ghorbani (2019) afirmam que para lidar com representações mais complexas, tem-se evoluído na criação de motores de busca semântica capazes de promover respostas adequadas a partir de perguntas mais complexas, para além do seria possível em sistemas de IR baseados em palavras-chave. Grande parte desses mecanismos de busca, mantém índices separados, porém interconectados através do conteúdo semântico e textual disponível nos documentos. Os algoritmos *k-nearest neighbor* (k-NN) representam uma classe com abordagens não paramétricas utilizadas para classificação de texto e associação. Estes localizam os pontos mais próximos em relação a um ponto de pesquisa, ou seja, o *embedding* gerado a partir da consulta do usuário, e representam uma forma simples e generalizada de busca através do vizinho mais próximo (do inglês *Nearest Neighbor* - NN), chamado também de busca por similaridade, busca por proximidade ou busca por itens relacionados. Apesar disso, a implementação da vizinhança mais próxima, tendo um espaço de atributos de alta

---

<sup>1</sup> Um token é entendido com uma unidade de informação indivisível em um texto, por exemplo, uma palavra ou o radical de uma palavra, uma pontuação, um símbolo, um número, uma sigla ou um padrão que semanticamente, ao ser decomposto, perderia seu sentido, por exemplo, o número de CPF.

dimensão, pode reduzir o desempenho visto que a distância entre os NNs pode ser elevada.

O objetivo da busca pelo(s) NN(s) é encontrar o(s) ponto(s) de dados ligado(s) à consulta no banco de dados, enquanto a busca aproximada pelo vizinho mais próximo (do inglês *Approximate Nearest Neighbor* - ANN) retorna os pontos de dados de distância equivalente ou menor entre os vizinhos. Em ambas as estratégias, a busca é normalmente realizada em uma coleção de pontos de dados específicos (geralmente produzidos aplicando-se o conceito de *embedding*), sendo o procedimento de organização do banco de dados independente do número de consultas. Quando o número de dimensões se torna extenso, geralmente excedendo o número 20, a busca em árvores e estruturas relacionadas pode exigir recursos significativos do banco de dados, afetando negativamente o desempenho computacional, mas ao mesmo tempo, beneficiando positivamente a precisão da recuperação. Logo, a busca aproximada pelo vizinho mais próximo tem capturado a atenção para problemas com dados que possuem alta dimensionalidade (Wang; Shen, 2022).

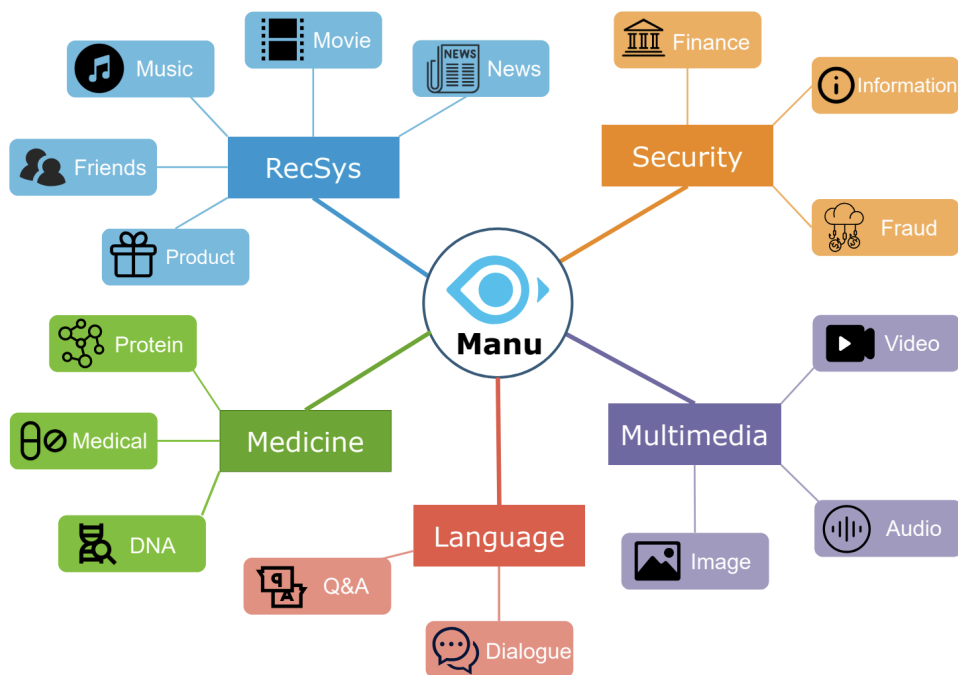
### **2.3.1 Banco de Dados Vetorial**

Para Taipalus (2024), um sistema de gerenciamento de banco de dados vetorial é um tipo especializado de sistema de gerenciamento de banco de dados que se concentra principalmente no gerenciamento eficiente de dados vetoriais de alta dimensão. Similarmente a outros tipos de sistemas de gerenciamento de banco de dados, espera-se que um SGBD direcionado a aplicação vetorial, seja um software funcional que possa gerenciar dados. O gerenciamento de dados inclui em geral a consulta e manipulação de dados, coleta de metadados, indexação, controle de acesso, backups, suporte para escalabilidade e interfaces com outros sistemas, como drivers de banco de dados, linguagens de programação, *frameworks* e sistemas operacionais.

O surgimento de LLMs para tarefas como recuperação de informações, juntamente com o crescimento de dados não estruturados que sustentam impulsionadores econômicos como comércio eletrônico e plataformas de recomendação, demanda novos sistemas de gerenciamento de banco de dados que possam oferecer capacidades tradicionais, como otimização de consultas,

transações, escalabilidade, tolerância a falhas e privacidade e segurança, mas para dados não estruturados (Pan *et al.*, 2023). Neste sentido, surgem os bancos de dados vetoriais que possibilitam o armazenamento e recuperação de qualquer unidade de informação, desde que exista uma forma de transformar um texto, uma imagem, um vídeo ou um áudio, em uma representação numérica na forma de *embeddings* (Guo *et al.*, 2022). A Figura 3 apresenta esta ideia a partir do banco de dados vetorial Manu<sup>®</sup>.

Figura 3 - Banco de dados vetorial



Fonte: Guo *et al.* (2022).

## 2.4 TRABALHOS CORRELATOS

Realizando pesquisas na literatura científica relacionada à área de Recuperação de Informação, mais especificamente de Busca Semântica e Sistema de Perguntas e Respostas, e juntamente com termos referentes à jurisprudências, foram encontrados e selecionados trabalhos referentes à LLMs. Para tal, as seguintes bases de dados científicas foram utilizadas para a pesquisa: Scopus<sup>®</sup>, Web of Science<sup>®</sup>, ScienceDirect<sup>®</sup>, IEEE Xplore<sup>®</sup>, ACM Digital Library<sup>®</sup>, Spring Link<sup>®</sup> e Periódicos CAPES<sup>®</sup>. Os artigos científicos selecionados foram publicados entre os anos de 2018 até início de 2024, escritos inteiramente em língua inglesa. A *string* de

busca a seguir, foi aplicada ao título resumo e palavras-chave sem restrição temporal: ("*jurisprudence*" OR "*precedent*") AND ("*legal*" OR "*law*") AND ("*information retrieval*" OR "*semantic search*") AND ("*large language model*" OR *embedding*).

A quantidade de artigos recuperados nas bases de dados especificadas anteriormente é apresentada por meio da Tabela 1.

Tabela 1 - Resultados da revisão literária

Base de Artigos Acadêmicos	Quantidade de Artigos Resultantes
ACM Digital Library®	0
Science Direct®	0
Scopus®	12
Web of Science®	1
IEEE Xplore®	0
Spring Link®	41
Periódicos CAPES® (busca por assunto)	5
<b>Todos os Resultados</b>	<b>59</b>

Fonte: Elaborado pelo autor (2024).

A tabela acima representa numericamente os resultados obtidos a partir da expressão de busca desenvolvida. De 59 artigos encontrados, 15 apresentaram títulos e resumos aderentes ao objetivo da pesquisa e forneceram o documento completo disponível para leitura. Os resumos foram lidos na íntegra, permitindo a análise e seleção na próxima etapa. Por fim, os 8 artigos com maior aderência ao assunto abordado nesta pesquisa são descritos.

Mandal *et al.* (2021), examinam 56 técnicas (incluindo 26 desenvolvidas pelos autores) para calcular a similaridade entre documentos de casos judiciais, foram elencadas diferentes formas de representações de texto e cálculo de similaridade entre essas representações de texto por meio de metodologias recentes. Entre os cinco melhores métodos de desempenho, três foram originados dos autores e os outros dois baseiam-se em métodos de recuperação de informações pré-existentes adaptados para a tarefa específica de recuperação de informações legais. Para documentos inteiros, utilizou-se a métrica de acurácia na

representação de textos para medir a taxa de similaridade entre os modelos propostos, tendo como resultado, as seguintes taxas numéricas: TFIDF (0.872), D2V (0.768), W2V (0.766), L2V (0.723), LDA (0.723), PScoreVect (0.723) e BERT (0.660).

Kim *et al.* (2024) fazem referência à Competição na Extração/Anexação de Informações Legais (do inglês *Competition on Legal Information Extraction/Entailment* - COLIEE), tendo como objetivo desafiar os participantes a realizar quatro tarefas distintas, incluindo recuperação de jurisprudência (Tarefa 1) e inferência (Tarefa 2), além de recuperação de lei estatutária (Tarefa 3) e inferência (Tarefa 4). Na tarefa de recuperação de jurisprudência (Tarefa 1), o objetivo é identificar casos legais que devem ser "notados", a partir de um conjunto de casos candidatos fornecidos. A abordagem adotada para essa tarefa se baseia em um modelo *transformers* nomeado de *sentence-transformers/all-mpnet-base-v2*, que gera uma representação numérica multidimensional de cada parágrafo dentro de cada caso. A partir disso, são calculadas as distâncias cosseno entre cada parágrafo de um caso de consulta e um caso candidato, sendo gerado um histograma. Essas distâncias são utilizadas para treinar um classificador binário, determinando se um caso de é relevante ao contexto jurídico. O melhor resultado foi alcançado com os seguintes parâmetros de pós-processamento: pontuação de confiança mínima de 0.80, número máximo de casos relevantes igual a 10, número máximo de casos notados repetidos igual a 3. O segundo melhor resultado apresentou parâmetros semelhantes (0,7, 9 e 2, respectivamente).

Já Naik e Patel (2024), enfatizam a importância da busca semântica em documentos de julgamento legal, enfatizando a combinação do reconhecimento de entidades nomeadas (do inglês *Named Entity Recognition* - NER) com a classificação semântica para melhorar a eficiência da busca. Utilizando conjuntos de dados baseados em documentos judiciais do Supremo Tribunal de Justiça da Índia, os autores aplicam a arquitetura de rede neural *Long-Short Term Memory* (LSTM) para NER e segmentação semântica, utilizando técnicas de *embeddings* para representação de texto. Os resultados foram avaliados utilizando métricas como *f1-score*, *recall* e *precision*, demonstrando que tanto a perda de treinamento quanto a de validação diminuem e se estabilizam em um ponto específico, o que sugere que o modelo não apresenta *overfitting* e *underfitting*.

De Martino, Pio e Ceci, 2022 definem o método chamado de Regularidades de Parágrafos em Decisões Judiciais (do inglês *Paragraph Regularities in Legal*



*Case Judgments* - PRILJ) como sendo utilizado para apoiar especialistas jurídicos durante a redação de documentos legais. O PRILJ adota uma abordagem de três passos, incluindo a fase de treinamento, durante a qual treina um modelo de incorporação de documentos a partir de  $Dt$ , que é capaz de representar documentos em um espaço de características semânticas e aprender  $k$  modelos de incorporação de parágrafos, um para cada *cluster*, sendo capazes de representar parágrafos em um espaço de características semânticas. Incorporação (do inglês *embedding*) de parágrafos, que explora tanto o modelo de *embedding* de documentos quanto os  $k$  modelos de *embedding* de parágrafos aprendidos durante a fase de treinamento. Finalizando as três fases com a identificação de regularidades de parágrafos que explora os clusters de documentos identificados, o modelo de incorporação de documentos e os  $k$  modelos de incorporação de parágrafos para avaliar, por meio de uma estratégia eficiente, a similaridade entre os parágrafos. Os resultados obtidos permitem concluir que o PRILJ baseado em Word2Vec e o modelo de dois passos e  $k = \sqrt{|Dt|} * 2$ , fornecem os melhores resultados gerais para a identificação de regularidades de parágrafos em julgamentos legais. Porém, o modelo de dois passos baseado em clustering implementado no PRILJ proporciona claras vantagens, uma vez que é capaz de modelar adequadamente os diferentes tópicos na coleção de documentos.

Zhebel, Devyatkin, Zubarev e Shochenkov (2023) identificam que a IR entre idiomas pode ser considerada um problema, pois idiomas diferentes usam terminologia específica em cada país. Incorporar uma consulta de pesquisa apropriada requer amplo conhecimento das especificidades da terminologia jurídica no idioma em questão, contudo, uma solução bem sucedida envolveria recuperar documentos tematicamente semelhantes para um documento de entrada. Desenvolvendo a ideia em utilizar um mediador à transição entre dois idiomas, os autores propuseram um método de estudo baseado no uso de *embeddings* como conceitos, aplicável a modelos de busca probabilística, como a função de classificação *Best Match 25* (BM25). Embora os resultados do método BM25 sejam aceitáveis, possui uma limitação quanto à natureza dos documentos jurídicos, sendo eles fragmentados e contendo tópicos dentro do mesmo documento, fazendo com que o sistema de busca perca documentos relevantes ao contexto. Propondo o uso de redes siamesas com o modelo de *embeddings* SentenceBERT, os documentos são divididos em fragmentos de 30 *tokens*, sendo posteriormente combinados em

*embeddings* de documentos. A configuração ideal de uma rede multilíngue que emprega o modelo SBERT necessita de um corpus paralelo anotado, em termos de referências e de sentenças. Os resultados apresentados mostram que o modelo SBERT\_BASELINE atingiu um *f1-score* de 0.79, com *precision* de 0.80 e *recall* de 0.78. Já o modelo SIAM\_SBERT\_FEED obteve um *f1-score* de 0.56, com *precision* de 0.59 e *recall* de 0.58, enquanto o modelo SIAM\_SBERT\_TR apresentou um *f1-score* de 0.54, com *precision* de 0.56 e *recall* de 0.57.

