CESAR SCHOOL

FLÁVIO JOSÉ FERREIRA JUNIOR

EXTRAÇÃO DE ENDEREÇOS EM PROCESSOS DE USUCAPIÃO COM LLMS

RECIFE

2026

FLÁVIO JOSÉ FERREIRA JUNIOR

EXTRAÇÃO DE ENDEREÇOS EM PROCESSOS DE USUCAPIÃO COM LLMS

Dissertação apresentada ao programa de Mestrado em Engenharia de Software do Centro de Estudos e Sistemas Avançados do Recife – CESAR School, como requisito para a obtenção do título de Mestre em Engenharia de Software.

Orientação: Prof. Anderson Tenório Sergio.

RECIFE

2026

FOLHA DE APROVAÇÃO

FLÁVIO JOSÉ FERREIRA JUNIOR

EXTRAÇÃO DE ENDEREÇOS EM PROCESSOS DE USUCAPIÃO COM LLMS

Trabalho aprovado em Recife: **31/03/2026.**

Professor: Prof. Anderson Tenório Sergio.
(CESAR SCHOOL)
Orientador(a)

Professor: Beatriz Leandro Bonafini
(CESAR SCHOOL)
Avaliador(a) Interno(a)

Professor: Hemir da Cunha Santiago
(UPE - POLI)
Avaliador(a) Externo(a)

RECIFE
2026

Dedico este trabalho à minha esposa Germana, companheira em todas as jornadas, pelo incentivo constante e compreensão nos momentos de ausência.

Às minhas filhas Ana Flávia e Laura, razão do meu esforço e inspiração para continuar aprendendo e ensinando.

Agradecimentos

Agradeço primeiramente a Deus, pela saúde e sabedoria que me permitiram chegar até aqui.

Ao meu orientador, Prof. Dr. Anderson Tenório Sérgio, pela orientação precisa, paciência e pelas valiosas contribuições que aprimoraram o rumo deste trabalho.

Aos colegas do Tribunal de Justiça de Pernambuco (TJPE) em especial a Sara Lima, pela colaboração na coleta e compreensão dos dados.

Aos professores do Programa de Mestrado Profissional em Engenharia de Software, pela excelência acadêmica e apoio institucional.

E à minha família, por todo o amor, compreensão e apoio incondicional durante este percurso.

RESUMO

A Inteligência Artificial (IA) tem se consolidado como vetor de modernização institucional no Poder Judiciário brasileiro, com impacto direto na eficiência e na celeridade processual. Esta dissertação investiga a aplicação de Grandes Modelos de Linguagem (LLMs) na extração automatizada de endereços de imóveis em processos de usucapião tramitando no Tribunal de Justiça de Pernambuco (TJPE), no contexto do programa institucional TJPE Moradia Legal e das metas do Conselho Nacional de Justiça (CNJ).

Por meio de técnicas de Reconhecimento de Entidades Nomeadas (NER), o estudo compara o desempenho de três modelos – ChatGPT, Gemini e DeepSeek – avaliando precisão, revocação, F1-score, latência e custo financeiro na identificação de componentes de endereço (logradouro, número, bairro, município, estado e CEP) em petições judiciais não estruturadas.

A metodologia adota abordagem quantitativa e experimental, com corpus de 100 processos reais do sistema PJe, previamente anonimizados em conformidade com a Lei Geral de Proteção de Dados (LGPD). Os prompts foram elaborados com técnica de zero-shot prompting e as saídas estruturadas em formato JSON, submetidas posteriormente a geocodificação via Google Maps para conversão dos endereços extraídos em coordenadas geográficas.

Os resultados demonstram que todos os modelos alcançaram revocação máxima (1,0), identificando integralmente os campos de endereço esperados. O ChatGPT obteve o melhor desempenho global em qualidade (acurácia de 45%, F1-score de 0,7552 e precisão média de 0,6067), seguido pelo DeepSeek (acurácia de 40%, F1-score de 0,7500) e pelo Gemini (acurácia de 39%, F1-score de 0,7234). Em termos operacionais, o Gemini apresentou a menor latência média (0,97 s), enquanto o DeepSeek destacou-se pelo menor custo financeiro (US$ 0,056 por consulta). Os testes estatísticos ANOVA e Mann-Whitney confirmaram ausência de diferença significativa entre os modelos (p-valor = 0,7627), indicando desempenho equivalente

para fins práticos. A taxa de conversibilidade geográfica situou-se em 55% para ChatGPT e DeepSeek, e 52% para o Gemini.

Como contribuição prática, desenvolveu-se o protótipo GeoLLM-Extractor, composto pelo módulo de extração automatizada integrado às APIs das LLMs e pelo Sistema de Gestão Moradia Legal (SGML), plataforma web de governança fundiária com visualização georreferenciada, validação humana e rastreabilidade dos dados. A pesquisa recomenda um fluxo de trabalho híbrido, no qual a automação por LLMs é complementada pela supervisão de profissionais do direito, e contribui para o avanço do uso responsável da IA no Judiciário, com potencial de replicação em outras unidades estaduais e em distintos ramos da Justiça que lidam com conflitos fundiários e territoriais.

Palavras-chave: Inteligência Artificial; LLM; NER; Usucapião; Poder Judiciário; Georreferenciamento.

ABSTRACT

Artificial Intelligence (AI) has emerged as a key driver of institutional modernization within the Brazilian Judiciary, contributing to greater efficiency, procedural speed, and quality of jurisdictional services. This dissertation investigates the application of Large Language Models (LLMs) for the automated extraction of property addresses in adverse possession (usucapião) cases processed by the Court of Justice of Pernambuco (TJPE), within the scope of the TJPE Moradia Legal institutional program and the performance targets set by the National Council of Justice (CNJ).

Using Named Entity Recognition (NER) techniques, three models – ChatGPT, Gemini, and DeepSeek – were compared across multiple dimensions: precision, recall, F1-score, response latency, and financial cost in identifying address components (street name, number, neighborhood, city, state, and postal code) from unstructured legal petitions.

A quantitative and experimental methodology was employed using a corpus of 100 real court cases anonymized from the PJe system in compliance with Brazilian data protection law (LGPD). Prompts were designed using a zero-shot prompting strategy, with outputs structured in JSON format and subsequently submitted to geocoding via Google Maps to convert extracted addresses into geographic coordinates.

All three models achieved maximum recall (1.0), fully recovering the expected address fields. ChatGPT delivered the best overall quality performance (45% accuracy, F1-score of 0.7552, average precision of 0.6067), followed by DeepSeek (40% accuracy, F1-score of 0.7500) and Gemini (39% accuracy, F1-score of 0.7234). In operational terms, Gemini achieved the lowest average latency (0.97 s), while DeepSeek stood out for the lowest financial cost (US$ 0.056 per query). ANOVA and Mann-Whitney statistical tests confirmed no significant difference among the models (p-value = 0.7627), indicating statistically equivalent performance. Geographic convertibility rates reached 55% for ChatGPT and DeepSeek, and 52% for Gemini.

As a practical contribution, the GeoLLM-Extractor prototype was developed, consisting of an automated extraction module integrated with LLM APIs and the Moradia Legal

Management System (SGML) – a web-based land governance platform featuring georeferenced visualization, human validation, and data traceability. The research recommends a hybrid workflow in which LLM-based automation is complemented by legal professional supervision and contributes to the responsible advancement of AI use in the Judiciary, with replication potential across other state courts and legal branches dealing with land and territorial disputes.

Keywords: Artificial Intelligence; LLM; NER; Usucapion; Judiciary; Georeferencing.

Lista de ilustrações

Lista de tabelas

Lista de gráficos

Lista de abreviaturas e siglas

Sigla	Significado
AI / IA	Artificial Intelligence / Inteligência Artificial
ANOVA	Analysis of Variance (Análise de Variância)
API	Application Programming Interface
CEP	Código de Endereçamento Postal
CNJ	Conselho Nacional de Justiça
CRF	Conditional Random Fields
GPT	Generative Pre-trained Transformer
HMM	Hidden Markov Model
IBGE	Instituto Brasileiro de Geografia e Estatística
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
JSON	JavaScript Object Notation
LGPD	Lei Geral de Proteção de Dados Pessoais
LSTMs	Long Short-Term Memory
NLP / PLN	Natural Language Processing / Processamento de Linguagem Natural
PJe	Processo Judicial Eletrônico
STF	Supremo Tribunal Federal
STJ	Superior Tribunal de Justiça
TF-IDF	Term Frequency – Inverse Document Frequency

SUMÁRIO

1 INTRODUÇÃO

A IA tem se consolidado como uma tecnologia transformadora em diversos setores da sociedade, inclusive no Poder Judiciário, sendo apontada como um dos principais vetores de modernização institucional, com impacto direto na eficiência, celeridade e qualidade da prestação jurisdicional (CNJ, 2019).

Entre suas vertentes mais avançadas, os Grandes Modelos de Linguagem (*Large Language Models* – LLMs) destacam-se por sua capacidade de compreender e gerar linguagem natural em escala, viabilizando aplicações que vão desde a análise textual até a automação de tarefas cognitivas complexas no âmbito do Processamento de Linguagem Natural (ARAUJO; SILVEIRA, 2025).

Dentre as técnicas associadas a esses modelos, o Reconhecimento de Entidades Nomeadas (*Named Entity Recognition* – NER) destaca-se como uma abordagem relevante para a extração automatizada de informações estruturadas a partir de textos não estruturados, inclusive no domínio jurídico, permitindo a identificação de entidades como pessoas, organizações, locais e referências normativas, além de elementos espaciais e temporais (ARAUJO; SILVEIRA, 2025). Essa técnica é amplamente empregada em conjunto com outros métodos de mineração textual, como classificação de documentos, sumarização automática e análise semântica.

No contexto do Judiciário brasileiro, o uso de ferramentas baseadas em IA generativa, como os LLMs, tem se expandido, ainda que de forma heterogênea. Relatório de pesquisa do CNJ aponta que aproximadamente metade dos magistrados e servidores já utilizaram ferramentas de IA generativa, com destaque para atividades de apoio técnico, elaboração de minutas e pesquisa jurídica (CNJ, 2024).

Apesar das limitações identificadas como: a possibilidade de imprecisões, ocorrência de respostas alucinatórias e a ausência de controle institucional sistematizado, a percepção geral é de que tais ferramentas contribuem para o

aumento da eficiência e da qualidade da prestação jurisdicional, desde que utilizadas de forma supervisionada e responsável (CNJ, 2024).

A aplicação de LLMs e técnicas de NER no domínio jurídico, especialmente para o reconhecimento automatizado de endereços em processos de usucapião, revela-se uma frente promissora de inovação tecnológica. Esses processos exigem, como condição essencial para seu regular trâmite, a correta delimitação e localização do imóvel objeto da pretensão possessória. A automação dessa identificação, por meio da extração de entidades como logradouro, número, bairro, cidade e estado, pode mitigar falhas humanas, acelerar a triagem documental e conferir maior segurança à instrução processual, conforme apontam estudos recentes sobre NER aplicado a textos jurídicos (ARAUJO; SILVEIRA, 2025).

O presente trabalho investiga a eficiência e a eficácia de LLMs amplamente utilizados no ecossistema tecnológico contemporâneo na tarefa de identificar endereços em textos jurídicos extraídos de ações de usucapião. Considera-se, para isso, o desafio intrínseco de aplicar modelos que não foram originalmente treinados em bases jurídicas especializadas e que, portanto, enfrentam limitações na interpretação do vocabulário técnico-formal do Direito.

A hipótese central reside na capacidade desses modelos, mesmo sem ajustes específicos, de reconhecer padrões sintáticos relacionados à localização formal do imóvel, possibilitando a extração precisa de informações essenciais à instrução processual. Além disso, busca-se explorar o potencial dessas ferramentas na identificação automatizada de coordenadas geográficas, permitindo o reconhecimento espacial do imóvel e sua eventual vinculação a áreas de interesse social, em consonância com as diretrizes institucionais de regularização fundiária promovidas pelo CNJ (CNJ, 2023a; CNJ, 2023b).

Conforme destaca o CNJ (2024), a utilização responsável de IAs generativas exige atenção permanente aos riscos de opacidade, vieses algorítmicos, erros factuais e ausência de mecanismos formais de validação. Por essa razão, esta pesquisa incorpora uma análise crítica dos limites e das boas práticas recomendadas

para o uso dessas tecnologias no âmbito do Poder Judiciário, em alinhamento com os marcos normativos e programáticos estabelecidos pelo CNJ.

Espera-se, com isso, contribuir para o avanço técnico e social do emprego de técnicas de NER na Justiça brasileira, especialmente no apoio a ações de regularização fundiária. Tais iniciativas beneficiam majoritariamente populações em situação de vulnerabilidade social, público central dos programas institucionais de regularização fundiária e governança territorial, como os programas “Solo Seguro” e “Solo Seguro – Favela” (CNJ, 2025a; CNJ, 2023a; CNJ, 2023b).

1.1 Motivação

- Motivação de Mercado

A motivação de mercado desta pesquisa fundamenta-se na necessidade de os órgãos do Poder Judiciário brasileiro atenderem às metas de qualidade da prestação jurisdicional estabelecidas pelo CNJ, especialmente no que se refere à eficiência, à celeridade processual e à modernização dos fluxos de trabalho institucionais. Nesse contexto, o estudo contribui diretamente para o alcance dos indicadores definidos na Meta 2 do CNJ, a qual estabelece como objetivo prioritário o julgamento dos processos mais antigos em tramitação no âmbito dos Tribunais de Justiça Estaduais.

Tal meta busca reduzir o congestionamento processual e eliminar o estoque de feitos de longa duração, por meio de metas anuais progressivas. Especificamente para o ano de 2025, a Meta 2 determina o julgamento da maior parte dos processos distribuídos até os anos de 2021 e 2022, tanto em primeiro quanto em segundo graus de jurisdição, incluindo a obrigatoriedade de julgamento de 100% dos processos pendentes há 15 anos ou mais. Essas diretrizes visam assegurar o cumprimento do princípio constitucional da razoável duração do processo e a melhoria da eficiência da prestação jurisdicional.

Sendo assim, uma das iniciativas adotadas pelo Tribunal de Justiça de Pernambuco consistiu na identificação de processos de usucapião cujos imóveis estejam localizados em áreas de interesse social, passíveis de priorização, agilização

processual e consequente encerramento do feito. Essa estratégia busca direcionar esforços institucionais para processos com elevado impacto social, ao mesmo tempo em que contribui para a redução do estoque processual e para o cumprimento das metas estabelecidas pelo CNJ.

O Painel de Monitoramento de Usucapião do CNJ consolida indicadores processuais sobre ações de regularização fundiária em todo o Poder Judiciário nacional, evidenciando o volume expressivo de processos pendentes e o impacto social das iniciativas de celeridade processual. Os dados disponibilizados por esse painel reforçam a justificativa institucional desta pesquisa, ao demonstrar que a automação da identificação e georreferenciamento de imóveis pode contribuir diretamente para as metas nacionais de julgamento e para a redução do acervo processual, em consonância com as diretrizes do programa TJPE Moradia Legal e as políticas do CNJ (CNJ, 2023a; CNJ, 2023b).

- Motivação Técnica

A motivação técnica desta pesquisa decorre das dificuldades associadas ao processamento automatizado de documentos jurídicos não estruturados, especialmente em processos de usucapião, nos quais informações essenciais sobre a identificação e a localização dos imóveis encontram-se dispersas em textos extensos, heterogêneos e com baixa padronização. Essa característica limita a extração manual e a utilização dessas informações em sistemas de apoio à gestão processual.

Nesse contexto, o NER apresenta-se como uma técnica fundamental para a transformação de texto jurídico em dados estruturados, possibilitando a identificação automática de elementos de endereço. Contudo, abordagens tradicionais de NER enfrentam limitações no domínio jurídico brasileiro, em razão da linguagem especializada, da ambiguidade semântica e da escassez de bases anotadas específicas (ARAUJO; SILVEIRA, 2025; FRANKENREITER; NYARKO, 2022).

Assim, a pesquisa se motiva tecnicamente pela necessidade de investigar e validar o uso de LLMs para a tarefa de NER em processos de usucapião, avaliando

sua precisão, consistência e viabilidade operacional, com vistas a subsidiar a aplicação dessas tecnologias em ambientes judiciais.

1.2 Problema

O Poder Judiciário, em especial no âmbito dos processos de usucapião, lida diariamente com um grande volume de documentos textuais não estruturados que contêm elementos essenciais à instrução processual, como endereços completos dos imóveis objeto da lide. A extração manual dessas informações é uma atividade lenta, repetitiva e suscetível a erros, impactando diretamente a eficiência das unidades judiciais e a celeridade processual.

Embora modelos avançados de Inteligência Artificial, como LLMs, tenham demonstrado potencial para tarefas de extração de informação, ainda não está claro qual é o desempenho real desses modelos ao processar textos jurídicos brasileiros, nem quais são os limites, custos e requisitos de supervisão humana necessários para sua adoção segura no ambiente judicial.

Diante desse cenário, o problema de pesquisa consiste em avaliar se LLMs são capazes de extrair, com precisão, consistência e baixo custo computacional, dados endereçáveis presentes em processos de usucapião tramitando no TJPE, bem como verificar quais modelos apresentam melhor desempenho e se podem ser integrados a um fluxo automatizado de apoio à atividade jurisdicional, mantendo níveis aceitáveis de precisão, latência, custo financeiro e interoperabilidade com sistemas como o PJe.

1.2.1 Objetivo geral

Esta pesquisa tem como objetivo principal analisar o desempenho de diferentes LLMs na tarefa de identificação de elementos relevantes em processos judiciais, com ênfase na extração de dados relacionados ao endereço de imóveis objeto de ações de usucapião. O foco recai sobre a aplicação de técnicas de NER para a extração automatizada de informações de localização a partir de textos jurídicos e a posterior conversão dos dados de endereços em coordenadas geográficas.

Sob a perspectiva jurídica, a identificação precisa do endereço de um imóvel é fundamental para a correta instrução dos processos de regularização fundiária, viabilizando o armazenamento estruturado de dados e contribuindo para a integração com sistemas geográficos e registrários. Tal organização da informação favorece a adoção de práticas mais eficientes na análise territorial e na tomada de decisões judiciais, fortalecendo a segurança jurídica e a efetividade das decisões.

Do ponto de vista tecnológico, a avaliação precisa do comportamento de diferentes LLMs quanto à sua capacidade de processamento, interpretação documentos jurídicos e o custo financeiro da utilização da ferramenta, oferece subsídios técnicos e comparativos para a escolha de ferramentas mais adequadas à tarefa de NER no contexto do Judiciário.

1.2.2 Objetivos específicos

1. Propor um modelo de avaliação para LLMs aplicados à tarefa de extração automatizada de texto em processos judiciais de usucapião, com foco na identificação de entidades nomeadas relacionadas a endereços de imóveis.

2. Investigar a possibilidade de conversão dos dados extraídos referentes a endereços em coordenadas geográficas, por meio de serviços de geocodificação, a fim de analisar a viabilidade técnica e a efetividade do uso de LLMs no processo de georreferenciamento de imóveis em ações de usucapião.

3. Implementar e documentar um protótipo de *software* denominado *GeoLLM-Extractor*, responsável por integrar a extração de endereços da base de dados do PJe com APIs externas de LLM e geocodificação a bases relacionais de apoio a tomada de decisão;

1.3 Justificativa

A pesquisa se justifica pela necessidade de automatizar a extração de informações essenciais em processos de usucapião, reduzindo etapas manuais repetitivas e ampliando a eficiência das unidades judiciais. Essa automação tende a

contribuir para a diminuição do tempo de tramitação e para o cumprimento do princípio da duração razoável do processo, previsto no art. 5º, inciso LXXVIII, da Constituição Federal.

Ao avaliar o desempenho de LLMs na tarefa de extração de endereços, o estudo fortalece o avanço técnico e institucional do uso responsável de IA no Judiciário, além de demonstrar a viabilidade de soluções integradas ao PJe para apoio às rotinas processuais.

1.4 Contribuições

A pesquisa contribui diretamente para o programa TJPE Moradia Legal, segundo (TJPE, 2025), o programa é responsável pela entrega de mais de 15 mil títulos de propriedade em 2024, e indiretamente, para políticas de regularização fundiária e inclusão social, pois acelera a identificação e localização de imóveis passíveis de titulação. Segundo Sara Lima em (TJPE, 2025), o programa não só garante a segurança jurídica, mas também contribui para a qualidade de vida e o desenvolvimento de cidades sustentáveis.

Além disso, reforça a ética e responsabilidade no uso de IA no setor público, alinhando-se às diretrizes de governança digital do CNJ. O domínio técnico do NER aplicado a documentos jurídicos representa, portanto, uma contribuição significativa à eficiência da atividade jurisdicional. Seu potencial de replicabilidade em diferentes tribunais do país reforça a relevância estratégica desta pesquisa, especialmente diante das diretrizes do CNJ voltadas à ampliação do uso de tecnologias para fins de regularização fundiária, conforme previsto em suas políticas de inovação e governança digital (CNJ, 2024).

A pesquisa resultou na implementação de um protótipo de *software* denominado *GeoLLM-Extractor*, projetado para integrar:

- Extração de entidades nomeadas (endereços) em textos jurídicos não estruturados;
- Comunicação com APIs externas de LLMs (ChatGPT, Gemini, DeepSeek) e

- Integração com serviços de geocodificação e bancos de dados relacionais.

