Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL

FACULDADE DE COMPUTAÇÃO

MESTRADO PROFISSIONAL EM COMPUTAÇÃO APLICADA

LIRIANE APARECIDA DA SILVA NOGUEIRA

MODELO DE INTELIGÊNCIA ARTIFICIAL PARA IDENTIFICAR MEDICAMENTOS EM PROCESSOS DE JUDICIALIZAÇÃO NO TRIBUNAL DE JUSTIÇA DO ESTADO DE MATO GROSSO DO SUL

Campo Grande - MS

Junho de 2025

Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul

LIRIANE APARECIDA DA SILVA NOGUEIRA

MODELO DE INTELIGÊNCIA ARTIFICIAL PARA IDENTIFICAR MEDICAMENTOS EM PROCESSOS DE JUDICIALIZAÇÃO NO TRIBUNAL DE JUSTIÇA DO ESTADO DE MATO GROSSO DO SUL

Dissertação apresentada para o curso de Mestrado Profissional em Computação Aplicada da Faculdade de Computação FACOM-UFMS. Área de concentração: Tecnologias Computacionais para Cidades Inteligentes. Linha de Pesquisa: Sistemas computacionais aplicados aos serviços públicos.

Orientador: Prof. Dr. Marcelo Augusto Santos Turine

Coorientadora: Dra. Joseliza Alessandra Vanzela Turine

Campo Grande - MS

Junho, 2025

TERMO DE APROVAÇÃO

LIRIANE APARECIDA DA SILVA NOGUEIRA

MODELO DE INTELIGÊNCIA ARTIFICIAL PARA IDENTIFICAR MEDICAMENTOS EM PROCESSOS DE JUDICIALIZAÇÃO NO TRIBUNAL DE JUSTIÇA DO ESTADO DE MATO GROSSO DO SUL

Dissertação apresentada para o curso de Mestrado Profissional em Computação Aplicada da Faculdade de Computação da Universidade Federal de Grosso do Sul, como requisito parcial da obtenção do título de Mestre. Área de concentração: Tecnologias Computacionais para Cidades Inteligentes. Linha de pesquisa: Sistemas computacionais aplicados aos serviços públicos

Campo Grande - MS, 18 de Junho de 2025.

COMISSÃO EXAMINADORA

_____

Prof. Dr. Marcelo Augusto Santos Turine

Universidade Federal de Mato Grosso do Sul - UFMS

_____

Prof. Dr. Bruno Magalhães Nogueira

Universidade Federal de Mato Grosso do Sul - UFMS

_____

Desembargador Dr. Odemilson Roberto Castro Fassa

Tribunal de Justiça do Estado de Mato Grosso do Sul - TJMS

À minha filha, meus pais, meus orientadores e a Deus.

AGRADECIMENTOS

Ao meu orientador Marcelo Turine e a co-orientadora Joseliza pelo incentivo incansável, pelas orientações tão valiosas e dedicação ao meu projeto.

À UFMS, em especial à FACOM, e também ao TJMS, pela oportunidade de desenvolver este projeto e adquirir conhecimentos tão transformadores.

À minha filha, pela paciência durante o tempo que me ausentei para cumprir com as obrigações de discente.

Ao Douglas, meu colega de curso, que me ajudou com muitas dúvidas sobre diversos assuntos no decorrer do mestrado e me incentivou a continuar quando quase desisti.

Aos meus amigos e colegas de trabalho, em especial ao Ademar, Diego, Eder, Gilliard, Gustavo, João Paulo, Marlon, Roberto e Roní, pelo incentivo e colaboração na construção do projeto.

A todos que me ajudaram, obrigada.

RESUMO

A judicialização da saúde no Brasil tem gerado desafios significativos para o Sistema Único de Saúde (SUS), especialmente em relação ao fornecimento de medicamentos. No Tribunal de Justiça do Estado de Mato Grosso do Sul (TJMS), o crescente volume de processos relacionados a demandas por medicamentos impacta as políticas públicas e exige soluções tecnológicas inovadoras. Com a consolidação do processo judicial eletrônico, há muitas informações a serem exploradas nos textos processuais que, ao serem extraídas e organizadas, tornam uma poderosa ferramenta de gestão e governança. Este trabalho tem como objetivo aplicar um modelo de inteligência artificial (IA) capaz de identificar automaticamente os medicamentos mencionados nas petições iniciais de processos judiciais, utilizando técnicas de mineração de textos e aprendizado de máquina com foco em documentos de processos judiciais redigidos em língua portuguesa. O modelo foi treinado com algoritmos baseados em *transformers*, em especial o BioBERTpt. Após gerar os dados, foi implementada uma plataforma computacional interativa intitulada MED-SUS-MS, que visualiza os dados da judicialização da saúde no TJMS a fim de contribuir para tomada de decisão nas políticas públicas de fornecimento de medicamentos no Estado de Mato Grosso do Sul. A metodologia envolveu a construção de um *dataset* anonimizado, a aplicação de técnicas de reconhecimento de entidades nomeadas (*Named Entity Recognition* – NER) e a avaliação do desempenho do modelo por meio de métricas como precisão, recall e F1-score. Os resultados demonstram a viabilidade técnica e a relevância institucional do modelo proposto, permitindo maior transparência, agilidade e fundamentação na formulação de políticas públicas de saúde. O modelo está alinhado às diretrizes e às resoluções do CNJ e TJMS, promovendo governança ética e segura no Judiciário, contribuindo diretamente para os Objetivos de Desenvolvimento Sustentável (ODS) 3 - Saúde e Bem-Estar e 16 - Instituições Eficazes. A pesquisa reforça o potencial da Inteligência Artificial como ferramenta estratégica para a desjudicialização da saúde, ao transformar dados judiciais em insumos qualificados para a gestão e governança pública.

Palavras-chave: inteligência artificial, reconhecimento de entidade nomeada, mineração de medicamentos, judicialização da saúde e sistema único de saúde.

ABSTRACT

The healthcare judicialization in Brazil has posed significant challenges to the Unified Health System (SUS), especially with respect to the provision of medicines. At the Court of Justice of the State of Mato Grosso do Sul (TJMS), the growing number of lawsuits related to demands for medicines impacts public policies and requires innovative technological solutions. With the consolidation of the electronic case files, there is a wealth of information to be explored in procedural texts which, once extracted and organized, becomes a powerful tool for management and governance. This study aims to apply an artificial intelligence (AI) model capable of automatically identifying medicines names mentioned in the initial petitions of judicial records, using text-mining and machine-learning techniques focused on case files documents written in Portuguese. The model was trained with transformer-based algorithms, particularly BioBERTpt. After generating the data, an interactive computational platform called MED-SUS-MS was implemented to visualize data related to the healthcare judicialization at TJMS, in order to support decision-making in public policies for medicines provision in the State of Mato Grosso do Sul. The methodology involved constructing an anonymized dataset, applying Named Entity Recognition (NER) techniques, and evaluating the model's performance using metrics such as precision, recall, and F1-score. The results demonstrate the technical feasibility and institutional relevance of the proposed model, enabling greater transparency, agility, and evidence-based formulation of public health policies. The model is aligned with the guidelines and resolutions of the National Council of Justice (CNJ) and TJMS, promoting ethical and secure governance within the Judiciary and directly contributing to Sustainable Development Goals (SDGs) 3 – Good Health and Well-being and 16 – Peace, Justice and Strong Institutions. This research reinforces the potential of artificial intelligence as a strategic tool for the healthcare de-judicialization by transforming judicial data into qualified inputs for public management and governance.

Keywords: artificial intelligence, named entity recognition, medicines mining, healthcare judicialization and unified health system.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

2VP - 2ª Vice-presidência

ACM - *Association for Computing Machinery* (Associação para Maquinaria da Computação)

ADABOOST - *Adaptive Boosting* (reforço adaptável)

ANVISA - Agência Nacional de Vigilância Sanitária

ART - Artigo

ASCII - *American Standard Code for Information Interchange* (Código Padrão Americano para o Intercâmbio de Informação)

AUC - *Area Under The Curve* (área sob a curva)

BioBERT - *Bidirectional Encoder Representations for Transformer*s (representações codificadoras bidirecionais de transformadores pré-treinado para área biomédica)

CAERD - Companhia de Águas e Esgotos de Rondônia

CDA - Certidão de Dívida Ativa

CNJ - Conselho Nacional de Justiça

CNN - *Convolutional Neural Network* (Rede Neural Convolucional)

COVID - (CO)rona (VI)rus (D)isease

DETRAN - Departamento de Trânsito do Estado de Mato Grosso do Sul

DPVAT - Danos Pessoais por Veículos Automotores Terrestres

FACOM - Faculdade de Computação

GRU - *Gated Recurrent Unit* (Unidades recorrentes fechadas)

HDBSCAN - *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (agrupamento espacial baseado em densidade hierárquica de aplicativos com ruído)

HTML - *HyperText Markup Language* (Linguagem de Marcação de Hipertexto)

IA - Inteligência Artificial

IEEE - Instituto de Engenheiros Eletricistas e Eletrônicos

KNN - *K-nearest neighbors* (k-vizinhos mais próximos)

LRC - *Legal Reading Comprehension* (Compreensão de Leitura Jurídica)

Lsa250 - *Latent Semantic Analisys* (análise semântica latente)

LSTM - *Long Short-Term Memory* (memória de longo prazo)

MS - Mato Grosso do Sul

NLP - *Natural Language Processing* (Processamento de Linguagem Natural)

ODS - Objetivos de Desenvolvimento Sustentável

OMS - Organização Mundial de Saúde

PDF - *Portable Document Format* (Formato Portátil de Documento)

PoS - *Parts of Speech* (partes do discurso)

RF - *Random Forest* (Floresta Aleatória)

RNN - *Recurrent Neural Network* (rede neural recorrente)

SBC - Sociedade Brasileira de Computação

STF - Supremo Tribunal Federal

STJ - Superior Tribunal de Justiça

SUS - Sistema Único de Saúde

SVM - *Support Vector Machine* (máquina de vetores de suporte)