Greco e Tagarelli (2023) destacam avanços na pesquisa em AI no contexto jurídico, especialmente em relação aos métodos baseados em *Transformer-based Language Models* (TLMs). A arquitetura *transformer* baseia-se em mecanismos de atenção a partir de dois modos, sendo a autoatenção, onde a importância da palavra é estimada na mesma sequência e atenção ponderada, onde a representação de entrada é usada para prever os *tokens* alvo de forma auto-regressiva. No contexto jurídico, um dos primeiros modelos baseados em *transformer* que obteve um impacto significativo na pesquisa em AI jurídica é o *Bidirectional Encoder Representations from Transformers* (BERT). Desde seu surgimento, houve um crescente interesse em adaptar BERT e modelos relacionados para tarefas específicas no domínio jurídico. Os autores citam a possibilidade de realização de ajustes no BERT utilizando os dados de treinamento da Tarefa 2 da Competição na Extração/Anexação de Informações Legais (do inglês *Competition on Legal Information Extraction/Entailment* - COLIEE). Utilizou-se o modelo em uma etapa de pós-processamento que combinava pontuações de similaridade de *token* e de frase nominal para identificar parágrafos relevantes. Esta estratégia conquistou o primeiro lugar na Tarefa 2 do COLIEE-2019, obtendo um *f1-score* de 0.70 no conjunto de testes oficial.

Sampath e Durairaj (2022) apresentam o método *Precedence Retrieval from Legal Documents Using Catch Phrases* (PReLCaP) para aprimorar a recuperação de precedentes no domínio jurídico, usando uma combinação de características de similaridade estatística, semântica e de subtração extraídas de frases-chave de documentos legais. O método possui uma rede neural com arquitetura de *encoder-decoder* e unidades recorrentes de LSTM para representar de forma concisa os documentos legais, facilitando a recuperação de casos anteriores. Dividido em três componentes, sendo extração de frases-chave, extração de características e classificação e ranqueamento, o método utiliza uma abordagem em

*pipeline* para a extração de frases-chave, extraindo características de similaridade de texto, como subtração, semântica e estatística. Os documentos são classificados e ranqueados com base nos valores de probabilidade e o desempenho é metrificado a partir do valor de Precisão Média (do inglês *Mean Average Precision* - MAP). Resultados mostram melhor desempenho ao combinar diferentes características, com MAP variando de 0.58 a 0.432 em uma escala de 1.

Para Mandal *et al.* (2022), existe a importância de incorporar informações específicas do documento em um modelo de marcação de sequência para melhorar a extração de frases-chave no contexto jurídico. Experimentos são conduzidos em um conjunto de documentos de casos da Suprema Corte Indiana, onde as frases-chave padrão são obtidas de um sistema de informação jurídica popular. Por mais que os documentos sejam extensos, possuem um número relativamente baixo de frases-chave. Para avaliar os métodos propostos, o conjunto de dados é dividido em 10 partes, seguindo uma abordagem de validação cruzada de 10 fases. Métodos não supervisionados, como *Phrases Scores legal* (PSlegal) e *MyScore*, são diretamente aplicados aos documentos, ao contrário dos métodos supervisionados, como *Keyword Extraction Algorithm* (KEA) e *Deep Convolutional Neural Networks* (DeepCNN), no qual são avaliados através da validação cruzada. Essa metodologia permite que cada documento seja utilizado para teste exatamente uma vez, garantindo uma avaliação precisa e robusta do desempenho dos modelos propostos.

A partir disso, o Quadro 1 apresenta de forma sintetizada as características gerais dos trabalhos correlatos utilizados na elaboração desta seção com o intuito de posicionar esta pesquisa em relação aos trabalhos avaliados.

Esta pesquisa situa-se na interseção das técnicas avançadas de NLP e DL para a recuperação de jurisprudências, destacando-se em relação aos trabalhos correlatos analisados na tabela acima. Em comparação com Mandal *et al.* (2021), que exploram uma variedade de métodos, esta pesquisa foca em modelos de linguagem pré-treinados como BERT e GPT para uma representação semântica das jurisprudências. Assim como Naik e Patel (2024) e De Martino, Pio e Ceci (2022), esta pesquisa utiliza metodologias avançadas de NLP para o reconhecimento de entidades jurídicas e extração de informações relevantes. Ao aplicar modelos de DL para capturar relações específicas e variações semânticas, esta pesquisa se alinha com as abordagens de Kim *et al.* (2024) e Greco e Tagarelli (2023), que também empregam técnicas baseadas em *transformers* para tarefas de incorporação de

textos jurídicos. De modo geral, esta pesquisa contribui na recuperação de informações jurídicas ao combinar *embeddings* e avanços das pesquisas correlatas mencionadas com foco na elaboração de defesas jurídicas.

Quadro 1 - Resultados da revisão literária

<b>Autor</b>	<b>Ano</b>	<b>Método</b>	<b>Conjunto de Dados</b>	<b>Métrica</b>	<b>Resultados</b>
Mandal et al. (2021)	2021	TFIDF, Doc2Vec, Word2Vec, L2V, LDA, PScoreVect, BERT	Documentos de casos judiciais	Acurácia	TFIDF (0.872), D2V (0.768), W2V (0.766), L2V (0.723), LDA (0.723), PScoreVect (0.723), BERT (0.660)
Kim et al. (2024)	2024	Transformers, Sentence-BERT, classificador binário	Competição COLIEE	F1-score	Melhores resultados com parâmetros específicos (não numéricos)
Naik e Patel (2024)	2024	LSTM, NER, embeddings	Documentos do Supremo Tribunal da Índia	F1-score, recall, precision	Melhores resultados quanto a eficiência de busca
De Martino, Pio e Ceci (2022)	2022	PRILJ, Word2Vec, clustering	Decisões judiciais	F1-score, precision, recall	Melhores resultados com modelo de dois passos baseado em clustering
Zhebel et al. (2023)	2023	SBERT, redes siamesas, BM25	Documentos jurídicos	F1-score, precision, recall	SBERT_BASELINE (F1-score 0.79, precision 0.80, recall 0.78)
Greco e Tagarelli (2023)	2023	BERT, Transformer-based Language Models	Dados do COLIEE	F1-score	Primeiro lugar na Tarefa 2 do COLIEE-2019 (F1-score 0.70)
Sampath e Durairaj (2022)	2022	PReLCaP, LSTM, encoder-decoder	Documentos legais	Mean Average Precision (MAP)	MAP varia de 0.432 a 0.58
Mandal et al. (2022)	2022	Incorporação baseada em DeepCNN	Documentos de casos da Suprema Corte Indiana	Validação cruzada de 10 fases	Avaliação de métodos não supervisionados (PSIlegal e MyScore) e supervisionados (KEA e DeepCNN) com validação cruzada, permitindo uma avaliação precisa dos modelos.

Fonte: Elaborado pelo autor (2024).

### 3 METODOLOGIA

#### 3.1 ENQUADRAMENTO METODOLÓGICO

A pesquisa é a atividade principal da metodologia. Assim, quanto à natureza, este trabalho caracteriza-se como uma pesquisa aplicada, pois, segundo Gerhardt e Silveira (2009), busca-se gerar conhecimentos com o intuito de aplicá-los de maneira prática, voltada à resolução de problemas específicos relacionados a verdades e interesses locais. A pesquisa aplicada destaca-se pelo seu interesse direcionado à aplicação efetiva, à utilização prática e as consequências resultantes dos conhecimentos obtidos (Gil, 2008). Além disso, pode ser considerada aplicada, uma vez que constrói uma parcela do conhecimento para ser utilizada em situações reais, visando auxiliar indivíduos e organizações no processo de tomada de decisões.

Neste contexto, este trabalho pode ser identificado como tecnológico, pois a pesquisa nesta área busca criar artefatos e desenvolver processos. O domínio fundamental da pesquisa tecnológica é a mudança, e as estratégias concebidas para efetuar tais alterações são reconhecidas como inovações. Cada processo de transformação demanda metodologias tecnológicas para sua respectiva gestão, incluindo a implementação de novas abordagens de trabalho (Ramos; Ramos; García-Peñalvo, 2019).

Em relação aos objetivos, o trabalho é categorizado como exploratório, visto que busca fornecer uma visão geral sobre um fato específico. Esse tipo de pesquisa é abordado principalmente quando o tema selecionado é pouco explorado, sendo assim desafiador formular hipóteses precisas e operacionais a respeito dele (Gil, 2008).

Segundo Gil (2008), uma parcela dos estudos exploratórios pode ser identificada como pesquisas bibliográficas. Sendo assim, quanto aos procedimentos, este trabalho se enquadra como uma pesquisa bibliográfica, já que se desenvolve baseado no levantamento de documentos já existentes, de forma predominantemente constituído por artigos científicos e livros. A pesquisa bibliográfica apresenta como vantagem principal possibilitar ao pesquisador abranger uma variedade de fatos de forma mais ampla do que seria normalmente possível através de outros métodos (Gil, 2008).

Este trabalho também está relacionado com a *Design Science* (DS), que tem como objetivo gerar conhecimento, não estando limitado à sua aplicação. Trata-se de uma disciplina científica com foco no *design*, destinada a desenvolver e conceber soluções que aprimorem os sistemas existentes, solucionem problemas ou mesmo criem artefatos que contribuam para um comportamento humano mais eficiente, tanto na sociedade quanto nas organizações (Dresch; Lacerda; Antunes Jr., 2015).

Dessa forma, a DS é considerada a base epistemológica para a *Design Science Research Methodology* (DSRM), metodologia responsável pela sistematização deste trabalho. Na seção seguinte, serão abordados os princípios e fundamentos da DSRM que orientam a metodologia a ser adotada no decorrer desta pesquisa.

### 3.2 DESIGN SCIENCE RESEARCH METHODOLOGY

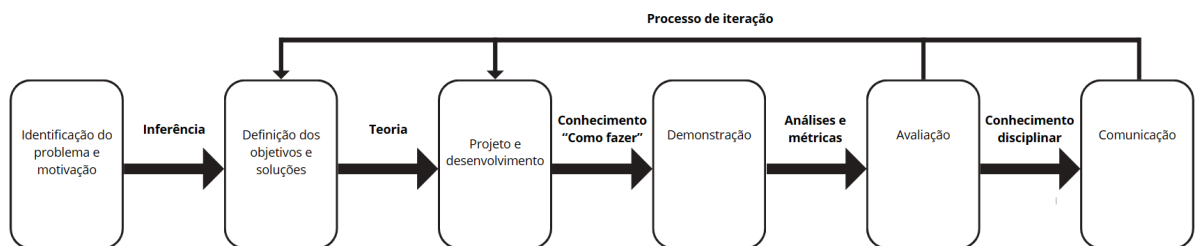
A DSRM incorpora princípios, práticas e procedimentos essenciais para a realização de pesquisas, cumprindo três objetivos: a) manter a consistência com a literatura existente; b) oferecer orientação aos pesquisadores; e c) proporcionar um modelo conceitual para a apresentação dos resultados (Peppers *et al.*, 2007). Para tal, se apropria do conceito de artefatos, estando intimamente ligado à condução de pesquisas em *Design Science* (DS) e em Sistemas de Informação (SI). Esses artefatos incluem, por exemplo, constructos, *frameworks*, modelos, métodos e instâncias de sistemas de informação (Hevner *et al.*, 2004).

A DSRM incorpora um processo meticuloso composto por seis atividades, conforme definido por Peppers *et al.* (2007). Conforme a Figura 4, a execução das atividades pode seguir uma sequência predeterminada ou adaptar-se às exigências específicas do projeto. A seguir, consta uma breve descrição de cada uma dessas atividades:

- Atividade 1 - Identificação do problema e motivação: concentra-se na definição de um problema de pesquisa específico e na justificativa da importância da solução proposta;
- Atividade 2 - Definição dos objetivos para uma solução: aqui são deduzidos os objetivos de uma solução com base na definição do problema e no entendimento do que é possível e viável realizar;

- Atividade 3 - Projeto e desenvolvimento: aborda a criação do artefato, incluindo a determinação das funcionalidades desejadas, a arquitetura e a subsequente criação do artefato real;
- Atividade 4 - Demonstração: concentra-se na apresentação do uso do artefato para resolver o problema. Isso envolve a utilização do artefato em experimentação, simulação, estudo de caso, prova ou outra atividade apropriada;
- Atividade 5 - Avaliação: refere-se à observação e medição do artefato para solucionar o problema, incluindo a comparação dos objetivos da solução com os resultados reais observados durante a demonstração;
- Atividade 6 - Comunicação: destaca-se pela divulgação do problema e sua importância, bem como pela apresentação do artefato produzido.

Figura 4 - Atividades para condução da DSRM



Fonte: Adaptado de Peffers *et al.* (2007, p. 54).

### 3.3 DESENVOLVIMENTO DA PESQUISA

Esta pesquisa tem uma abordagem aplicada e tecnológica, com objetivos exploratórios centrados na proposição de uma ferramenta para a recuperação de informações a partir de jurisprudências apresentadas em formato de dados não estruturados, de forma textual. Considera, principalmente, aspectos semânticos na explicitação do conhecimento. Para alcançar esse propósito, será adotada a DSRM, conforme proposta por Peffers *et al.* (2007). As atividades utilizadas para abordar e resolver o problema são evidenciadas no Quadro 2.

Quadro 2 - Atividades da pesquisa

Atividades	Descrição
Identificação do problema e motivação	- Realização de pesquisa bibliográfica voltada à elaboração da introdução/contextualização do trabalho, assim como sua problemática e definição da pergunta de pesquisa.
Definição dos objetivos para uma solução	- Elaboração dos objetivos, geral e específicos, de modo que a pergunta de pesquisa deste trabalho possa ser respondida.
Projeto e desenvolvimento	- Propor e desenvolver um método destinado à recuperação jurisprudências levando em conta aspectos semânticos e de explicitação do conhecimento para o auxílio na elaboração de defesas em processo jurídicos. A instanciação do método ocorre por meio da disponibilização de um sistema, na forma de protótipo.
Demonstração	- Demonstrar o protótipo utilizando cenário de estudo no domínio jurídico.
Avaliação	- Avaliar a aplicabilidade do protótipo por meio de cenário de estudo com mensuração por meio de alguma métrica que possibilite discutir os resultados obtidos.
Comunicação	- Comunicação à comunidade acadêmica dos resultados obtidos através deste trabalho e publicação de artigo científico.

Fonte: adaptado de Peffers *et al.* (2007, p. 54).