Essa arquitetura híbrida configura uma aplicação prática de engenharia de *software* orientada por dados, baseada em inteligência artificial, que integra técnicas de PLN, consumo de APIs e *pipelines* de processamento de dados. Tal abordagem contribui para o avanço da engenharia de *software* aplicada a sistemas inteligentes no domínio jurídico, ao demonstrar a viabilidade da automação de tarefas complexas a partir de bases textuais não estruturadas. A integração de técnicas de IA às práticas jurídicas vem de encontro a necessidade de promoção de celeridade defendida pelo CNJ, e sobretudo pela sociedade. Conforme depoimento do presidente do TJPB, desembargador Fred Coutinho:

> "... antes, a coleta de dados exigia deslocamentos e muito tempo dos servidores; agora, tudo é feito automaticamente. É mais uma ferramenta de inteligência artificial que veio para ficar e que mostra o compromisso do Tribunal com a inovação e a eficiência" (TJPB, 2025).

Não obstante às contribuições técnicas à engenharia de *software*, esta pesquisa contribuiu com a publicação do artigo "Tecnologia de Mineração de Texto no Direito: o caso da identificação de endereços em processos de usucapião"[1] no I Congresso Internacional de Regularização Fundiária & VIII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação, realizado em setembro de 2025, consolidando a transversalidade do tema inteligência artificial.

O artigo encontra-se disponível em acesso aberto pelo DOI: https://doi.org/10.5281/zenodo.17155538. O conteúdo do artigo está integralmente incorporado a esta dissertação, sendo apresentado de forma expandida e aprofundada nos capítulos de Fundamentação Teórica, Metodologia e Análise dos Resultados. O texto integral do artigo publicado é reproduzido no Apêndice B desta dissertação.

[1] Disponível em: https://doi.org/10.5281/zenodo.17155538

1.5 Estrutura da Dissertação

Esta dissertação está organizada em seis capítulos, além dos elementos pré-textuais e pós-textuais, estruturados de forma a conduzir o leitor da contextualização do problema à análise dos resultados e à conclusão. A organização adotada busca assegurar uma progressão lógica entre os fundamentos teóricos, a metodologia empregada, a implementação da solução proposta e a análise crítica dos achados da pesquisa.

O Capítulo 1 – Introdução apresenta o contexto geral da pesquisa, abordando a motivação de mercado e a motivação técnica, a definição do problema de pesquisa, o objetivo geral e os objetivos específicos, a justificativa, as contribuições do estudo e a estrutura da dissertação.

O Capítulo 2 – Fundamentação Teórica discute os principais conceitos que sustentam o trabalho, incluindo o Processamento de Linguagem Natural no domínio jurídico, a evolução das técnicas de PLN, os fundamentos dos modelos baseados em Transformers, o Reconhecimento de Entidades Nomeadas, as métricas de avaliação utilizadas e a análise de trabalhos e pesquisas relacionadas ao tema.

O Capítulo 3 – Metodologia de Pesquisa descreve o processo metodológico adotado, detalhando o fluxo da pesquisa, a construção do corpus de dados a partir do PJe do TJPE, a definição do conjunto de referência (ground truth), os modelos de linguagem avaliados, os critérios de inferência automatizada e as métricas utilizadas para a análise comparativa dos resultados. A seção 3.6 – Análise dos Resultados apresenta a análise comparativa do desempenho dos modelos, incluindo as métricas de qualidade e operacionais, a análise de erros de classificação, os testes estatísticos ANOVA e Mann-Whitney, a análise de conversibilidade dos dados e a comparação com a literatura.

O Capítulo 4 – Protótipo do Produto Desenvolvido apresenta o software implementado no contexto da pesquisa, descrevendo a arquitetura da solução, o fluxo de execução da ferramenta, a integração com APIs de LLMs e serviços de

geocodificação, bem como os mecanismos de armazenamento, normalização e comparação dos dados extraídos.

O Capítulo 5 – Limitações do Estudo discute os principais condicionantes metodológicos que devem ser considerados na interpretação dos resultados, abordando aspectos relativos à amostragem, à anotação do conjunto de referência, à geocodificação e à mensuração da latência.

O Capítulo 6 – Conclusão sintetiza os principais resultados obtidos ao longo da pesquisa, destacando os achados centrais relacionados ao desempenho dos modelos avaliados, suas implicações práticas e o atendimento aos objetivos propostos, e aponta perspectivas de trabalhos futuros em outros contextos do Poder Judiciário.

Por fim, as Referências reúnem as obras, artigos científicos, documentos normativos e demais fontes utilizadas na fundamentação teórica e metodológica da dissertação, conforme as normas vigentes. O Apêndice A – Documentação Técnica do Script de Avaliação apresenta a descrição detalhada das funções, parâmetros, dependências e estrutura de dados do script corpus_analysis_v3.py, utilizado no pipeline de avaliação NER da pesquisa, incluindo a análise do arquivo SQL_Peticao.sql que compõe o corpus primário. O Apêndice B reproduz integralmente o artigo “Tecnologia de Mineração de Texto no Direito: o Caso da Identificação de Endereços em Processos de Usucapião”, publicado no I Congresso Internacional de Regularização Fundiária & VIII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação (Recife, 2025), que constituiu a prova de conceito que originou esta dissertação.

Essa organização visa assegurar uma compreensão progressiva e integrada dos aspectos técnicos, jurídicos e sociais da pesquisa, promovendo a interlocução entre teoria e prática com vistas à aplicabilidade dos resultados no contexto do Poder Judiciário.

2 FUNDAMENTAÇÃO TEÓRICA

A extração automática de informações estruturadas em documentos jurídicos exige compreender tanto os fundamentos técnicos do PLN quanto os elementos característicos da linguagem jurídica. Como enfatizam Finatto e Macohin (2023), o Direito é uma prática intensamente mediada pela linguagem, e seus produtos — petições, despachos, sentenças e pareceres — são essencialmente textos construídos segundo convenções discursivas próprias. A análise automática desses conteúdos requer modelos capazes de lidar com estruturas complexas, termos técnicos, polissemia e variações estilísticas presentes nas práticas forenses.

Nesse contexto, modelos avançados de PLN, em especial os LLMs, têm sido aplicados a tarefas cada vez mais sofisticadas, incluindo sumarização, classificação jurídica, predição de decisões e extração de entidades (Katz et al., 2023; Modi, 2023). A utilização dessas tecnologias no Poder Judiciário integra os esforços nacionais de modernização e transformação digital, alinhados às diretrizes do CNJ.

Esta seção apresenta a base conceitual necessária para sustentar a pesquisa, abrangendo a evolução do PLN, técnicas tradicionais, transformadores, modelos de linguagem, definição formal de NER, *prompting*, e fundamentos jurídicos associados ao contexto da usucapião.

2.1 Processamento de Linguagem Natural no Domínio Jurídico

Conforme apresentado na seção anterior, o PL constitui um campo interdisciplinar voltado à análise computacional da linguagem humana. No domínio jurídico, sua aplicação envolve desafios adicionais decorrentes do vocabulário técnico especializado, das estruturas sintáticas extensas e da heterogeneidade estilística dos documentos processuais, exigindo modelos capazes de lidar com ambiguidade semântica e dependências contextuais complexas.

Frankenreiter e Nyarko (2022) observam que, embora modelos modernos sejam capazes de identificar padrões linguísticos, ainda carecem de compreensão profunda sobre argumentação jurídica, demandando supervisão humana.

2.2 Evolução das Técnicas de PLN

A evolução do PLN acompanha, de maneira direta, os avanços dos paradigmas da IA. Onde inicialmente predominavam as abordagens simbólicas, baseadas em regras explícitas e estruturas formais produzidas por especialistas.

Contudo, com o aumento da capacidade computacional e a ampliação dos *corpora* textuais disponíveis, os modelos estatísticos passaram a ocupar um papel central, possibilitando a aprendizagem de padrões linguísticos a partir da análise da frequência de palavras em grandes volumes de dados.

Essa transição também marcou uma mudança no modo de representar a linguagem. Técnicas esparsas, como *Bag-of-Words* e TF-IDF, tornaram-se amplamente utilizadas para tarefas de classificação, sumarização e recuperação da informação. Posteriormente, os métodos de representação distribuída, como *Word2Vec* e *GloVe*, permitiram a criação de vetores densos capazes de preservar relações semânticas entre palavras, constituindo a base para modelos neurais mais sofisticados.

Caseli e Nunes (2023) observam que essa trajetória histórica do PLN está alinhada à própria transformação dos paradigmas da IA ao longo das décadas. As autoras ressaltam ainda que:

> “o PLN tem acompanhado a evolução de paradigmas da IA: simbólico, estatístico e neural. Porém, diante da insuficiência de uma única abordagem, ganham espaço os paradigmas híbridos, que combinam principalmente o simbólico com um dos demais, garantindo alguma explicitação do conhecimento e, consequentemente, alguma explicabilidade dos passos seguidos pelos algoritmos” (CASELI; NUNES, 2023, p. 5).

A consolidação das arquiteturas neurais profundas culminou no surgimento dos modelos baseados em *Transformers*, os quais representam um marco significativo no

avanço das técnicas de processamento de linguagem natural e aprendizado profundo. Seu mecanismo de *self-attention* permite capturar dependências longas, paralelizar o processamento e escalar modelos para bilhões de parâmetros, tornando-os a base dos LLMs.

Esses avanços estabeleceram um novo padrão tecnológico para tarefas contemporâneas de compreensão e geração de linguagem natural.

- ***Transformers***

A principal inovação do *Transformer*, conforme descrito por Vaswani et al. (2017), reside na eliminação completa da recorrência presente nas redes neurais tradicionais, como as *Recurrent Neural Networks* (RNNs) e *Long Short-Term Memory* (LSTMs). Nessas arquiteturas recorrentes, a informação é processada de forma sequencial, *token* a *token*, o que impõe limitações à paralelização e dificulta a captura de dependências de longo alcance.

O *Transformer* substitui esse processamento temporal pelo mecanismo de *self-attention*, que permite ao modelo avaliar simultaneamente todas as relações entre os *tokens* de entrada, identificando quais partes do texto são mais relevantes para cada palavra, independentemente da distância entre elas. Essa mudança estrutural torna o treinamento altamente paralelizável e escalável, reduzindo o custo computacional e permitindo a construção de modelos mais profundos e expressivos.

Segundo Brown et al. (2020), essa arquitetura viabilizou a expansão dos grandes modelos de linguagem, como o GPT, ao permitir o aprendizado contextual em larga escala, enquanto Ariai e Demartini (2025) destacam que a atenção global proporcionada pelo *Transformer* foi determinante para a evolução de sistemas capazes de compreender e gerar linguagem com precisão quase humana. Dessa forma, a eliminação da recorrência representa um marco na engenharia dos modelos de linguagem, estabelecendo as bases para as arquiteturas modernas que impulsionam o atual estado da arte em PLN.

O *Transformer* é composto por blocos de codificadores e decodificadores empilhados, nos quais cada camada realiza operações de *multi-head* e *self-attention* em redes totalmente conectadas (*feed-forward*). Um de seus diferenciais é a capacidade de paralelizar o treinamento, tornando-se altamente escalável e eficiente em tarefas de tradução, resumo e compreensão de linguagem (Vaswani et al., 2017).

A *self-attention* permite que o modelo foque seletivamente em partes diferentes da entrada, gerando representações contextuais ricas para cada *token*. Essa abordagem facilitou a construção de modelos extremamente profundos e expressivos, pavimentando o caminho para os grandes modelos de linguagem que viriam a seguir, como GPT.

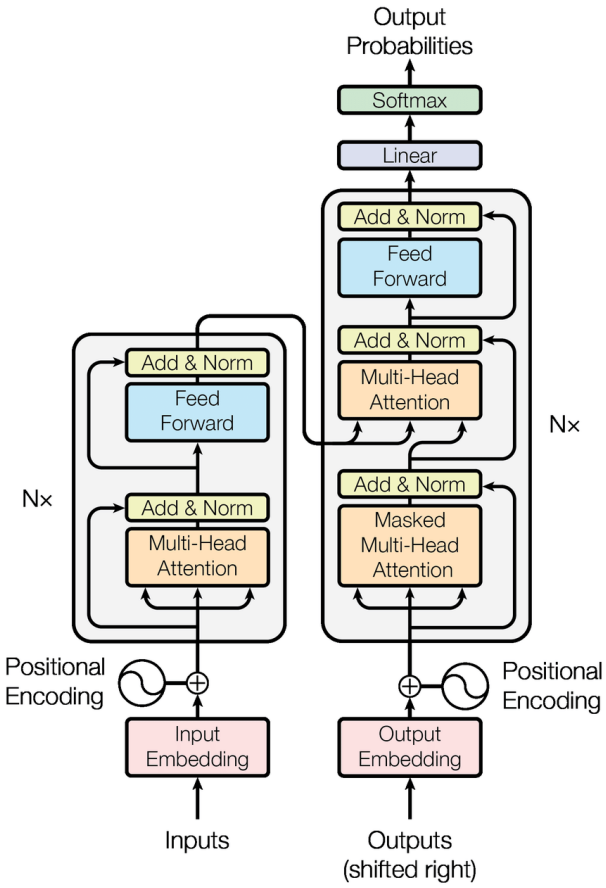
A Figura 1 apresenta a arquitetura do modelo *Transformer*, proposta por Vaswani et al. (2017), que serve de base para os LLMs. A arquitetura é composta por dois blocos principais: o *encoder* e o *decoder*, formados por camadas empilhadas repetidas N vezes.

O *encoder* é responsável por processar a sequência de entrada por meio de mecanismos de *self-attention* e redes *feed forward*, permitindo capturar dependências contextuais entre os *tokens*. O *decoder*, por sua vez, incorpora um mecanismo de *masked self-attention*, que garante a geração autorregressiva da saída, além de uma atenção cruzada com as representações do encoder.

Em ambos os blocos, os *tokens* são representados por *embeddings* combinados com codificação posicional, assegurando a preservação da ordem da sequência. As conexões residuais e a normalização (*Add & Norm*) contribuem para a estabilidade do treinamento. Ao final do processo, uma camada linear seguida da função *Softmax* gera a distribuição de probabilidades dos *tokens* de saída.

Essa arquitetura elimina a necessidade de recorrência, permitindo maior paralelização e melhor modelagem de dependências de longo alcance, sendo fundamental para o desempenho dos LLMs utilizados nesta pesquisa em tarefas de NER.

Figura 1 – Modelo Arquitetural Transformers



Fonte: Vaswani et al. (2017)

2.3 Reconhecimento de Entidades Nomeadas (NER)

O NER é uma tarefa fundamental do PLN cuja finalidade é identificar automaticamente, em textos não estruturados, segmentos linguísticos que correspondem a entidades semanticamente relevantes. Trata-se de um processo que converte expressões da linguagem natural em categorias formais, servindo como mecanismo de estruturação de dados textuais para análises posteriores.

Conforme explica o DataCamp (2023), o NER integra o campo da extração de informações e possui como função central localizar entidades nomeadas e classificá-las em categorias predefinidas, como pessoas, organizações, localizações, valores numéricos ou expressões temporais. De forma indireta, o portal enfatiza que essa tarefa atua como uma ponte entre informações textuais não estruturadas e

representações organizadas, permitindo que os dados extraídos sejam utilizados em aplicações analíticas, preditivas ou de recuperação da informação.

Por sua relevância descritiva, cabe destacar a definição apresentada pelo DataCamp, que afirma:

> “Reconhecimento de Entidades Nomeadas (NER) é uma subtarefa da extração de informações em Processamento de Linguagem Natural (PLN) que classifica entidades nomeadas em categorias pré-definidas, como nomes de pessoas, organizações, locais, códigos médicos, expressões temporais, quantidades, valores monetários e mais.” (DATACAMP, 2023, n.p.)

Essa definição, reforça que o NER atua não apenas como técnica de identificação, mas também como processo estruturante, permitindo que grandes volumes de textos sejam convertidos em dados interpretáveis por sistemas computacionais. A partir dessa perspectiva, o reconhecimento de entidades nomeadas se mostra indispensável em cenários de alta variabilidade textual, como documentos jurídicos, nos quais informações essenciais não aparecem padronizadas.

O funcionamento do NER envolve um conjunto de etapas estruturadas, entre as quais se incluem a tokenização, a identificação de candidatos a entidades, a classificação semântica, a análise contextual e o pós-processamento. Segundo o DataCamp (2023), esses componentes são determinantes para garantir que o sistema consiga interpretar relações sintáticas e semânticas presentes no texto, reduzindo ambiguidades e elevando a precisão das classificações.

Os métodos utilizados para NER evoluíram ao longo do tempo, partindo de abordagens baseadas em regras e dicionários, passando por técnicas estatísticas como HMM e CRF, aproximando-se das metodologias de aprendizado de máquina supervisionado e, mais recentemente, das arquiteturas de aprendizagem profunda, como redes neurais recorrentes e *Transformers*. A literatura também reconhece o avanço de metodologias híbridas, que combinam técnicas simbólicas e estatísticas para ampliar a abrangência e robustez da extração.

No contexto desta pesquisa, o NER desempenha um papel metodológico essencial, pois possibilita a identificação automatizada dos componentes de endereço presentes nos autos dos processos de usucapião analisados. Esse procedimento estabelece as bases estruturadas necessárias para a comparação de desempenho entre modelos de linguagem e para a avaliação das métricas propostas.

2.4 Métricas de Avaliação

Para a avaliação de modelos de IA no reconhecimento de entidades nomeadas, foram definidas métricas operacionais e de desempenho. As métricas operacionais visam avaliar a viabilidade da operação do modelo, especialmente com grandes volumes de dados. Indicadores como latência e custo operacional impactam diretamente a escolha do modelo, influenciando a eficiência e a escalabilidade do sistema em ambientes produtivos. A latência, por exemplo, é fundamental para garantir respostas rápidas, enquanto o custo operacional afeta a sustentabilidade econômica da solução (SILVA, 2023).

Por outro lado, as métricas de desempenho avaliam a precisão e a corretude dos modelos na identificação e classificação das entidades nomeadas. Elas garantem a qualidade dos dados extraídos e a eficácia da detecção das informações relevantes.

No contexto do NER, as métricas técnicas mais utilizadas são a precisão, que mede a proporção de entidades corretamente identificadas entre as previstas; a revocação, que avalia a proporção de entidades relevantes que foram efetivamente detectadas; e o F1-score, a média harmônica entre precisão e revocação, proporcionando uma visão equilibrada do desempenho do modelo (OLIVEIRA; PEREIRA, 2022).

Além disso, é importante destacar que a escolha das métricas deve considerar o equilíbrio entre desempenho técnico e impacto operacional. Melhorias em métricas como precisão e revocação nem sempre se traduzem em ganhos práticos ou econômicos, sendo necessário alinhar métricas técnicas com indicadores operacionais para garantir a viabilidade do modelo em produção (COSTA; ALMEIDA, 2024).

Para a comparação estatística entre múltiplos modelos, a literatura recomenda o uso de testes de hipóteses adequados à natureza dos dados. O teste ANOVA (Analysis of Variance) é amplamente utilizado para comparar médias de três ou mais grupos independentes, pressupondo normalidade e homocedasticidade da distribuição. Quando tais pressupostos não são satisfeitos, recomenda-se o uso de alternativas não paramétricas, como o teste de Mann-Whitney, que compara a distribuição de dois grupos independentes sem assumir normalidade (COSTA; ALMEIDA, 2024). Nesta pesquisa, ambos os testes foram aplicados para verificar a existência de diferenças estatisticamente significativas entre os modelos avaliados, adotando-se o nível de significância convencional de 5% ($\alpha = 0,05$).

2.5 Trabalhos e Pesquisas Relacionados ao Tema

De forma semelhante ao movimento observado na academia, os órgãos de direção e planejamento do Poder Judiciário brasileiro têm empreendido estudos e esforços voltados à melhoria da prestação jurisdicional. Nesse contexto, acredita-se que a IA pode contribuir significativamente para ampliar a capacidade institucional no tratamento do acervo processual e na otimização dos serviços prestados aos jurisdicionados (SOUZA; RODRIGUES, 2021, p. 11).

De acordo com Gomes et al. (2024), a adoção de soluções baseadas em IA no Judiciário visa enfrentar os desafios decorrentes do congestionamento processual, promover maior eficiência administrativa e ampliar o acesso à justiça.

Seguindo essa tendência, o uso da IA difundiu-se amplamente entre os tribunais brasileiros. A primeira fase da pesquisa “Tecnologia Aplicada à Gestão dos Conflitos no Âmbito do Poder Judiciário com Ênfase em Inteligência Artificial”, conduzida pela Fundação Getúlio Vargas em junho de 2020, identificou 72 projetos de IA em funcionamento no Poder Judiciário. Em atualização posterior, o levantamento apontou 64 projetos distribuídos em 47 tribunais, além da Plataforma Sinapses, desenvolvida pelo CNJ (SOUZA; RODRIGUES, 2021, p. 12).

Entre os agentes de Inteligência Artificial atualmente utilizados no Poder Judiciário, destacam-se:

- O sistema Victor, iniciado em 2017 por meio de parceria entre o STF e a Universidade de Brasília (UnB), é baseado em técnicas tradicionais de Processamento de Linguagem Natural e Aprendizado de Máquina, como vetorização textual (*bag-of-words* e TF-IDF) e classificadores supervisionados (regressão logística, SVM). Seu objetivo é aumentar a celeridade e a precisão na classificação de peças processuais e temas de repercussão geral.

- O sistema Athos, do STJ, utiliza algoritmos de Aprendizado de Máquina supervisionado aplicados ao PLN, com predominância de abordagens clássicas de classificação textual. Entre as técnicas empregadas destacam-se a vetorização de textos (como *bag-of-words* e TF-IDF) e classificadores supervisionados, a exemplo de regressão logística, máquinas de vetores de suporte (SVM) e modelos estatísticos equivalentes. Esses algoritmos são aplicados à identificação de temas repetitivos, classificação de documentos e apoio à gestão de precedentes.

- O Sinapses, plataforma do CNJ, é um ambiente para desenvolvimento e compartilhamento de modelos de Inteligência Artificial no Judiciário, baseado em Aprendizado de Máquina supervisionado e PLN. A plataforma suporta algoritmos de classificação e regressão, com uso de técnicas como vetorização textual (*bag-of-words* e TF-IDF) e classificadores supervisionados (por exemplo, regressão logística, SVM e modelos estatísticos equivalentes), com foco na padronização, governança e apoio à decisão humana.