TF-IDF - *Term Frequency–Inverse Document Frequency* (frequência do termo–inverso da frequência)

TJMS - Tribunal de Justiça do Estado de Mato Grosso do Sul

TJBA - Tribunal de Justiça do Estado da Bahia

TJBALABJUS - Laboratório de Inovação e Inteligência do Poder Judiciário do Estado da Bahia

TJMA - Tribunal de Justiça do Estado do Maranhão

TJMS - Tribunal de Justiça do Estado de Mato Grosso do Sul

TJRO - Tribunal de Justiça do Estado de Rondônia

TPU - Tabela Processual Unificada

TRE-BA - Tribunal Regional Eleitoral da Bahia

TRT9 - Tribunal Regional do Trabalho da 9ª Região

UFMS - Universidade Federal de Mato Grosso do Sul

ULMFiT - *Universal Language Model Fine-tuning* (Ajuste fino do modelo de linguagem universal)

UMAP - *Uniform Manifold Approximation and Projectio*n (Aproximação e Projeção de Coletores Uniformes)

VGG - *Visual Geometry Group* (Grupo de Geometria Visual)

XGBoost - *eXtreme Gradient Boosting* (reforço de gradiente extremo)

XLNet - Técnica de Pré-treinamento

SUMÁRIO

CAPÍTULO 1

INTRODUÇÃO

1.1. Considerações Iniciais

Nos últimos anos, o Judiciário brasileiro enfrenta desafios para garantir celeridade e eficiência diante da alta demanda processual e recursos escassos. A adoção de tecnologias digitais tem sido essencial para modernizar a Justiça, promovendo avanços significativos na gestão processual, na comunicação entre os atores jurídicos e no acesso da população aos serviços judiciais. O Conselho Nacional de Justiça (CNJ), entidade de supervisão e transparência do sistema judiciário brasileiro, por meio de suas resoluções, programas e diretrizes estratégicas, tem desempenhado um papel central na indução dessa transformação digital, estimulando tribunais de todo o país a adotarem soluções tecnológicas inovadoras que melhorem o desempenho institucional e a qualidade da prestação jurisdicional.

O CNJ publica, anualmente, o "Relatório Justiça em Números" que detalha e oferece transparência sobre a performance, finanças, acesso à justiça, e uma variedade de indicadores processuais do Poder Judiciário. De acordo com o relatório de 2024 [CNJ 2024a], última versão disponibilizada, o judiciário encerrou 2023 com 83,8 milhões de processos pendentes, 7,4% a mais que os 77,1 milhões registrados cinco anos antes. A força de trabalho contou com um aumento modesto de menos

de 1% no número de magistrados, totalizando 18.265, e um decréscimo 1% no quadro de servidores, totalizando 275.581, o que não correspondeu proporcionalmente ao aumento de demandas.

O Tribunal de Justiça do Estado de Mato Grosso do Sul (TJMS) terminou o ano de 2023 com 1.109.764 processos pendentes, refletindo a tendência nacional. Com 219 magistrados e 5.258 servidores, o TJMS enfrenta desafios equivalentes. Esses dados impulsionam os tribunais brasileiros a buscar inovações em tecnologia da informação e comunicação, visando melhorar o desempenho na redução dos casos pendentes de julgamento. Desde 2005, o TJMS vem implementando processos eletrônicos em todas as suas unidades judiciais, o que possibilitou a automação de tarefas repetitivas e a aceleração do trâmite processual.

A busca por eficiência na gestão dos processos judiciais é constante, reforçada por um princípio que deve ser obedecido pela administração pública em cumprimento à Constituição Federal Brasileira [Brasil 1988] , disposto no artigo 37. Por conseguinte, os membros e a administração do TJMS têm buscado o apoio na evolução da ciência para acompanhar o ritmo de produtividade perante o aumento do volume de trabalho a fim de ser mais eficiente. Diante da maturidade alcançada nas fases iniciais supramencionadas, o TJMS está avançando em soluções de análise e de reflexão dos seus dados, sobretudo naqueles disponíveis em textos de documentos que compõem os processos judiciais, que estão disponíveis em sua totalidade no formato digital, por meio da utilização de modelos de inteligência artificial para mineração desses textos.

A despeito das inúmeras iniciativas reportadas pelo CNJ, presentes na SINAPSES [de Justiça 2022], a plataforma nacional de armazenamento e distribuição de modelos de inteligência artificial para o Judiciário brasileiro, criada pela Resolução nº 332/2010 do CNJ [CNJ 2020], tem muitas informações processuais não exploradas ainda. Especificamente, a análise detalhada dos dados por meio da leitura e interpretação de textos processuais judiciais que oferecem possibilidades amplas.

Neste contexto, este trabalho tem como objetivo extrair informações relevantes das petições iniciais em processos de judicialização da saúde, com foco na identificação dos medicamentos pleiteados junto ao Sistema Único de Saúde (SUS). O objetivo é subsidiar políticas públicas que contribuam para a redução da judicialização nessa área. Além disso, a proposta está alinhada aos Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030 da Organização das Nações Unidas - ONU [das Nações Unidas ONU 2015], contribuindo diretamente para o ODS 16 - Paz, Justiça e Instituições Eficazes, ao mesmo tempo em que apoia o

ODS 3 - Saúde e Bem-Estar, gerando impacto social relevante por meio da promoção do acesso equitativo à saúde e da melhoria da governança pública.

A necessidade de contribuir com os ODS's, sob a perspectiva de soluções tecnológicas, encontra fundamento na busca pela redução do tempo de espera por medicamentos necessários à garantia do acesso à saúde, a serem fornecidos pelo SUS, previsto no artigo 6º da Constituição Federal de 1988 [Brasil 1988], seja com a previsibilidade de demanda, possível através do conhecimento dos medicamentos requeridos por meio da justiça nas petições iniciais, seja pela desjudicialização deles, com a busca de resolução de forma administrativa daqueles medicamentos mais demandados, reduzindo assim, os casos pendentes de julgamento nesta área do direito.

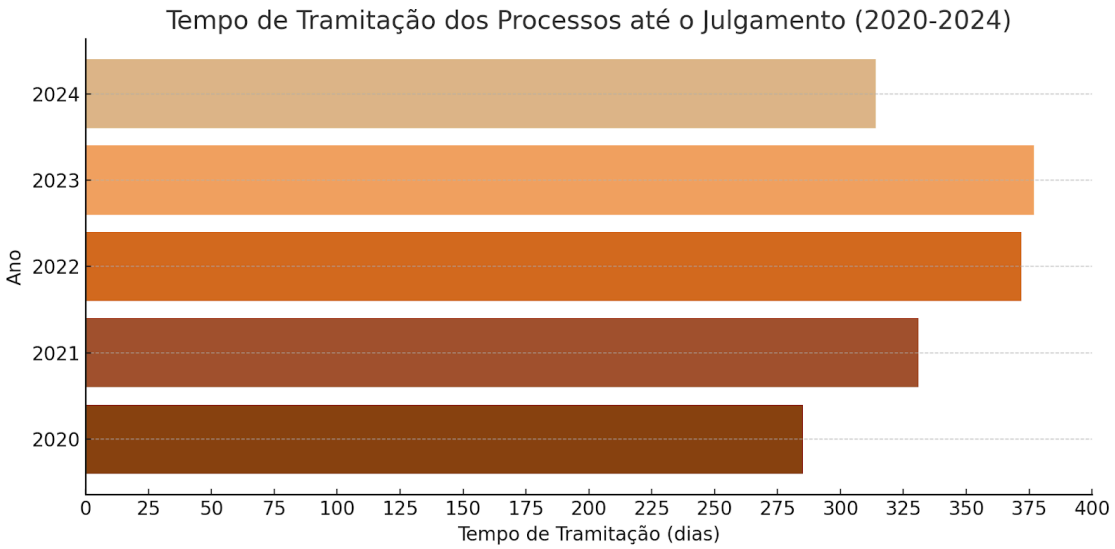
> Art. 6º São direitos sociais a educação, a saúde, o trabalho, o lazer, a segurança, a previdência social, a proteção à maternidade e à infância, a assistência aos desamparados, na forma da Constituição Federal do Brasil [Brasil 1988].

Em 2023, o número de processos relacionados ao fornecimento de medicamentos pelo sistema público de saúde, conforme os códigos 12484, 12492, 12493, 12494, 12495 e 12496 da Tabela Processual Unificada do CNJ, foi de 49,3 mil, segundo dados consolidados no Painel de Estatísticas Processuais de Direito à Saúde [CNJ 2023b]. Nesse mesmo período, o tempo médio de espera por decisão definitiva nesses casos alcançou 425 dias. Com o objetivo de aprimorar a gestão judiciária e qualificar a extração de dados estatísticos, o CNJ padronizou a nomenclatura dos processos por meio da Resolução nº 46/2007 [DE JUSTIÇA 2007], estabelecendo a estrutura das chamadas árvores de classificação. Esse modelo organiza os processos inicialmente por sua classe processual — ou seja, o tipo de procedimento judicial, e, em seguida, por assunto, que corresponde ao tema discutido na demanda.

Essa sistematização permite agrupar e quantificar ações relacionadas ao acesso à saúde, além de distinguir subtemas específicos, como o fornecimento de medicamentos e o tratamento médico-hospitalar, proporcionando maior precisão na análise estatística e no desenvolvimento de políticas públicas.

Além da classificação padronizada, os dados estatísticos também permitem mensurar o tempo médio de tramitação dos processos. Um dos principais desafios na área da saúde é justamente a morosidade no atendimento judicial, especialmente em casos de urgência médica.