### 3.3.1 Identificação do problema e motivação

O problema e motivação foram identificados a partir da revisão bibliográfica e elaboração da seção de trabalhos correlatos, por meio da qual se procurou encontrar uma lacuna de pesquisa no contexto das Tecnologias de Informação e Comunicação. Após a pesquisa, revelou-se a ausência de trabalhos com foco na recuperação de jurisprudências do estado de Santa Catarina utilizando busca



semântica, com o objetivo de auxiliar profissionais do domínio jurídico na obtenção de resultados aproximados às sutilezas linguísticas.

### **3.3.2 Definição dos objetivos**

A definição dos objetivos deste trabalho está fundamentada em uma análise dos trabalhos correlatos, com base em uma abordagem estratégica para o suporte ao método proposto e sua avaliação em determinado cenário de estudo. Essa contextualização com a literatura existente fornece um ponto de partida e também uma base para a contribuição específica do presente trabalho.

### **3.3.3 Projeto e desenvolvimento**

O artefato deve ser o principal resultado entregável desta atividade. Neste trabalho é representado por um método voltado à recuperação semântica de jurisprudências, bem como a representação do conhecimento na forma de *embeddings* e suporte à elaboração de defesas em processos jurídicos, objetivando apoiar a tomada de decisão por parte de determinado defensor jurídico. As atividades voltadas ao desenvolvimento do artefato, são apresentadas nas seções a seguir.

#### **3.3.3.1 Coleta de dados**

Os dados das jurisprudências utilizadas foram selecionados a partir da base de jurisprudências catarinense conforme Figura 5. O acesso facilitado possibilitou a escolha, visto a representatividade no contexto jurídico. O site permite acesso a mais de três milhões de jurisprudências e possui seis níveis de abrangência, envolvendo os acórdãos do tribunal de justiça, acórdãos do conselho da magistratura, acórdãos das turmas recursais e de uniformização, despachos das vice-presidências, decisões monocráticas do tribunal de justiça e decisões monocráticas das turmas recursais.

Figura 5 - Tela do site Jurisprudência Catarinense

**JURISPRUDÊNCIA CATARINENSE**

PODER JUDICIÁRIO de Santa Catarina

**Busca avançada**

Abrangência da busca:  Ementa  Inteiro Teor

Pesquisar em:

- Acórdãos do Tribunal de Justiça
- Acórdãos do Conselho da Magistratura
- Acórdãos das Turmas Recursais e de Uniformização
- Despachos das Vice-Presidências
- Decisões Monocráticas do Tribunal de Justiça
- Decisões Monocráticas das Turmas Recursais

Procurar Resultados

- com todas as palavras
- com a expressão
- com qualquer uma das palavras
- sem as palavras
- em que entre a palavra e a palavra existam até termos

Ementa

N.º processo  Período entre

Classe

Relator

Juiz Prolator

Comarca

Fonte: Tribunal de Justiça Catarinense - Jurisprudências (2023).

Visto que o *website* disponibilizado acima não oferece a captura das jurisprudências por meio de serviços externos como APIs Rest, realizou-se o desenvolvimento de um *web scraping* com o intuito de extrair informações específicas de jurisprudências como: número do processo, relator, origem, órgão julgador, entre outras e, após isso, persistiu-se os dados em um banco de dados relacional PostgreSQL®.

### 3.3.3.2 Transformação dos dados

Após a coleta dos dados, o passo seguinte envolve a sumarização e criação da representação vetorial densa (*embedding*) das jurisprudências catarinenses. Geralmente, as jurisprudências coletadas na etapa anterior apresentam uma quantidade significativa de *tokens*, excedendo o limite estabelecido pelo LLM utilizado neste artigo (*juridics/bertlaw-base-portuguese-sts-scale*) para a geração dos *embeddings*. Isso cria uma limitação, o que torna necessário resumir ou sumarizar o texto da jurisprudência.

Desta forma, o modelo *t5-portuguese-legal-summarization*, disponível na plataforma Hugging Face® conforme Figura 6, identifica trechos relevantes do texto jurídico e realiza a sumarização. Este modelo utilizado contribui para a redução da quantidade de *tokens* na jurisprudência em questão.

Figura 6 - Sumarização do texto jurídico

Fonte: Hugging Face (2023).

Após a sumarização da jurisprudência, o modelo responsável pela representação matemática do texto jurídico constrói o vetor denso correspondente. Ao concluir a geração dos vetores das jurisprudências, realiza-se a criação de arquivos no formato JSON contendo os respectivos vetores densos gerados pelo LLM.

### 3.3.3.3 Indexação e busca semântica

Com os dados das jurisprudências devidamente processados e armazenados em arquivos no formato JSON, a implementação da indexação de vetores torna-se uma etapa significativa para a organização eficiente das informações processadas. O objetivo principal dessa indexação é permitir o armazenamento estruturado dos vetores derivados dos textos jurídicos e facilitar a

criação de uma estrutura de busca semântica, especialmente adaptada a um banco de dados vetorizado específico.

Para iniciar o processo, optou-se pela utilização do banco de dados Qdrant<sup>®</sup> para o armazenamento dos vetores associados aos textos jurídicos. Nesse contexto, as jurisprudências são estruturadas em coleções no Qdrant<sup>®</sup>, cada uma configurada com parâmetros específicos, como o tamanho do vetor e a métrica utilizada na recuperação dos vetores.

A coleção principal, que representa o domínio específico das jurisprudências, adota um tamanho de vetor de 512 dimensões. A respectiva dimensão está relacionada ao modelo de linguagem utilizado no projeto, refletindo a caracterização semântica dos textos jurídicos. A distância aritmética é calculada a partir da similaridade semântica entre os vetores, sendo geralmente empregada a equação do cosseno. Esse cálculo é essencial para estabelecer a proximidade semântica entre diferentes jurisprudências.

Após a indexação em conjunto com a criação de coleções no Qdrant<sup>®</sup> a busca semântica pode ser realizada. Para tal, por meio do LLM utilizado no trabalho, o texto de interesse é transformado em um *embedding* específico servindo de entrada para a consulta no banco de dados vetorial. Isso possibilita a realização de buscas contextualizadas e semânticas, retornando um conjunto de respostas caracterizadas como as jurisprudências mais próximas. Na sequência, por meio da escolha pelo especialista das jurisprudências mais adequadas, estas são encaminhadas ao LLM para a geração de defesas.

#### **3.3.4 Demonstração**

A atividade de demonstração do protótipo resultante do método proposto, que tem objetivo recuperar jurisprudências através do aprendizado profundo, envolve um processo abrangente de projeto e desenvolvimento. Soma-se a isso a necessidade de integração de técnicas avançadas, desde a seleção e coleta de dados até a indexação e conversão dos mesmos, proporcionando uma solução baseada na análise e recuperação de jurisprudências catarinenses, bem como na elaboração de defesas. O uso de tecnologias como *web scraping*, sumarização e indexação semântica e geração de textos são destacados a partir da seção 3.3.3.

### 3.3.5 Avaliação

O fluxo de desenvolvimento, apresentado na seção 3.3.3 do trabalho, expõe alguns materiais empregados na pesquisa. Entre as ferramentas utilizadas, algumas das quais não mencionadas nas seções anteriores, incluem o *framework* construído em PHP<sup>®</sup> denominado Hyperf<sup>®</sup>, responsável pela estrutura fundamental do *web scraping*, em conjunto com o banco de dados relacional PostgreSQL<sup>®</sup>. A linguagem Python<sup>®</sup> também é utilizada para modelar os dados não estruturados, utilizando o modelo de sumarização *t5-portuguese-legal-summarization* e o modelo *juridics/bertlaw-base-portuguese-sts-scale* para a conversão dos dados em representações matemáticas, ou seja, vetores densos (*embeddings*). Posteriormente, outra ferramenta empregada é o banco de dados vetorial Qdrant<sup>®</sup>, responsável pelo armazenamento e recuperação dos dados transformados na etapa anterior. Para a geração de texto, mais especificamente de defesas de processos jurídicos, utiliza-se o LLM GPT-3.5-turbo<sup>®</sup> da OpenAI<sup>®</sup>.

Por fim, para a avaliação, o estudo propõe uma abordagem fundamentada na correlação de *Spearman*, onde o coeficiente é empregado para medir a intensidade da relação entre duas variáveis por meio de uma função monótona (Spearman, 1904). Com a definição e delineamento da estratégia de avaliação da pesquisa, o objetivo é envolver advogados em um cenário específico de comparação. Nesse contexto, a ordenação de um número pré-definido de documentos jurídicos é realizada tanto pelos advogados quanto pelo protótipo desenvolvido, resultando na obtenção de um coeficiente numérico através da correlação de *Spearman*.

### 3.3.6 Comunicação

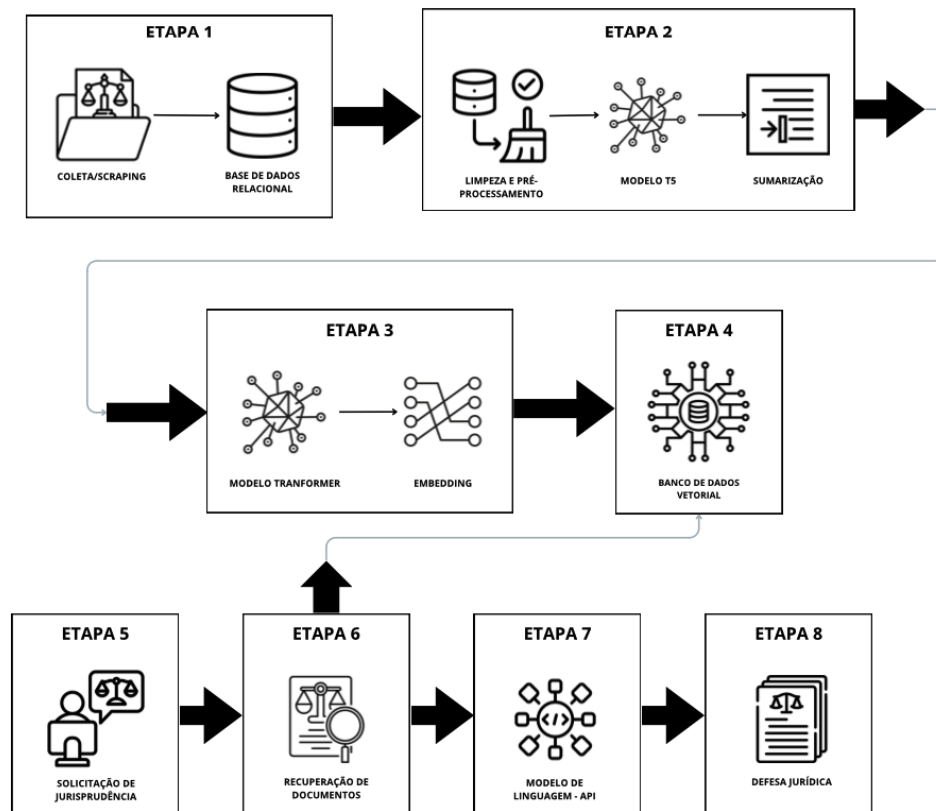
A comunicação das análises e dos resultados desta pesquisa será conduzida por meio de publicação deste trabalho em formato de artigo em periódico ou congresso de domínio tecnológico ou jurídico. Ao término da pesquisa, será elaborado um documento final de trabalho de conclusão de curso que consolida os principais achados e conclusões alcançados. A intenção é compartilhar percepções e contribuições desta pesquisa com a comunidade acadêmica e profissional,

promovendo um diálogo construtivo no campo das Tecnologias de Informação e Comunicação aplicadas no âmbito jurídico.

## 4 MÉTODO PROPOSTO

Este capítulo apresenta, de forma detalhada, as etapas e o funcionamento do método proposto neste trabalho. Utilizando uma base de dados relacional contendo jurisprudências coletadas por meio de *scraping*, o processo envolve a aplicação do modelo de sumarização. Isso é feito com o objetivo de ampliar a capacidade de geração de *embeddings* por meio de LLMs com arquitetura *transformers*. Os *embeddings* resultantes são integrados ao banco de dados vetorial, permitindo que usuários realizem solicitações de jurisprudências. Essas solicitações são convertidas em *embeddings* e enviadas ao banco de dados vetorial, que recupera os documentos mais próximos aos *embeddings* gerados. A partir dos documentos recuperados, a defesa jurídica é elaborada também com auxílio de LLM, acessado via API Rest. A Figura 7, apresenta em linhas gerais o método proposto, contendo as etapas necessárias para realização deste trabalho.

Figura 7 - Etapas do método proposto

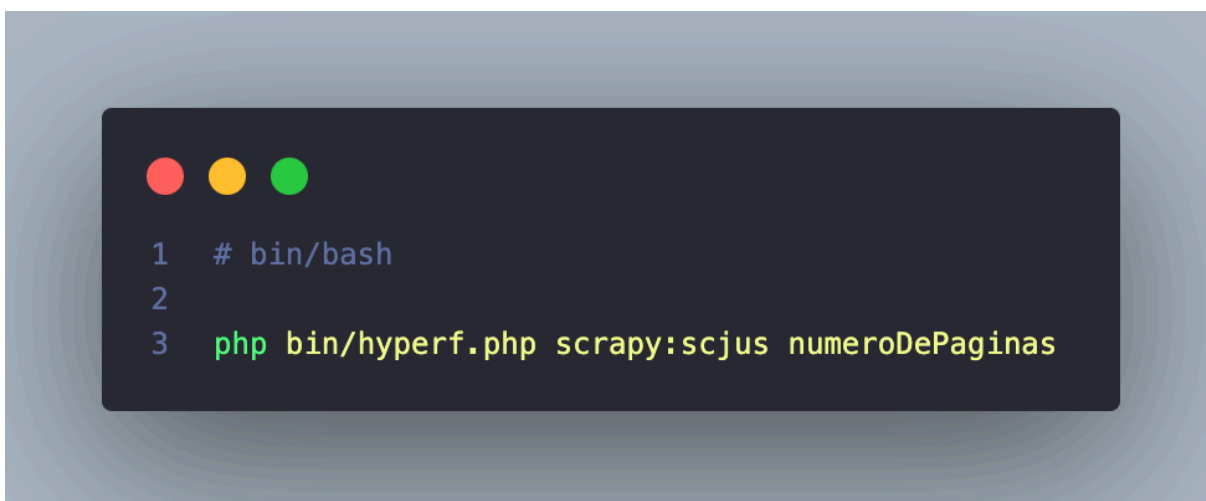


Fonte: Elaborado pelo autor (2024).