No âmbito estadual, destaca-se:

- O sistema ELIS, desenvolvido pelo TJPE), emprega algoritmos de Aprendizado de Máquina supervisionado aplicados ao PLN, com predominância de técnicas clássicas de classificação textual, como vetorização de textos (*bag-of-words* e TF-IDF) e classificadores supervisionados, a exemplo de regressão logística, máquinas de vetores de suporte (SVM) e modelos estatísticos equivalentes.
- O ARENITA, do TJPB, é uma iniciativa de IA aplicada ao apoio à atividade jurisdicional, voltada principalmente à triagem, classificação e análise de documentos processuais, caracterizado como uma solução baseada em PLN e Aprendizado de Máquina supervisionado, empregando técnicas clássicas de

classificação textual para apoiar tarefas como identificação de classes processuais, assuntos, peças relevantes e organização do fluxo processual, sem substituir a decisão humana.

O artigo “Inteligência Artificial, Gestão de Documentos e Vulnerabilidade Socioambiental: Experiências, Desafios e Tendências para Regularização Fundiária no Brasil”, apresentado no I Congresso Internacional de Regularização Fundiária & VIII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação (Recife/PE, 2025), é diretamente relevante para esta pesquisa. O estudo parte da constatação de que a regularização fundiária no Brasil produz uma volumosa base de dados jurídico-agrários, cuja autenticação, preservação e interoperabilidade são determinantes para a segurança da posse e a mitigação de vulnerabilidades socioambientais (Arana et al., 2025).

Os autores adotam uma abordagem interdisciplinar, integrando Arquivologia, Ciência da Informação, Computação e Ciências da Terra, e relatam uma experiência prática no âmbito do projeto TED INCRA/UFPR, na qual técnicas de classificação textual supervisionada (TF-IDF e regressão logística) alcançaram acurácia de 95% e F1-score de 0,94. Os resultados demonstram a viabilidade de soluções híbridas, combinando curadoria humana e IA, para aprimorar a prova documental, reduzir a morosidade administrativa e evitar sobreposições de registros fundiários, destacando a necessidade de governança, ética e conformidade com a LGPD.

O trabalho também discute riscos associados à transformação digital, apontando que a ausência de políticas públicas, infraestrutura tecnológica e mecanismos de preservação pode intensificar desigualdades fundiárias. Nesse contexto, propõe a adoção de auditoria algorítmica e repositórios arquivísticos digitais, como o Hipátia (IBICT), para assegurar autenticidade, integridade e rastreabilidade documental.

Conclui-se que o estudo de Arana et al. (2025) oferece uma perspectiva complementar à presente dissertação, ao evidenciar empiricamente o potencial da IA para a eficiência processual, a qualidade probatória e a redução da vulnerabilidade socioambiental na gestão documental fundiária.

3 METODOLOGIA DE PESQUISA

A relevância dessa tarefa transcende a simples extração de texto, pois para a conversão dos dados de endereço em coordenadas de geolocalização por ferramentas como *Google Maps* e a biblioteca *Python GeoPandas* exige-se que o endereço seja o mais completo e preciso possível. Endereços incompletos ou mal formatados podem resultar em imprecisões significativas, impossibilitando a localização ou, pior, indicando um local incorreto do imóvel.

O objetivo metodológico é determinar a eficácia comparativa dos modelos e validar a viabilidade operacional da ferramenta no contexto jurídico real.

3.1 Privacidade dos Dados e Conformidade com a LGPD

Esta pesquisa utiliza petições iniciais de ações de usucapião provenientes do TJPE. Os documentos são submetidos a modelos de LLMs. O objetivo é exclusivamente a extração e a estruturação de informações de endereços de imóveis. Não há qualquer finalidade decisória.

A aplicação da IA adotada neste estudo é de natureza instrumental. Os LLMs atuam apenas como ferramenta de apoio analítico. Não ocorre substituição da atividade jurisdicional humana. Tampouco há interferência no mérito dos processos.

Do ponto de vista jurídico, o tratamento dos dados pessoais encontra respaldo na LGPD. A base legal adotada é o cumprimento de obrigação legal e a execução de políticas públicas, conforme os artigos 7º, inciso II, e 23 da Lei nº 13.709/2018 (BRASIL, 2018). Os dados analisados já integram regularmente o PJe.

Conforme argumenta Freitas (2025), a LGPD não impede o uso de sistemas de IA pelo Poder Judiciário. A compatibilidade é possível quando são observados os princípios da finalidade, da necessidade, da segurança e da transparência. Também é essencial a manutenção do controle humano significativo sobre os resultados produzidos por sistemas automatizados

No âmbito metodológico, o tratamento dos documentos foi limitado ao estritamente necessário. A análise concentrou-se apenas nos trechos textuais relacionados à localização do imóvel. Dados sensíveis irrelevantes ao escopo da pesquisa não foram explorados. Essa abordagem está alinhada aos princípios da minimização e da adequação previstos no artigo 6º da LGPD.

Os resultados gerados pelos LLMs foram submetidos à validação humana. Essa etapa permitiu verificar a consistência das informações extraídas. Também contribuiu para mitigar riscos associados à automação. Dessa forma, preservou-se a rastreabilidade do tratamento dos dados.

A metodologia adotada observa, ainda, as diretrizes da Resolução CNJ nº 615/2025. Essa norma estabelece parâmetros para o uso ético, seguro e responsável da IA no Poder Judiciário (CNJ, 2025b). A pesquisa se enquadra no uso de IA como apoio à gestão da informação judicial.

No contexto do TJPE, a extração automatizada de endereços em ações de usucapião possui relevância prática. A técnica contribui para a organização do acervo processual. Também viabiliza a integração com bases de dados georreferenciadas. Além disso, oferece suporte técnico a políticas públicas de regularização fundiária.

Dessa forma, o uso de LLMs nesta pesquisa mostra-se juridicamente legítimo. A abordagem é metodologicamente adequada. A técnica é compatível com a LGPD e com a governança de IA estabelecida pelo CNJ. Ao mesmo tempo, promove ganhos de eficiência analítica no tratamento de dados processuais do TJPE.

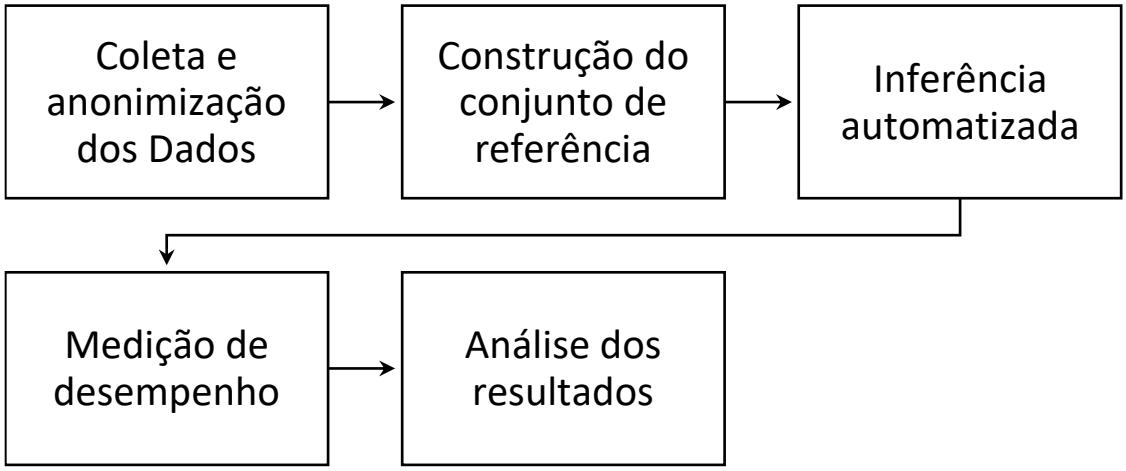
3.2 Processo de Pesquisa

O macrofluxo das etapas realizadas no processo desta pesquisa é apresentado na Figura 2, a qual sintetiza a sequência metodológica empregada. Na primeira etapa, procedeu-se à extração de dados da base do PJe, considerando os processos da classe Usucapião distribuídos no período de 02 de janeiro a 31 de dezembro de 2024, totalizando 2.587 registros. A partir desse conjunto, foram selecionados aleatoriamente 100 processos para compor a amostra de análise.

Na segunda etapa, os 100 processos selecionados foram analisados com o objetivo de identificar e registrar os endereços dos imóveis constantes nos autos. Esses endereços serviram como referência comparativa para a avaliação dos resultados obtidos por meio da inferência automatizada dos LLMs.

Em seguida, os processos foram submetidos ao algoritmo de análise automatizada desenvolvido nesta pesquisa, o qual processou os textos e extraiu os endereços correspondentes em três modelos distintos de linguagem natural previamente definidos. Os resultados foram armazenados em uma base tabular estruturada, possibilitando o cálculo das métricas qualitativas e quantitativas adotadas, incluindo os indicadores de desempenho de extração e as estatísticas descritivas de acurácia e eficiência. Esse procedimento permitiu comparar, de forma objetiva, o desempenho das LLMs frente à extração manual, validando os parâmetros definidos na metodologia proposta.

Figura 2 - Macrofluxo do processo de pesquisa



Fonte: Autor (2026).

O processo é caracterizado da seguinte forma:

- Coleta e Anonimização dos Dados

Os documentos extraídos foram armazenados em uma estrutura de banco de dados independente, organizada segundo os metadados relevantes à mensuração da eficácia do processo de extração automatizada.

A anonimização dos documentos processuais seguiu protocolo padronizado, consistindo nas seguintes etapas: (i) substituição dos nomes das partes (autor, réu e representantes legais) por identificadores genéricos (ex.: "PARTE A", "PARTE B"); (ii) supressão de números de CPF, RG e outros documentos de identificação pessoal; (iii) remoção de dados de contato como telefone e e-mail; e (iv) preservação integral dos dados de endereço do imóvel, uma vez que esses constituem o objeto de extração da pesquisa e não caracterizam dado pessoal sensível no contexto de processo judicial público. O protocolo foi aplicado de forma manual pelo autor, com revisão de uma amostra aleatória de 10% dos documentos para verificação de completude, em conformidade com os princípios da minimização e da necessidade previstos no art. 6º da LGPD (BRASIL, 2018).

Os arquivos processuais encontram-se em formato textual e em linguagem natural, conforme ilustrado na Figura 3, estando disponíveis na base de dados do PJe nos formatos PDF e HTML. Essa característica reforça o desafio técnico de processamento de linguagem natural, uma vez que os conteúdos são heterogêneos, apresentam estrutura textual livre e ausência de padronização entre os autos.

Conforme evidenciado nas Figuras 3(a), 3(b) e 3(c), os endereços dos imóveis estão distribuídos nos textos das peças processuais sem formatação específica ou estrutura delimitada, o que dificulta a identificação automatizada por métodos tradicionais. A estrutura que melhor expressa o modelo ideal de estrutura de endereço é o apresentado nas Figuras 3(a), 3(b) e 3(c), onde os elementos que compõem a identificação da localização se encontram completos.

Esse contexto constitui o ambiente experimental de análise dos LLMs, permitindo avaliar sua capacidade de compreender e extrair informações jurídicas não estruturadas em comparação com a identificação manual realizada.

Figura 3(a) - Trecho de Petição Inicial de Processo de usucapião.

I. DOS SUBSTRATOS FÁTICOS

Ab initio, cumpre trazer à baila as circunstâncias fáticas consumadas até aqui e que, consequentemente, fundamentam o direito em questão:

O imóvel objeto desta usucapião encontra-se em Tamandaré -PE, Rua do Forte, Nº 10, LT. 10, Q E, centro, em zona urbana. O valor venal constante no Cadastro Imobiliário da Prefeitura do Município em questão de [redacted] conforme DOC. anexo.

O autor é filho do Senhor [redacted] comprador legal do referido imóvel em comento conforme instrumento Particular de compra e venda tendo adquirido o referido imóvel dos senhores [redacted], tendo o Sr. [redacted] adquirido da [redacted] conforme certidão de inteiro teor.

Figura 3(b).

Em face de aquisição de uma casa, edificada em alvenaria, situada na Rua Antônio Antunes nº 64, Bairro Alto do Rio Branco, nesta Cidade de Sertânia-PE, **CEP: 56.600-000**, esse imóvel foi edificado, **há mais de 40 (quarenta) anos**, estando na posse da Autora, há mais de **25 (vinte e cinco) anos**, época do falecimento de sua mãe [redacted], ocorrido no 01 de março de 1996, desde esse tempo que reside neste im óvel, limitando na **Frente**: com o leito da Rua Antônio Antunes, aos **Fundos**: com o loteamento de Severino Veras, **Lado Direito**: com a casa nº 120, **Lado Esquerdo**: com a casa nº 110, área total de 124,30m², área construída 63,30m², composta de 01 (uma) área, na entrada da casa, 01 (uma) sala, 03 (três) quartos, 01 (uma) cozinha, 01 (um) banheiro, e, um quintal todo cercado, na época o imóvel foi adquirido a do [redacted].

Figura 3(c).

AÇÃO DE USUCAPIÃO EXTRAORDINÁRIA

Do imóvel situado na rua jornalista Célio Meira, nº 341, bairro Cajá, na cidade de Vitória de Santo Antão/PE, CEP 55610-065, em desfavor de [redacted], portador do CPF [redacted] residente e domiciliado no Loteamento [redacted] nesta cidade, o que faz com os fundamentos de fato e de direito adiante expostos. Pelos fundamentos de fato e de direito a seguir expostos.

Fonte: Autor (2026).

- Construção do Conjunto de Referência

O conjunto de referência foi constituído a partir da seleção aleatória de 100 processos, realizada pelo autor sem aplicação de critérios específicos de inclusão ou exclusão, servindo apenas para representar uma amostra inicial do acervo analisado.

O processo de validação manual consistiu nas seguintes etapas: (i) o anotador acessou cada processo de usucapião no sistema PJe-TJPE individualmente; (ii) realizou leitura integral da petição inicial, identificando manualmente os componentes do endereço do imóvel (logradouro, número, bairro, cidade, estado e CEP) conforme

o texto original; (iii) registrou os valores identificados em planilha estruturada, que serviu como base de comparação (ground truth); e (iv) nos casos de endereço ausente ou incompleto no documento, o campo correspondente foi registrado como nulo. Os critérios de anotação foram: prevalecer sempre o dado explicitamente mencionado no texto da petição, sem inferências ou complementações externas; e, em caso de endereços múltiplos no mesmo documento, considerar o endereço do imóvel objeto da ação, descartando endereços de partes, testemunhas ou outros imóveis citados. Esses critérios foram aplicados de forma uniforme a todos os 100 processos da amostra.

A validação manual dos endereços foi realizada por um único anotador, prática recorrente em estudos exploratórios e aplicados de NER quando há definição prévia de critérios objetivos de anotação. Considerando que o objetivo da validação foi exclusivamente a construção do *ground truth* para avaliação comparativa dos modelos, e não a produção ou ajuste dos dados extraídos, a ausência de múltiplos anotadores não compromete a consistência dos resultados, uma vez que todos os registros foram avaliados sob os mesmos critérios e condições.

Reconhece-se, contudo, que a ausência de múltiplos anotadores e de uma medida de concordância inter-anotador (como o Kappa de Cohen) constitui uma limitação metodológica do estudo, especialmente para os campos de endereço que admitem mais de uma interpretação válida (como bairros com variações nominais). Essa limitação é discutida na seção de Considerações Finais e deve ser endereçada em trabalhos futuros por meio da constituição de um painel de anotação com pelo menos dois avaliadores independentes e cálculo formal do índice de concordância.

O conjunto de referência (*ground truth*), previamente construído e validado, foi utilizado como base de comparação direta para a avaliação do desempenho dos modelos de linguagem, permitindo o cálculo objetivo das métricas de precisão, revocação e F1-score para cada componente do endereço extraído.

Para assegurar uma análise contextualmente precisa, o *corpus* de texto é composto exclusivamente por documentos jurídicos autênticos de processos de usucapião. Esta seleção direcionada é vital, pois a linguagem, a estrutura e a

terminologia presentes nesses documentos são altamente especializadas e distintas de outros domínios textuais.

Para garantir a máxima qualidade e consistência, foi utilizada uma ferramenta de anotação desenvolvida pelo autor, que facilita o processo de identificação, registro e comparação dos dados. Um alto índice de concordância válida a clareza da definição de entidade adotada e a consistência da marcação humana, assegurando a confiabilidade do *ground truth*.

- Inferência Automatizada Utilizando *ChatGPT*, *Gemini* e *DeepSeek*

Esta etapa constituiu o núcleo experimental da pesquisa, onde se avaliou a capacidade de três modelos de LLMs *ChatGPT, Gemini e DeepSeek*, para extrair endereços completos de documentos jurídicos de usucapião do sistema PJe-TJPE.

A seleção dos três modelos de linguagem de grande escala *ChatGPT, Gemini e DeepSeek*, para compor o núcleo experimental desta pesquisa fundamentou-se em critérios técnicos, metodológicos e de relevância no cenário atual de soluções baseadas em inteligência artificial. A opção por múltiplos modelos buscou garantir uma análise comparativa abrangente, capaz de refletir diferentes arquiteturas, abordagens e níveis de maturidade tecnológica presentes no ecossistema contemporâneo de LLMs.

Primeiramente, adotou-se como critério a representatividade no estado da arte, selecionando modelos amplamente reconhecidos pelo desempenho em tarefas de processamento de linguagem natural, especialmente aquelas voltadas à compreensão textual e à extração de informações estruturadas. O ChatGPT consolidou-se como referência acadêmica e industrial; o Gemini destaca-se pela integração com ecossistemas de busca e capacidades multimodais; e o DeepSeek foi incluído por representar uma abordagem emergente orientada à eficiência computacional e ao baixo custo por *token*, aspecto relevante para aplicações em larga escala no setor público.

Também se considerou o critério de acessibilidade e disponibilidade prática, selecionando modelos que oferecem interfaces de uso estáveis, documentação consolidada e facilidade de integração em fluxos de trabalho experimentais.

Por fim, a seleção dos três modelos buscou refletir um cenário realista de tomada de decisão institucional, oferecendo uma análise comparativa que dialoga diretamente com os desafios concretos da Justiça brasileira na adoção de tecnologias de IA. Assim, a escolha dos modelos foi orientada por critérios de relevância tecnológica, aplicabilidade prática e viabilidade de uso, permitindo que a avaliação fosse abrangente e alinhada ao objetivo central desta pesquisa: verificar a capacidade das LLMs de extrair endereços completos de documentos judiciais de usucapião do PJe de forma acurada, eficiente e replicável.

A tarefa de inferência exigiu a criação de *prompts* otimizados com instruções explícitas para o reconhecimento da entidade ENDEREÇO_IMÓVEL. Cada modelo recebeu entradas padronizadas em formato *JSON*, contendo os campos esperados (rua, número, bairro, cidade, estado, CEP). Essa abordagem, denominada *zero-shot prompting*[2] , foi selecionada para testar a capacidade intrínseca de generalização dos modelos, sem treinamento adicional no domínio jurídico.

[2] Segundo DataCamp (2024), zero-shot prompting é uma técnica na qual um modelo de IA recebe uma tarefa sem nenhum exemplo prévio ou treinamento específico sobre essa tarefa.

- Medição de Desempenho

As métricas operacionais avaliam o comportamento computacional dos modelos, considerando aspectos como tempo de resposta, custo de execução e viabilidade de uso em larga escala.

A seguir, detalharemos cada uma das métricas utilizadas e sua operacionalização na ferramenta computacional de avaliação.

I. Latência

A latência mede o tempo que leva desde o momento em que a requisição é enviada para a API do LLM até o momento em que a resposta completa é recebida. É essencialmente o "tempo de espera" para obter a resposta do modelo medido em segundos.

Os experimentos foram executados em um computador com processador Intel Core i7-12700H, 16 GB de memória RAM DDR5 e sistema operacional Windows 11 Home, conectado à internet via rede cabeada com banda de 300 Mbps. Cabe ressaltar que a latência medida nesta pesquisa reflete o tempo total de resposta da API, o qual é influenciado por fatores como carga nos servidores do provedor, congestão de rede e tamanho do documento processado, e não apenas pelo tempo de inferência do modelo. Essa limitação deve ser considerada na interpretação comparativa dos resultados de latência entre os modelos.

Para o cálculo da latência registra-se o tempo inicial antes da chamada à API do LLM. Após a resposta ser recebida, o tempo final é registrado. A latência é então calculada como a diferença, apresentada conforme a Equação 1:

$$\text{latência} = \text{tempo final} - \text{tempo inicial} \qquad \text{Eq. 1}$$

Em situações de processamento em lote, latências menores significam que mais documentos podem ser processados em menos tempo, otimizando a eficiência operacional.

II. Custo Financeiro

O custo financeiro mede o valor monetário gasto ao interagir com a API do LLM. As APIs de LLMs geralmente cobram com base no uso de *tokens* – tanto os *tokens* enviados no *prompt* (entrada) quanto os *tokens* gerados na resposta (saída).

O custo é medido pela informação contida no objeto *usage* da resposta da API do LLM. Este objeto contém *prompt_tokens* (tokens de entrada) e *completion_tokens*

(tokens de saída). O cálculo utiliza taxas de referência por milhão de *tokens* para entrada (*input_cost_per_million*) e saída (*output_cost_per_million*), conforme valores especificados na Tabela 1. A fórmula de cálculo é apresentada conforme a definida pela Equação 2:

$$\text{Custo} = \left(\frac{PT}{1.000.000} * TE\right) + \left(\frac{CT}{1.000.000} * TS\right) \qquad \text{Eq. 2}$$

Onde:

PT: *Prompt_Tokens*

TE: Taxa_Entrada

CT: *Completion_Tokens*

TS: Taxa_Saída

A avaliação dos custos é fundamental para gerenciar os gastos com APIs, especialmente em projetos que envolvem um grande volume de requisições, e principalmente em órgãos públicos onde é essencial a racionalização e o controle do erário.