De acordo com o Painel de Estatísticas Processuais de Direito à Saúde [CNJ 2025], o tempo médio entre o ajuizamento e o julgamento de ações envolvendo fornecimento de medicamentos foi de até 377 dias em 2023, caindo para 314 dias em 2024, no cenário nacional. No âmbito do Tribunal de Justiça de Mato Grosso do Sul (TJMS), entretanto, esse tempo chegou a 420 dias, evidenciando a necessidade de medidas que agilizem a resposta judicial em situações que impactam diretamente a saúde dos cidadãos.



Fonte: Painel de Estatísticas Processuais de Direito à Saúde do CNJ, mai. 2025.

Gráfico 1: Tempo de tramitação até o julgamento - CNJ, em âmbito nacional.

A judicialização da saúde é um tema sensível no Poder Judiciário, sendo motivo de preocupação das organizações públicas e privadas e tem causado grande repercussão social [Lucena, 2021]. O direito à saúde, previsto no art. 196 da Constituição Federal, impõe ao Estado a obrigação de garantir acesso universal a serviços de promoção, proteção e recuperação da saúde. O desafio está em equilibrar o direito individual com o interesse coletivo, evitando conflitos e assegurando equidade [Carline, 2014].

Segundo TURINE [2018], a relevância da inclusão da saúde como direito fundamental é um importante marco constitucional, um elemento para a realização do princípio democrático, pois os direitos fundamentais nascem e se desenvolvem com as Constituições que os reconheceram e asseguraram, como direito social materialmente fundamental. O direito à saúde deve ser visto nas vertentes negativa, como direito de exigir do Estado ou de terceiros abstenção de ato que cause prejuízo à saúde, e positiva, como direito a exigir do Estado que preste serviços para

prevenir, promover ou tratar doenças, estas vistas com base no conceito da Organização Mundial da Saúde (OMS) de que saúde é o bem estar físico, mental e social, e não somente ausência de enfermidades.

A atuação do Poder Público nas ações de saúde pública ocorrem com base no SUS, que consiste em uma rede regionalizada e hierarquizada, com diretrizes de organização ditadas também pela Constituição Federal, em seu art. 198, quais sejam a descentralização, o atendimento integral e a participação da comunidade.

A organização e funcionamento do SUS está regulada pela Lei nº 8080/90, que, em seu art. 4º, define que o SUS é constituído pelo conjunto de ações e serviços de saúde prestados por órgãos e instituições públicas federais, estaduais e municipais, vinculadas à Administração direta e indireta, bem como fundações mantidas pelo Poder Público.

A formulação da política de medicamentos integra o campo de atuação do SUS e baseia-se em protocolos clínicos, que são documentos que definem critérios para diagnóstico de doenças e agravos à saúde, tratamentos recomendados, medicamentos indicados, posologias, controle clínico e resultados terapêuticos esperados, a serem seguidos pelos gestores do sistema.

Com base nesses protocolos e nas diretrizes terapêuticas, são definidos os medicamentos a serem utilizados em cada etapa da evolução da doença, levando-se em conta fatores como perda de eficácia, intolerância, reações adversas relevantes e outras condições clínicas associadas.

Cabe ainda ao SUS, conforme os arts. 19-N e 19-O da Lei nº 8.080/90, avaliar a eficácia, segurança, efetividade e custo-efetividade dos medicamentos indicados para cada estágio da enfermidade, assegurando que as decisões terapêuticas sigam critérios técnicos e de racionalidade econômica..

Uma vez estabelecido o protocolo clínico, a responsabilidade pelo fornecimento dos medicamentos cabe à União, aos Estados ou aos Municípios. No entanto, a incorporação, exclusão ou alteração de medicamentos, produtos e procedimentos no âmbito do SUS é atribuição do Ministério da Saúde, com o suporte técnico da Comissão Nacional de Incorporação de Tecnologias no SUS (CONITEC), conforme previsto no art. 19-Q da Lei nº 8.080/90.

A CONITEC baseia suas decisões em evidências científicas relativas à eficácia, acurácia, efetividade e segurança dos medicamentos. Embora não seja o órgão competente para o registro ou autorização de uso de fármacos, que é da Anvisa, cabe à comissão decidir sobre a incorporação dessas tecnologias ao SUS,

conforme os critérios estabelecidos na legislação vigente.

Para isso, são realizadas avaliações econômicas comparativas entre os medicamentos em análise e as tecnologias já incorporadas, com base em processos que observam os princípios da participação social e da transparência.

Segundo SCHULZE (2018), a tarefa do magistrado na temática de judicialização da saúde é norteada por dilemas diários:

> ... diante de casos difíceis (hard cases) que precisam de uma definição, pois há uma pessoa que possui uma patologia, há uma prescrição médica e há um tratamento disponível (ainda que sem efetividade, eficácia e eficiência) em algum lugar do mundo. Para alguns, isso é suficiente para a procedência do pedido. Para outros, é preciso muito mais, como a comprovação do sucesso da providência buscada, com a demonstração de que o custo é suportável socialmente, sem provocar colapso no Sistema de Saúde (que precisa ser compreendido em uma perspectiva mais ampla e não isoladamente). Portanto, as escolhas trágicas são inerentes à Judicialização da Saúde.

No campo da judicialização de medicamentos, a fim de promover um critério mais igualitário das decisões, o Superior Tribunal de Justiça estabeleceu, para o âmbito do SUS, os critérios para deferimento da concessão de medicamentos, no julgamento do recurso especial julgado sob a sistemática de recurso repetitivo. Como resultado, o Tema 106 firmou a obrigatoriedade do poder público fornecer medicamentos não incorporados em atos normativos ao SUS na presença, cumulativa, de três requisitos.

Com base no julgamento, deve restar comprovado, por laudo médico fundamentado e circunstanciado, expedido pelo médico que assiste o paciente, de que o medicamento é imprescindível ou necessário, ante a ineficácia para o tratamento da moléstia dos fármacos fornecidos pelo SUS. Este primeiro requisito deve estar cumulado aos outros dois estabelecidos no mesmo tema 106, incapacidade financeira de arcar com o custo do medicamento prescrito e existência de registro do medicamento na ANVISA, observados os usos autorizados pela agência. Com relação aos medicamentos sem registro na ANVISA e aos medicamentos de alto custo, há ainda os temas 6 e 1161 do STF.

A judicialização em relação aos medicamentos demonstra uma falha na

execução da política pública a ser aperfeiçoada pelo Poder Judiciário. Porém, ainda há a lacuna de medicamentos não registrados na Agência Nacional de Vigilância Sanitária – ANVISA, bem como de medicamentos não incorporados ao SUS, que necessitam de análise dos gestores públicos.

Assim, a proposta deste trabalho é revisar a literatura de Inteligência Artificial aplicada ao sistema judiciário a fim de propor um modelo e um painel computacional para identificar os medicamentos requisitados ao SUS pleiteados por meio de processos judiciais ao Poder Judiciário do Estado de Mato Grosso do Sul. Os resultados da pesquisa poderão ser utilizados para a melhoria das políticas públicas de fornecimento de medicamentos para a sociedade.

1.2. Motivações e Objetivos

De acordo com a Nota Técnica 02/2022 [TJMS 2022], emitida pelo Centro de Inteligência do TJMS, existe uma problematização da judicialização da saúde que é antiga e que representa uma parcela expressiva de processos ajuizados entre os anos de 2015 e 2020, correspondendo a mais de um milhão de processos novos em todo o judiciário, sobretudo os relacionados à aquisição de medicamentos de alto custo não previsto nas listas oficiais de fármacos do SUS.

Este estudo motivou o presente trabalho para pesquisar e propor uma solução, utilizando inteligência artificial, para auxiliar o TJMS no diagnóstico da judicialização da saúde, com informações detalhadas sobre a natureza dos medicamentos pleiteados para viabilizar levantamentos qualitativos e melhoria das políticas públicas do Estado de Mato Grosso do Sul.

A pesquisa também se alinha à política de inovação do CNJ que incentiva a implementação de ferramentas como o Processo Judicial Eletrônico (PJe), a automação de rotinas cartorárias, o uso da inteligência artificial para triagem e análise de processos, além da ampliação do atendimento virtual ao cidadão. Tais iniciativas evidenciam o potencial das tecnologias digitais na promoção de uma Justiça mais acessível, econômica e eficiente.

Compreender os impactos dessa transição tecnológica é essencial para garantir que a inovação atenda às exigências legais e operacionais, mas também fortaleça a construção de um Judiciário mais moderno, justo e orientado ao cidadão.

Nesse contexto, objetiva-se uma atuação do Poder Judiciário mais efetiva na garantia do acesso universal e igualitário à saúde, de forma que o acesso à justiça a

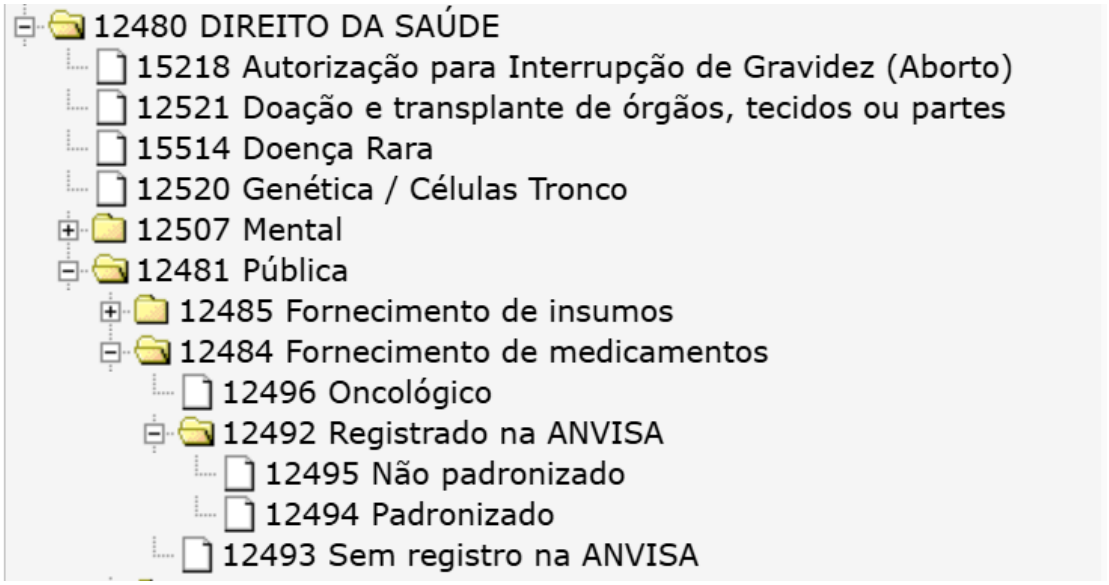
todos seja regido por instituições eficazes, responsáveis e inclusivas. Promover igualdade de atendimento ao usuário final, olhando as portas de entrada e saída do sistema de justiça, oportuniza uma efetiva e ágil prestação de acesso à justiça na pretensão do usuário.