#### 4.1 ETAPA 1: COLETA DE DADOS

Conforme mencionado na seção 3.3.3.1, os dados são recuperados da base de jurisprudências catarinense, podendo ser acessados de forma pública por meio de um site. As informações recuperadas com o uso da técnica de *web scraping* são processadas pela linguagem de programação PHP® com o auxílio do *framework* Hyperf®. O *scraping* realiza um *request* ao site, adicionando dinamicamente o número da página de ordenação de 50 resultados, ou seja, após recuperar 50 jurisprudências da primeira página, de forma dinâmica o processo irá incrementar a página no qual os dados devem ser recuperados. O *scraping* irá recuperar as jurisprudências de acordo com o número de páginas definidas no comando da Figura 8.

Figura 8 - Comando *scraping*

A terminal window with a dark background and three colored window control buttons (red, yellow, green) at the top left. The terminal shows three lines of text: a shell prompt, a blank line, and a PHP command. The command is highlighted in yellow and green.

```
1 # bin/bash
2
3 php bin/hyperf.php scrapy:scjus numeroDePaginas
```

Fonte: Elaborado pelo autor (2024).

O *framework* possui uma estrutura de conexão com o banco de dados PostgreSQL®, possibilitando a criação de *migrations* e *scripts* automatizados. A Figura 9 apresenta de forma detalhada quais colunas das jurisprudências são armazenadas no banco de dados PostgreSQL®.



Figura 9 - Tabela das jurisprudências

```
1  <?php
2
3  use Hyperf\Database\Schema\Schema;
4  use Hyperf\Database\Schema\Blueprint;
5  use Hyperf\Database\Migrations\Migration;
6
7  class CreateJurisprudenciaTable extends Migration
8  {
9      /**
10     * Run the migrations.
11     */
12     public function up(): void
13     {
14         Schema::create('jurisprudencia', function (Blueprint $table) {
15             $table->bigIncrements('id');
16             $table->string('processo')->nullable(true);
17             $table->string('relator')->nullable(true);
18             $table->string('origem')->nullable(true);
19             $table->string('orgao_julgador')->nullable(true);
20             $table->string('julgado_em')->nullable(true);
21             $table->string('classe')->nullable(true);
22             $table->string('texto', 10485760)->nullable(true);
23         });
24     }
25
26     /**
27     * Reverse the migrations.
28     */
29     public function down(): void
30     {
31         Schema::dropIfExists('jurisprudencia');
32     }
33 }
```

Fonte: Elaborado pelo autor (2024).

Os dados inseridos na tabela “jurisprudência” são exportados no formato JSON a partir da seleção das colunas *id* e *texto*.

## 4.2 ETAPA 2: LIMPEZA E SUMARIZAÇÃO

O JSON exportado na primeira etapa é submetido a uma função de limpeza na Etapa 2 com auxílio da linguagem Python®, que visa remover as tags “&nbsp;” contidas na coluna *texto*. Posteriormente, ocorre a sumarização das jurisprudências, conforme detalhado na seção 3.3.3.2. A sumarização é essencial devido à quantidade significativa de *tokens* presentes nas jurisprudências, que excedem o

limite estabelecido pelo modelo *transformers* utilizado neste estudo (*juridics/bertlaw-base-portuguese-sts-scale*) para a geração de *embeddings*. Dessa forma, utilizamos o modelo *t5-portuguese-legal-summarization*, disponível na plataforma Hugging Face®, para identificar e resumir os trechos relevantes do texto jurídico, contribuindo para reduzir a quantidade de tokens.

As jurisprudências são sumarizadas e adicionadas em um arquivo JSON em lotes de 1000 unidades, abrangendo diferentes arquivos, para facilitar a criação de *checkpoints* durante o processo. Isso se deve ao fato de que a sumarização é executada no Google Colab® e requer aproximadamente 60 minutos para processar 1000 jurisprudências. A Figura 10 ilustra como os dados são processados e exportados.

Figura 10 - Sumarização e processamento

```
1 from transformers import AutoTokenizer, AutoModelWithLMHead
2 import json, os
3
4 tokenizer = AutoTokenizer.from_pretrained('stjiris/t5-portuguese-legal-summarization')
5 model = AutoModelWithLMHead.from_pretrained('stjiris/t5-portuguese-legal-summarization',
6 return_dict=True)
7
8 model.to('cuda:0')
9
10 def generate_summarization_jurisprudencia(jurisprudencia):
11     inputs = tokenizer.encode("summarize: " + jurisprudencia,
12                             return_tensors='pt',
13                             max_length=1000,
14                             truncation=True).to('cuda:0')
15     summary_ids = model.generate(inputs, max_length=400, min_length=100,
16                                length_penalty=5., num_beams=2)
17
18     output = tokenizer.decode(summary_ids[0])
19     return output.replace("summarize: ", "")
20
21 file = open(file='/dataset-jurisprudencias.json')
22
23 file_content = json.load(file)
24 texto_resumido = []
25 texto_temporario = []
26
27 quantidade_arquivo_id = 100
28
29 for content in file_content:
30     texto_temporario.append({'id': content['id'], 'texto': content['texto']})
31     if content['id'] % quantidade_arquivo_id == 0:
32         if os.path.exists('/summarization-dataset_' + str(content['id']) + '.json') == False:
33             for texto in texto_temporario:
34                 output = generate_summarization_jurisprudencia(texto['texto'])
35                 texto_resumido.append({'id': texto['id'], 'texto': output})
36             with open('/summarization-dataset_' + str(content['id']) + '.json', 'w') as f:
37                 json.dump(texto_resumido, f)
38             texto_resumido = []
39             texto_temporario = []
```

Fonte: Elaborado pelo autor (2024).

### 4.3 ETAPA 3: GERAÇÃO DOS *EMBEDDINGS*

Na Etapa 3 é realizada a geração de *embeddings* a partir da leitura do arquivo JSON previamente gerado que contém um conjunto de dados com as colunas *id* e *texto*. Após a leitura e carga do referido arquivo, o código itera sobre cada item desse conteúdo. Para cada item, o texto associado a uma chave específica é extraído e um modelo *transformers* nomeado como *juridics/bertlaw-base-portuguese-sts-scale* é utilizado para gerar um *embedding* correspondente. Os *embeddings* gerados são adicionados a uma lista chamada '*embedding*'. Após o cálculo e armazenamento de todos os *embeddings* nessa lista, a biblioteca *numpy* é utilizada para salvá-la em um arquivo denominado '*embeddings.npy*', sendo este um formato eficiente para o armazenamento de arrays do tipo *numpy*.

### 4.4 ETAPA 4: BANCO DE DADOS VETORIAL

A Etapa 4 consiste na inserção dos resultados da geração dos *embeddings* na ferramenta de banco de dados vetorial Qdrant<sup>®</sup>, conforme indicado na seção 3.3.3.3. Os parâmetros são configurados no *client\_qdrant* com utilização da linguagem Python<sup>®</sup> para a criação ou recriação de uma coleção de vetores, definindo especificamente a métrica de distância e o tamanho dos vetores. Os *embedding*, previamente calculados e armazenados no arquivo '*embeddings.npy*', são carregados, juntamente com os dados das jurisprudências sumarizadas do arquivo '*summarization.json*'.

Após os parâmetros e arquivos carregados, ocorre o *upload* dos vetores e dados de sumarização para a coleção no Qdrant Cloud<sup>®</sup>. Essa etapa do método visa disponibilizar os resultados do processamento de texto em uma estrutura adequada para consultas e análises, permitindo operações como busca aproximada por similaridade semântica.

### 4.5 ETAPA 5: SOLICITAÇÃO DA JURISPRUDÊNCIA

Nesta Etapa, o objetivo é possibilitar uma experiência de busca ao usuário dentro da aplicação. O mecanismo de busca recebe as solicitações dos usuários e

realiza a consulta por meio do processo de busca semântica. De maneira geral, os profissionais do direito se utilizam de uma interface de consulta de maneira que possam informar demandas de pesquisas.

#### 4.6 ETAPA 6: RECUPERAÇÃO DE DOCUMENTOS

A fase de recuperação de documentos desempenha um papel vinculado à Etapa 4.4, sendo responsável por recuperar documentos relevantes que foram previamente inseridos no banco de dados vetorial. Quando um profissional do campo jurídico realiza uma solicitação de busca através de *prompts*, o modelo *transformers* converte a sentença em um *embedding* por meio do modelo *juridics/bertlaw-base-portuguese-sts-scale*. Esse *embedding* é comparado com os *embeddings* dos documentos armazenados no banco de dados vetorial, utilizando o cálculo de similaridade do cosseno (Jones; Furnas, 1987) (Equação 1).

$$similaridade = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Após a conclusão da pesquisa, o banco de dados vetorial apresenta os  $n$  resultados mais similares, sendo  $n$  um parâmetro arbitrário, mas que em função da complexidade da operação de consulta não deve ser elevado. Em geral, valores na casa das centenas podem ser utilizados. Os resumos das jurisprudências mais relevantes são então apresentados, proporcionando ao profissional jurídico o acesso aos documentos recuperados. Essa fase permite que o profissional avalie e selecione os resumos mais relevantes ao contexto solicitado.

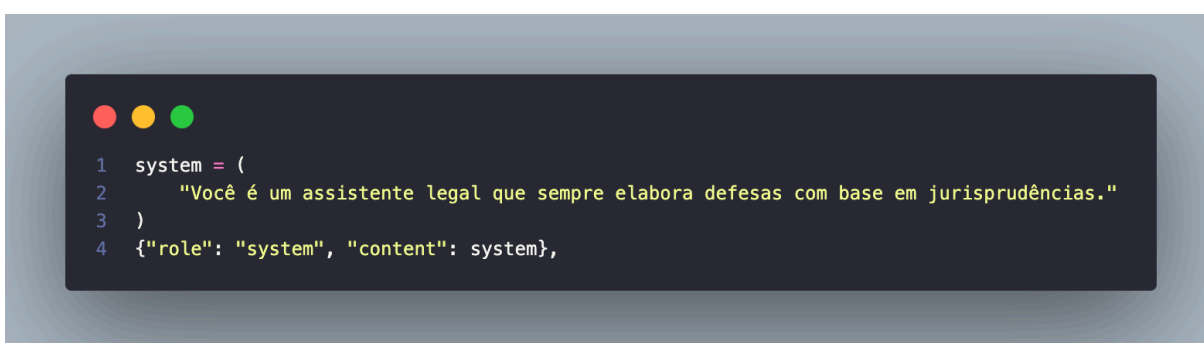
A seleção realizada pelo profissional durante essa etapa determina quais resumos serão utilizados nas etapas posteriores do processo. A precisão nesta etapa é fundamental para que apenas informações relevantes e úteis sejam consideradas durante a análise jurídica.

#### 4.7 ETAPA 7: MODELO DE LINGUAGEM

Nesta etapa os documentos jurídicos previamente recuperados, analisados e selecionados pelo profissional jurídico servem de base para a elaboração da defesa legal. Isto ocorre utilizando o LLM GPT-3.5 Turbo. Este modelo, disponível como serviço em nuvem, oferece soluções para algumas tarefas, entre elas, a geração de textos. A integração requer conformidade com regras específicas para aproveitar seu potencial como assistente especializado em um determinado assunto, contextualizando as demandas jurídicas de forma eficiente, podendo também agilizar a abordagem de redação da defesa legal.

A Figura 11 apresenta o formato utilizado para inferir a geração de um conteúdo de forma específica e direcionado por meio do serviço GPT-3.5 Turbo.

Figura 11 - Configuração das regras



```
1 system = (  
2     "Você é um assistente legal que sempre elabora defesas com base em jurisprudências."  
3 )  
4 {"role": "system", "content": system},
```

Fonte: Elaborado pelo autor (2024).

#### 4.8 ETAPA 8: DEFESA JURÍDICA

Além da definição da regra jurídica direcionada ao LLM na etapa anterior, é necessário nesta etapa definir os parâmetros requeridos pela função `client.chat.completions.create()`, como `max_tokens`, `temperature` e `prompt`. O primeiro informa a quantidade máxima de tokens na geração do texto, o segundo um valor contínuo indicando o grau de liberdade que o LLM possui para criar texto e o terceiro, a demanda do usuário. A Figura 12 apresenta a função do assistente jurídico que realiza a elaboração de defesas baseadas em jurisprudências, limitando a geração em 1500 palavras e com criatividade regular. Ou seja, a geração de texto estará limitada ao conjunto de jurisprudências previamente selecionadas e a um

parâmetro de criatividade que permita gerar texto com qualidade. Estes dois elementos têm como função minimizar um problema presente nos LLMs atuais, a alucinação.

Um *prompt* é caracterizado como uma instrução ou comando em linguagem natural que uma pessoa utiliza para se comunicar com um modelo de linguagem. Esse comando pode ser usado para realizar tarefas, tais como, como traduzir textos e gerar código de programação de forma direta e sem intermediários complexos (Cheng *et al.*, 2024).

Figura 12 - Definição do *prompt* para o assistente jurídico

```
1 import openai
2
3 system = (
4     "Você é um assistente legal que sempre elabora defesas com base em jurisprudências."
5 )
6
7 async def gerar_defesa(request: Request):
8     body = await request.json()
9     prompt = body.get('prompt', '')
10
11     result = openai.ChatCompletion.create(
12         model="gpt-3.5-turbo",
13         messages=[
14             {"role": "system", "content": system},
15             {"role": "user", "content": prompt},
16         ],
17         max_tokens=1500,
18         temperature=0.5,
19     )
20
```

Fonte: Elaborado pelo autor (2024).

A criação do *prompt*, no qual indicará as jurisprudências que devem ser utilizadas como referência pelo LLM, ocorre via linguagem JavaScript®. O *prompt* indica ao LLM a pergunta inicial, concatenado das jurisprudências recuperadas, analisadas e selecionadas pelo profissional jurídico. A Figura 13 apresenta o processo de criação do *prompt*, agrupado em duas fases, sendo elas a introdução e contexto do *prompt*.

Figura 13 - Criação do *prompt*

```
1  function createPrompt(question, jurisprudencias) {
2    let intro = `Baseado na pergunta inicial:
3    "${question}" e nas seguintes jurisprudências,
4    elabore uma argumentação detalhada e analítica
5    para uma defesa jurídica:`;
6
7    let content = jurisprudencias.join("\n\n");
8
9    return `${intro}\n\n${content}`;
10 }
```

Fonte: Elaborado pelo autor (2024).