Os valores implementados nesta pesquisa foram coletadas no site da empresa responsável pelo LLM (com data de referência de 02 de agosto de 2025), apresentados na Tabela 1, com valores expressos em dólar americano.

Tabela 1 - Custo[3] em dólar dos modelos utilizados

Modelo	Custo de Entrada (US$)	Custo de Saída (US$)
ChatGPT	0,60	2,40
Gemini	0,35	1,05
DeepSeek	0,27	1,10

Fonte: Autor (2026)

[3] As versões específicas utilizadas nos experimentos foram: ChatGPT (gpt-4o-mini-2024-07-18), Gemini (gemini-1.5-flash) e DeepSeek (deepseek-chat v2). Os valores de custo foram coletados nos respectivos sites oficiais de cada provedor em 02 de agosto de 2025.

- Medição de Qualidade

As métricas de qualidade avaliam a correção das extrações realizadas pelos modelos, mensurando o grau de correspondência entre as entidades identificadas e o ground truth estabelecido.

I. Precisão

A precisão mede a proporção de identificações positivas corretas (componentes de endereço extraídos corretamente) em relação ao total de identificações positivas feitas pelo modelo (todos os componentes de endereço que o modelo extraiu, corretos ou incorretos). Em outras palavras, "dos que o modelo disse que eram, quantos realmente eram?".

O cálculo desta métrica é obtido pela Equação 3:

$$\text{Precisão} = \frac{TP}{TP+FP} \qquad \text{Eq. 3}$$

Onde:

TP: *True_Positive* (Verdadeiros Positivos)

FN: *False_ Positive* (Falsos Positivos)

II. Revocação

A revocação mede a proporção de identificações positivas corretas (componentes de endereço extraídos corretamente) em relação ao total de identificações positivas que “deveriam” ter sido feitas (todos os componentes de endereço presentes no *ground truth*). Em outras palavras, "dos que realmente eram, quantos o modelo conseguiu encontrar?".

O cálculo desta métrica está expresso na Equação 4:

$$\text{Revocação} = \frac{TP}{TP+FN} \qquad \text{Eq. 4}$$

Onde:

TP: *True_Positive* (Verdadeiros Positivos)

FN: *False_Negative* (Falsos Negativos)

III. F1-Score

O *F1-Score* consiste na média harmônica entre as métricas de Precisão e Revocação, sintetizando em um único valor o equilíbrio entre ambas. Essa medida é particularmente relevante em cenários nos quais é necessário simultaneamente garantir a correta identificação dos elementos relevantes (alta precisão) e a capacidade do modelo de recuperar a totalidade dessas ocorrências (alta revocação). Assim, um *F1-Score* elevado indica que o modelo apresenta desempenho robusto, conciliando níveis satisfatórios de precisão e revocação na tarefa avaliada.

A Equação 5 demonstra a expressão de cálculo da métrica:

$$\text{F1-Score} = 2 * \left(\frac{Precisão * Revocação}{Precisão + Revocação}\right) \qquad \text{Eq. 5}$$

IV. Acurácia

A acurácia mede a proporção total de classificações corretas realizadas pelo modelo em relação ao conjunto completo de entidades avaliadas, considerando tanto as identificações positivas corretas quanto as identificações corretas de ausência de entidades. Em outras palavras, responde à questão: *"do total de decisões tomadas pelo modelo, quantas estavam corretas?"*

O cálculo desta métrica é obtido pela Equação 6:

$$\text{Acurácia} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Eq. 6}$$

Onde:

TP: *True_Positive* (Verdadeiros Positivos)
FN: *False_Negative* (Falsos Negativos)
TN: *True_Negative* (Verdadeiros Negativos)
FN: *False_Negative* (Falsos Negativos)

- Processo de Medição

O processo de medição visa assegurar que os componentes mínimos do endereço sejam extraídos de forma a permitir sua conversão em coordenadas geográficas e o efetivo georreferenciamento do imóvel. Para isso, os modelos são avaliados a partir de métricas de desempenho operacional — latência e custo

financeiro — e de métricas de qualidade — Precisão, Revocação, F1-Score e acurácia.

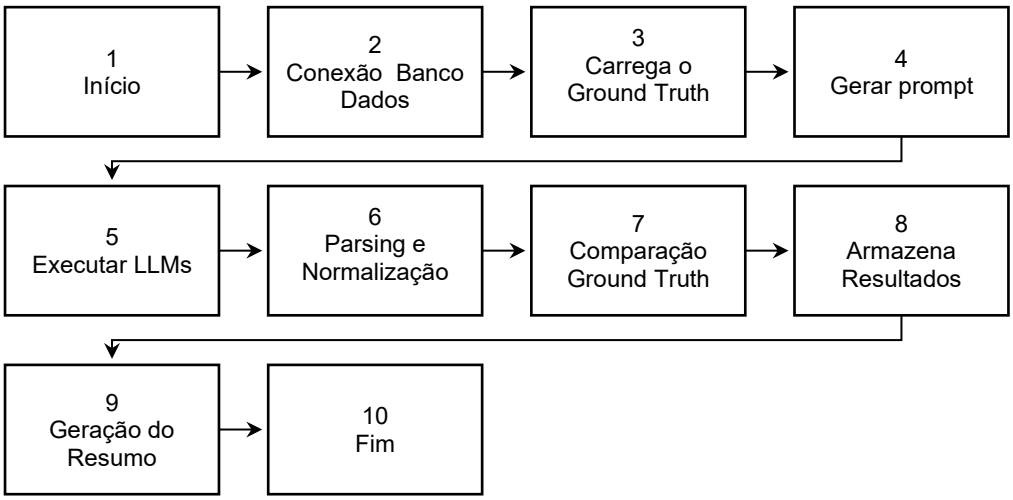
Cada entidade de endereço (tais como logradouro, número ou CEP) será considerada corretamente identificada quando apresentar correspondência exata com o valor previamente definido no *ground truth*. Nos casos em que a entidade não estiver presente no documento original, a identificação somente será classificada como correta se o modelo indicar explicitamente a sua ausência. Qualquer divergência em relação ao ground truth, seja por extração incorreta ou por omissão não justificada, implicará a classificação da entidade como incorreta, uma vez que afeta a consistência semântica dos dados utilizados nas etapas subsequentes de geolocalização.

3.3 Ferramenta de Inferência da Pesquisa

O *software* implementa um *pipeline* completo de NER com três Modelos de Linguagem (LLMs), *ChatGPT, Gemini e DeepSeek* voltado à extração automática de endereços de imóveis em petições de usucapião.

O processamento automatizado desenvolvido nesta pesquisa segue um fluxo composto por dez etapas que vão da a inicialização dos modelos de linguagem até a consolidação estatística dos resultados. Cada etapa foi estruturada para garantir padronização, rastreabilidade e reprodutibilidade dos experimentos. Conforme mostrado na Figura 4.

Figura 4 - Fluxo de Execução da Ferramenta de Pesquisa



Fonte: Autor (2026).

- **Etapa 1 – Inicialização dos Modelos de Linguagem (LLMs)**

O sistema inicia configurando os clientes responsáveis por acessar cada modelo de linguagem:

  - *ChatGPT*, via API da *OpenAI*;
  - *Gemini*, por meio da classe *google.generativeai.GenerativeModel();*
  - *DeepSeek*, utilizando o *endpoint* específico (base_url="https://api.deepseek.com/v1").

Essas instâncias são armazenadas em um dicionário unificado (LLM_CONFIGS), permitindo gerenciamento centralizado e chamadas paralelas. **Resultado:** ambiente inicial configurado, apto a executar inferências uniformes entre os diferentes modelos.

- **Etapa 2 – Conexão com o Banco de Dados**

A aplicação abre a conexão com o repositório onde está armazenado o *corpus* de documentos da pesquisa. Após o carregamento, todas as operações são tratadas com fechamento seguro da sessão, garantindo integridade transacional. **Resultado:** acesso controlado ao conjunto de dados utilizado como base experimental.

- **Etapa 3 – Carrega o Ground Truth**

Os registros do banco — contendo os campos id, rua, número, bairro, cidade, estado, cep e ds_modelo_documento — são lidos e transformados em objetos estruturados contendo:

  - o texto jurídico original do documento;
  - o endereço validado manualmente (*ground truth*);
  - metadados associados.

Cada item é armazenado em memória para processamento subsequente. **Resultado:** corpus estruturado, com textos e endereços previamente validados para comparação.

- **Etapa 4 – Geração do *Prompt* Jurídico Padronizado**

Para garantir consistência entre os modelos, o sistema gera um *prompt* contendo instruções explícitas sobre como extrair o endereço e o formato JSON requerido, incluindo campos obrigatórios (rua, número, bairro, cidade, estado e cep). O *prompt* também fornece exemplos e orientações de conformidade.

Resultado: comando padronizado capaz de induzir respostas comparáveis entre os diferentes LLMs.

O texto completo dos prompts utilizados para cada modelo de linguagem, incluindo as instruções de extração, os exemplos de saída esperada e as orientações de conformidade, encontra-se reproduzido integralmente no Apêndice A deste documento. Adicionalmente, os códigos-fonte, os parâmetros das funções utilizadas nos scripts de avaliação e a documentação técnica complementar estão disponíveis no repositório digital da pesquisa (ver nota de rodapé 3), permitindo a reprodução integral dos experimentos.

- **Etapa 5 – Execução das Inferências nos LLMs**

Cada modelo é acionado individualmente. Durante a execução, o sistema registra:

  - o texto retornado pelo modelo;
  - o custo estimado da requisição;
  - o tempo de latência entre envio e resposta.

Um mecanismo de tratamento de erros detecta falhas de quota, formatação ou conexão, reenviando automaticamente a requisição quando necessário. **Resultado:** conjunto de respostas padronizadas, com métricas de custo e tempo associadas a cada inferência.

Os principais parâmetros utilizados nas chamadas às APIs foram padronizados para garantir comparabilidade entre os modelos: temperatura (temperature) = 0,0 para maximizar o determinismo das respostas; número máximo de tokens de saída (max_tokens) = 500; e formato de resposta definido como JSON estruturado. O parâmetro de temperatura igual a zero elimina a aleatoriedade na geração do texto, assegurando que a mesma entrada produza sempre a mesma saída, o que é

essencial para a reprodutibilidade dos experimentos. Esses parâmetros foram aplicados de forma idêntica nos três modelos avaliados. A documentação completa dos parâmetros e dos scripts de avaliação encontra-se disponível no repositório da pesquisa (DOI: https://doi.org/10.5281/zenodo.17155538).

- **Etapa 6 – *Parsing* e Normalização do JSON Retornado**

A resposta do modelo é analisada para identificar o bloco JSON, mesmo quando encapsulado em *Markdown*. O sistema:

  - extrai o conteúdo;
  - normaliza nomes de campos;
  - corrige variações de formatação;
  - remove caracteres não numéricos do CEP;
  - registra eventuais erros de formatação.

Resultado: dicionário limpo e estruturado contendo os seis componentes normalizados do endereço.

- **Etapa 7 – Comparação com o *Ground Truth***

Cada componente extraído (rua, número, bairro, cidade, estado, cep) é comparado diretamente com o ground truth após normalização semântica. O sistema calcula:

  - Precisão
  - Revocação
  - F1-Score
  - Acurácia

O algoritmo trata automaticamente equivalências regionais, como “PE” ↔ “Pernambuco”.

Resultado: métricas detalhadas por documento e por modelo, além de diagnóstico das divergências encontradas.

- **Etapa 8 – Armazenamento dos Resultados Individuais e Agregados**

  O sistema grava cada análise na tabela, incluindo:
  - endereço original e extraído;
  - métricas individuais;
  - custo e latência da requisição;
  - timestamp da execução.

Em seguida, consolida os resultados gerais de cada modelo armazenando médias, totais e indicadores globais (acurácia, precisão, revocação, F1-score, latência média e custo médio).

Resultado: rastreabilidade completa de todas as execuções e histórico comparativo entre os modelos.

- **Etapa 9 – Geração do Resumo Estatístico Final**

A aplicação exporta todo o dicionário de resultados para o arquivo **analysis_results.json**, preservando acentuação UTF-8 e a estrutura hierárquica do experimento.
Resultado: arquivo consolidado pronto para auditoria, replicação e visualização em dashboards de análise.

- **Etapa 10 - Encerramento do Fluxo**

O processo é finalizado com o registro de logs automáticos — armazenados em banco de dados relacional — garantindo controle de versão, auditabilidade e reprodutibilidade integral dos experimentos.

3.4 Formato de Saída Esperado

Para viabilizar a validação da extração e a integração com ferramentas de geolocalização, a saída dos LLMs para a entidade ENDEREÇO_IMÓVEL será estruturada em formato *JSON* e armazenada em tabela de banco de dados. Esse formato permite a análise comparativa dos resultados, a geração de tabelas e gráficos sumarizados e a avaliação individual dos componentes do endereço (rua, número, bairro, cidade, estado, CEP e complemento, quando aplicável), cuja integridade é essencial para a precisão do georreferenciamento. A estrutura *JSON* é apresentada na Figura 5.

Figura 5 – Exemplo de saída em formato JSON

```
{
   "text_original": "Rua da Aurora, 456, apto 101, Boa Vista, Recife, Pernambuco, 50030-000",
    "rua": "Rua da Aurora",
    "numero": "456",
    "complemento": "apto 101",
     "bairro": "Boa Vista",
     "cidade": "Recife",
     "estado": "PE",
     "cep": "50030000"
   }
```

Fonte: Autor (2026).

3.5 Reconhecimento de Entidades Nomeadas e sua Avaliação Qualitativa

A avaliação qualitativa de modelos em tarefas de NER baseia-se tipicamente na identificação de categorias textuais e na comparação entre entidades previstas e entidades verdadeiras.

No caso específico desta pesquisa, cada componente do endereço é tratado como uma “entidade”, permitindo que as métricas tradicionais sejam aplicadas naturalmente.

Contudo, a métrica de Verdadeiros Negativos (TN) não é usual nesta tarefa, pois não existe uma lista explícita de “não entidades”. Isso significa que TN não contribui para o cálculo das métricas adotadas.

Todas as métricas foram calculadas sobre uma mesma base de comparação:

- Cada documento produz até seis comparações (uma para cada campo do endereço).
- A soma dessas comparações fornece o total de ocorrências avaliadas para cada modelo.

Sendo assim, definem-se:

a) Verdadeiros Positivos (TP) - Campo corretamente identificado pelo modelo com o valor exato do ground truth.

Exemplo:

Ground truth: "Rua da Paz" → Modelo: "Rua da Paz" → TP

b) Falsos Positivos (FP) - Campo identificado pelo modelo, porém com valor errado ou que não existe no documento.

Exemplo:

Ground truth: "Rua da Paz" → Modelo: "Rua da Esperança" → FP

c) Falsos Negativos (FN) - Campo presente no ground truth, mas não identificado pelo modelo.

Exemplo:

Ground truth: nº 55

Modelo: ‘‘número não encontrado’’, campo vazio ou omisso → FN

d) Verdadeiros Negativos (TN) - Não aplicável ao problema, pois não há categorias explícitas de “não endereço” a serem mapeadas.

Em decorrência da inaplicabilidade do TN ao contexto deste estudo, o cálculo da Acurácia (Equação 6) adota TN = 0, o que implica que essa métrica mensura exclusivamente a proporção de campos corretamente identificados em relação ao total de campos avaliados, sem considerar a correção da ausência de entidades. Essa

restrição deve ser levada em conta na interpretação dos valores de acurácia reportados.

3.6 Análise dos Resultados

Nesta etapa são descritos os procedimentos adotados para a análise comparativa do desempenho dos modelos avaliados. A análise baseia-se na aplicação de métricas definidas anteriormente e testes estatísticos definidos, com o objetivo de garantir consistência metodológica e permitir a comparação objetiva entre as abordagens consideradas.

Para complementar a análise descritiva, foram realizados testes estatísticos (ANOVA e Mann-Whitney) para verificar se existiam diferenças estatisticamente significativas nas médias de precisão entre os modelos.

3.6.1 Análise das Métricas

Na primeira parte da análise foram avaliadas as métricas de qualidade que demonstram a eficiência do processo de extração de dados, conforme apresentado na Tabela 2.

Tabela 2 - Principais métricas de desempenho de cada LLM.

Modelo	Acurácia	Precisão Média	Revocação Média	F1 Média	Latência Média (s)	Custo Médio
ChatGPT	45.0%	0.6067	1.0	0.7552	1.6838	0.1154
Gemini	39.0%	0.5667	1.0	0.7234	0.9706	0.0730
DeepSeek	40.0%	0.6000	1.0	0.7500	6.5354	0.0564

Fonte: Autor (2026)

Os resultados indicam que o *ChatGPT* apresentou o melhor desempenho global em termos de qualidade da extração, alcançando 45,0% de acurácia, o maior valor entre os modelos avaliados. Esse resultado indica que, proporcionalmente, o *ChatGPT* obteve o maior número de acertos exatos quando comparado ao total de campos avaliados. O *DeepSeek* apresentou acurácia de 40,0%, enquanto o *Gemini* obteve 39,0%, valores próximos entre si, porém inferiores ao do *ChatGPT*.

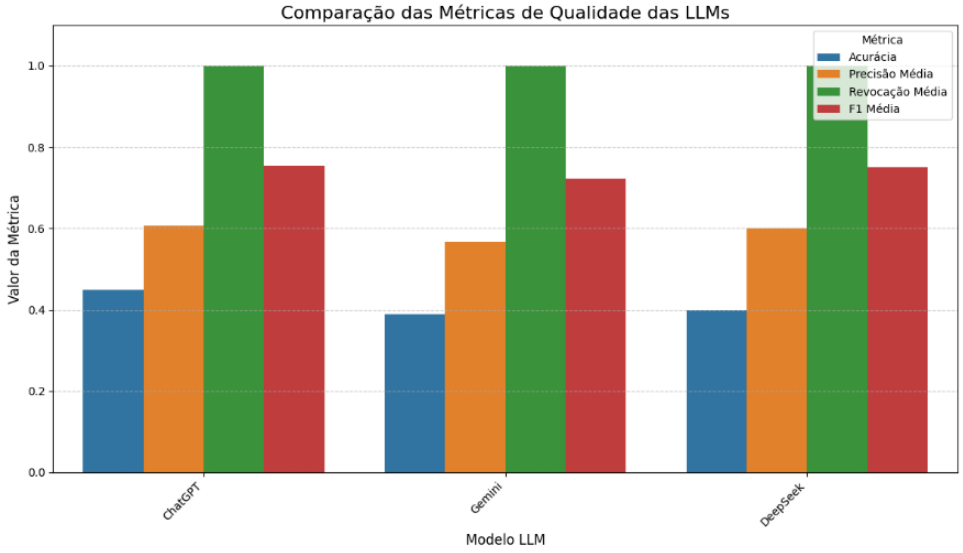
Em relação à precisão média, o *ChatGPT* também se destacou, com valor de 0,6067, seguido de perto pelo *DeepSeek* (0,6000). O *Gemini* apresentou a menor precisão (0,5667), indicando maior incidência de extrações incorretas ou parcialmente inconsistentes. Esses resultados evidenciam que, embora todos os modelos sejam capazes de identificar os campos esperados, há diferenças relevantes quanto à exatidão textual das informações extraídas.

A revocação média foi igual a 1,0 para os três modelos, indicando que todos conseguiram recuperar integralmente os campos de endereço esperados, sem ocorrência de omissões. Esse comportamento revela uma forte tendência à maximização da cobertura da informação, característica desejável no contexto jurídico, no qual a ausência de dados relevantes pode comprometer etapas posteriores do fluxo processual.

Como consequência da revocação máxima, o *F1-score* médio refletiu predominantemente as variações de precisão. O *ChatGPT* apresentou o maior *F1-score* (0,7552), seguido pelo *DeepSeek* (0,7500) e pelo *Gemini* (0,7234). Esses resultados indicam que o *ChatGPT* alcançou o melhor equilíbrio entre identificação completa dos campos e correção das informações extraídas.

O Gráfico 1 apresenta a comparação dos indicadores de qualidade entre as LLMs analisadas, evidenciando visualmente a similaridade entre elas.

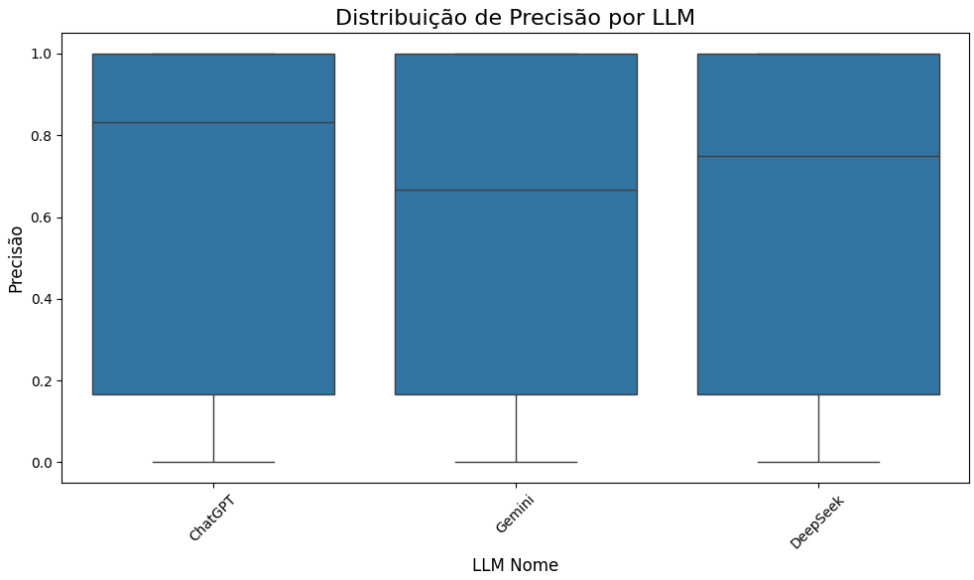
Gráfico 1 - Comparação das Métricas de Qualidade das LLMs



Fonte: Autor (2026)

O Gráfico 2 evidencia que a distribuição da precisão entre os modelos é semelhante, apresentando medianas próximas, o que indica que, de forma geral, os modelos exibem comportamento consistente quanto ao nível central de exatidão das extrações, sem predominância significativa de um modelo sobre os demais no desempenho típico observado.