Diante das fragilidades das políticas públicas na promoção do acesso à saúde e à justiça, este trabalho tem como objetivo contribuir para a identificação em tempo hábil das demandas de fornecimento de medicamentos pelo SUS, por intervenção do judiciário, a fim de contribuir com ações que visem a agilidade do atendimento ao cidadão. E para isso, propõe responder às hipóteses que motivam este estudo:

a) É possível identificar os medicamentos demandados em textos jurídicos na língua portuguesa usando algoritmos pré-treinados?

b) A assertividade das técnicas disponíveis na literatura, nacional e internacional, é suficiente para identificar medicamentos demandados em textos jurídicos na língua portuguesa com segurança?

Além da contribuição almejada relativamente à identificação dos fármacos demandados atualmente, outros benefícios poderão ser gerados a partir dos dados obtidos por classificação textual dinâmica utilizando soluções baseadas em aprendizado de máquina, que a classificação atual existente está limitada ao enquadramento na Tabela Processual Unificada - TPU do CNJ [TPU CNJ] na Figura 1.



Fonte: Sistema de Gestão de Tabelas Processuais Unificadas do CNJ.

Figura 1: Recorte dos Assuntos da Tabela Processual Unificada - CNJ.

Como pode ser observado na Figura 1, na padronização existente não é possível especificar nome de medicamentos por serem dinâmicos e, supostamente, haver constantemente inovações no mercado. Com análise das principais demandas formuladas e concedidas, o SUS tem um importante instrumento diagnóstico para implementação de políticas públicas, de forma que possa ser universalizado e igualitário o acesso à saúde, evitando impactos na sua gestão por demandas não programadas. Pode-se com a presente classificação, que é o foco deste estudo, racionalizar a prestação do TJMS, relativamente às demandas de saúde, bem como, auxiliar as políticas públicas de saúde, visando melhor implementação que permita ajustes financeiros necessários.

Após a validação do modelo de classificação de medicamentos desenvolvido para o TJMS, ele poderá ser disponibilizado a outros tribunais por meio da plataforma nacional SINAPSES [de Justiça 2022], repositório oficial dos modelos de inteligência artificial homologados pelo CNJ. Essa iniciativa visa otimizar tempo e recursos públicos, promovendo a cooperação e o compartilhamento de soluções tecnológicas no âmbito do Judiciário.

Para embasar a proposta, o estudo iniciou-se com uma pesquisa exploratória voltada à identificação de trabalhos e modelos pré-existentes relacionados à judicialização da saúde, tanto no CNJ quanto na plataforma SINAPSES. Verificou-se, contudo, a ausência de soluções similares entre os modelos disponíveis, o que reforça a originalidade e a relevância da iniciativa.

Desta forma, o presente trabalho tem os seguintes objetivos:

Objetivo Geral

Propor um modelo de inteligência artificial aplicado à mineração de textos jurídicos, com foco na identificação automática de medicamentos solicitados ao SUS em processos de judicialização da saúde no TJMS, visando subsidiar a gestão pública e a formulação de políticas de saúde mais eficientes e sustentáveis.

Objetivos Específicos

1. Realizar revisão da literatura sobre técnicas de *Named Entity Recognition* (NER) aplicadas à identificação de medicamentos em textos não estruturados, especialmente em língua portuguesa, com foco na promoção da inovação tecnológica responsável;

2. Construir um dataset anonimizado e estruturado a partir das petições iniciais de processos judiciais da área da saúde no TJMS, respeitando princípios éticos e legais de privacidade e proteção de dados;

3. Implementar, treinar e avaliar modelos de aprendizado de máquina para extrair os medicamentos solicitados ao SUS, com foco na eficiência e transparência dos processos;

4. Propor um painel interativo intitulado MED-SUS-MS para visualizar os medicamentos judicializados, contribuindo para a governança digital da saúde e o acesso igualitário à informação;

5. Analisar os impactos dos dados extraídos para fomentar a desjudicialização da saúde em Mato Grosso do Sul e a formulação de políticas públicas de saúde alinhadas aos 17 Objetivos de Desenvolvimento Sustentável da ONU, promovendo equidade no acesso aos medicamentos e mitigando o impacto orçamentário da judicialização; e

6. Promover parcerias institucionais entre universidade, poder judiciário e gestores públicos, fortalecendo novas pesquisas aplicadas e a inovação aberta para o desenvolvimento regional.

1.3. Organização do Texto

Este trabalho está organizado em quatro capítulos. No Capítulo 1 foi apresentada a contextualização do trabalho, as motivações e os objetivos principais. No Capítulo 2 será apresentado um panorama sobre a literatura existente acerca da aplicação de inteligência artificial no judiciário, o estado da arte em extração de medicamentos utilizando mineração de textos. No Capítulo 3 serão apresentados os resultados alcançados na pesquisa, destacando as técnicas aplicadas ao trabalho e a metodologia. As conclusões e os possíveis trabalhos futuros a partir da presente pesquisa são apresentados no Capítulo 4.

CAPÍTULO 2

REVISÃO DA LITERATURA DE INTELIGÊNCIA ARTIFICIAL APLICADA AO SISTEMA JUDICIÁRIO

2.1. Considerações Iniciais

O Conselho Nacional de Justiça (CNJ) tem incentivado os órgãos do Poder Judiciário Brasileiro a aplicar Inteligência Artificial (IA) para promover maior agilidade e coerência nos seus processos de trabalho, tendo como requisitos: a ética, a transparência e a governança na produção e uso dessas aplicações.

A trajetória do uso da IA no âmbito do CNJ iniciou-se formalmente com a publicação da Resolução CNJ nº 332/2020, que estabeleceu as primeiras diretrizes éticas e técnicas para o uso de IA no Poder Judiciário brasileiro. Essa norma foi um marco inicial para orientar os tribunais na adoção de tecnologias emergentes, voltadas à melhoria da gestão processual e à promoção da eficiência jurisdicional.

Com a rápida evolução dos sistemas de IA, especialmente os modelos de linguagem e as soluções generativas, o CNJ identificou a necessidade de revisar e atualizar seu marco regulatório. Assim, foi instituído o Grupo de Trabalho por meio da Portaria CNJ nº 338/2023 a fim de propor uma nova regulamentação alinhada às transformações tecnológicas e às melhores práticas internacionais.

Após um processo participativo com audiências públicas e contribuições da sociedade civil, magistrados, especialistas e instituições públicas e privadas, o CNJ

aprovou por unanimidade, em fevereiro de 2025, o Ato Normativo nº 0000563-47.2025.2.00.0000, que resultou na promulgação da Resolução CNJ nº 615/2025, que atualiza e amplia significativamente o escopo da anterior, estabelecendo um conjunto robusto de princípios, diretrizes e requisitos para o desenvolvimento, uso, governança, auditabilidade e transparência de soluções de IA no Judiciário. Entre seus destaques estão:

- A obrigatoriedade da supervisão humana em todas as fases do ciclo de vida das soluções de IA;

- A classificação dos sistemas conforme grau de risco (baixo ou alto) e a exigência de auditorias regulares;

- A criação do Comitê Nacional de Inteligência Artificial do Judiciário, responsável por monitorar, revisar e implementar diretrizes de uso da tecnologia;

- O fortalecimento da Plataforma Sinapses, que centraliza o registro, a catalogação e o controle das aplicações de IA no Judiciário; e

- A promoção de princípios como transparência, explicabilidade, justiça decisória, segurança de dados, diversidade e letramento digital.

A nova Resolução também proíbe o uso de IA para decisões judiciais autônomas e para finalidades que envolvam vieses discriminatórios, ranking de pessoas com base em características sociais ou emocionais, ou riscos à privacidade. Reforça ainda o respeito à Lei Geral de Proteção de Dados, ao devido processo legal e à dignidade humana, além de prever mecanismos de *accountability* (responsabilização, prestação de contas), com relatórios públicos, supervisão participativa e auditorias permanentes.

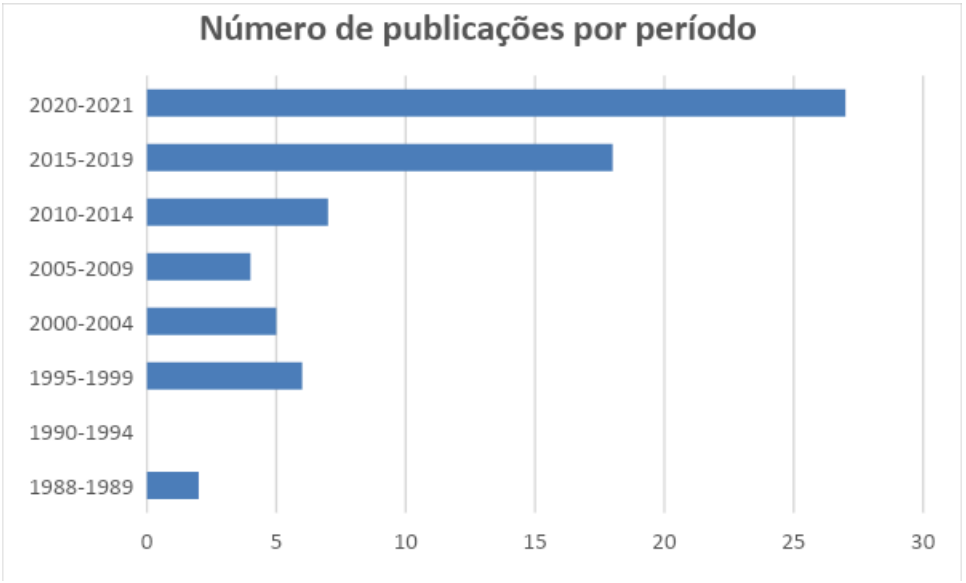
Assim, o CNJ consolida-se como uma referência internacional na regulação ética e segura da IA no Poder Judiciário, promovendo uma transformação digital orientada pela confiança, responsabilidade e inclusão.