Como resultado deste processo é produzido um texto que, conforme já mencionado, leva em consideração um *prompt* base e um conjunto de jurisprudências selecionadas, criando assim um *container* de raciocínio do LLM com o intuito de reduzir a geração de conteúdo equivocado (Figura 14).

Figura 14 - Fragmento de texto gerado pelo LLM

```

1  prompt = '''
2  Baseado na pergunta inicial:
3  "Ação de reembolso de honorários médicos em plano de saúde resulta em sentença de improcedência.
4  Autor diagnosticado com carcinoma buscou consultas e cirurgias com médico particular,
5  apesar de existir rede credenciada. O contrato não prevê reembolso para livre escolha de profissionais.
6  Falha no atendimento não comprovada. Legislação e contrato não obrigam reembolso em casos de prestadores
7  particulares." e nas seguintes jurisprudências, elabore uma argumentação
8  detalhada e analítica para uma defesa jurídica:
9
10 Apelação No 5023954-66.2019.8.24.0038/SC RELATOR: Juiz ALEXANDRE MORAIS DA ROSA APELANTE: AGEMED SAUDE
11 LTDA EM LIQUIDACAO EXTRAJUDICIAL (RÉU) APELADO: PROCTOGIN SERVICOS MEDICOS EIRELI (AUTOR) RELATÓRIO
12 Trata-se de recurso de apelação em que figuram como apelante AGEMED SAUDE LTDA. EM LIQUIDACAO EXTRAJUDICIAL e
13 apelada PROCTOGIN SERVICOS MEDICOS EIRELI, interposto contra sentença proferida pelo
14 juízo de origem nos autos n. 5023954-66.2019.8.24.0038.
15 '''
16
17
18 result = client.chat.completions.create(
19     model="gpt-3.5-turbo",
20     messages=[
21         {"role": "system", "content": "Você é um assistente legal que sempre elabora defesas com base em jurisprudências."},
22         {"role": "user", "content": prompt}
23     ],
24     max_tokens=1500,
25     temperature=0.5
26 )
27
28 result.choices[0].message.content
29 '''
30 Com base nas jurisprudências apresentadas, é possível argumentar que a decisão de improcedência na ação de reembolso
31 de honorários médicos em plano de saúde está correta. No caso em questão, o autor buscou consultas e cirurgias
32 com um médico particular, mesmo havendo uma rede credenciada disponível no plano de saúde. Além disso, o contrato
33 não previa reembolso para livre escolha de profissionais e a falha no atendimento não foi comprovada.
34 Na Apelação No 5023954-66.2019.8.24.0038/SC, a sentença refletiu a prestação de serviços médicos realizados
35 dentro da rede credenciada da operadora de plano de saúde, não havendo previsão contratual para reembolso em
36 casos de escolha de prestadores particulares. Portanto, a decisão de improcedência seguiu a legislação e o contrato
37 estabelecido entre as partes. Já na Apelação No 5013400-65.2019.8.24.0008/SC, o caso envolveu a falta de
38 fornecimento do contrato de condições gerais do plano de saúde para o titular, o que pode configurar uma
39 irregularidade por parte da operadora. No entanto, a dificuldade para agendamentos médicos não é suficiente
40 para justificar o reembolso de honorários médicos realizados fora da rede credenciada, especialmente
41 se não houve comprovação de falha no atendimento ou de impossibilidade de atendimento dentro da rede. Portanto,
42 com base nas jurisprudências apresentadas e nos argumentos expostos, é possível concluir que a decisão de
43 improcedência na ação de reembolso de honorários médicos em plano de saúde está de acordo com a legislação
44 vigente e os contratos estabelecidos entre as partes.
45 '''

```

Fonte: Elaborado pelo autor (2024).



## 5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

### 5.1 CENÁRIO DE ESTUDO

O cenário de estudo deste trabalho envolve a recuperação de jurisprudências a partir do site Jurisprudência Catarinense (conforme Seção 3.3.3.1), levando em consideração a descrição, servindo de auxílio aos profissionais jurídicos. Como conjunto de dados, foram utilizados os *embeddings* gerados conforme descrição da Seção 4.3 e inseridos no banco de dados vetorial, conforme descrito na Seção 4.4. Através de uma entrada de determinado questionamento referente à área jurídica, jurisprudências são retornadas pelo protótipo de buscas desenvolvido, oferecendo uma interface gráfica para que especialistas jurídicos possam avaliar a ordenação das jurisprudências. Essa avaliação foi realizada utilizando o método de correlação de ordenação de *Spearman*. Os resultados de cada busca serão apresentados com o objetivo de avaliar a semelhança entre o processamento do banco de dados vetorial e a análise efetuada por profissionais jurídicos.

### 5.2 IMPLEMENTAÇÃO DO MÉTODO PROPOSTO

O método proposto neste trabalho, foi desenvolvido através da linguagem de programação Python® visto que esta possui um conjunto extenso de bibliotecas que facilitam atividades de implementação em cenários de Aprendizado de Máquina, Inteligência Artificial e Análise de Dados, como a própria biblioteca *SciPy*. Com foco nesta biblioteca, utilizou-se o cálculo de *Spearman* para medir a correlação entre a ordenação da aplicação e de profissionais jurídicos.

O protótipo desenvolvido permite a realização de buscas e a recuperação de jurisprudências de maneira ordenada pela relevância das mesmas. Todavia, a ordem do protótipo é registrada para fins de cálculo posterior, sendo então o resultado reordenado e enviado para avaliação do usuário, que, neste contexto, é um profissional jurídico. Este então, organiza as jurisprudências de acordo com o *prompt* fornecido e o sistema armazena a nova ordem, ou seja, a ordem elaborada pelo usuário. Os registros das ordenações são organizados em um arquivo de valores separados por vírgulas (CSV) no formato " $x_n$ ", " $y_n$ ", sendo " $x$ " a ordem fornecida pela aplicação e " $y$ " a ordem definida pelo especialista jurídico. A Figura 15 apresenta um

exemplo de quatro consultas (*prompts*), com as ordenações do sistema (x1 a x4) e as ordenações de determinado especialista (y1 a y4).

Figura 15 - Ordenação dos resultados

```

1  "x1", "x2", "x3", "x4", "y1", "y2", "y3", "y4"
2  3, 1, 4, 2, 9, 5, 7, 6
3  5, 2, 6, 1, 4, 3, 9, 8
4  7, 4, 9, 5, 8, 6, 2, 10
5  1, 5, 8, 6, 7, 10, 3, 4
6  6, 8, 2, 9, 6, 8, 4, 1
7  8, 9, 1, 4, 2, 1, 8, 7
8  2, 3, 5, 10, 10, 9, 6, 3
9  9, 6, 7, 3, 1, 4, 10, 9
10 4, 10, 10, 7, 5, 2, 1, 2
11 10, 7, 3, 8, 3, 7, 5, 5
12

```

Fonte: Elaborado pelo autor (2024).

A partir das ordenações combinadas entre o sistema (x) e o especialista (y) para cada *prompt*, utiliza-se a função *spearmanr()* da biblioteca *SciPy* para determinar a correlação de ordenação. A Figura 16 demonstra um exemplo de correlação de valores ordenados entre a aplicação e o profissional jurídico.

Figura 16 - Correlação de *Spearman*

```

1  from scipy.stats import spearmanr
2
3  ordenacao_original = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
4  ordenacao_advogado = [8, 5, 10, 6, 9, 1, 3, 4, 2, 7]
5
6  correlacao, p_valor = spearmanr(ordenacao_original, ordenacao_advogado)
7
8  print(f"Correlação de Spearman: {correlacao}")
9  print(f"Valor-p: {p_valor}")

```

Fonte: Elaborado pelo autor (2024).

### 5.3 APRESENTAÇÃO DOS RESULTADOS

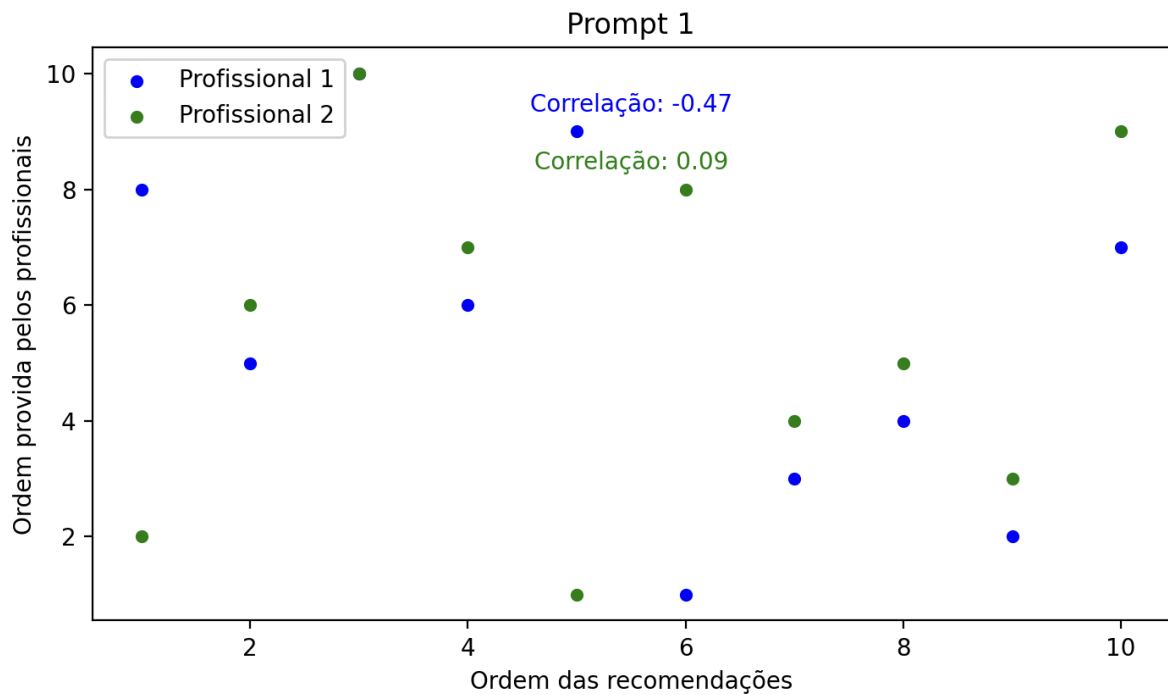
Para avaliação do método proposto, juntamente com o cenário principal descrito na seção 5.1, foram estabelecidos alguns cenários específicos com o objetivo de verificar e explicitar, a partir de jurisprudências recuperadas, a ordenação realizada pela aplicação e por profissionais do meio jurídico, ou seja, verificar o grau de correlação entre as ordenações efetuadas entre ambos os agentes. Foram realizadas ordenações de jurisprudências recuperadas por dois profissionais jurídicos. Cada profissional levou em torno de 2 horas para realizar a ordenação de quatro consultas (*prompts*), cada uma contendo 10 jurisprudências. Além disso, um dos profissionais jurídicos também realizou a avaliação da ferramenta de elaboração de defesa.

Para o primeiro cenário de estudo, o *prompt* utilizado está relacionado com a improcedência de uma ação de reembolso de honorários médicos em plano de saúde, descrito como "*Ação de reembolso de honorários médicos em plano de saúde resulta em sentença de improcedência. Autor diagnosticado com carcinoma buscou consultas e cirurgias com médico particular, apesar de existir rede credenciada. O contrato não prevê reembolso para livre escolha de profissionais. Falha no atendimento não comprovada. Legislação e contrato não obrigam reembolso em casos de prestadores particulares.*".

Após a análise da correlação obtida pela avaliação realizada pelo primeiro profissional jurídico produziu um índice **-0.47**. Este que valor sugere que a correlação entre o resultado do protótipo e o primeiro profissional jurídico é altamente negativa. Ou seja, houve uma discordância entre as ordenações nesta primeira análise. Por outro lado, a correlação do segundo profissional jurídico em relação à ordenação provida pelo protótipo resultou no valor de **0.09**, indicando neutralidade, ou seja, não havendo concordância ou discordância. A Figura 17 demonstra o mapeamento das ordenações das 10 jurisprudências retornadas e a ordem atribuída por cada profissional. Por exemplo, para a jurisprudência na primeira ordem, o "Profissional 1" atribuiu a 8ª posição, enquanto o "Profissional 2" atribuiu a 2ª posição. Já para a segunda ordem, o "Profissional 1" atribuiu a 6ª posição, enquanto o "Profissional 2" atribuiu a 5ª posição. Assim, percebe-se uma ordenação mais precisa efetuada pelo "Profissional 2", apesar do resultado não ser adequado para os dois avaliadores. Este resultado pode ter sido influenciado por

diversos fatores que serão abordados adiante. Vale mencionar que no caso da jurisprudência na terceira ordem os dois profissionais atribuíram a 10ª posição, ocorrendo a sobreposição dos pontos.

Figura 17 - Correlação para o *prompt* 1 considerando os dois profissionais

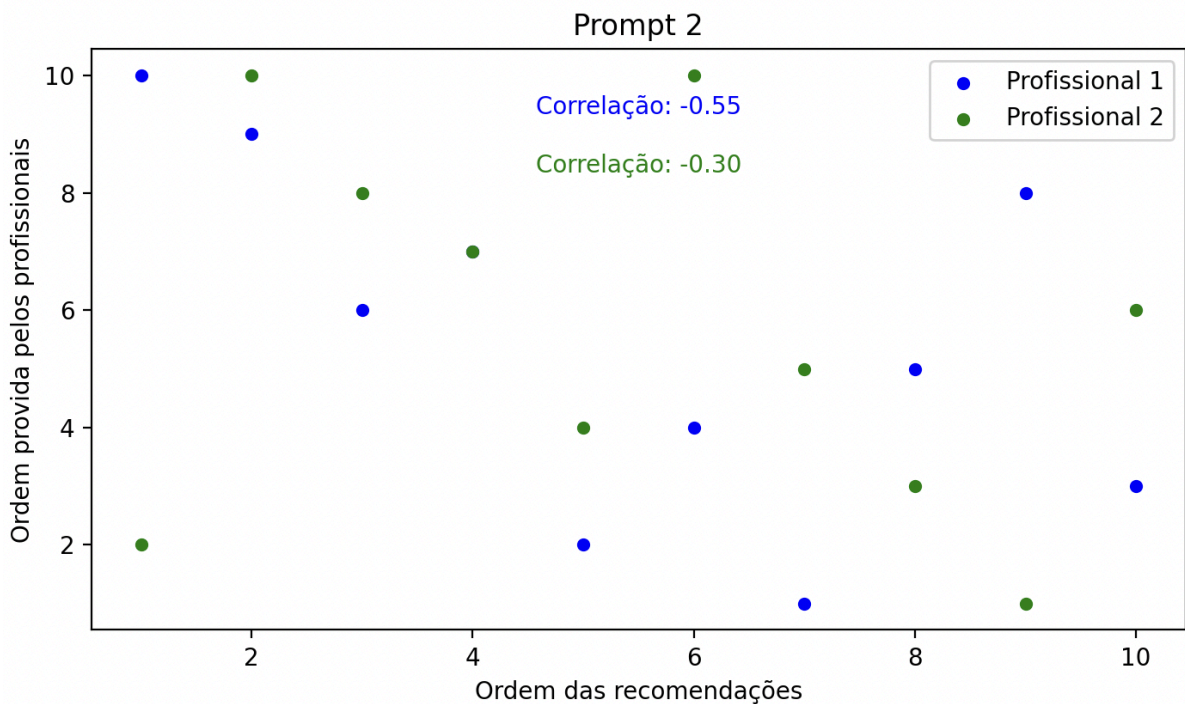


Fonte: Elaborado pelo autor (2024).