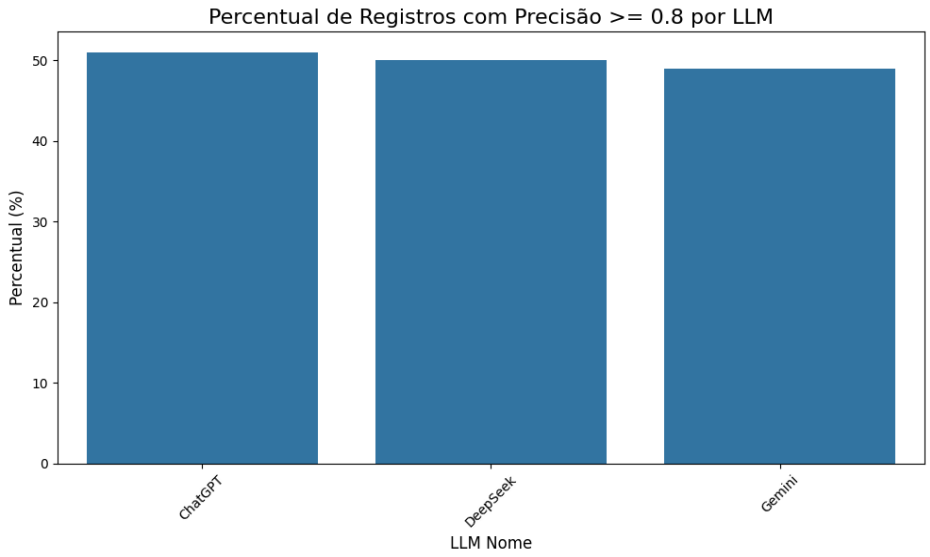
Gráfico 2 - Distribuição de Precisão por LLM



Fonte: Autor (2026)

A análise da precisão de alto desempenho (precisão ge0.8) também aponta para um desempenho equilibrado, como ilustra o Gráfico 3.

Gráfico 3 - Percentual de Registros com Precisão ge0.8 por LLM



Fonte: Autor (2026)

O indicador "Percentual de Registros com Precisão ≥0.8 por LLM" é uma métrica complementar para avaliar o desempenho de alta qualidade dos modelos. Ele mede a proporção de respostas em que a precisão foi considerada alta, ou seja, em que pelo menos 80% das informações extraídas estavam corretas.

Essa proximidade nos resultados reafirma a conclusão da análise estatística, que não encontrou diferença estatisticamente significativa entre os modelos em termos de qualidade e acurácia.

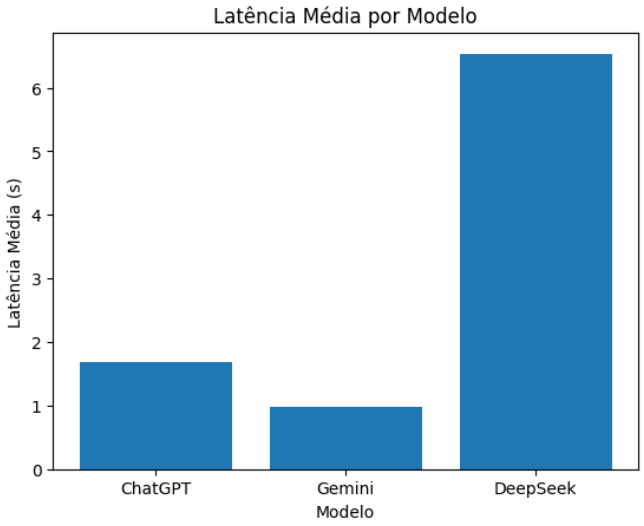
3.6.2 Análise dos Erros de Classificação

A inspeção manual de uma amostra representativa dos registros incorretamente classificados pelos modelos revelou padrões recorrentes de erro que merecem destaque. O primeiro padrão refere-se à confusão entre logradouro e bairro em endereços incompletos ou redigidos de forma não convencional. Por exemplo, em documentos que descrevem o imóvel como "localizado na Ilha de Deus, Recife-PE", todos os modelos identificaram "Ilha de Deus" tanto como logradouro quanto como bairro, resultando em Falso Positivo para um dos campos. O segundo padrão observado é a omissão ou preenchimento incorreto do CEP, especialmente em petições mais antigas, nas quais o código postal não é mencionado. Nesses casos, o ChatGPT e o Gemini tenderam a inferir um CEP aproximado com base no bairro e

cidade, gerando Falso Positivo, enquanto o DeepSeek optou com mais frequência por reportar o campo como ausente, resultando em Falso Negativo, porém semanticamente mais conservador. O terceiro padrão diz respeito a abreviações e variações regionais, como "Av." versus "Avenida" e "R." versus "Rua", que, após normalização semântica pelo algoritmo de avaliação, foram corretamente tratadas na maioria dos casos, mas ainda geraram divergências em situações de abreviação não padronizada. Esses padrões de erro evidenciam que a acurácia dos modelos é sensível à qualidade e completude do texto original, reforçando a necessidade de validação humana no fluxo de trabalho proposto.

Quanto ao desempenho operacional (Gráfico 4), observam-se diferenças relevantes entre os modelos. O *Gemini* apresentou a menor latência média (0,9706 s), sendo o mais rápido, seguido pelo *ChatGPT* (1,6838 s). Já o *DeepSeek* registrou latência significativamente superior (6,5354 s), o que pode limitar seu uso em aplicações de tempo quase real ou em cenários de processamento em larga escala.

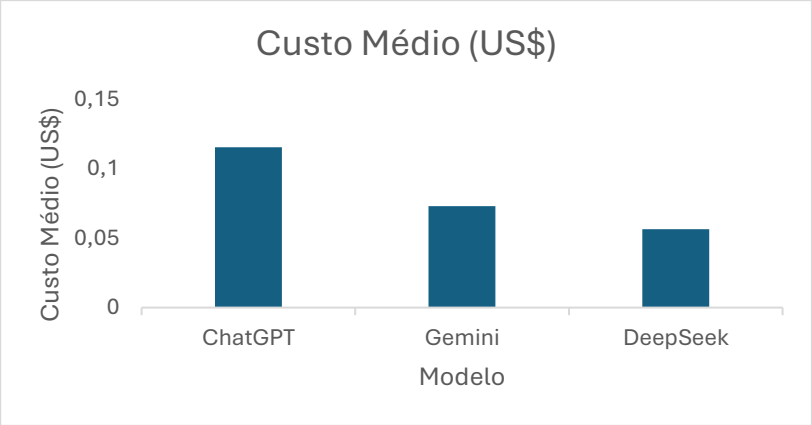
Gráfico 4 – Latência média por LLM



Fonte: Autor (2026).

Quanto ao custo financeiro médio, o *DeepSeek* mostrou-se a alternativa mais econômica (US$ 0,0564), seguido pelo *Gemini* (US$ 0,0730). O *ChatGPT* apresentou o maior custo financeiro (US$ 0,1154), refletindo uma troca direta entre qualidade da extração e custo financeiro. O Gráfico 5 apresenta a variação do comparativo custo total.

Gráfico 5 – Custo Médio (US$) por LLM



Fonte: Autor (2026).

3.6.3 Teste ANOVA e Mann-Whitney

A análise estatística formal confirmou a ausência de diferenças significativas. O teste ANOVA apresentou um p-valor de 0,7627, e os testes de *Mann-Whitney* (comparação par a par) também resultaram em p-valores acima de 0,05, corroborando a conclusão de que não há um vencedor estatístico claro em termos de acurácia entre os modelos.

Do ponto de vista prático, a equivalência estatística entre os modelos implica que a decisão de adoção institucional não deve ser orientada exclusivamente pelo desempenho em qualidade de extração, uma vez que as diferenças observadas não são estatisticamente significativas. Nesse cenário, os critérios determinantes para a escolha do modelo mais adequado ao ambiente do TJPE devem considerar os atributos operacionais: (i) quando a prioridade é o menor custo financeiro, o DeepSeek apresenta vantagem clara (US$ 0,056 por consulta, frente a US$ 0,073 do Gemini e US$ 0,115 do ChatGPT); (ii) quando a prioridade é a menor latência — relevante em aplicações de processamento em tempo quase real — o Gemini é a escolha mais adequada (0,97 s em média); e (iii) quando o contexto exige o melhor equilíbrio entre qualidade e robustez, o ChatGPT apresenta a maior acuácia e F1-score, ainda que com maior custo. Para ambientes públicos com restrições orçamentárias e grande volume de processamento, como o TJPE, recomenda-se a adoção do DeepSeek ou Gemini, com validação humana das extrações para garantir confiabilidade jurídica.

a) Teste ANOVA

O teste ANOVA foi utilizado para comparar as médias da métrica de precisão entre os três modelos de linguagem. A Estatística F (0,271142) e o p-valor (0,762697) são os resultados-chave deste teste, apresentados na Tabela 3.

Tabela 3 - Comparação ANOVA

Teste	Estatística F	p-valor
ANOVA	0,271142	0,762697

Fonte: Autor (2026).

O p-valor é a probabilidade de se observar os dados ou dados mais extremos, assumindo que a hipótese nula seja verdadeira (neste caso, a hipótese de que não há diferença entre as médias). O nível de significância comumente utilizado em estudos acadêmicos é de 0,05.

Neste caso, como o p-valor (0,7627) é maior que 0,05, a conclusão é que não há uma diferença estatisticamente significativa entre as médias de precisão dos modelos. Isso indica que, do ponto de vista estatístico, o desempenho dos três LLMs em termos de precisão pode ser considerado similar.

b) Teste *Mann-Whitney*

O teste de *Mann-Whitney* é uma alternativa não paramétrica ao teste t de *Student* e é utilizado para comparar a precisão de dois grupos independentes (os pares de LLMs, neste caso). A hipótese nula é que não há diferença na distribuição de precisão entre os pares de modelos, como observado na Tabela 4.

Tabela 4 - Comparativo entre as LLMs no Teste MANN-Whitney

Comparação	Estatística U	p-valor
ChatGPT vs Gemini	5344,0	0,380320
ChatGPT vs DeepSeek	5121,5	0,757236
Gemini vs DeepSeek	4777,0	0,572365

Fonte: Autor (2026).

A interpretação dos resultados é a seguinte:

- *ChatGPT vs Gemini*: O p-valor foi de 0,380320. Como este valor é maior que 0,05, não há diferença estatisticamente significativa entre a precisão do *ChatGPT* e do *Gemini.*
- *ChatGPT vs DeepSeek*: O p-valor foi de 0,757236. Sendo maior que 0,05, não há diferença estatisticamente significativa entre a precisão do *ChatGPT* e do *DeepSeek*.
- *Gemini vs DeepSeek*: O p-valor foi de 0,572365. Este valor também é maior que 0,05, indicando que não há diferença estatisticamente significativa entre a precisão do *Gemini* e do *DeepSeek*.

3.6.4 Análise de conversibilidade dos dados

Esta seção apresenta a análise da conversibilidade dos endereços extraídos pelos modelos de linguagem em coordenadas geográficas, etapa fundamental para a validação do uso desses dados em aplicações de visualização espacial, análise territorial e apoio à tomada de decisão no contexto jurídico-administrativo.

A conversibilidade refere-se à capacidade de um endereço textual estruturado (logradouro, número, bairro, cidade, estado e CEP) ser corretamente interpretado por um serviço de geocodificação — neste estudo, a ferramenta *Google Maps* — resultando na atribuição de coordenadas geográficas válidas (latitude e longitude). Essa etapa é essencial para permitir a plotagem espacial dos imóveis e a análise de sua distribuição geográfica no território do estado de Pernambuco.

A Tabela 5 sintetiza os resultados obtidos para os três modelos avaliados, considerando um conjunto de 100 registros para cada modelo.

Tabela 5 - Análise de conversibilidade dos dados em coordenadas geográficas

Modelo	Número de Registros Convertidos	% Registros Convertidos
ChatGPT	55	55%
Gemini	52	52%
DeepSeek	55	55%

Fonte: Autor (2026)

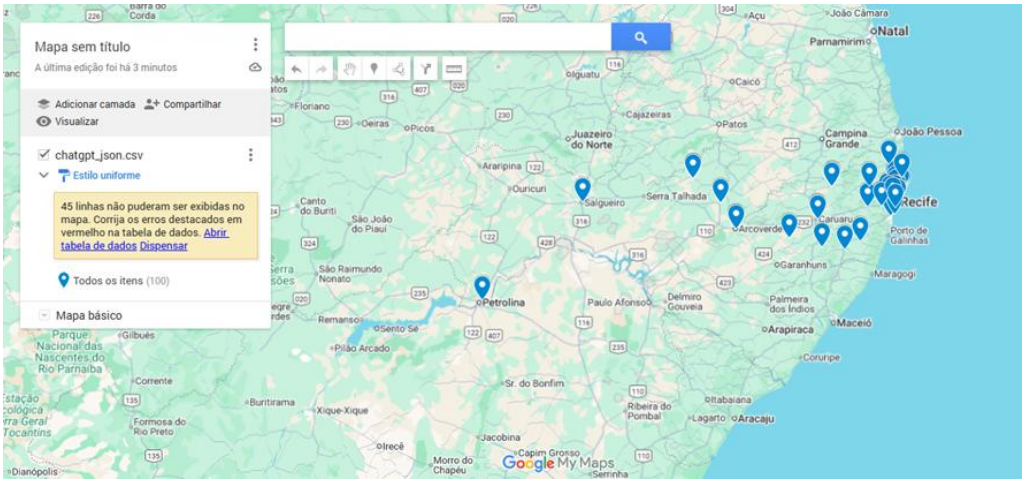
Os resultados indicam que, embora os modelos apresentem desempenhos semelhantes, há diferenças sutis na taxa de conversão, refletindo variações na qualidade, completude e padronização dos endereços extraídos, fatores diretamente relacionados ao sucesso da geocodificação.

a) ChatGPT

No caso do *ChatGPT*, dos 100 registros submetidos ao processo de geocodificação, 55 foram convertidos com sucesso, correspondendo a uma taxa de conversibilidade de 55%. Esse resultado demonstra que mais da metade dos endereços extraídos apresentava estrutura mínima suficiente para permitir a identificação espacial pelo Google Maps.

A Figura 6 ilustra a distribuição espacial dos imóveis convertidos no território do estado de Pernambuco. Observa-se uma concentração maior de pontos em áreas urbanas, o que sugere que endereços localizados em regiões com parâmetros de endereço mais bem definidos tendem a apresentar maior taxa de conversão.

Figura 6 - Geolocalização feita com dados do ChatGPT



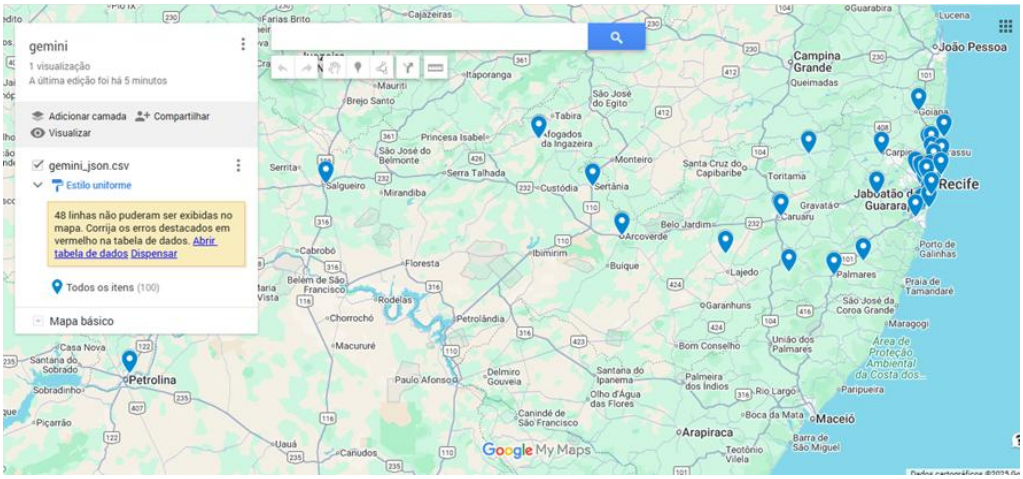
Fonte: Autor (2026).

b) Gemini

Os dados extraídos pelo *Gemini* apresentaram uma taxa de conversibilidade ligeiramente inferior, com 52 registros convertidos, o que corresponde a 52% do total analisado. Isso indica que 48 endereços não puderam ser geocodificados, possivelmente em razão de inconsistências como ausência de parâmetros relevantes para a localização do imóvel.

A Figura 7 apresenta a distribuição geográfica dos registros convertidos. Assim como no caso do *ChatGPT*, observa-se maior concentração de pontos em áreas urbanas, reforçando a dependência da geocodificação em relação à qualidade da base cartográfica e à padronização dos dados de entrada.

Figura 7 - Geolocalização feita com dados do Gemini



Fonte: Autor (2026)

c) DeepSeek

O *DeepSeek* apresentou uma taxa de conversibilidade de 55%, resultado equivalente ao observado para o *ChatGPT*. Dos 100 registros analisados, 55 foram convertidos com sucesso, conforme ilustrado na Figura 8.

Esse desempenho sugere que, embora o *DeepSeek* apresente diferenças em outros aspectos avaliados ao longo da pesquisa, sua capacidade de gerar endereços compatíveis com processos de geocodificação automática é comparável à do *ChatGPT*, ao menos no conjunto de dados analisado.

Figura 8 - Geolocalização feita com dados do DeepSeek

Fonte: Autor (2026).

A análise qualitativa dos resultados de conversibilidade evidencia que o desempenho dos modelos de linguagem não deve ser interpretado apenas a partir dos percentuais de conversão, mas sobretudo à luz da qualidade semântica e estrutural dos endereços extraídos. Embora *ChatGPT* e *DeepSeek* tenham apresentado a maior taxa de conversibilidade (55%), esse resultado não implica, necessariamente, superioridade absoluta na geração de dados geoespaciais prontos para uso, mas indica maior frequência de endereços com nível mínimo de completude e coerência sintática exigido pelos serviços de geocodificação.

Observou-se que os registros convertidos com sucesso tendem a apresentar padronização mais consistente dos campos de logradouro, município e estado, bem como menor incidência de ambiguidades linguísticas, como abreviações não convencionais, variações ortográficas ou ausência de elementos essenciais (especialmente número do imóvel e bairro).

Por outro lado, os registros não convertidos, independentemente do modelo, frequentemente continham informações incompletas, descrições genéricas de localização ou referências contextuais comuns em petições judiciais, mas inadequadas para interpretação automática por ferramentas cartográficas.

Do ponto de vista qualitativo, os resultados também revelam que a geocodificação é altamente sensível ao contexto urbano. Endereços situados em áreas metropolitanas ou municípios com melhor cobertura cartográfica apresentaram maior probabilidade de conversão, enquanto imóveis localizados em zonas rurais, loteamentos irregulares ou áreas com nomenclatura informal mostraram maiores taxas de insucesso. Esse comportamento evidencia que parte das limitações observadas não decorre exclusivamente do desempenho dos modelos de linguagem, mas da interação entre a qualidade do texto extraído e as restrições das bases geoespaciais disponíveis.

No caso específico do *Gemini*, a taxa de conversibilidade ligeiramente inferior (52%) sugere maior incidência de variações na estrutura textual dos endereços, o que, qualitativamente, pode indicar maior flexibilidade linguística, porém com impacto negativo na interoperabilidade com sistemas que exigem dados rigidamente estruturados.

Já o desempenho equivalente entre *ChatGPT* e *DeepSeek* aponta que ambos conseguem produzir saídas mais alinhadas às expectativas de sistemas de geocodificação, ainda que com limitações semelhantes no tratamento de casos ambíguos ou incompletos.

3.6.5 Comparação com a Literatura

Os resultados obtidos nesta pesquisa guardam relação com os achados de estudos prévios sobre NER em domínios especializados. Araujo e Silveira (2025), ao compararem BERT e ChatGPT em tarefas de NER no domínio jurídico brasileiro, identificaram que modelos de linguagem de grande escala superam abordagens tradicionais na identificação de entidades em textos forenses, especialmente quando não há corpus anotado para ajuste fino — cenário diretamente compatível com o presente estudo. Arana et al. (2025), por sua vez, reportaram acurácia de 95% e F1-score de 0,94 em tarefa de classificação textual aplicada à regularização fundiária com técnicas clássicas supervisionadas (TF-IDF e regressão logística). A diferença expressiva em relação aos valores de acurácia obtidos nesta pesquisa (39–45%) deve ser interpretada com cautela: aquele estudo realizou classificação binária ou de categorias fechadas, enquanto o presente estudo realiza extração de entidades em texto livre, tarefa consideravelmente mais complexa e suscetível a variações ortográficas, abreviações e ausência de padronização nos documentos. Além disso, Ariai e Demartini (2025) destacam que a ausência de treinamento específico no domínio jurídico representa uma limitação intrínseca dos LLMs de propósito geral, o que corrobora os níveis de precisão observados e reforça a necessidade de supervisão humana no fluxo de trabalho proposto.