Neste contexto, diversas soluções tecnológicas de IA no Brasil estão sendo homologadas pelo CNJ e disponíveis na plataforma SINAPSES, onde a maioria delas trata de mineração de textos. Seguindo nesse raciocínio, este trabalho tem como objetivo minerar informações nos processos judiciais a fim de identificar os

medicamentos pleiteados ao SUS, especificamente, na petição inicial, que é o primeiro documento de um processo judicial, onde consta o pedido formulado pelo usuário do sistema de justiça.

A capacidade de extrair informações precisas de textos não estruturados é uma das fronteiras mais desafiadoras da computação atual. Nesse contexto, o reconhecimento de entidades nomeadas (*Named-Entity Recognition* - NER) surge como uma área de investigação científica, especialmente no que se refere à identificação de nomes de medicamentos em registros eletrônicos no domínio da justiça. Este trabalho visa explorar o estado da arte, investigando metodologias, técnicas e ferramentas que se destacaram no reconhecimento e classificação de entidades nomeadas relacionadas a medicamentos.

Segundo [Oliveira et al. 2022b], a evolução da IA na justiça iniciou-se com publicações em 1988, mas sem expressão no campo científico, somando apenas 24 publicações até o ano de 2014. Após 2015, foram 45 publicações, conforme distribuição apresentada no Gráfico 2.



Fonte: Elaborado pela autora com base em informações de [Oliveira et al. 2022b].

Gráfico 2 - Número de publicações por período.

Segundo o estudo, entre os anos de 2015 e 2019 houve uma polarização dos sistemas de apoio à decisão e de temas emergentes como o acesso à justiça. A partir de 2021, temas relacionados a ciências de dados e forense digital surgiram. Constatou-se que a literatura sobre IA mudou desde os anos 1990, quando se constituiu como tema motor, para se tornar um *cluster* em transição entre a temática

motora e a básica nos anos 2020-2021, com pulverização para os temas “classificação”, “viés”, “lei” e “inteligência artificial na lei”. Destacou que há poucas ou nenhuma publicação sobre o assunto nos países da América Latina. Recomendou a exploração do uso da IA na esfera judicial para melhorar o bem-estar social, concluindo que mais estudos são desejados sobre o uso da IA para permitir ou facilitar o acesso à justiça para pessoas carentes ou estratos sociais marginalizados ou jurisdições em condições de desigualdades.

Oliveira ressaltou a importância de usar modelos de IA explicativos em vez de modelos de “caixa preta”, a fim de respeitar o princípio da transparência, que é um dos pilares fundamentais da prestação jurisdicional.

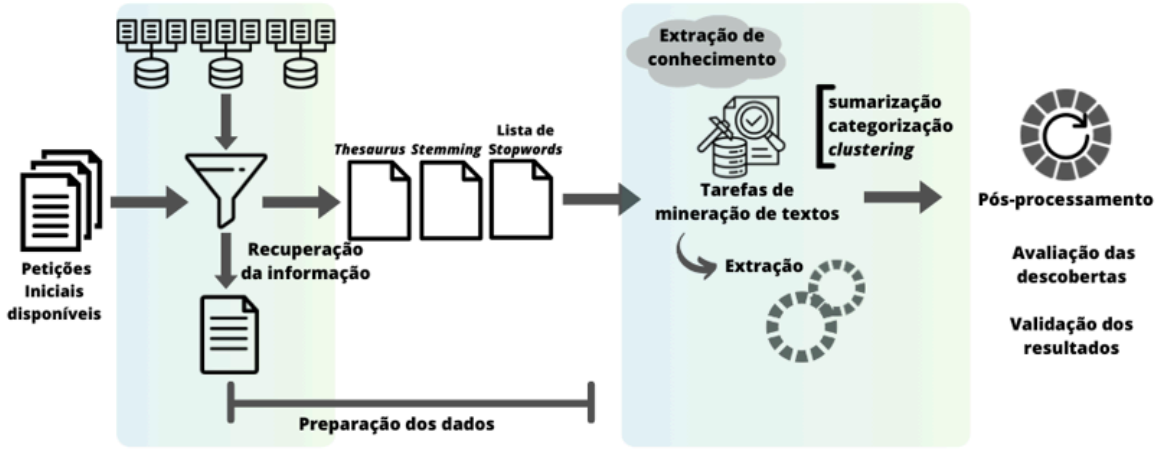
Buscando seguir tais recomendações, neste trabalho será apresentada uma revisão de literatura que foca não apenas a eficácia dos métodos em extrair informações relevantes de textos não estruturados, mas também sua aplicabilidade em campos específicos como biomedicina e jurídico, onde a precisão, transparência e a confiabilidade das informações extraídas são cruciais.

Considerando a diversidade e a complexidade dos dados, o estudo investiga técnicas de mineração de texto, com um interesse particular na eficiência de tais métodos para o idioma português. Além da performance, aspectos como o agrupamento das informações, facilidade de implementação e disponibilização dos resultados também serão analisados, visando uma compreensão abrangente que possa guiar futuras aplicações práticas na identificação de entidades nomeadas em textos para identificação de nomes de medicamentos. Este esforço de pesquisa busca, portanto, contribuir significativamente para o avanço das tecnologias de informação na área jurídica, melhorando a precisão e a eficiência na identificação de informações críticas em registros eletrônicos.

2.2. Inteligência Artificial e Mineração de Textos

A pesquisa textual requer a utilização da tecnologia de mineração de textos para extrair conhecimento de dados não-estruturados e que, de acordo com [Rezende, 2005] é uma das cinco técnicas ou tecnologias para aquisição de conhecimento que ainda é o gargalo para utilização de sistemas inteligentes. As outras quatros técnicas são aquisição de conhecimento manuais, baseadas em entrevistas, acompanhamento ou modelos; aquisição semiautomáticas, baseadas em teorias cognitivas ou modelos existente; aprendizado de máquina por indução de regras a partir de exemplos catalogados; e mineração de dados para extrair regras e comportamentos a partir de análise de grande massa de dados.

A mineração de textos passa pela etapa de preparação dos dados, extração por meio de sumarização, categorização ou clusterização, e, por fim, avaliação das descobertas e validação dos resultados, conforme o esquema da Figura 2, inspirada em [Rezende, 2005].



Fonte: Elaborado pela autora com base em [Rezende, p. 339 , 2005].

Figura 2: Etapas da Mineração de Textos.

Neste contexto, a abordagem dos dados textuais utilizada neste trabalho será por meio da análise semântica, que, de acordo com [Rezende, 2005] utiliza técnicas e fundamentos de processamento de linguagem natural, levando em consideração a estrutura do texto e o contexto.

De acordo com Rodríguez et al. (2020), o Processamento de Linguagem Natural (NLP) é uma subárea da IA que se concentra no desenvolvimento de sistemas computacionais capazes de interpretar fala e texto de maneira semelhante aos humanos. Este estudo demonstra como o NLP pode ser utilizado para automatizar o Reconhecimento de Entidades Nomeadas (NER), como medicamentos, em uma petição judicial.

O NER identifica elementos específicos em textos, como nomes de pessoas, lugares e organizações, e é essencial para várias aplicações de processamento de linguagem natural, como respostas automáticas, resumos de textos e traduções automáticas. Inicialmente, sistemas NER tiveram sucesso ao usar engenharia humana para desenvolver características e regras específicas. Recentemente, o uso de aprendizado profundo, que utiliza representações vetoriais contínuas e processamento não linear, aperfeiçoou o desempenho desses sistemas, e é nesta

abordagem que este estudo irá atuar. [Li et al. , 2020].

Para avaliar a qualidade dos modelos, será utilizada a métrica F1 score, uma variação entre 0 e 1, sendo que valores mais próximos de 1 indicam melhor desempenho. Ela é calculada como a média harmônica entre a precisão e o recall do modelo. A precisão é a proporção de resultados verdadeiros positivos em relação ao total de resultados positivos (verdadeiros positivos mais falsos positivos). O recall, por sua vez, é a proporção de verdadeiros positivos em relação ao total de resultados relevantes (verdadeiros positivos mais falsos negativos).

Neste Capítulo será apresentada uma revisão de literatura para identificar estudos de aplicação de IA capazes de extrair elementos de textos jurídicos, mormente de modelos treinados para textos em língua portuguesa do Brasil. Além disso, objetiva identificar a pré-existência de aplicações para reconhecimento de entidades nomeadas em textos jurídicos que seja capaz de extrair informações mais detalhadas sobre a natureza dos pleitos de acesso à saúde pública judicializados na Justiça Estadual Brasileira, utilizando como recorte o Tribunal de Justiça de MS.

2.3. Protocolo de revisão

Os objetivos da revisão de literatura nesta temática são:

- Identificar trabalhos de reconhecimento de entidade nomeada aplicáveis à identificação de nomes de medicamentos em textos não estruturados a fim de descobrir os métodos e técnicas com melhores desempenho nessa tarefa; e
- Identificar a pré-existência de aplicações que sejam capazes de extrair informações mais detalhadas sobre a natureza dos pleitos de medicamentos por meio do acesso à saúde pública judicializados na justiça estadual brasileira.