No segundo cenário de estudo, o *prompt* tratou de uma disputa sobre a cobertura parcial de cirurgia ortognática, sendo "O caso envolve disputa entre segurado e plano de saúde pela cobertura parcial de ortognática. O médico solicitou os procedimentos negados pela junta médica da seguradora. Alegação de cerceamento de defesa pela seguradora. Necessidade de prova pericial para esclarecimento técnico. Essencial garantir equidade na resolução.". Após a análise dos dados a correlação do primeiro profissional jurídico alcançou o valor **-0.55**, seguido do segundo profissional jurídico, com o valor **-0.30**, indicando uma inversão mais acentuada nas ordens. A ordenação ajustada pelos profissionais jurídicos divergiu da ordem proposta inicialmente, apontando para uma diferença na avaliação da importância das jurisprudências. Este resultado apresenta os casos de disputa técnica sobre a necessidade de procedimentos médicos, onde a percepção dos profissionais jurídicos sobre a necessidade de prova pericial e equidade na

resolução pode ser diferente comparada à ordenação do protótipo conforme apresentado pela Figura 18. A correlação negativa sugere que o protótipo precisa incorporar as nuances destes casos para fornecer resultados melhores alinhados com a prática jurídica. Vale mencionar que no caso da jurisprudência na quarta ordem os dois profissionais atribuíram a 7ª posição, ocorrendo a sobreposição dos pontos.

Figura 18 - Correlação para o *prompt 2* considerando os dois profissionais

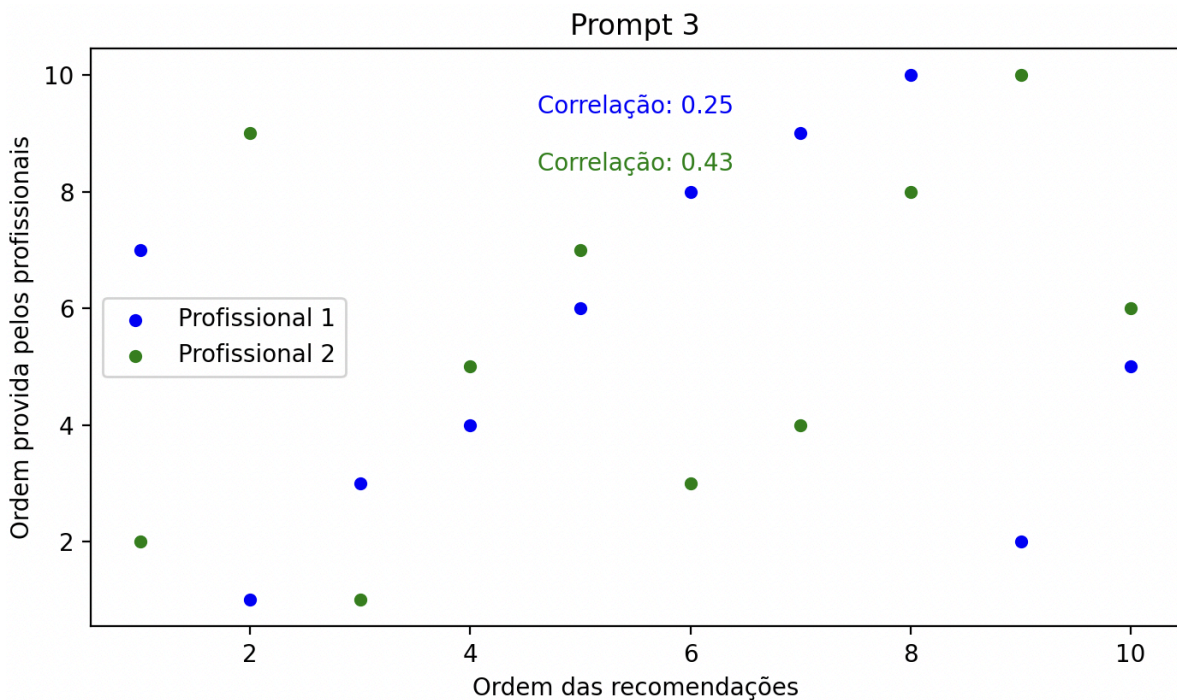


Fonte: Elaborado pelo autor (2024).

Já no terceiro cenário de estudo, o prompt diz respeito a uma ação de obrigação de custear terapias multidisciplinares relacionadas ao transtorno do espectro autista, sendo "*Ação de obrigação de fazer visa custeio de terapias multidisciplinares para transtorno do espectro autista. Tutela provisória indeferida. O autor busca atendimento fora da rede credenciada do plano. Não há evidências de que dentro da rede não haja prestadores aptos a fornecer os atendimentos necessários.*". Neste caso, atingiu-se uma correlação de **0.25** relacionada à avaliação do primeiro profissional jurídico e **0.43** na avaliação do segundo profissional jurídico. Estes valores sugerem um nível de concordância interessante, com destaque para o segundo profissional. Esta situação pode ser atribuída à especificidade dos serviços requeridos e a interpretação das evidências disponíveis

sobre a adequação dos prestadores de serviços dentro da rede credenciada. A correlação positiva indicada na Figura 19, demonstra que a ordenação das jurisprudências sugerida pelo protótipo demonstra um resultado que pode ser aceito como parcialmente aderente aos providos pelos profissionais. A melhora nos resultados em relação aos cenários anteriores pode ser observada nas sequências 3, 4 e 8, visto que pelo menos um dos profissionais sugeriu a mesma sequência.

Figura 19 - Correlação para o *prompt* 3 considerando os dois profissionais

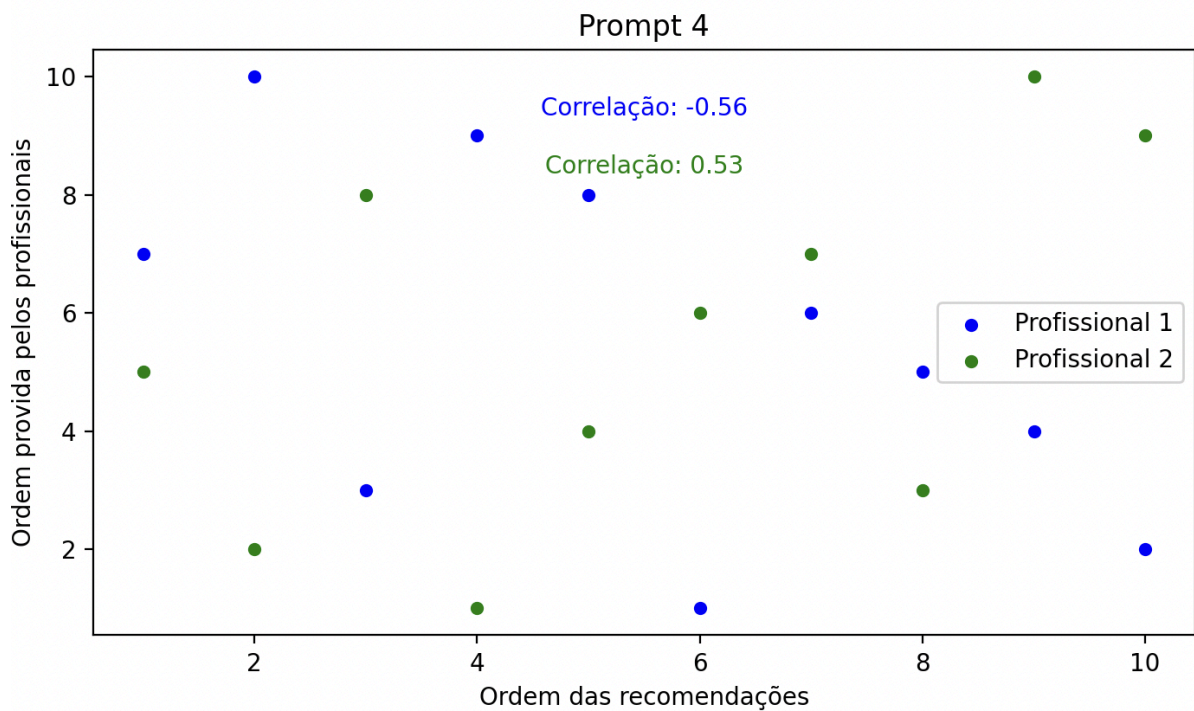


Fonte: Elaborado pelo autor (2024).

Por fim, o quarto cenário de estudo, contém um *prompt* relacionado à cobertura de procedimentos para cirurgia ortognática, sendo "*Ação contra plano de saúde resulta em sentença favorável ao paciente. A Seguradora contesta cobertura de procedimentos e materiais para cirurgia ortognática, negados pela junta médica. Disputa técnica levanta questões sobre a necessidade dos itens requisitados. Alegação de cerceamento de defesa pela seguradora. Prova pericial necessária para esclarecer aspectos técnicos.*". Para o primeiro profissional o índice ficou em **-0.56**, sendo a pior correlação entre todos os casos de estudo. Este resultado evidencia uma discrepância considerável entre a ordenação do protótipo e a avaliação do profissional, indicando uma inversão no objeto esperado. Porém a avaliação do segundo profissional jurídico, apresenta um contraponto de avaliação,

pois o respectivo valor é equivalente a **0.53**, sendo o maior número positivo entre todos os casos de estudo. A discrepância entre as correlações, indica que os profissionais jurídicos responsáveis possuem avaliações distintas quanto aos casos onde disputas técnicas e provas periciais são o objetivo central. Este valor de correlação obtido a partir da avaliação do segundo profissional pode ser observado na Figura 20 e nas sequências 2, 6 e 7, e ficando muito próximo nas sequências 5, 9 e 10.

Figura 20 - Correlação para o *prompt* 4 considerando os dois profissionais



Fonte: Elaborado pelo autor (2024).

A Tabela 2 demonstra o cenário geral da avaliação do trabalho, informando a média de avaliação dos profissionais em cada uma das correlações apresentadas anteriormente. Foi possível entender que a visão de ambos os profissionais, nos *prompts* 1 e 4, diferem entre si. A média apresentou resultados negativos nos dois primeiros *prompts*, seguido de um resultado positivo e outro próximo ao valor nulo.

Tabela 2 - Resultados da média de correlação entre os profissionais

<i>Prompt</i>	Profissional 1	Profissional 2	Média
1	-0.47	0.09	-0.19
2	-0.55	-0.30	-0.42
3	0.25	0.43	0.34
4	-0.56	0.53	-0.01

Fonte: Elaborado pelo autor (2024).

A disparidade avaliativa na maioria dos cenários de estudo anteriores e a baixa taxa de correlação pode ser discutida a partir de fatores humanos. A competência jurídica é peculiar entre ambos os profissionais, inferindo-se uma compreensão de mundo distinta em relação aos *prompts* avaliados. O tempo de avaliação empregado também é determinante para nuances específicas da jurisprudência serem devidamente ordenadas. A quantidade de profissionais avaliadores, influencia na média de correlação geral de cada um dos *prompts* avaliados. A pesquisa contou com o auxílio de dois profissionais jurídicos, porém caso obtivesse um maior número de especialistas envolvidos, poderia captar uma série de avaliações semelhantes entre si, proporcionando uma visão mais clara e precisa das tendências interpretativas no campo jurídico.

As limitações técnicas também são fundamentais para a discussão dos resultados de busca das jurisprudências. O banco de dados vetorial possui 5000 documentos relacionados à área da saúde, na forma de *embeddings*, reduzindo a quantidade de jurisprudências aderentes aos *prompts* sugeridos. Outro fator determinante refere-se à sumarização da jurisprudência realizada pelo modelo *t5-portuguese-legal-summarization*, na qual demonstrou limitações quanto ao desempenho, pois impossibilita altas cargas de trabalho devido ao custo computacional. O modelo *transformers juridics/bertlaw-base-portuguese-sts-scale* utilizado para a geração de *embeddings*, também possui limitações relacionadas à quantidade de *tokens* processados, ou seja, nuances específicas do texto jurídico podem não serem captadas e processadas a depender do tamanho da jurisprudência que será sumarização. Além disso, algoritmos de ordenação disponibilizados pelo banco de dados vetorial Qdrant®, como distância euclidiana, distância de *manhattan* e produto dos pontos, não foram explorados nesta pesquisa.



Além dos cenários de estudo relacionados à ordenação das jurisprudências, foi realizado um quinto cenário de estudo, analisado a partir da geração da defesa jurídica por um dos profissionais. A avaliação ocorreu após o uso do protótipo de elaboração de defesas, utilizando-se do *prompt* do primeiro cenário de estudo como entrada para a recuperação de jurisprudências similares.

O protótipo implementado a partir do método proposto possibilita que determinado profissional do direito realize interações por meio de sentenças e textos (designadas como *prompt*) ao invés de palavras-chave, como é comum em sistemas de busca (Figura 21). Com isso, o protótipo converte o *prompt* em um *embedding* e o utiliza para enviar ao banco de dados vetorial que, por meio de comparações vetoriais, retorna os vetores (textos) semanticamente mais semelhantes (Figura 22). Os resultados são então disponibilizados ao usuário que deve selecionar as jurisprudências mais aderentes à sua consulta inicial. Com a seleção realizada, o protótipo interage com o LLM informando os textos que servirão de base para a geração da defesa (Figura 23).