Sob a perspectiva aplicada, os resultados qualitativos reforçam que o uso de LLMs para extração e georreferenciamento de endereços no contexto do Judiciário deve ser compreendido como apoio ao trabalho humano, e não como substituição integral. A conversibilidade parcial observada indica que soluções baseadas exclusivamente em extração automática podem introduzir vieses espaciais, privilegiando áreas urbanas formalizadas e marginalizando regiões com menor padronização cadastral.

Assim, a análise qualitativa confirma a viabilidade técnica do uso de modelos de linguagem como componente de um *pipeline* híbrido, no qual a extração automatizada é complementada por validação, enriquecimento e normalização dos dados. Essa abordagem é particularmente relevante em iniciativas institucionais que demandam confiabilidade, transparência e rastreabilidade, como aquelas associadas à regularização fundiária e à modernização dos fluxos processuais no âmbito do Poder Judiciário.

4 PROTÓTIPO DO PRODUTO DESENVOLVIDO

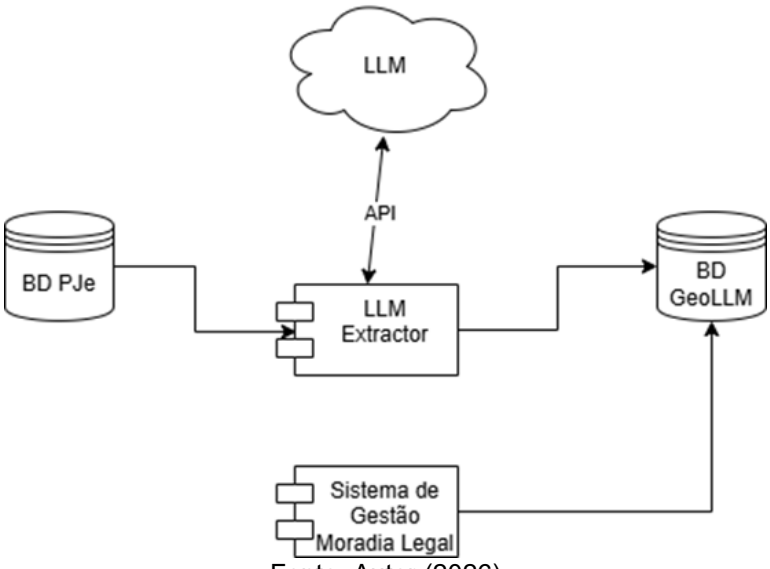
A materialização do objeto da pesquisa de mestrado foi a construção do protótipo *GeoLLM-Extractor*.

Esta aplicação foi concebida como uma ferramenta gerencial de integração entre métodos de extração de entidades nomeadas e a conversão em coordenadas de geolocalização de endereços, associado a uma área de visualização dedicada a tomada de decisões.

Para garantir a transparência metodológica e possibilitar a reprodutibilidade dos experimentos, os artefatos técnicos desenvolvidos neste estudo — incluindo códigos-fonte, *scripts* de avaliação, conjuntos de dados derivados e documentação complementar — foram disponibilizados em um repositório digital de apoio, mantido pelo autor[4].

A visão completa e o detalhamento técnico da estrutura tecnológica do *GeoLLM-Extractor* são formalmente apresentados por meio do diagrama arquitetônico, conforme ilustrado na Figura 9.

Figura 9 - Arquitetura da Solução GeoLLM Extractor



Fonte: Autor (2026)

[4] Repositório digital contendo os artefatos técnicos e a documentação de suporte da pesquisa, disponível para acesso público em: https://drive.google.com/drive/folders/1k1XYCdN6EIllsJvqTEOOSR-kg9-GyUGC

A solução é composta de dois módulos *LLM_Extractor* e o Sistema de Gestão Moradia Legal, com funcionalidades descritas a seguir:

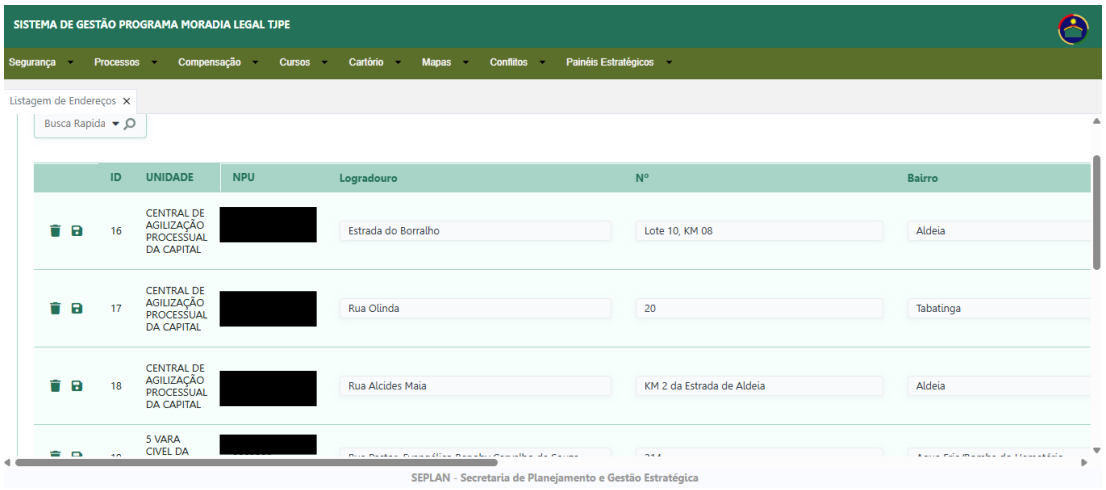
O *LLM_Extractor* é uma aplicação escrita na linguagem Python que seleciona os processos de usucapião da base de dados do PJe e processa os documentos interagindo com a solução de LLM via API para a extração dos dados de endereço dos documentos, salvando diretamente em outra base de dados exclusiva para a solução GeoLLM Extractor.

O Sistema de Gestão Moradia Legal (SGML) se estabelece como uma ferramenta de governança fundiária, materializando o fluxo de trabalho híbrido proposto, conforme Figura 12. Desenvolvido como uma aplicação web, seu papel fundamental é promover a interação e a governança dados referentes a regularização fundiária no Estado de Pernambuco, consolidando informações de endereços que foram processados e estruturados na base de dados *GeoLLM*.

O SGML é um centro de inteligência de negócios no que se refere a regularização fundiária. Ele oferece a gestão completa dos dados, transformando informações brutas em conclusões acionáveis por meio de ferramentas de visualização georreferenciada. A plataforma disponibiliza:

Relatórios Analíticos e Sintéticos detalhados para profundidade técnica, como visto na Figura 10.

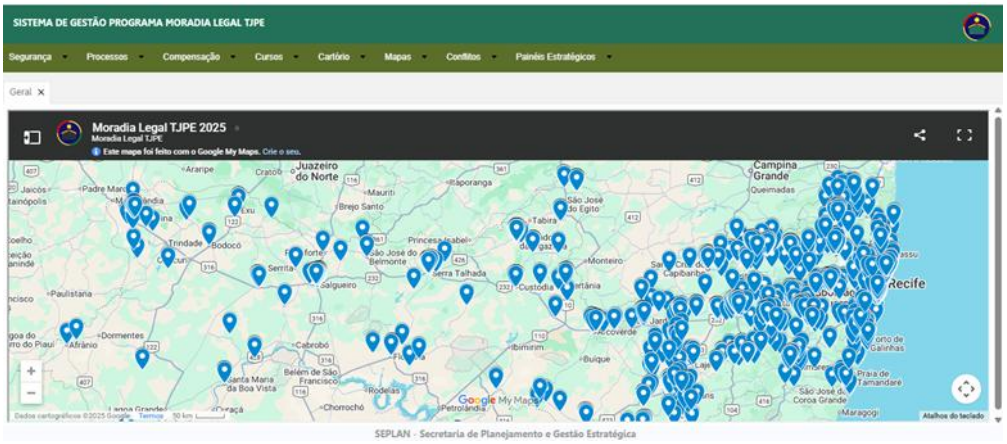
Figura 10 – Tela do Software para os registros de endereços



Fonte: Autor (2026)

Painéis de Mapas com georreferenciamento dos imóveis usucapiendo, essenciais para o acompanhamento gerencial e estratégico, apresentado na Figura 11.

Figura 11 – Tela do Software com a geolocalização dos imóveis



Fonte: Autor (2026)

Entre suas funcionalidades essenciais, o sistema eleva a capacidade de gestão ao permitir:

Validação e Enriquecimento de Dados: Permitir a interação direta para validar, corrigir e enriquecer os registros de endereço extraídos.

Visualização Geográfica Interativa: Apresentação imediata dos endereços em mapas, facilitada pela integração com o *Google Maps*, conforme Figura 12.

Acompanhamento Processual Integrado: Capacidade de associar e rastrear o status de regularização de cada endereço vinculado a um processo.

Filtros e Buscas Avançadas: Implementação de mecanismos de busca e filtros baseados em metadados geográficos e processuais.

Rastreabilidade e Segurança: Geração de trilhas de auditoria para rastrear modificações e garantir a integridade e conformidade da informação.

Dessa forma, o SGML consolida a pesquisa, convertendo a eficiência da automação (*GeoLLM*) em uma ferramenta de governança fundiária decisiva para todos os agentes do Poder Judiciário que atuam na área de regularização fundiária.

5 LIMITAÇÕES DO ESTUDO

O presente estudo apresenta limitações que devem ser consideradas na interpretação dos resultados. Em primeiro lugar, a amostra de 100 processos corresponde a 3,8% do universo de 2.587 processos distribuídos no período analisado, e sua seleção foi realizada sem critérios de estratificação por grau de complexidade do endereço, zona urbana ou rural, ou período de distribuição. Essa amostragem foi condicionada pelos dados disponibilizados pelo órgão parceiro e pelo custo financeiro do processamento via APIs de LLMs, o que limita a generalização institucional dos resultados para o universo completo de processos do TJPE. Em segundo lugar, a anotação do conjunto de referência foi realizada por um único analista, sem medida formal de concordância inter-anotador, o que pode introduzir viés humanístico nas comparações, embora os critérios objetivos de anotação adotados mitiguem parcialmente esse risco. Em terceiro lugar, a geocodificação via Google Maps introduz uma variável externa ao desempenho dos LLMs, pois falhas na conversão podem decorrer de limitações da base cartográfica — como ausência de CEP, ambiguidade urbana ou nomenclatura informal — e não necessariamente de erros na extração do endereço pelo modelo. Por fim, a latência medida reflete o tempo total de resposta das APIs em ambiente externo, sendo influenciada por fatores não controlados como carga dos servidores e variações de conectividade, o que limita a comparabilidade direta entre os modelos nessa dimensão.

6 CONCLUSÃO

A análise dos resultados aponta para uma conclusão dual: embora a utilização de LLMs para a tarefa de identificação de endereços seja tecnicamente viável, ela se revela insuficiente como uma solução autônoma para aplicações críticas.

Os modelos demonstram elevada eficiência na extração de dados, evidenciada pela revocação máxima (1,0) e pelas baixas latências, especialmente no Gemini. Contudo, esse desempenho operacional não se traduz em confiabilidade jurídica, uma vez que a precisão média em torno de 0,6 e a acurácia inferior a 50% são insuficientes para o contexto judicial, no qual erros na identificação de endereços podem comprometer a validade do processo.

Diante desse gap entre eficiência e rigor, a abordagem recomendada é a implementação de um fluxo de trabalho híbrido assistido por IA. Neste modelo, os LLMs são integrados como ferramenta de triagem inicial e extração de endereços, otimizando o tempo de processamento. Contudo, a validação humana por um profissional do direito é indispensável para conferir a precisão final exigida. Essa abordagem não só capitaliza a eficiência da automação, mas também assegura a segurança e o rigor necessários para a tarefa, mitigando os riscos associados à baixa acurácia.

Essa necessidade de supervisão humana torna-se ainda mais evidente quando se considera a taxa de conversibilidade geográfica: aproximadamente metade dos endereços extraídos pôde ser geocodificada com sucesso, indicando que a completude e a padronização das informações são fatores críticos para aplicações de georreferenciamento. Esse resultado reforça a necessidade de estratégias complementares, como validação semântica, enriquecimento de dados e supervisão humana, especialmente em contextos de endereçamento informal.

O desenvolvimento do software GeoLLM-Extractor, fruto direto desta pesquisa, materializa essa abordagem e atende de forma estratégica à demanda do NUREF – Núcleo de Regularização e Demandas Fundiárias do TJPE. O sistema tem como

escopo aprimorar a gestão de processos e a geolocalização dos endereços, permitindo ainda a integração com plataformas externas para identificação de Zonas Especiais de Interesse Social (ZEIS). Dessa forma, consolida-se como uma solução completa, capaz de atender de forma ágil às solicitações de informações provenientes de unidades internas e externas do TJPE, com possibilidade real de replicação em outras unidades estaduais do Poder Judiciário.

Por fim, os resultados obtidos abrem espaço para trabalhos futuros em outras áreas do Poder Judiciário que lidam com conflitos fundiários e territoriais. Destacam-se, entre elas, a identificação automatizada de endereços de imóveis em processos de reintegração de posse, ações possessórias e desapropriações; a análise espacial de áreas de conflito urbano e rural; o mapeamento de conflitos fundiários coletivos; e o cruzamento de dados judiciais com bases geográficas para apoio à mediação, prevenção de litígios e formulação de políticas públicas. Tais extensões reforçam o caráter transversal da abordagem proposta e seu potencial de replicabilidade em diferentes ramos da Justiça.

7 REFERÊNCIAS

ARIAI, Farid; MACKENZIE, Joel; DEMARTINI, Gianluca. Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges. **ACM Computing Surveys**, v. 58, n. 6, art. 163, p. 1–37, dez. 2025. DOI: 10.1145/3777009.

ARAUJO, Geovane; SILVEIRA, Raquel. **Análise comparativa do BERT e ChatGPT no reconhecimento de entidades nomeadas do domínio jurídico**. Revista Eletrônica de Iniciação Científica em Computação, 2025, v. 23, p. 1. DOI: 10.5753/reic.2025.3903.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Diário Oficial da União: Brasília, DF, 15 ago. 2018.

BROWN, T. et al. **Language Models are Few-Shot Learners. OpenAI**, 2020. arXiv:2005.14165.

CONSELHO NACIONAL DE JUSTIÇA. **Inteligência Artificial no Poder Judiciário Brasileiro**. Brasília: CNJ, 2019. Disponível em: https://bibliotecadigital.cnj.jus.br/jspui/bitstream/123456789/98/1/Intelig%c3%aancia%20Artificial%20no%20Poder%20Judiciario%20Brasileiro.pdf. Acesso em: 27 janeiro 2026.

CONSELHO NACIONAL DE JUSTIÇA. **Portaria nº 4, de 30 de janeiro de 2025a.** Disponível em: **https://atos.cnj.jus.br/files/original1639122025020567a39430abf69.pdf.** Acesso em: 27 janeiro 2026.

CONSELHO NACIONAL DE JUSTIÇA. **Provimento nº 144, de 25 de abril de 2023a.** Disponível em: https://atos.cnj.jus.br/files/original152742202304266449 42eee2ea6.pdf. Acesso em: 27 janeiro 2026.

CONSELHO NACIONAL DE JUSTIÇA. **Provimento nº 158, de 5 de dezembro de 2023b.** Disponível em: https://atos.cnj.jus.br/files/original22225220231207657245bc65439.pdf. Acesso em: 27 janeiro 2026.

CONSELHO NACIONAL DE JUSTIÇA. **O uso da inteligência artificial generativa no Poder Judiciário brasileiro: relatório de pesquisa**. Brasília: CNJ, 2024. Disponível em: https://www.cnj.jus.br. Acesso em: 11 jul. 2025.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Resolução nº 615**, **de 11 de março de 2025b.** Dispõe sobre a governança, o uso ético e responsável da Inteligência Artificial no âmbito do Poder Judiciário. Diário da Justiça Eletrônico: Brasília, DF, 2025.

COSTA, João; ALMEIDA, Maria. **Avaliação integrada de modelos de IA: métricas técnicas e operacionais.** Revista Brasileira de Inteligência Artificial, São Paulo, v. 10, n. 2, p. 45-60, abr. 2024.

DATACAMP. ***What is Named Entity Recognition (NER)?***. DataCamp, 2023. Disponível em: https://www.datacamp.com/blog/what-is-named-entity-recognition-ner. Acesso em: 30 janeiro 2026.

DATACAMP. **Prompting de disparo zero: Exemplos, teoria e casos de uso**. DataCamp, 29 jul. 2024. Disponível em: https://www.datacamp.com/pt/tutorial/zero-shot-prompting. Acesso em: 30 janeiro 2026.

FINATTO, Maria José Bocorny; MACOHIN, Aline. PLN no Direito. In: FINATTO, Maria José Bocorny (org.). **Processamento de Linguagem Natural com foco no português do Brasil**. [S. l.]: GLOBE, 2023. Cap. 22. Disponível em: https://brasileiraspln.com/livro-pln/1a-edicao/. Acesso em: 15 maio 2025.

FRANKENREITER, Jens; NYARKO, Julian. **Natural Language Processing in Legal Tech**. SSRN, 2022. Disponível em: https://ssrn.com/abstract=4027030.

FREITAS, Elisa Lauren Mangini. **A proteção de dados pessoais no uso de Inteligência Artificial pelo Poder Judiciário: um estudo à luz da Resolução CNJ nº 615/2025 e da Lei Geral de Proteção de Dados (LGPD).** 2025. Trabalho de Conclusão de Curso (Graduação em Direito) – Universidade Federal de Santa Catarina, Florianópolis, 2025.

GOMES, Jordânia de Sousa et al. **Uso da Inteligência Artificial no Poder Judiciário Brasileiro: Um Ensaio Teórico**. 2024.

KATZ, Daniel Martin et al. **Natural Language Processing in the Legal Domain**. Singapore Management University, 2023. Disponível em: https://ink.library.smu.edu.sg/sol_research/4527. Acesso em: 14 maio 2025.

MODI, Hiral. **Leveraging Natural Language Processing for Legal Research: Trends and Future Directions.** SSRN, 2023. DOI: 10.2139/ssrn.4514036.

OLIVEIRA, Lucas; PEREIRA, Ana. Métricas de desempenho para reconhecimento de entidades nomeadas: uma revisão. Journal of Computational Linguistics, Rio de Janeiro, v. 8, n. 1, p. 12-25, jan. 2022.

SILVA, Roberto. **Impacto das métricas operacionais na escolha de modelos de IA para grandes volumes de dados.** Revista de Tecnologia e Sistemas, Belo Horizonte, v. 15, n. 3, p. 78-90, jul. 2023.

TRIBUNAL DE JUSTIÇA DE PERNAMBUCO (TJPE). **TJPE usará inteligência artificial para agilizar processos de execução fiscal no Recife**. Publicado em 20 nov. 2018. Disponível em: https://portal.tjpe.jus.br/comunicacao/-/asset_publisher/ubhL04hQXv5n/content/tjpe-usara-inteligencia-artificial-para-agilizar-processos-de-execucao-fiscal-no-recife. Acesso em: 15 maio 2025.

TRIBUNAL DE JUSTIÇA DA PARAÍBA (TJPB). **TJPB lança ferramenta de IA que automatiza coleta de dados de divórcios em parceria com o IBGE**. Publicado em 18 nov. 2025. Disponível em: https://www.tjpb.jus.br/noticia/tjpb-lanca-ferramenta-de-ia-que-automatiza-coleta-de-dados-de-divorcios-em-parceria-com-o

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. **Attention is all you need.** In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS), 2017. Disponível em: https://arxiv.org/pdf/1706.03762.pdf. Acesso em: 14 maio 2025.

APÊNDICE A

Documentação Técnica do Script de Avaliação

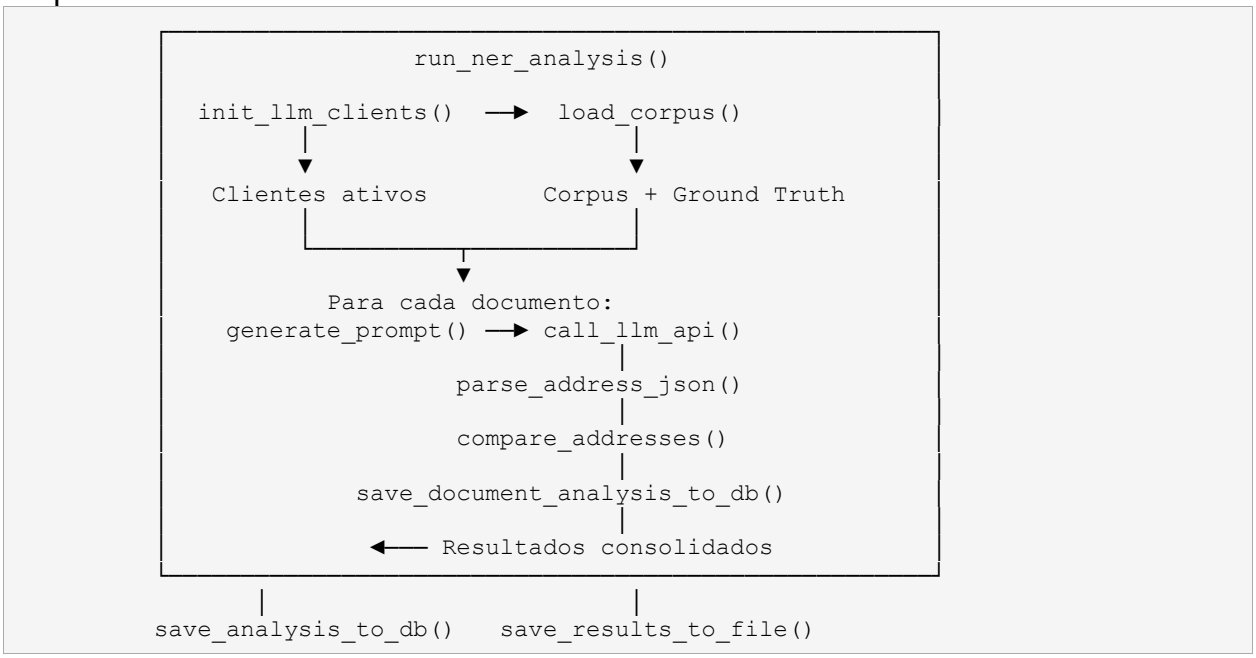
1. Visão Geral do Script

O script corpus_analysis_v3.py implementa o pipeline completo de avaliação de Modelos de Linguagem de Grande Escala (LLMs) para a tarefa de Reconhecimento de Entidades Nomeadas (NER) aplicada à extração de endereços em processos judiciais de usucapião. O código foi desenvolvido no contexto da pesquisa de mestrado realizada junto ao Tribunal de Justiça de Pernambuco (TJPE) e integra diretamente a base de dados do sistema PJe.