Serão pesquisadas referências de estudos da área de computação que tratam do Reconhecimento de Entidades Nomeadas (Named-Entity Recognition - NER) em textos não estruturados, no formato eletrônico, aplicáveis à identificação de nomes de medicamentos, com a finalidade de descobrir os métodos e técnicas com melhores desempenho, utilizando como base inicial de conteúdo sobre o assunto: os artigos “Drug Name Recognition: Approaches and Resources” [Liu et al., 2015], “Drug name recognition in biomedical texts: a machine-learning-based method” [He et al., 2014], “Using Snomed to recognize and index chemical and drug mentions.” [López Úbeda et al., 2019] e o livro Sistemas Inteligentes: Fundamentos e Aplicações de [Rezende 2005].

Dessas referências, espera-se uma visão consistente sobre os métodos e técnicas que obtiveram os melhores resultados em efetivamente extrair e interpretar informações sobre medicamentos em textos, considerando as nuances do domínio jurídico ou biomedicina. Serão considerados os resultados no contexto de pesquisadores da área de NLP que desenvolveram aplicações para tarefa de reconhecimento de entidades nomeadas em textos não estruturados de registros eletrônicos sobre medicamentos, comparando diferentes métodos e algoritmos entre si. Busca-se como resultado a eficiência dos métodos de identificação de medicamentos, sobretudo para o idioma português, facilidade de implementação e de disponibilização do resultado.

Para nortear a pesquisa, foram elaboradas as seguintes questões:

- Quais são os métodos mais eficazes da atualidade para reconhecimento de nomes de medicamentos em documentos textuais no mundo?
- Quais são os métodos mais eficazes da atualidade para reconhecimento de nomes de medicamentos em documentos da língua portuguesa?
- Quais desafios são enfrentados no processo de identificação de nomes medicamentos em documentos não estruturados, sobretudo na língua portuguesa?
- Existem estudos ou aplicações que tratam da extração em documentos de petições judiciais de nomes de medicamentos pleiteados à saúde pública na justiça brasileira?

Com base nos artigos pesquisados sobre o assunto, foram determinadas as palavras-chave, dando origem a "*String* de busca" (Tabela 1), que foram criadas na língua inglesa, de modo a alcançar também os estudos publicados nesse idioma.

Keyword	***Synonyms***	***Related to***
named entity recognition	*medication Identification* *drug recognition* *drug names* *entity recognition? named?* *NER* *DNER*	*population*
natural language processing	*NLP* *deep learning* *generative AI* *generative artificial intelligence*	*intervention*

Keyword	***Synonyms***	***Related to***
drug	*pharmaco* *medicine* *biomedical text?*	*population*
Search String		
(((("named entity recognition" OR "Medication Identification" OR "drug recognition" OR "drug names" OR "entity recognition? named?" OR NER OR DNER) AND *("natural language processing" OR NLP OR Deep learning OR "Generative AI" OR "Generative artificial intelligence")) AND (medicine OR drug OR pharmaco or "Biomedical text?"))*		

Fonte: Elaborado pela autora.

Tabela 1: *String* de Busca.

A busca por artigos foi realizada nas bibliotecas: ACM Digital Library (http://portal.acm.org), IEEE Digital Library (http://ieeexplore.ieee.org), Portal de Periódicos da CAPES (https://www-periodicos-capes-gov-br.ez51.periodicos.capes.gov.br/), Scopus (http://www.scopus.com), além do acervo pré-existente construídos com buscas realizadas de forma aleatória extraídos do Google Scholar, ACL Anthology (https://aclanthology.org/), Science Direct (https://www.sciencedirect.com/) e SBC (https:// www.sbc.org.br/), no período de 2021 a 2024.

Para análise da lista inicial, foram estabelecidos, arbitrariamente, os seguintes critérios de inclusão e exclusão:

Critérios de Inclusão:

- Critério 1: Serão incluídos trabalhos publicados nos últimos 3 anos (2021 - 2024) porque o judiciário e a inteligência artificial avançaram muito nesse período;
- Critério 2: Serão incluídos trabalhos sobre identificação de nomes de medicamentos em documentos do domínio jurídico ou de biomedicina, pois é o nosso objetivo principal para o desenvolvimento da aplicação;
- Critério 3: Serão incluídos trabalhos relativos a tratamentos de dados pessoais em documentos do domínio jurídico ou de biomedicina, que temos que preservar os dados pessoais por conta da Lei Geral de Proteção de Dados;

- Critério 4 : Serão incluídos trabalhos sobre revisões de literatura de reconhecimento de entidade nomeada.

Critérios de Exclusão:

- Critério 1: Serão excluídos trabalhos duplicados;
- Critério 2: Serão excluídos trabalhos de outras áreas diferentes de Ciência da Computação;
- Critério 3: Serão excluídos trabalhos não revisados;
- Critério 4: Serão excluídos trabalhos publicados em idiomas diferentes do inglês e português ou aqueles de reconhecimento de entidades nomeadas treinados para caracteres específicos de outro idioma, como o chinês;
- Critério 5: Serão excluídos trabalhos sobre COVID-19, diagnósticos médicos, reações adversas de medicamentos e outros títulos não correlacionados à identificação do nome de medicamentos em documentos digitais;
- Critério 6: Os estudos que não apresentarem alguns quesitos indispensáveis de qualidade serão, a princípio, desconsiderados;
- Critério 7: Serão excluídos trabalhos que não estiverem no formato de documento de texto, html ou pdf;
- Critério 8: Serão excluídos trabalhos que obtiverem nota abaixo de 5.0 conforme os critérios de qualidade estabelecidos no planejamento desta revisão; e
- Critério 9: Para a base de dados de periódicos da Capes, serão incluídos apenas os 50 melhores resultados, assim definidos pela própria ferramenta de busca dela, excluindo automaticamente os demais.

Além dos critérios de inclusão e exclusão, foram definidos critérios de qualidade, supostamente importantes para selecionar estudos que agreguem para o desenvolvimento de aplicações de classificação de textos jurídicos brasileiros, dotados de linguagem natural e própria do domínio jurídico:

- Critério 1: O estudo se aplica a textos jurídicos brasileiros?

- Critério 2: O estudo apresenta os algoritmos utilizados para o desenvolvimento da aplicação?
- Critério 3: Os algoritmos utilizados possuem F1-Score satisfatório? (Sim: >=0.8; Parcialmente: 0.8> e <0.7)
- Critério 4: O estudo se aplica a identificação de medicamentos em documentos com dados não estruturados?
- Critério 5: O estudo apresenta a metodologia utilizada para o desenvolvimento da aplicação?
- Critério 6: A amostra de dados (Dataset) utilizada no estudo é representativa (Sim: acima de 3000 registros; Parcialmente: acima de 2000)?
- Critério 7: Possui informações sobre a infraestrutura computacional necessária para rodar a aplicação?
- Critério 8: O estudo apresenta o desempenho dos algoritmos utilizados?
- Critério 9: O estudo apresenta a quantidade de registros do dataset utilizado?
- Critério 10: O estudo aborda classificação multi-rótulos?
- Para esses critérios, foram definidas respostas e pontuações, estabelecendo pontuação máxima de 10.0 e pontuação de corte de 5.0 (Tabela 2):

Descrição	Pontuação
Sim	1.0
Parcialmente	0.5
Não	0.0

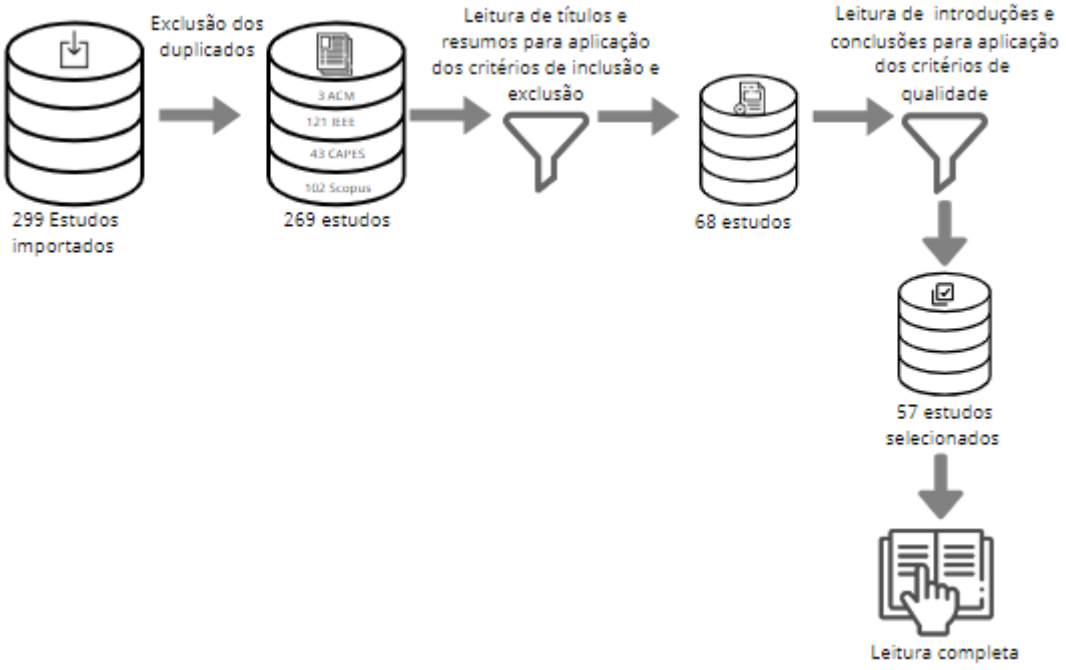
Fonte: Elaborado pela autora.

Tabela 2: Pontuação.

2.4. Condução e resultados da revisão

Concluída a definição do protocolo, foram importados os artigos/estudos das bibliotecas supramencionadas, totalizando 299 produções intelectuais. Após a

exclusão dos duplicados, restaram 3 estudos da ACM, 121 da IEEE, 43 do Portal de Periódicos da CAPES e 102 da Scopus, conforme fluxo ilustrado na Figura 3.