Figura 21 - *Prompt* de entrada para a consulta às jurisprudências

The screenshot shows the JurisChat interface. At the top, it says "JurisChat" and "Chat Jurisprudências". Below that, a user query (U) is displayed: "Ação de reembolso de honorários médicos em plano de saúde resulta em sentença de improcedência. Autor diagnosticado com carcinoma buscou consultas e cirurgias com médico particular, apesar de existir rede credenciada. O contrato não prevê reembolso para livre escolha de profissionais. Falha no atendimento não comprovada. Legislação e contrato não obrigam reembolso em casos de prestadores particulares." Below the query, search results (G) are shown under the heading "Jurisprudências encontradas:". Three results are listed, each starting with a square bullet point and containing a detailed summary of a legal case, including the case number, parties, and the court's decision.

**U** Ação de reembolso de honorários médicos em plano de saúde resulta em sentença de improcedência. Autor diagnosticado com carcinoma buscou consultas e cirurgias com médico particular, apesar de existir rede credenciada. O contrato não prevê reembolso para livre escolha de profissionais. Falha no atendimento não comprovada. Legislação e contrato não obrigam reembolso em casos de prestadores particulares.

**G** Jurisprudências encontradas:

- Apelação No 0305585-56.2018.8.24.0075/SC APELADO: UNIMED DE TUBARAO COOPERATIVA DE TRABALHO MEDICO DA REG RELATÓRIO Centro de Análises Laboratoriais Ltda. apelou da sentença que, nos autos da "tutela de urgência para exibição de documentos e manutenção de convênio" ajuizada contra Unimed de Tubarão Cooperativa de Trabalho Médico da Região da Amurel, julgou improcedentes os pedidos formulados, condenando a autora ao pagamento das despesas processuais e honorários advocatícios, estes fixados em R\$ 1.500,00 (evento 28, SENT97). A apelante sustenta que: a) celebrou com a operadora de plano de saúde ré contrato de prestação de serviços laboratoriais e de análises clínicas (n. 046/2012); b) enquanto vigente o pacto, recebeu notificação, na qual a ré informava o desinteresse na manutenção do vínculo contratual sem explicar o porquê; c) tenta desde o ano de 2018 obter com a ré cópia do contrato na via administrativa, sem sucesso; d) apenas após o ingresso com a presente ação o plano de saúde apresentou os documentos outrora solicitados na esfera extrajudicial; e) ante a resistência da parte ré na apresentação do instrumento, incumbe a ela o pagamento dos ônus sucumbenciais; f) jamais recebeu cópia de eventual autorização por parte da Agência Nacional de Saúde (ANS) para fins de descredenciamento, conforme determina o art. 17, § 4º, da Lei n. 9.656/1998; g) ao contrário do que concluiu o magistrado singular, não há prova de que houve a substituição do fornecedor por outro equivalente,
- Apelação No 5023954-66.2019.8.24.0038/SC RELATOR: Juiz ALEXANDRE MORAIS DA ROSA APELANTE: AGEMED SAUDE LTDA EM LIQUIDACAO EXTRAJUDICIAL (RÉU) APELADO: PROCTOGIN SERVICOS MEDICOS EIRELI (AUTOR) RELATÓRIO Trata-se de recurso de apelação em que figuram como apelante AGEMED SAUDE LTDA. EM LIQUIDACAO EXTRAJUDICIAL e apelada PROCTOGIN SERVICOS MEDICOS EIRELI, interposto contra sentença proferida pelo juízo de origem nos autos n. 5023954-66.2019.8.24.0038. Em atenção aos princípios da celeridade e economia processual [CR, art. 5º, inciso LXXVIII], adoto o relatório da sentença como parte integrante deste acórdão, por refletir com fidelidade o trâmite processual na origem: PROCTOGIN SERVICOS MEDICOS EIRELI ajuizou ação monitória em face de AGEMED SAUDE S/A, ambos qualificados, alegando, em síntese, que: a) durante os anos de 2018 e 2019, prestou serviços médicos nas especialidades de ginecologia e obstetria, através da realização de consultas, exames e cirurgias aos usuários dos planos de saúde comercializados pela ré; b) após a prestação dos serviços, o pedido de pagamento é realizado por lote
- Apelação No 5013400-65.2019.8.24.0008/SC RELATOR: Desembargador RICARDO FONTES APELANTE: DALVA SILVEIRA DE OLIVEIRA (AUTOR) APELADO: AGEMED PLANOS DE SAUDE LTDA, todos qualificados. Alega a parte autora que é titular de plano de saúde (na modalidade coletivo empresarial), desde meados de setembro de 2012, identificado pelo no 98001.8087.013241.00-5.94, vinculado a Cooperuso Saúde, arcando com os custos mensais de R\$ 709,90, porém,

faça uma nova pergunta

Fonte: Elaborado pelo autor (2024).

Figura 22 - Seleção das das jurisprudências para servirem de entrada para a defesa jurídica

**JurisChat**

**G** Jurisprudências encontradas:

- Ação de Obrigação de Fazer c/c pedido de Tutela Antecipada de Urgência" em face de Unimed Criciúma Cooperativa Trabalho Médico Região Carbonífera, dizendo-se usuária de um de seus planos de saúde, contratado de forma particular. Disse que em decorrência de insistentes dores em sua coluna cervical procurou atendimento médico especializado, o qual lhe prescreveu tratamento a base de antibióticos, encaminhando-lhe para avaliação e tratamento médico com mastologista, já que as dores difusas na coluna teriam relação com o tamanho das mamas da autora. Mesmo diagnóstico foi atestado por segundo médico ortopedista e, diante da ausência de sucesso do tratamento medicamentoso e fisioterápico, a parte autora consultou-se com médico cirurgião plástico, o qual atestou que a mesma "necessita passar pelo procedimento de mastoplastia redutora, por apresentar problemas relacionados com o tamanho da mama" (p. 3 da inicial). Acrescentou que diante das prescrições médicas, encaminhou ao plano de saúde requerido a solicitação para realização do procedimento indicado, o que foi negado na seara administrativa, sob o fundamento de expressa exclusão contratual e porque tal procedimento cirúrgico não constava no rol da Agência Nacional de Saúde - ANS. Assim, como a ré não autorizou a cirurgia, aforou a presente ação e requereu a antecipação de tutela, para que a requerida autorize a realização do procedimento cirúrgico com cobertura integral e nos moldes indicados pelo médico assistente, com a confirmação da medida, ao final. Fez os demais requerimentos de praxe, valorou a causa e juntou documentos. O pedido de tutela foi indeferido na decisão do evento 4. Citada, a requerida ofereceu resposta em forma de contestação (evento 11). Não apresentou preliminares. No mérito, asser
- ✓ RECURSO CÍVEL No 5014860-61.2021.8.24.0091/SC RELATOR: Juiz de Direito DAVIDSON JAHN MELLO RECORRENTE: UNIMED GRANDE FLORIANÓPOLIS - COOPERATIVA DE TRABALHO MÉDICO (RÉU) RECORRIDO: RENATO BOTTER MAIO (AUTOR) RELATÓRIO Relatório dispensado, a teor do art. 38 da Lei n. 9.099/95. VOTO Trata-se de recurso inominado interposto por UNIMED GRANDE FLORIANÓPOLIS - COOPERATIVA DE TRABALHO MÉDICO com o propósito de obter a reforma da sentença que julgou parcialmente procedentes os pedidos iniciais. Em síntese, a recorrente sustenta que não pode ser obrigada a arcar com os custos do tratamento médico em questão (radiofrequência - ablação percutânea de tumor) por que há limitação contratual e o procedimento não está previsto no rol da ANS (Agência Nacional de Saúde). Ao final, postula o provimento do reclamo. Pois bem. O autor é portador de enfermidade denominada bócio não-tóxico multinodular e possui a prescrição médica para a realização de "radiofrequência - ablação percutânea de tumor" (Evento 1; EXMMED6), procedimento que a operadora de planos de saúde insurgente se nega a custear. Convém destacar que o contrato firmado pelas partes é, sim, regulamentado pela Lei n. 9.656/98, e está, portanto, submetido às regras do Código de Defesa do Consumi
- ✓ Ação Declaratória de Nulidade de Cláusula Contratual cumulada com Obrigação de Fazer e Pedido de Tutela de Urgência n. 5015346-45.2020.8.24.0038, em face de Bradesco Saúde S/A, perante a 4ª Vara Cível da comarca de Joinville. A lide restou assim delimitada, consoante relatório da sentença da lavra do magistrado Fernando Seara Hickel (evento 30): Lucas dos Santos Rengel, devidamente representado por seus genitores Juliana Testoni dos Santos Rengel e Fernando Borges Rengel, ajuizou "Ação Declaratória de Nulidade de Cláusula Contratual c/c Obrigação de Fazer e pedido de tutela de urgência" em face de Bradesco Saúde S/A, ambos devidamente qualificados. Aduziu, em síntese, que é beneficiário de plano de saúde ofertado pela ré, mediante o pagamento de parcelas mensais. De acordo com laudo médico, emitido em 24.01.2020, o autor apresenta "padrão comportamental e desenvolvimento sugestivo de autismo", sendo necessária a realização de "terapias de estimulação neurossensorial precoce com fonoaudiólogo e terapeuta ocupacional, que deverão ser mantidas por período prolongado e indeterminado". O autor começou a realizar sessões de terapia ocupacional, bem como de fonoaudiologia. Apesar disso, a partir da 13ª consulta a ré se nega a continuar adimplindo o tratamento, sob a justificativa de cumprimento das "diretrizes de utilização do Anexo II da RN428/ANS". Requereu, assim: a) concessão de tutela de urgência para determinar à ré que aprove o tratamento médico, bem como reembolse as respectivas despesas médicas
- Apelação No 0305595-56.2018.8.24.0075/SC APELADO: UNIMED DE TUBARÃO COOPERATIVA DE TRABALHO MEDICO DA REG RELATÓRIO Centro de Análises Laboratoriais Ltda. apelou da sentença que, nos autos da "tutela de urgência para exibição de documentos e manutenção de convênio" ajuizada contra Unimed de Tubarão Cooperativa de Trabalho Médico da Região do Amurel, julgou improcedentes os pedidos formulados, condenando a autora ao pagamento das despesas processuais e honorários advocatícios, estes fixados em R\$ 1.500,00 (evento 28, SENT97). A apelante sustenta que: a) celebrou com a operadora de plano de saúde ré contrato de prestação de serviços laboratoriais e de análises clínicas (n. 046/2012); b) enquanto vigente o pacto, recebeu notificação, na qual a ré informava o desinteresse na manutenção do vínculo contratual sem explicar o porquê; c) tenta desde o ano de 2018 obter com a ré cópia do contrato na

Faça uma nova pergunta

Fonte: Elaborado pelo autor (2024).

Figura 23 - Elaboração da defesa jurídica a partir das jurisprudências selecionadas

**JurisChat**

(ANS). Réplica no Evento 28. Instadas a se manifestarem sobre eventual produção de outras provas (Evento 29), as partes peticionaram nos Eventos 34 e 35. Por fim, vieram-me os autos

- Apelação No 5013400-65.2019.8.24.0008/SC RELATOR: Desembargador RICARDO FONTES APELANTE: DALVA SILVEIRA DE OLIVEIRA (AUTOR) APELADO: AGEMED PLANOS DE SAUDE LTDA, todos qualificados. Alega a parte autora que é titular de plano de saúde (na modalidade coletivo empresarial), desde meados de setembro de 2012, identificado pelo no 98001.8087.013241.00-5.94, vinculado a Cooperuso Saúde, arcando com os custos mensais de R\$ 709,90, porém, nunca recebeu uma cópia do contrato de condições gerais do plano. Nada obstante o regular pagamento, relata que de fevereiro/2019 em diante tem encontrado dificuldades para efetuar agendamentos médicos, em razão dos prestadores alegarem que "não possuem mais convênio com a empresa Ré". Que em contato com a ré por intermédio da central de relacionamento da empresa, assim como pelo 0800, não obteve sucesso. Também se deslocou até o estabelecimento físico da demandada, onde foi informada que em razão de reformas na infraestrutura os atendimentos e agendamentos estariam sendo realizados somente pelo 0800. Alega que pela dificuldade em utilizar o plano de saúde contratado, busca alternativamente atendimento médico no SUS e consultórios particulares, o que tem onerado ainda mais suas finanças. Menciona que em 17/09/19 teve que efetuar um pagamento no valor de R\$ 180,00, referente à consulta médica (cardiologista) e no dia 09/10/2019, outro pagamento no valor de R\$ 193,00, referente a exame ergométrico computadorizado, ambos na Cardíoprime, que fica nas dependências do Hospital Santa Catarina, o qual supostamente é um dos prestadores conveniados à

GERAR DEFESA JURÍDICA

**G** Trata-se de recurso cível interposto pela UNIMED GRANDE FLORIANÓPOLIS - COOPERATIVA DE TRABALHO MÉDICO, a fim de reformar a sentença que julgou parcialmente procedentes os pedidos iniciais da ação de reembolso de honorários médicos. A recorrente alega que não pode ser obrigada a arcar com os custos do tratamento médico em questão, uma radiofrequência - ablação percutânea de tumor, devido à limitação contratual e à ausência de previsão do procedimento no rol da ANS. No entanto, é importante ressaltar que, conforme jurisprudência consolidada, a ausência de previsão do procedimento no rol da ANS não é determinante para a negativa de reembolso de honorários médicos em casos específicos. O Tribunal de Justiça de Santa Catarina, por exemplo, tem entendido que a negativa de cobertura de tratamento médico prescrito por profissional de saúde, ainda que não conste no rol da ANS, pode configurar abusividade por parte da operadora de plano de saúde. No caso em questão, o autor da ação foi diagnosticado com carcinoma e buscou consultas e cirurgias com um médico particular, mesmo havendo rede credenciada disponível no plano de saúde. O contrato firmado entre as partes não prevê reembolso para livre escolha de profissionais de saúde, e a falha no atendimento não foi comprovada. Diante disso, a jurisprudência tem entendido que, nos casos em que o contrato não prevê reembolso para tratamentos realizados por prestadores particulares e não há comprovação de falha no atendimento da rede credenciada, a negativa de reembolso por parte da operadora de plano de saúde é legítima. Portanto, com base na jurisprudência citada e nas circunstâncias do caso em análise, é possível argumentar que a sentença de improcedência da ação de reembolso de honorários médicos está em conformidade com a legislação e com o contrato firmado entre as partes. Assim, o recurso interposto pela UNIMED GRANDE FLORIANÓPOLIS - COOPERATIVA DE TRABALHO MÉDICO deve ser julgado improcedente.