O script executa, em sequência, as seguintes etapas: (1) inicialização dos clientes de API das LLMs; (2) carregamento do corpus e do conjunto de referência (ground truth) a partir do banco de dados MySQL; (3) geração de prompts padronizados; (4) chamada às APIs das LLMs; (5) análise e comparação dos endereços extraídos; e (6) persistência dos resultados individuais e consolidados no banco de dados.

2. Arquitetura e Fluxo de Execução

O diagrama abaixo representa, de forma simplificada, o fluxo de execução do script:



3. Documentação das Funções Principais

3.1 init_llm_clients()

init_llm_clients
`def init_llm_clients() -> None`

Inicializa os clientes de acesso às APIs dos modelos de linguagem configurados no dicionário LLM_CONFIGS. Na versão atual do script, apenas o cliente OpenAI (ChatGPT) é instanciado ativamente; os clientes do Gemini e do DeepSeek estão implementados como blocos comentados, reservados para ativação futura.

Parâmetro / Retorno	Tipo	Descrição
(sem parâmetros)	—	Função sem argumentos de entrada.
Retorno	`None`	Atualiza o campo 'client' de cada entrada em LLM_CONFIGS diretamente.

Dependência: Requer que as chaves de API estejam corretamente definidas em LLM_CONFIGS antes da chamada.

3.2 load_corpus()

load_corpus
`def load_corpus() -> list[dict]`

Estabelece conexão com o banco de dados MySQL da instância local e carrega os registros da view vw_verdade, que reúne o texto dos documentos jurídicos e os respectivos dados de endereço validados manualmente. Apenas registros com verdade = '1' são selecionados, garantindo que somente os documentos integrantes do conjunto de referência (ground truth) sejam processados.

Parâmetro / Retorno	Tipo	Descrição
(sem parâmetros)	—	Utiliza credenciais fixas de conexão ao banco de dados local.
Retorno	`list[dict]`	Lista de dicionários, cada um contendo: id, text (texto do documento) e ground_truth_address (endereço de referência com os campos rua, numero, bairro, cep, cidade e estado).

Estrutura de cada elemento retornado:

```
{
    'id': <int>,
    'text': <str>,           # Texto completo da petição
    'ground_truth_address': {
        'rua': <str>,
        'numero': <str>,
        'bairro': <str>,
        'cep': <str>,
        'cidade': <str>,
        'estado': <str>
    }
}
```

3.3 generate_prompt()

generate_prompt
def generate_prompt(text_content: str) -> str

Constrói o prompt de instrução que será enviado ao modelo de linguagem. O prompt adota a estratégia zero-shot, fornecendo ao modelo apenas a instrução da tarefa e o texto do documento, sem exemplos de treinamento adicionais. A saída é solicitada obrigatoriamente no formato JSON, com os seis campos de endereço esperados. Campos não encontrados devem ser retornados como string vazia.

Parâmetro / Retorno	Tipo	Descrição
text_content	str	Texto completo do documento jurídico a ser analisado pelo modelo.
Retorno	str	String contendo o prompt formatado, pronto para envio à API do LLM.

Formato JSON esperado na resposta do modelo:

```
{
    "rua": "Rua Exemplo",
    "numero": "45",
    "bairro": "Bairro Legal",
    "cidade": "Cidade Feliz",
    "estado": "PE",
    "cep": "99999999"
}
```

Estratégia de prompting: Zero-shot — o modelo recebe apenas a instrução e o texto, sem exemplos anteriores de entrada/saída.

3.4 calculate_openai_cost()

calculate_openai_cost

```
def calculate_openai_cost(usage: object) -> float
```

Calcula o custo financeiro estimado de cada chamada à API da OpenAI com base no consumo de tokens reportado no objeto usage da resposta. O cálculo distingue tokens de entrada (prompt_tokens) e tokens de saída (completion_tokens), aplicando tarifas distintas para cada categoria.

Parâmetro / Retorno	Tipo	Descrição
usage	object	Objeto retornado pela API da OpenAI contendo prompt_tokens, completion_tokens e total_tokens.
Retorno	float	Custo estimado em dólares americanos (USD) para a chamada realizada.

Fórmula de cálculo:

```
custo = (prompt_tokens / 1.000.000) × input_cost_per_million
      + (completion_tokens / 1.000.000) × output_cost_per_million


# Tarifas aplicadas (modelo gpt-4o):
#   input_cost_per_million  = US$ 5,00
#   output_cost_per_million = US$ 15,00
```

3.5 call_llm_api()

call_llm_api

```
def call_llm_api(llm_name: str, prompt: str) -> tuple[str, float, float]
```

Realiza a chamada à API do modelo de linguagem especificado, mensura a latência de resposta e retorna o texto gerado pelo modelo, o custo estimado e o tempo decorrido. A função suporta múltiplos modelos por meio de desvio condicional, embora apenas o ChatGPT esteja ativo na versão atual. A resposta da OpenAI é solicitada no formato JSON estruturado (response_format: json_object).

Parâmetro / Retorno	Tipo	Descrição
llm_name	str	Identificador do modelo a ser chamado: 'ChatGPT', 'Gemini' ou 'DeepSeek'.
prompt	str	Texto do prompt gerado por generate_prompt().
Retorno [0]	str	Texto da resposta gerada pelo modelo (string JSON).
Retorno [1]	float	Custo estimado da chamada em USD.
Retorno [2]	float	Latência total da chamada em segundos (tempo de resposta da API).

Medição da latência:

```
start_time = time.time()
# ... chamada à API ...
end_time = time.time()
latency = end_time - start_time   # valor em segundos
```

3.6 parse_address_json()

parse_address_json

```
def parse_address_json(json_string: str) -> dict
```

Realiza o parsing da string JSON retornada pelo modelo de linguagem, normalizando as chaves para letras minúsculas e garantindo a presença de todos os seis campos de endereço esperados. Campos ausentes na resposta do modelo são preenchidos com None. A função trata explicitamente erros de decodificação JSON, que podem ocorrer quando o modelo retorna texto malformado.

Parâmetro / Retorno	Tipo	Descrição
json_string	str	String JSON retornada pela API do modelo de linguagem.
Retorno	dict	Dicionário com os campos rua, numero, bairro, cidade, estado e cep. Em caso de erro, retorna dicionário com chave 'error' descrevendo a falha.

Campos garantidos na saída (campos ausentes recebem None):

```
required_fields = ['rua', 'numero', 'bairro', 'cidade', 'estado', 'cep']
```

3.7 compare_addresses()

compare_addresses

```
def compare_addresses(extracted: dict, ground_truth: dict) -> tuple[bool, str]
```

Compara campo a campo o endereço extraído pelo modelo com o endereço de referência (ground truth), após normalização de ambos para letras minúsculas e remoção de espaços extras. Um endereço é considerado correto apenas quando todos os campos presentes no ground truth correspondem exatamente aos campos extraídos. Divergências são registradas com o valor esperado e o valor extraído para cada campo.

Parâmetro / Retorno	Tipo	Descrição
extracted	dict	Endereço extraído pelo modelo, retornado por parse_address_json().
ground_truth	dict	Endereço de referência carregado da base de dados por load_corpus().
Retorno [0]	bool	True se todos os campos correspondem exatamente; False caso contrário.
Retorno [1]	str	'Correspondência exata' se correto; lista de divergências detalhadas se incorreto.

Lógica de comparação:

```
# Normalização antes da comparação:
extracted_norm    = {k: str(v).strip().lower() for k, v in extracted.items()}
ground_truth_norm    =    {k:    str(v).strip().lower()    for    k,    v    in
ground_truth.items()}


# Verificação campo a campo:
for key in ground_truth_norm:
    if extracted_norm.get(key) != ground_truth_norm.get(key):
        is_match = False
        details.append(f"Campo '{key}' difere...")
```

Critério de acurácia: A função implementa correspondência exata (exact match). Um único campo divergente — mesmo que por diferença de formatação — resulta em is_match = False, o que explica a acurácia global inferior a 50% mesmo com revocação máxima.

3.8 save_document_analysis_to_db()

save_document_analysis_to_db
`def save_document_analysis_to_db(llm_name: str, analysis_data: dict) -> None`

Persiste os dados de análise de cada documento individual na tabela tb_analises_documentos do banco de dados MySQL. Cada registro armazena o ground truth, o endereço extraído, o resultado da comparação, os detalhes de divergências, a latência e o custo da chamada, junto com o timestamp de execução.

Parâmetro / Retorno	Tipo	Descrição
llm_name	str	Nome do modelo que gerou a extração ('ChatGPT', 'Gemini', 'DeepSeek').
analysis_data	dict	Dicionário com document_id, ground_truth, extracted_address, is_correct, comparison_details, latency e cost.
Retorno	None	Sem retorno. Persiste diretamente no banco de dados.

Campos inseridos na tabela tb_analises_documentos:

```
documento_id          → ID do processo
llm_nome              → Nome do modelo utilizado
ground_truth_json     → Endereço de referência (JSON serializado)
endereco_extraido_json → Endereço extraído pelo modelo (JSON serializado)
correto               → Resultado da comparação (booleano)
detalhes_comparacao   → Descrição das divergências por campo
latencia              → Tempo de resposta da API (segundos)
custo                 → Custo estimado em USD
data_analise          → Timestamp de execução (datetime.now())
```

3.9 run_ner_analysis()

run_ner_analysis
`def run_ner_analysis() -> dict`

Função central do script. Orquestra a execução completa do pipeline de avaliação NER: inicializa os clientes, carrega o corpus, itera sobre cada documento e modelo configurado, acumula os resultados individuais e calcula as métricas consolidadas de acurácia, latência média e custo total para cada LLM processado.

Parâmetro / Retorno	Tipo	Descrição
(sem parâmetros)	—	Utiliza as configurações globais de LLM_CONFIGS e o corpus carregado do banco de dados.
Retorno	dict	Dicionário indexado pelo nome do LLM, contendo: individual_results (lista), total_cost, average_latency, accuracy, correct_extractions e total_documents.

Estrutura do dicionário de retorno:

```
{
  'ChatGPT': {
    'individual_results': [ ... ],  # lista de análises por documento
    'total_cost':         <float>,  # custo acumulado em USD
    'average_latency':    <float>,  # latência média em segundos
    'accuracy':           <float>,  # % de correspondências exatas
    'correct_extractions':<int>,
    'total_documents':    <int>
  }
}
```

Controle de taxa: Um delay de 0,1 segundo (time.sleep(0.1)) é aplicado entre chamadas consecutivas para respeitar os limites de requisições por minuto das APIs.

3.10 save_analysis_to_db()

save_analysis_to_db
def save_analysis_to_db(results: dict) -> None

Salva os resultados consolidados de cada modelo de linguagem na tabela tb_resultados do banco de dados MySQL. Cada linha da tabela corresponde a uma execução de um modelo, registrando as métricas agregadas de acurácia, latência média, custo total e volume de documentos processados, junto com o identificador do prompt utilizado.

Parâmetro / Retorno	Tipo	Descrição
results	dict	Dicionário retornado por run_ner_analysis(), indexado pelo nome do LLM.
Retorno	None	Sem retorno. Persiste diretamente na tabela tb_resultados.

Campos inseridos na tabela tb_resultados:

```
ia              → Nome do modelo (ex: 'ChatGPT')
prompt          → Identificador do modelo de prompt utilizado (ex: '1')
acuracia        → Percentual de acurácia (exact match)
latencia_media  → Latência média em segundos
custo_total     → Custo total acumulado em USD
total_documentos → Número de documentos processados
```

3.11 save_results_to_file()

save_results_to_file
def save_results_to_file(results: dict, filename: str = 'analysis_results.json') -> None

Exporta os resultados completos da análise para um arquivo JSON local, preservando todos os detalhes individuais por documento e por modelo. O arquivo é serializado com indentação de 4 espaços e codificação UTF-8, garantindo a legibilidade de caracteres especiais do português. Na versão atual do script, esta função está comentada no bloco __main__ e não é executada por padrão.

Parâmetro / Retorno	Tipo	Descrição
results	dict	Dicionário retornado por run_ner_analysis().
filename	str	Nome do arquivo de saída. Padrão: 'analysis_results.json'.
Retorno	None	Sem retorno. Grava o arquivo no diretório de trabalho atual.

4. Constantes e Configurações Globais

O script utiliza as seguintes constantes e estruturas de configuração definidas no escopo global:

openai	1.30.x	pip (externa)	Cliente oficial da API OpenAI. Fornece a classe OpenAI e o método chat.completions.create().
google-generativeai	0.5.x	pip (externa)	SDK do Google para acesso ao Gemini. Expõe genai.configure() e genai.GenerativeModel(). Inativa no script atual.

mysql-connector-python	8.4.0	pip (externa)	Driver oficial MySQL para Python. Utilizado em load_corpus(), save_document_analysis_to_db() e save_analysis_to_db().
json	stdlib	Python stdlib	Serialização e desserialização de estruturas JSON em todas as funções de persistência e parsing.
time	stdlib	Python stdlib	Medição de latência via time.time() em call_llm_api() e controle de delay entre chamadas (time.sleep).
datetime	stdlib	Python stdlib	Geração de timestamp de execução (datetime.now()) inserido em cada registro de tb_analises_documentos.
os	stdlib	Python stdlib	Importado no script; não utilizado diretamente na versão atual (reservado para leitura de variáveis de ambiente).

7. Estrutura de Dados: SQL_Peticao.sql

7.1 Visão Geral do Arquivo

O arquivo SQL_Peticao.sql contém o script de carga da base de dados primária da pesquisa. Trata-se de um único comando INSERT em lote direcionado à tabela tb_usucapiao do banco de dados usucapiao, populando os registros de petições iniciais de ações de usucapião extraídos do sistema PJe do Tribunal de Justiça de Pernambuco (TJPE), referentes ao período de 02 de janeiro a 31 de dezembro de 2024. O arquivo possui 22,6 MB, 41.142 linhas e contém 2.587 registros, abrangendo documentos nos formatos PDF (application/pdf) e HTML (text/html).

7.2 Tabela de Destino: tb_usucapiao

Cada linha do INSERT corresponde a um documento de petição inicial vinculado a um processo de usucapião. A tabela abaixo descreve os oito campos presentes no script, seu tipo inferido e sua função no pipeline de análise:

Campo	Tipo inferido	Descricao e papel no pipeline
ds_processo_documento	VARCHAR	Tipo do documento no PJe. Valor constante em todos os registros: Peticao Inicial (Outras).
id_processo	BIGINT	Identificador unico do processo no PJe. Chave de juncao com as demais tabelas do banco usucapiao. Utilizado como document_id no pipeline NER.
ds_extensao	VARCHAR	MIME type do documento. application/pdf: 1.691 registros (65,4%); text/html: 591 registros (22,8%). Documentos PDF tem ds_modelo_documento = NULL nesta etapa.
ds_modelo_documento	LONGTEXT/NULL	Campo central do pipeline NER: texto da peticao em HTML. Mapeado para text em load_corpus(). Presente em 523 registros text/html (total: 18,5 MB de conteudo). PDFs tem este campo NULL.
nm_arquivo	VARCHAR/NULL	Nome do arquivo PDF anexado ao processo no PJe. Presente apenas em documentos PDF. Documentos HTML tem este campo NULL. Nao utilizado diretamente no pipeline NER.
dt_inclusao	DATETIME	Data e hora de inclusao do documento no PJe, com precisao de milissegundos. Periodo: 2024-01-02 a 2024-12-31. Utilizado para delimitacao temporal do corpus.
ds_classe_judicial	VARCHAR	Classe processual registrada no PJe. Valor unico em todos os registros: Usucapiao. Confirma a homogeneidade do corpus quanto a natureza da acao.
ds_assunto_principal	VARCHAR	Modalidade da acao: Extraordinaria (1.421, 54,9%), Ordinaria (440, 17,0%), Especial/Constitucional (330, 12,7%), Lei 6.969/1981 (48, 1,9%). Registros com nomes de arquivo no campo indicam anomalia de dados do PJe.

7.3 Estatisticas do Corpus

Indicador	Valor
Total de registros no arquivo	**2.587**
Documentos PDF (application/pdf)	**1.691 (65,4%)**
Documentos HTML com texto (text/html)	**591 (22,8%) — 523 com conteudo**
Periodo coberto	**2024-01-02 a 2024-12-31**

Mes com maior volume de registros	Outubro/2024 (269 registros)
Modalidade predominante	Usucapiao Extraordinaria (1.421 — 54,9%)
Amostra selecionada para o experimento NER	100 processos (3,8% do total)

7.4 Relacao com o Script corpus_analysis_v3.py

A tabela tb_usucapiao, carregada por este script SQL, e a fonte primaria de dados do pipeline NER. A funcao load_corpus() acessa essa tabela pela view vw_verdade, filtrando os registros com verdade = '1' — os 100 processos validados manualmente que compõem o ground truth. O campo ds_modelo_documento e o texto enviado ao modelo via generate_prompt(), enquanto id_processo serve como document_id em tb_analises_documentos e tb_resultados. Os 1.691 documentos PDF tem ds_modelo_documento = NULL nesta tabela, razao pela qual o corpus experimental foi restrito a documentos text/html.

APÊNDICE B

Artigo Publicado: Tecnologia de Mineração de Texto no Direito: o Caso da Identificação de Endereços em Processos de Usucapião

TECNOLOGIA DE MINERAÇÃO DE TEXTO NO DIREITO: O CASO DA IDENTIFICAÇÃO DE ENDEREÇOS EM PROCESSOS DE USUCAPIÃO

FLÁVIO JOSÉ FERREIRA JUNIOR[1]
[1] TJPE / Cesar School – flavio.junior@tjpe.jus.br

RESUMO – O artigo explora o uso de mineração de texto no setor jurídico, com foco na identificação de endereços em processos de usucapião. A pesquisa utiliza o Processamento de Linguagem Natural (PLN) para estruturar informações extraídas de petições iniciais, otimizando a análise documental. Com 10 petições iniciais extraídas da base de dados do TJPE, analisadas no modelo GPT-4.o-mini, foi possível extrair dados completos em 80% dos casos, evidenciando a eficiência e apresentando as limitações da tecnologia.

No contexto jurídico, destaca-se a relevância da usucapião para a regularização fundiária e a necessidade de automação no Processo Judicial Eletrônico (PJE). A aplicação de IA pode reduzir o tempo de tramitação de processos e contribuir para a eficiência jurisdicional, com isso alinhando-se às diretrizes do Conselho Nacional de Justiça (CNJ) para a modernização da Justiça.

ABSTRACT – This paper explores the use of text mining in the legal sector, focusing on the identification of addresses in adverse possession cases. The research employs Natural Language Processing (NLP) to structure information extracted from initial petitions, optimizing document analysis. Using 10 initial petitions sourced from the TJPE database and analyzed with the GPT-4.o-mini model, complete data extraction was achieved in 80% of cases, demonstrating the technology's efficiency while also highlighting its limitations.

In the legal context, the study underscores the significance of adverse possession for land regularization and the necessity of automation in the Electronic Judicial Process (PJe). The application of AI can reduce case processing time and enhance judicial efficiency, thereby aligning with the National Justice Council's (CNJ) guidelines for modernizing the judiciary.

INTRODUÇÃO

O avanço acelerado das tecnologias de inteligência artificial (IA) tem transformado profundamente os processos de trabalho em diversas áreas do conhecimento humano, promovendo maior eficiência, automação e precisão. No setor jurídico, marcado pela intensa produção de documentos textuais e informações não estruturadas, a aplicação de IA apresenta um potencial significativo para otimizar a análise e o processamento de dados, reduzindo atrasos e aumentando a eficácia das decisões.

Entre as tecnologias de IA, destaca-se o processamento de linguagem natural (PLN), definido por Guimarães (2019) como "um campo interdisciplinar que se concentra em permitir que computadores compreendam, interpretem e gerem linguagem humana". Essa tecnologia permite que máquinas analisem e interpretem textos de maneira semelhante ao raciocínio humano, alinhando-se ao conceito de interação homem-máquina delineado por Alan Turing (1950). A mineração de texto, uma aplicação prática do PLN, tem se mostrado uma ferramenta poderosa para extrair padrões, estruturar informações e gerar insights a partir de grandes volumes de dados textuais.

A área jurídica enfrenta desafios significativos no processamento de documentos como petições, contratos e decisões judiciais. Nesse contexto, a mineração de texto aplicada ao direito facilita a identificação de padrões, a extração de informações e a classificação de processos. Um exemplo prático é o sistema Elis, desenvolvido pelo Tribunal de Justiça de Pernambuco (TJPE), que automatiza a análise de certidões de dívida ativa (CDAs) e petições iniciais. Segundo o TJPE (2025), o Elis reduz prazos de análise de 18 meses para apenas 15 dias, evidenciando sua eficiência e impacto positivo na prestação jurisdicional.