Fonte: Elaborado pela autora.

Figura 3: Processo de seleção de produções científicas do protocolo.

Os resultados da revisão de literatura apresentam vários modelos de IA ou técnicas, identificados por siglas. Para facilitar a compreensão da leitura, os modelos ou métodos mencionados serão brevemente descritos a seguir:

a) **BART** (*Bidirectional and Auto-Regressive Transformers*): modelo de linguagem pré-treinado combina as vantagens dos modelos bidirecionais, como o BERT, com os auto-regressivos, como o GPT. É pré-treinado com introdução de ruído nos textos e aprendizado de reconstrução, tornando-o robusto para tarefas de compreensão e geração de linguagem natural. No contexto biomédico, BART tem se destacado em tarefas como extração de interações medicamentosas [Zaikis et al., 2022].

b) **BERT (***Bidirectional Encoder Representations for Transformers***):** representações

codificadoras bidirecionais de transformadores (*transformer*) aprendem as relações contextuais entre palavras em um texto considerando tanto o contexto à esquerda quanto à direita. Trata-se de um algoritmo de aprendizado profundo, de código aberto, desenvolvido pelos cientistas do *Google IA Language* em 2018, que usa a técnica de redes neurais para o processamento de linguagem natural, que é uma subárea da inteligência artificial. O modelo BERT foi projetado para pré-treinar representações bidirecionais profundas a partir de textos não rotulados, utilizando duas tarefas não supervisionadas: a modelagem de linguagem mascarada (MLM) e a predição da próxima sentença (NSP). Essas técnicas permitem que o BERT capture melhor as nuances do contexto de cada palavra, o que o torna eficaz em diversas aplicações de processamento de linguagem natural, como análise de sentimento, tradução de linguagem e respostas a perguntas [Devlin et al., 2018], [Jacob et al., 2020] e [QuantPedia, 2021].

c) **BERT-BILSTM-CRF:** arquitetura de rede neural combinada usada em tarefas de PLN, especialmente para tarefas de reconhecimento de entidades nomeadas (NER), etiquetagem de sequência e outras tarefas de sequenciamento de texto. De acordo com [Chen et al. 2023], a arquitetura BERT-Bi-LSTM-CRF é utilizada para melhorar o reconhecimento de entidades nomeadas em registros médicos chineses. A abordagem utiliza o algoritmo BERT para gerar representações contextuais detalhadas das palavras, seguido por uma camada Bi-LSTM que captura dependências sequenciais em ambas as direções, e um CRF para assegurar a coerência das etiquetas preditas.

d) **BERT-BILSTM-MULATT-CRF:** arquitetura híbrida para reconhecimento de entidades nomeadas (NER) que combina o modelo BERT para extração de características contextuais, BiLSTM para capturar dependências sequenciais bidirecionais, uma camada de Atenção Multi-Head para melhorar a precisão na identificação de entidades, especialmente abreviações e palavras polissêmicas, e CRF para garantir que as previsões de etiquetas sejam consistentes e estruturadas [Liu et al. 2023]. Uma combinação de técnicas que resulta em um desempenho superior na tarefa de NER em textos bioquímicos.

e) **BERT-CRF:** arquitetura que combina o modelo pré-treinado BERT para extração de representações contextuais das palavras com uma camada de *Conditional Random Fields* (CRF) que modela as dependências entre as entidades [Ge et al., 2022]. Como explicado anteriormente, o modelo BERT é usado para gerar representações contextuais detalhadas das palavras, enquanto a camada CRF é usada para garantir que as previsões de etiquetas sejam consistentes e estruturadas, melhorando a precisão na tarefa de reconhecimento de entidades

nomeadas em textos médicos.

f) **BIOALBERT:** versão especializada do modelo ALBERT (A Lite BERT), que é uma versão otimizada do modelo BERT, que melhora a eficiência de treinamento e inferência ao reduzir significativamente o número de parâmetros no domínio biomédico. Este modelo é treinado em um grande corpus biomédico para capturar representações contextuais específicas dessa área. O BioALBERT utiliza estratégias de redução de parâmetros e uma função de perda auto-supervisionada que modela a coerência entre sentenças para aprimorar o aprendizado das representações contextuais [Naseem et al., 2021].

g) **BioBERT:** versão adaptada do BERT para o domínio biomédico, treinada em domínio especializado como o PubMed e PubMed Central [Lee et al., 2019]. Demonstrou resultados superiores em tarefas de reconhecimento de entidades biomédicas (doenças, medicamentos, genes), com precisão significativamente maior do que modelos genéricos [Kolpakov et al., 2023].

h) **BioBERT + biLSTM + CRF:** arquitetura híbrida projetada para melhorar o reconhecimento de entidades nomeadas (NER) em textos biomédicos [Çelikmasat et al., 2022]. O modelo BioBERT é uma versão adaptada do modelo BERT, pré-treinada em bases biomédicas, que captura de maneira eficaz as especificações e terminologias da área médica. Ao integrar o modelo BioBERT, que fornece representações contextuais ricas, com uma camada biLSTM, que captura dependências sequenciais bidirecionais, e uma camada CRF, que assegura a consistência das previsões de sequência, essa arquitetura se destaca na extração de informações biomédicas complexas. Estudos mostram que o BioBERT+biLSTM+CRF supera significativamente outras abordagens, demonstrando melhorias notáveis no desempenho de NER em diversos conjuntos de dados biomédicos.

i) **BioNER** (*Biomedical Named Entity Recognition*)**:** processo essencial na extração de informações biomédicas de textos não estruturados, como artigos científicos e registros clínicos [Tho et al., 2023]. Diferentemente dos métodos tradicionais, que dependem de regras manuais ou dicionários biomédicos, o modelo BioNER utiliza modelos de aprendizado profundo, como redes neurais recorrentes bidirecionais (BiLSTM) combinadas com modelos de campos aleatórios condicionais (CRF). Esses modelos são aprimorados ainda mais com o uso de transformadores pré-treinados em grandes corpora biomédicos, como o PubMedBERT, que oferece representações contextuais robustas das sequências de entrada. Ao incorporar camadas adicionais de BiLSTM e CRF, o modelo captura informações contextuais detalhadas, além de garantir a consistência no reconhecimento de entidades

biomédicas.

j) **BOND-RoBERTa:** modelo híbrido de reconhecimento de entidades nomeadas (NER) que combina o modelo BERT pré-treinado com o modelo RoBERTa, otimizando ainda mais com treinamento supervisionado e autoaprendizado [Bajaj et al., 2022]. A abordagem BOND, inicialmente, adapta o modelo pré-treinado utilizando rótulos distantes e, em uma segunda etapa, aplica autoaprendizado para refinar as previsões e eliminar rótulos distantes usando regras feitas à mão e bases de conhecimento. Esta combinação melhora a performance do modelo em tarefas de NER, sobretudo em contextos de domínio específico como a saúde pública e o uso de drogas.

k) **CLAMP** (*Clinical Language Annotation, Modeling, and Processing*): modelo de processamento de linguagem natural (NLP) desenvolvido em 2018, especificamente projetado para extrair e codificar entidades nomeadas a partir de textos clínicos [Shah-Mohammadi et al., 2022]. CLAMP fornece um ambiente de desenvolvimento interativo que permite aos usuários construir pipelines de processamento de linguagem natural personalizados para diversas aplicações clínicas. Ele utiliza uma arquitetura baseada em pipeline que inclui componentes como tokenizador, etiquetador de partes do discurso (POS), identificador de seções, reconhecedor de entidades nomeadas (usando modelos LSTM), classificador de assertividade, reconhecedor de atributos, mapeador de conceitos e reconhecedor temporal. CLAMP se destaca por sua capacidade de vincular as entidades extraídas a identificadores de conceitos normalizados de bancos de dados médicos como RxNorm e ICD-10-CM, tornando-se uma ferramenta eficaz e eficiente para a pesquisa e prática clínica.

l) **CNN-BiLSTM:** abordagem híbrida para o reconhecimento de entidades nomeadas (NER) em textos biomédicos, combinando redes neurais convolucionais (CNN) e redes neurais recorrentes bidirecionais de longo curto prazo (BiLSTM) [Fudholi et al., 2022].

m) **En core sci lg da biblioteca SCISPACY:** modelo "en_core_sci_lg" da biblioteca scispaCy é uma ferramenta poderosa para o processamento de textos científicos [Curdin et al., 2022]. Modelo projetado para fornecer representações de alta qualidade das palavras no contexto de textos científicos e biomédicos, o que melhora significativamente o desempenho em tarefas de reconhecimento de entidades nomeadas (NER). O "en_core_sci_lg" é otimizado para capturar características em nível de palavra e caractere, automaticamente, facilitando a análise e a extração de informações relevantes em documentos científicos complexos.

n) **GatorTron:** modelo de linguagem desenvolvido para a extração de informações contextuais sobre medicamentos a partir de narrativas clínicas. Treinado com mais de 90 bilhões de palavras, incluindo mais de 80 bilhões de anotações clínicas da *University of Florida Health*, o GatorTron utiliza a arquitetura baseada em transformadores BERT para identificar menções de medicamentos, classificar eventos relacionados a mudanças de medicamentos e categorizar contextos como negação, temporalidade e certeza. Este modelo teve desempenho superior em tarefas de extração de informações clínicas, superando modelos pré-treinados em bases menores [Chen et al., 2022].

o) **Heterogeneous Graph Neural Network (GNN):** técnica de aprendizado profundo aplicada à extração conjunta de entidades nomeadas e relações (JNERE) [Nojoo Kambar et al., 2022]. Este método representa palavras e relações como diferentes tipos de nós em um grafo. Durante o treinamento, os vetores de *embedding* (palavras) são atualizados usando técnicas de passagem de mensagens interativas. Isso permite que as representações das palavras incorporem informações contextuais e de relações, melhorando a precisão na identificação de entidades e na extração de relações em textos. Esta abordagem é particularmente eficaz para tarefas que envolvem a detecção de interações complexas e de longo alcance entre entidades textuais.