Faça uma nova pergunta

Fonte: Elaborado pelo autor (2024).

Após os testes do protótipo, o profissional jurídico avaliou positivamente a ferramenta, destacando a simplicidade e o auxílio que o mesmo pode promover na área legal, conforme especificou através de um relato formal "*O aplicativo proposto é uma ferramenta que visa simplificar o processo de criação de defesas jurídicas para advogados e outros profissionais do direito. Ele utiliza algoritmos para analisar jurisprudências relevantes de diversos tribunais e gerar argumentos sólidos para responder às perguntas propostas ao aplicativo pelo profissional para utilização em casos concretos. A aplicação a meu ver oferece uma solução promissora para advogados que desejam otimizar o processo de criação de defesas jurídicas. Com melhorias contínuas e feedback dos usuários, ela tem o potencial de se tornar uma ferramenta indispensável no mundo jurídico.*".

## 6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Nesta pesquisa é enfatizada a importância da jurisprudência, possuindo papel fundamental para assegurar a convicção e previsibilidade nos processos judiciais, promovendo a igualdade entre os jurisdicionados, ao mesmo tempo em que ressalta o respeito, a sabedoria acumulada e a experiência dos magistrados. Porém, a busca de jurisprudências a partir de um contexto específico, servindo de suporte à elaboração de defesa de um processo jurídico por parte de um profissional legal, apresenta limitações quando não abrange o contexto semântico.

Para tal, este trabalho teve como objetivo propor e desenvolver um método voltado à recuperação de jurisprudências como suporte à elaboração de defesas em processos jurídicos. Para a construção deste documento, pesquisas bibliográficas foram realizadas voltadas à análise de jurisprudências, precedentes, técnicas de NLP e LLMs. Das pesquisas elencadas na seção de trabalhos correlatos não foram encontrados trabalhos com proposta similar à apresentada neste trabalho. O método proposto consiste em 8 etapas, começando pela coleta, limpeza e sumarização, passando pela transformação de jurisprudências em *embeddings*. Posteriormente, constitui a base de dados vetorial que possibilita a execução das próximas etapas relacionadas à solicitação e consulta das jurisprudências. Na última etapa, jurisprudências recuperadas devem ser selecionadas, possibilitando assim a geração de uma sugestão de defesa jurídica.

Nesta pesquisa, um conjunto de jurisprudências disponíveis no Tribunal de Justiça de Santa Catarina foi coletado para compor a base de dados (Seção 3.3.3.1). Deste conjunto de dados, foram consideradas jurisprudências relacionadas à área de saúde, sendo que cada jurisprudência foi sumarizada pelo modelo *t5-portuguese-legal-summarization* e, posteriormente, transformada em um vetor denso para armazenamento. Para criação dos vetores densos das jurisprudências utilizou-se o LLM *juridics/bertlaw-base-portuguese-sts-scale*. Por fim, as jurisprudências recuperadas de acordo com determinado *prompt* servem de entrada para a geração de proposta de defesa jurídica utilizando-se do LLM GPT-3.5-Turbo®.

A partir do conjunto de dados, foram apresentados quatro cenários voltados à avaliação do nível de concordância entre a apresentação de jurisprudências providas pelo protótipo e a determinação da ordem pelos profissionais, ambos analisados na Seção 5.3. Em cada um dos quatro cenários, o método proposto foi

avaliado considerando a correlação de ordenação de 10 jurisprudências efetuada tanto pela aplicação (protótipo) quanto pelos especialistas jurídicos. Concentraram-se assim, na identificação de possíveis correlações de ordenações entre o protótipo desenvolvido e os profissionais jurídicos. A avaliação dos profissionais jurídicos foi positiva somente no terceiro *prompt*, não obtendo concordâncias consistentes nos demais cenários, identificando um entendimento jurídico distinto entre ambos profissionais.

Os resultados destes quatro cenários foram ainda analisados a partir da média das correlações de ordenações. As avaliações realizadas pelo profissional 1, na maioria dos casos obteve valor negativo, enquanto o profissional 2, avaliou positivamente a maioria dos casos, definindo a média como sendo negativa em três dos quatro cenários específicos. Isto pode denotar entendimentos bem distintos das jurisprudências entre os profissionais, mas também possíveis pontos de melhoria no processo executado para a instanciação do método. Vale mencionar ainda que, para análises mais confiáveis, mais profissionais jurídicos deveriam participar da avaliação.

Adicionalmente, a etapa de geração de texto para suportar a elaboração de defesas foi avaliada de maneira informal, pois não foram aplicados quaisquer métodos ou questionários. Todavia, o intuito foi coletar a percepção inicial de um dos profissionais sobre a funcionalidade e qualidade do texto produzido. O profissional, após a realização do texto, declarou que o projeto “... *tem potencial de se tornar uma ferramenta indispensável no mundo jurídico*”.

Apesar das limitações das avaliações realizadas, os resultados demonstram potencial para continuidade da pesquisa e desenvolvimento com foco na utilização de AI, DL e LLMs na criação de ferramenta que auxilie profissionais da área jurídica na elaboração de defesas.

Neste sentido, outras possibilidades podem ser exploradas para trabalhos futuros. A fim de obter similaridades mais elevadas, é possível utilizar um modelo superior ao *juridics/bertlaw-base-portuguese-sts-scale* visto que seu treinamento é baseado no projeto SBERT que contém limitações quanto a quantidade de *tokens* de entrada quando comparado à LLMs como GPT® e Gemini®. O modelo pode também sofrer o ajuste fino (*fine-tuning*) a partir do conjunto de jurisprudências coletadas. Mais especificamente, quanto ao banco de dados vetorial, foi utilizado o ranqueamento baseado no cosseno, exemplificado na Etapa 4.6. Porém, com a

evolução do trabalho, é possível implementar e testar outras formas de ranqueamento de jurisprudências, por exemplo, baseados em distâncias como a euclidiana ou *manhattan*. Além disso, no contexto das jurisprudências, é possível viabilizar buscas com escopo delimitado por meio de um método híbrido, ou seja, mesclando a busca semântica e a busca por palavras-chave. O método proposto pode ser entendido como um sistema de Resposta a Perguntas (do inglês *Question Answering* - QI). Entretanto, com os devidos ajustes, pode ser transformado em um sistema de *ChatBot* adicionando a capacidade de considerar nos *prompts* às interações prévias dos profissionais jurídicos.

Mostra-se relevante ainda a criação de uma aplicação *web* ou aplicativo de recuperação de jurisprudências e geração de defesa jurídica, tendo como base o protótipo apresentado na Seção 5.3. Por fim, pensando em termos de empreendedorismo, a disponibilização para o público poderia viabilizar a coleta de diferentes *feedbacks* pensando nas melhorias e, em momento futuro, na oferta de serviço especializado.

## REFERÊNCIAS

- AVELINO, M. T. Conceitos Fundamentais do Sistema de Precedentes e Fundamentação das Decisões Judiciais. **Revista Acadêmica da Faculdade de Direito do Recife**, v. 91, n. 2, p. 151, 2020.
- BIDINOTTO, V. O. B.; MARTINS, R. C. N. Precedente Judicial como Ferramenta de Combate à Padronização Decisória no novo Código de Processo Civil. **Revista da Faculdade de Direito da Universidade Federal de Uberlândia**, v. 49, n. 2, p. 391–407, 2022.
- BHATTACHARYA, P.; GHOSH, K.; PAL, A.; GHOSH, S. Legal case document similarity: You need both network and text. **Information Processing & Management**, v. 59, n. 6, p. 103069, 2022.
- CASTELFRANCHI, C. Alan Turing's "computing machinery and intelligence". **Topoi: an international review of philosophy**, v. 32, n. 2, p. 293–299, 2013.
- CHENG, Yu; CHEN, Jieshan; HUANG, Qing; *et al.* Prompt sapper: a llm-empowered production tool for building ai chains. **ACM Transactions on Software Engineering and Methodology**, v. 33, n. 5, p. 124:1-124:24, 2024.
- CNJ (Conselho Nacional de Justiça). **Justiça em Números 2022: Judiciário julgou 26,9 milhões de processos em 2021**. 2022.
- DE MARTINO, G., PIO, G; CECI, M. PRILJ: an efficient two-step method based on embedding and clustering for the identification of regularities in legal case judgments. **Artificial Intelligence Law**, v. 30, p. 359–390, 2022.
- DRESCH, A.; LACERDA, D. P.; ANTUNES JR., J. A. V. **Design science research: a method for science and technology advancement**. [s. l.]: Springer, 2015.
- FERRAZ, A. C. C.; BARBOSA, V. Efeitos Vinculantes e Direitos Fundamentais: Questionamentos ainda Necessários. **Revista de Estudos e Pesquisas Avançadas do Terceiro Setor**, v. 3, n. 2, p. 168, 2017.
- FERNANDES, W. P. D.; *et al.* Extracting value from Brazilian Court decisions. **Information Systems**, v. 106, p. 101965, 2022.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.
- GERHARDT, T.; SILVEIRA, D. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.
- GRECO, C. M.; TAGARELLI, A. Bringing order into the realm of Transformer-based language models for artificial intelligence and law. **Artificial Intelligence Law**, 2023.
- GUO, Rentong; LUAN, Xiaofan; XIANG, Long; *et al.* Manu: A Cloud Native Vector Database Management System. 2022.

HEVNER, A. R.; MARCH, S. T.; PARK, J.; RAM, S. Design Science in Information Systems Research. **MIS Quarterly**, v. 28, n. 1, p. 75, 2004.

HOU, J.; COSMA, G.; FINKE, A. Advancing continual lifelong learning in neural information retrieval: definition, dataset, framework, and empirical evaluation. 2023.

JAIN, D.; BORAH, M. D.; BISWAS, A. Summarization of legal documents: Where are we now and the way forward. **Computer science review**, v. 40, n. 100388, p. 100388, 2021.

JONES, William P.; FURNAS, George W. Pictures of relevance: A geometric analysis of similarity measures. **Journal of the American Society for Information Science**, v. 38, n. 6, p. 420–442, 1987.

KIM, M.; RABELO, J.; BABIKER, H. K. B.; RAHMAN, M. A.; GOEBEL, R. Legal Information Retrieval and Entailment Using Transformer-based Approaches. **The Review of Socionetwork Strategies**, v. 18, n. 1, p. 101–121, 2024.

KRAFT, D. H.; COLVIN, E. Introduction to Information Retrieval. Em: **Fuzzy Information Retrieval**. Cham: Springer International Publishing, 2017.

LASHKARI, F.; BAGHERI, E.; GHORBANI, A. A. Neural embedding-based indices for semantic search. **Information processing & management**, v. 56, n. 3, p. 733–755, 2019.

LEUNG, E. S. M. The jurisprudence and administration of legal interpreting in Hong Kong (1966–2016). **International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique**, v. 32, n. 1, p. 95–116, 2019.

MA, Y.; WU, Y.; AI, Q.; LIU, Y.; SHAO, Y.; ZHANG, M.; MA, S. Incorporating structural information into legal case retrieval. **ACM transactions on information systems**, v. 42, n. 2, p. 1–28, 2024.

MANDAL, A.; GHOSH, K.; GHOSH, S.; MANDAL, S. Unsupervised approaches for measuring textual similarity between legal court case reports. *Artificial Intelligence and Law*. [S. l.]: Springer Science and Business Media LLC, 2021.

MANDAL, A.; GHOSH, K.; GHOSH, S.; MANDAL, S. A sequence labeling model for catchphrase identification from legal case documents. *Artif Intell Law* **30**, 325–358(2022).

MUMCUOĞLU, E.; ÖZTÜRK, C. E.; OZAKTAS, H. M.; KOC, A. Natural language processing in law: Prediction of outcomes in the higher courts of Turkey. **Information processing & management**, v. 58, n. 5, p. 102684, 2021.

NAIK, V.; K, R.; PATEL, P. Enhancing Semantic Searching of Legal Documents Through LSTM-Based Named Entity Recognition and Semantic Classification. **International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique**, 2024.



PAN, J. J.; WANG, J.; LI, G. Survey of Vector Database Management Systems. 2023.

PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A.; CHATTERJEE, S. A Design Science Research Methodology for Information Systems Research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45–77, 2007.

RAMOS, H. A. D. C.; RAMOS, P. A. D. C.; GARCÍA-PEÑALVO, F. J. **Technological research methodology to manage organizational change**. Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality. **Anais...**New York, NY, USA: ACM, 2019.

SAMPATH, K.; DURAIRAJ, T. PReLCaP : Precedence Retrieval from Legal Documents Using Catch Phrases. *Neural Process Lett* **54**, 3873–3891 (2022).

SANSONE, C.; SPERLÍ, G. Legal Information Retrieval systems: State-of-the-art and open issues. **Information systems**, v. 106, n. 101967, p. 101967, 2022.

SCHELEDER, A. F. P.; NOSCHANG, P. G.. Precedentes e Jurisprudência no Direito Brasileiro: Uma Distinção Necessária a Partir dos Sistemas Jurídicos de Common Law e Civil Law. **Revista da Faculdade de Direito da UFMG**, n. 72, 2018.

SMILKOV, D; THORAT, N.; NICHOLSON, C.; REIF, E.; VIÉGAS, F. B.; WATTENBERG, M.. Embedding Projector: Interactive Visualization and Interpretation of Embeddings. 2016.

SPEARMAN, C. The proof and measurement of association between two things. **The American journal of psychology**, v. 15, n. 1, p. 72, 1904.

TAIPALUS, T. Vector database management systems: Fundamental concepts, use-cases, and current challenges. **Cognitive Systems Research**, v. 85, p. 101216, 2024.

WANG, J.; SHEN, J. Fast spectral analysis for approximate nearest neighbor search. **Machine learning**, v. 111, n. 6, p. 2297–2322, 2022.

YANG, J.; *et al.* Harnessing the power of llms in practice: A survey on chatgpt and beyond. **ACM Transactions on Knowledge Discovery from Data**, v. 18, n. 6, p. 1-32, 2024.

ZHAO, W.; *et al.* A survey of large language models. **arXiv preprint arXiv:2303.18223**, 2023.

ZHEBEL, V. V.; DEVYATKIN, D. A.; ZUBAREV, D. V.; SOCHENKOV, I. V. Approaches to Cross-Language Retrieval of Similar Legal Documents Based on Machine Learning. **Scientific and Technical Information Processing**, v. 50, n. 5, p. 494–499, 2023.