Além disso, o TJPE (2015) destaca diversas aplicações da IA no setor jurídico, como triagem de processos judiciais, previsão de reincidência criminal, análise de contratos e identificação de advocacia predatória. Essas ferramentas não apenas

melhoram a produtividade e a celeridade, mas também liberam os operadores do direito para se concentrarem em atividades estratégicas.

O processo judicial é o instrumento pelo qual um cidadão (ou grupo de pessoas) requer formalmente ao Poder Judiciário que julgue determinada causa.

Atualmente o processo judicial tramita de forma eletrônica pelo sistema Processo Judicial Eletrônico (PJE), ferramenta desenvolvida pelo Conselho Nacional de Justiça (CNJ), utilizado em 13 tribunais estaduais. No entanto, esse sistema ainda apresenta limitações importantes no que diz respeito à estruturação dos dados de identificação do endereço dos imóveis, impossibilitado a automação de atividade e decisões, especialmente em processos de usucapião.

As informações relacionadas ao endereço são inseridas no texto da petição inicial do processo de usucapião. Essa falta de organização dificulta o acesso rápido e preciso a dados, condição essencial para a análise de dados e automatização de processos operacionais.

Este artigo explora o potencial da mineração de texto para analisar petições iniciais em processos judiciais de usucapião que tramitam pelo PJE, com foco na identificação e estruturação de informações sobre o endereço do imóvel.

O objetivo é demonstrar como essa tecnologia pode contribuir para a análise de dados, contribuindo para a tomada de decisões estratégicas, ações que visem a celeridade dos processos, como por exemplo mutirões de regularização fundiária, contribuindo para a melhora da qualidade do serviço jurisdicional, promovendo a eficiência operacional e subsidiando os auxiliando magistrados com informações mais assertivas.

Vale salientar que não é a pretensão desse artigo propor a utilização de tecnologias de IA em substituição a decisão do magistrado, mas avaliar o uso de tecnologias de inteligência artificial como apoio ao processo jurisdicional.

Destacamos também a iniciativa em desenvolvimento pelo Supremo Tribunal Federal (STF), que utiliza tecnologias de mineração de texto para otimizar o

processamento de informações judiciais. Segundo o Conselho Nacional de Justiça (CNJ, 2024), o projeto envolve uma inteligência artificial capaz de receber dados de processos judiciais e gerar resumos contendo informações das instâncias percorridas pelo processo.

A modernização do sistema judicial é uma prioridade do Conselho Nacional de Justiça (CNJ), que destaca a eficiência como um fator essencial para atender à crescente demanda por Justiça no Brasil.

METODOLOGIA

FUNDAMENTAÇÃO TEÓRICA

A composição de uma peça jurídica consiste, tradicionalmente, na demanda inicial e nos dispositivos jurídicos anexos ao processo. Durante décadas, esses documentos foram elaborados e organizados fisicamente, em papel. Com a informatização dos processos judiciais, essa dinâmica foi transposta para o meio digital, mantendo, no entanto, o mesmo *modus operandi*. A anexação de documentos passou a ser realizada virtualmente, transformando o processo eletrônico em um registro de dados predominantemente semiestruturados.

Apesar de facilitar a gestão logística, esse modelo apresenta limitações significativas, especialmente no que se refere à análise de dados processuais. A ausência de estruturação impede a extração eficiente de informações relevantes que poderiam melhorar a celeridade processual e a eficácia das decisões judiciais.

Nesse contexto, a mineração de dados se mostra uma ferramenta promissora. Ao estruturar informações de processos judiciais, possibilita a identificação de padrões e tendências, auxiliando na formulação de políticas públicas e na tomada de decisões com maior grau de assertividade. Sua aplicação no Direito, especialmente no campo da regularização fundiária, é estratégica para otimizar operações e atender às demandas crescentes por justiça eficiente e acessível.

Em destaque são apresentadas as tecnologias empregadas neste trabalho, bem como o conceito de usucapião para melhor entendimento da discussão envolta no artigo.

MINERAÇÃO DE TEXTO E PROCESSAMENTO DE LINGUAGEM NATURAL

A mineração de texto pode ser definida como o processo ou prática de examinar grandes coleções de recursos escritos para gerar novas informações. Segundo a definição apresentada pelo *Oxford Dictionary*, trata-se de “o processo ou prática de examinar grandes coleções de recursos escritos para gerar novas

informações”. Essa técnica visa identificar informações relevantes contidas em textos, analisando padrões linguísticos e transformando-os em dados estruturados.

De acordo com Fernandes (2019), o Processamento de Linguagem Natural (PLN; em inglês, *Natural Language Processing*) é uma área de estudo da Ciência da Computação que investiga as capacidades e limitações das máquinas em compreender a linguagem humana. O PLN combina conhecimentos das áreas de Computação e Linguística, com o propósito de, por meio do estudo do funcionamento do cérebro humano, aproximar os computadores da linguagem natural dos seres humanos.

O Processamento de Linguagem Natural (PLN) utiliza, segundo Caelen e Blete (2024), uma classe de algoritmos conhecida como aprendizado profundo *(deep learning*, em inglês). Esses algoritmos, inspirados na estrutura do cérebro humano, são chamados de redes neurais artificiais.

O PLN é um componente essencial na mineração de texto, realizando análises linguísticas que permitem às máquinas "lerem" textos de forma eficiente. Essa área utiliza metodologias diversas para lidar com as ambiguidades inerentes à linguagem humana, como sumarização automática, marcação de classes gramaticais, desambiguação, extração de entidades e de relações, além da compreensão e reconhecimento de linguagem natural (expert.ai, 2024).

O desafio da programação de linguagem natural é a ambiguidade dos termos linguísticos e a diversificação termos utilizados para expressar o mesmo significado. Não obstante a utilização de expressões em latim, tão comum no jargão jurídico.

GPT-4

Entre as ferramentas modernas que aplicam conceitos avançados de PLN está o GPT-4, desenvolvido pela OpenAI. Segundo Stefanovskyi (2024), o GPT é um *Large Language Model* (LLM) projetado para entender e gerar textos semelhantes à linguagem humana com base em entradas fornecidas. Demonstra grande utilidade no processo de mineração de texto.

Segundo Caelen (2024):

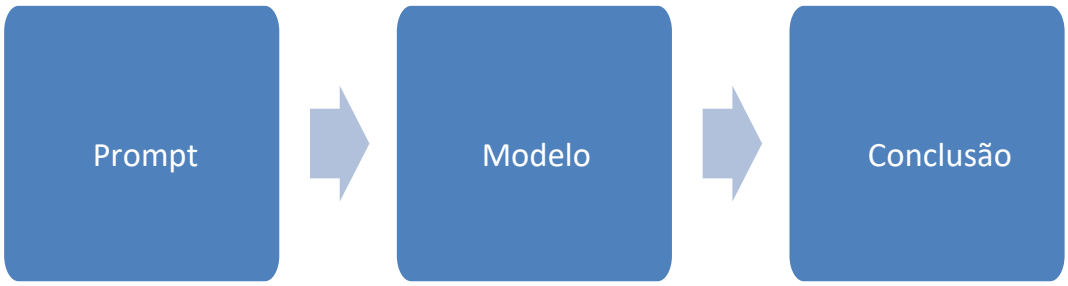
> O GPT-4 foi baseado em um tipo específico de algoritmo de aprendizado profundo chamado transformadores. Os transformadores são como máquinas de leitura. Eles prestam atenção em diferentes partes da frase ou bloco de texto para entender seu contexto e produzir uma resposta coerente. Também conseguem entender a ordem das palavras em uma frase e seu contexto. (CAELEN, 2024)

Apoiar a extração de entidades e relações: Sua capacidade de compreensão contextual permite identificar entidades específicas e suas conexões em textos extensos e não estruturados.

Fornecer respostas contextualizadas: O modelo consegue responder a consultas específicas, adaptando as respostas ao contexto fornecido pelo usuário, tornando-o uma ferramenta versátil na análise de dados textuais.

O processo de utilização do modelo LLM é apresentado pela Figura 1.

Figura 1 - Diagrama de processo.



Fonte: O autor (2025).

De acordo com Caelen e Blete (2024), os modelos da família GPT recebem um *prompt* como entrada e, como resposta, geram um texto, em um processo conhecido como conclusão de texto.

Dessa forma, GPT-4 apresenta-se como uma poderosa ferramenta de instrumentalização na mineração de texto, alinhando-se às práticas de PLN para otimizar a análise e interpretação de informações em grande escala.

Direito à Moradia e o Processo de Usucapião

O direito à moradia digna ganhou relevância mundial após a Segunda Guerra Mundial, sendo enfatizado na Declaração Universal dos Direitos Humanos, no artigo 25, parágrafo 1º, que reconhece o direito de todos a um padrão de vida adequado. No Brasil, essa garantia foi incorporada à Constituição Federal de 1988, como um direito fundamental no artigo 6º e reforçado nos termos do artigo 5º, que assegura a dignidade humana como um dos pilares constitutivos.

Nesse contexto, a regularização fundiária surge como um importante instrumento para concretizar o direito à moradia, especialmente diante da expressiva informalidade habitacional no Brasil, que compromete a dignidade das pessoas. Como enfatizam Freiria e Dosso (2021), “morar irregularmente é o mesmo que navegar em permanente insegurança”, citando o jurista José Renato Nalini.

Destaca Bazilio (2022):

> A usucapião é uma forma de aquisição da propriedade imóvel, sendo que pode ser requerida em caso em que o possuidor seja residente do imóvel, em regra, fazendo uso constante por mais de 10 anos (o prazo pode variar conforme o tipo de usucapião), sem qualquer reclamação do proprietário original. Portanto, o possuidor que requerer os direitos de posse da propriedade não pode tê-los

> obtido de forma violenta, clandestina ou ilegal. (BAZILIO, 2022)

A usucapião, por sua vez, desempenha um papel central na regularização fundiária. Segundo Costa (1999), tratasse de um modo autônomo de aquisição de propriedade, aplicável tanto a bens móveis quanto imóveis, desde que se cumpram os requisitos legais relacionados à posse qualificada pelo período estipulado em lei. A origem etimológica do termo deriva do latim *usus* (posse) e *capio, capere* (tomar, adquirir), ou seja, adquirir pela posse prolongada.

Conforme Freiria e Dosso (2021), a usucapião apresenta-se como uma ferramenta jurídica para transformar situações de ocupação informal em regularidade, permitindo que indivíduos e famílias conquistem a titularidade do imóvel que ocupam, promovendo, assim a dignidade e a segurança jurídica.

O crescimento desordenado das cidades tem provocado o fenômeno da ocupação de áreas sem a devida formalização legal, criando um cenário em que o poder público não consegue atender às necessidades mínimas exigidas pela sociedade moderna. O acesso ao saneamento básico e à energia elétrica são condições fundamentais para garantir uma vida digna.

A formalização da posse da moradia é o primeiro passo para dar visibilidade ao contingente populacional que vive na obscuridade da informalidade.

Segundo Lobo (2024):

> Também há de se destacar as políticas habitacionais inadequadas e a especulação imobiliária. Quando áreas nas quais comunidades se assentam de forma irregular tornam-se mais valorizadas, passam a figurar como objeto de desejo dos especuladores imobiliários, o que faz com que a pressão pela expulsão dos assentados

> aumente drasticamente e se efetive, na maior parte das vezes. (LOBO, 2024)

ESTUDO DE CASO: APLICAÇÃO DA MINERAÇÃO DE TEXTO EM PROCESSOS DE USUCAPIÃO

Segundo Ueno e Soares (2024), na jornada da inteligência artificial (IA), a análise de provas de conceito (*Proofs of Concept*, POCs), os testes de ferramentas e a adoção de projetos piloto podem ser estratégias eficazes antes de escalar um projeto de grande magnitude.

Para demonstrar a aplicação da mineração de texto no contexto de identificação de endereços em processos de usucapião, desenvolvemos uma aplicação que analisou 10 processos de usucapião selecionados de forma aleatória arquivados na base de dados do TJPE.

Com destaque para os desafios da mineração de texto, Alura (2024) enfatiza que:

> ... um ponto importante é a ambiguidade como característica comum na linguagem natural. Por isso, temos uma etapa de desambiguação, na qual o objetivo é tentar resolver as ambiguidades e determinar o significado mais apropriado com base no contexto. Assim, podemos realizar a extração de informações específicas, como nomes de pessoas, datas, localizações, eventos importantes etc. (ALURA, 2024)

A interação com o modelo foi realizada por meio da linguagem de programação Python, utilizando o ambiente de desenvolvimento web *Google Colab* para execução do código, a biblioteca *OpenAI* e o modelo GPT-4o-mini. O objetivo foi extrair, de

maneira estruturada, os dados relacionados aos imóveis, incluindo logradouro, número, cidade, estado e CEP.

Segundo Caelen (2024), a *OpenAI* disponibiliza o GPT-4 e o *ChatGPT* como serviços. Isso implica que os usuários não têm acesso direto ao código-fonte dos modelos, nem podem executá-los em seus próprios servidores. Desta forma é necessário a utilização de (API) para a iteração com o modelo GPT, exigindo do usuário uma chave de acesso.

O processo de extração dos dados do endereço do imóvel foi conduzido por meio do seguinte prompt: “Atue como analista judiciário e analise o texto a seguir e identifique a localização do imóvel a ser usucapido de forma estruturada separando os dados em: logradouro ou loteamento, número, bairro, cidade, CEP e estado.”

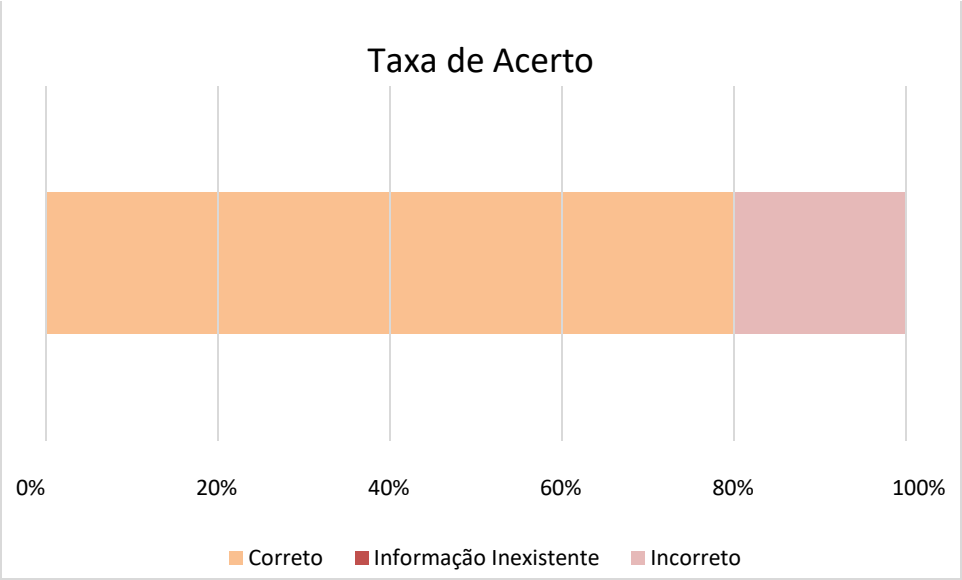
Mediante a definição do prompt foi anexado o texto da petição inicial do processo de usucapião e submetido a análise do modelo GTP.

RESULTADOS

Após a execução da mineração de texto sobre as 10 petições iniciais, os resultados foram categorizados em três grupos: (1) Identificação Correta; (2) Informação Inexistente e (3) Dados incorretos. Sendo assim a classificação correta se refere as informações constantes na petição inicial com os dados do imóvel usucapiendo e a mesma selecionada pela IA. Para a classificação de informação inexistente se aplica a definição que a IA não conseguiu definir o endereço do imóvel. Contudo, a classificação de dados incorretos compreende total divergência entre os dados de endereço constantes na petição e os dados definidos pela IA para o imóvel.

Esse detalhamento permitiu avaliar com precisão a eficiência do modelo e identificar limitações específicas, conforme Figura 2.

Figura 2 - Resultado da Extração dos Endereços



Fonte: O autor (2025).

Endereços identificados de forma correta: Em 8 dos 10 processos, os dados de logradouro, número, cidade, estado e CEP foram extraídos de forma como definido na petição inicial.

Endereços incorretos: Em 2 casos, os endereços estavam descritos nas petições, porém o modelo não conseguiu interpretá-los de forma estruturada, identificando outro endereço constante na petição como sendo o do imóvel usucapiendo.

Informação Inexistente: Não houve ocorrência de casos em que a aplicação não encontrou dados referentes a endereço de imóvel.

CONCLUSÃO

Os resultados demonstram a eficiência do modelo na extração de informações estruturadas a partir de textos jurídicos não estruturados. Ao analisar os dois casos em que a aplicação não conseguiu definir de forma correta o endereço do imóvel objeto da ação verificou-se que se tratava de imóveis localizados na zona rural em que a definição da localização do imóvel estava baseada em coordenadas geográficas e pontos de referência.

Este estudo serviu como prova de conceito para a utilização da tecnologia de mineração de texto, comprovando que o emprego da tecnologia impacta significativamente o processo de tomada de decisão auxiliando os operadores do Direito na busca por eficiência.

Apesar do aparente sucesso em 80% dos casos as limitações identificadas em 20% dos exemplos destacam a importância de fornecer textos bem formatados e claros, além de possíveis ajustes no prompt para melhorar o desempenho do modelo.

Esse estudo de caso evidencia o potencial da mineração de texto como ferramenta para agilizar e otimizar a análise documental especialmente em processos de usucapião onde a definição clara do objeto de pesquisa, no caso os dados o endereço do imóvel usucapiendo, facilita a análise computacional, oferecendo soluções práticas para a organização e padronização de dados que servirão de base para iniciativas de ações visando promover justiça social.

As limitações identificadas no estudo abrem caminhos para pesquisas futuras incluindo o desenvolvimento de algoritmos mais robustos e a criação de padrões textuais nos sistemas de tramitação eletrônica. A integração dessas soluções com sistemas o Processo Judicial Eletrônico (PJE) tem o potencial de revolucionar a administração da Justiça no Brasil.

Desta forma, evidenciamos que a adoção de tecnologia de inteligência artificial não deve ser encarada como uma inovação futurística, mas como uma ferramenta real de apoio ao poder judiciário que possibilita transformar a base de conhecimento

em um ativo estratégico capaz de apoiar tomadas de decisões e promover um judiciário operacionalmente mais eficiente.

PROPOSTA DE TRABALHO FUTURO

Este trabalho propõe o desenvolvimento de um algoritmo para a extração automática de endereços contidos em processos de usucapião. Como extensão dessa iniciativa, será criada uma base de dados estruturada que integrará os metadados dos processos, os dados de endereço e as respectivas coordenadas geográficas. O objetivo principal é possibilitar o mapeamento preciso dos imóveis localizados em áreas de interesse social ou em situação de risco, contribuindo para a formulação de políticas públicas voltadas à regularização fundiária e ao planejamento urbano.

Como benefício colateral, este trabalho também colabora para a desjudicialização dos processos de regularização fundiária em imóveis situados em zonas de interesse social (ZEIS), especialmente quando a regularização segue uma tramitação extrajudicial, evitando com isso a judicialização desnecessária de um processo de usucapião. Além disso, contribui para a agilização dos processos de usucapião já judicializados, tornando sua tramitação mais célere e auxiliando de forma significativa no alcance das metas de qualidade propostas pelo Conselho Nacional de Justiça (CNJ).

REFERÊNCIAS

ALURA. **O que é PLN?** Disponível em: https://www.alura.com.br/artigos/o-que-e-pln. Acesso em: 13 jan. 2025.

BAZILIO, JAQUELINE FERNANDES. **Como funciona o processo de usucapião judicial**. Disponível em: https://www.jusbrasil.com.br/artigos/como-funciona-o-processo-de-usucapiao-judicial/1369531396. Acesso em: 13 jan. 2024.

BRASIL. CONSTITUIÇÃO (1988**). Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, 1988.

CAELEN, O.; BLETE, M.-A. **Desenvolvendo aplicativos com GPT-4 e ChatGPT**. São Paulo: Novatec, 2024.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Programa Justiça 4.0 divulga resultados de pesquisa sobre IA no Judiciário brasileiro**. Disponível em: https://www.cnj.jus.br/programa-justica-4-0-divulga-resultados-de-pesquisasobre-ia-no-judiciario-brasileiro/. Acesso em: 13 jan. 2025.

COSTA, DILVANIR JOSÉ DA. **Usucapião: conceitos e fundamentos jurídicos**. Disponível em: https://www2.senado.leg.br/bdsf/bitstream/handle/id/524/r143-25.PDF. Acesso em: 13 jan. 2024.

EXPERT.AI. **Natural Language Processing and Text Mining**. Disponível em: https://www-expertai.translate.goog/blog/natural-language-processing-and-text-mining/. Acesso em: 13 jan. 2024.

FREIRIA, R. C.; DOSSO, T. C. **Direito Agrário**. 3. ed. Salvador: Juspodivm, 2021.

FERNANDES, Vitor. **Processamento de Linguagem Natural**. Disponível em: https://www.each.usp.br/petsi/jornal/?p=2577. Publicado em: 12 ago. 2019. Acesso em: 13 jan. 2024.

NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. [S.l.]: Nações Unidas, 1948. Disponível em: https://www.un.org/. Acesso em: 13 jan. 2024.

STEFANOVSKYI, OLEKSANDR. **Como o ChatGPT da OpenAI funciona**. Intelliarts. Disponível em: https://intelliarts-com.translate.goog/authors/oleksandr-stefanovskyi/. Acesso em: 13 jan. 2024.

UENO, GISELE; SOARES, JÚLIO CÉSAR. **A Jornada de IA como Estratégia no Escritório Tributário**. In: COELHO, Alexandre Zavaglia; BARBOSA, Maria Juliana do P. (Org.). Inteligência Artificial Aplicada aos Serviços Jurídicos. São Paulo: Thomson Reuters, 2024.