p) **IOBHI:** arquitetura neural conjunta projetada para o reconhecimento de entidades nomeadas biomédicas e normalização de entidades [Noh et al., 2021]. Usa o modelo SciBERT para codificar sentenças e emprega uma abordagem desacoplada para a classificação IOB e a tipificação semântica. O classificador IOB é colocado no final da rede, permitindo que as representações processadas de nomes e tipos sejam usadas para identificar segmentações de menções. Esta arquitetura inclui uma camada CRF biLSTM que processa os vetores concatenados de nome e tipo para etiquetar IOB, enquanto a projeção de nome e a camada de correspondência bilinear calculam as interações entre as representações de tokens e os embeddings de nomes de conceitos predefinidos.

q) **MT-CGAN ou Modelo de NER Guiado por Texto de Multigranularidade Baseado em CGAN**: modelo projetado para o reconhecimento de entidades nomeadas em textos de Medicina Tradicional Chinesa (TCM) [Ma et al., 2022]. Este modelo aborda o desafio da escassez de dados anotados em larga escala, comum na extração de conhecimento de textos TCM. A estrutura do MT-CGAN inclui um codificador de recursos de texto de multigranularidade (MTFE) que extrai informações semânticas e gramaticais de diferentes dimensões dos textos. O modelo utiliza uma Rede Adversária Generativa Condicional (CGAN), onde o

gerador cria sequências de rótulos de entidades que são refinadas pelo discriminador. A inclusão de sementes de diferentes tipos de texto TCM como entrada melhora a precisão do modelo na tarefa de NER. Resultados experimentais mostraram que o MT-CGAN supera outros métodos baseados em redes profundas, especialmente em bases anotadas de pequena escala, devido à sua capacidade de manter valores F1 estáveis e alta precisão na extração de entidades.

r) **Multi-DTR:** estrutura de aprendizado profundo multitarefa projetada para o reconhecimento de nomes de medicamentos (DNER) e normalização de nomes de medicamentos (DNEN) em textos biomédicos [Jin et al., 2022]. Este modelo combina representações de caracteres extraídos por CNNs, vetores de palavras sensíveis ao contexto adquiridos por meio do ELMo, e palavras biomédicas pré-treinadas incorporadas em uma arquitetura BiLSTM-CRF. A abordagem multitarefa permite que DNER e DNEN se apoiem mutuamente, melhorando a generalização do modelo. Os experimentos realizados destacaram sua eficácia na extração de informações de texto biomédico.

s) **MUNCHABLES-Stack:** modelo de aprendizado auxiliar que visa melhorar o reconhecimento de entidades nomeadas (NER) utilizando múltiplos conjuntos de dados de treinamento auxiliares. Ao invés de usar um único conjunto de dados auxiliares, MUNCHABLES-Stack aplica uma sequência de aprendizados auxiliares, ajustando o modelo principal com cada conjunto de dados. Isso permite que o modelo aprenda a partir de diversas fontes de dados, melhorando a performance geral no reconhecimento de entidades químicas, biomédicas e científicas [Watanabe et al., 2022].

t) **Parallel-BINER:** modelo proposto para o reconhecimento de entidades nomeadas biomédicas, que usa uma implementação paralela de camadas BiLSTM e CRF para segmentação de palavras e rotulagem de sequências [Asghari et al., 2022]. Essa abordagem treina simultaneamente duas seções de aprendizado, combinadas em uma camada de concatenação, resultando em uma melhoria significativa na precisão e no F1-Score em comparação com outros modelos de última geração, como o BioBERT. O modelo Parallel-BINER se destaca por seu desempenho consistente em diferentes conjuntos de dados biomédicos, demonstrando eficácia na detecção de nomes de medicamentos em publicações acadêmicas e notas clínicas.

u) **RNN** (Rede Neural Recorrente): um tipo de rede neural que se destaca no processamento de tarefas de rotulagem de sequência devido à sua capacidade de reter informações de séries temporais passadas, o que é útil em tarefas de processamento de linguagem natural (NLP) [Ali et al., 2022]. As RNNs são eficazes para o reconhecimento de entidades nomeadas (NER) que envolvem a identificação

e classificação de várias entidades em um texto. Diferentes arquiteturas de RNN, como a LSTM bidirecional, têm demonstrado um desempenho superior em modelos de NER, melhorando a eficiência ao integrar informações contextuais passadas e futuras. Estas redes foram amplamente aplicadas em várias implementações, mostrando-se eficientes na modelagem de dados sequenciais para diferentes tipos de NER.

v) **RoBERTa (*Robustly optimized BERT approach*)**: uma variante do modelo BERT, projetada para aprimorar o desempenho em tarefas de processamento de linguagem natural [Pandy et al., 2023]. Desenvolvido por Conneau et al., o RoBERTa aprimora o BERT original ao treinar com mais dados e por mais tempo, removendo o mascaramento de tokens dinâmicos e alterando a sequência de hiperparâmetros. O modelo é treinado em um conjunto de dados maior e ajustado para ser mais robusto e eficaz, especialmente em tarefas que envolvem reconhecimento de entidades nomeadas (NER) e outras aplicações em linguagens de baixo recurso. Pandy et al. (2023) demonstraram a eficácia do uso de RoBERTa para extrair informações de textos médicos, melhorando a precisão na identificação de nomes de medicamentos e dosagens.

w) **SciBERT:** uma variante do modelo BERT, projetada especificamente para lidar com textos científicos [Mathu et al., 2023]. Desenvolvido com o objetivo de melhorar a análise de textos biomédicos e de outras áreas científicas, SciBERT foi treinado em um grande corpus de artigos acadêmicos para captar melhor a linguagem e os termos técnicos usados nesses textos. Mathu et al. (2023) relatam que o SciBERT supera outros modelos BERT na tarefa de sumarização de textos científicos, o que o torna ideal para a extração de entidades nomeadas em dados biomédicos. O modelo consegue identificar e classificar com precisão entidades de drogas em textos médicos, mesmo em condições de tempo real e com dados não estruturados.

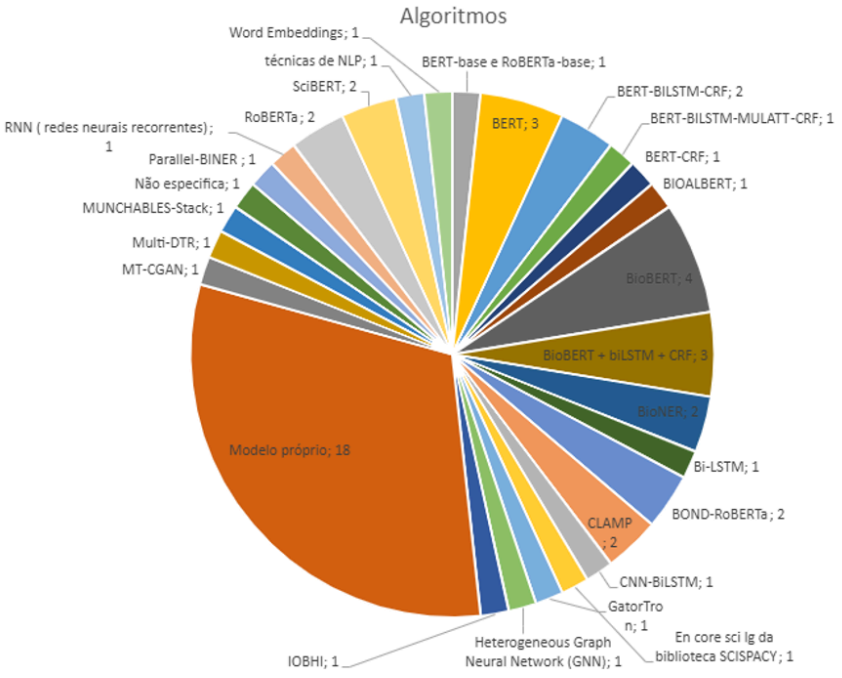
Por fim, da revisão realizada, dezesseis estudos criaram seus próprios modelos, sendo que destes, seis usaram como base o modelo BERT e nove o LSTM, e um não foi informado. Além disso, consultando o repositório [NLP-Progress 2023] de acompanhamento do progresso no Processamento de Linguagem Natural (*Natural Language Processing* - NLP), especificamente para reconhecimento de entidades nomeadas, os quatro modelos apresentados com melhores resultados, com base no F1 score, são: CL-KL [Wang et al., 2021], CrossWeigh + Flair [Wang et al., 2019], Flair embeddings [Akbik et al., 2019], BiLSTM-CRF+ELMo [Peters et al., 2018], respectivamente.

2.5. Reconhecimento de entidades nomeadas para extração de medicamentos

No contexto do reconhecimento de entidades nomeadas (NER), é essencial considerar os estudos específicos do domínio biomédico devido à sua formatação e linguagem específicas. Este campo exige abordagens especializadas para lidar com a terminologia complexa e a estrutura textual característica dos textos biomédicos, sobretudo quando essa linguagem está inserida num contexto jurídico. A identificação precisa de nomes de medicamentos e outras entidades biomédicas é estratégica para a extração de informações relevantes das petições judiciais de judicialização da saúde, contribuindo significativamente para o entendimento dos perfis das demandas. Desses estudos, é apresentado a seguir uma síntese para embasar a metodologia aplicada neste trabalho.

Dos cinquenta e oito estudos analisados, foram encontrados vinte e sete modelos diferentes com melhores desempenhos para extração de entidades nomeadas em textos biomédicos. Destes, trinta e um estão relacionados de alguma forma com o modelo BERT e dezesseis com LSTM. Dezoito estudos criaram seus próprios modelos, sendo nove baseados em LSTM, cinco baseados em BERT, um com base nestes dois e dois não informados.

No Gráfico 3 é ilustrado o total de utilização de cada modelo apresentado na revisão da literatura.

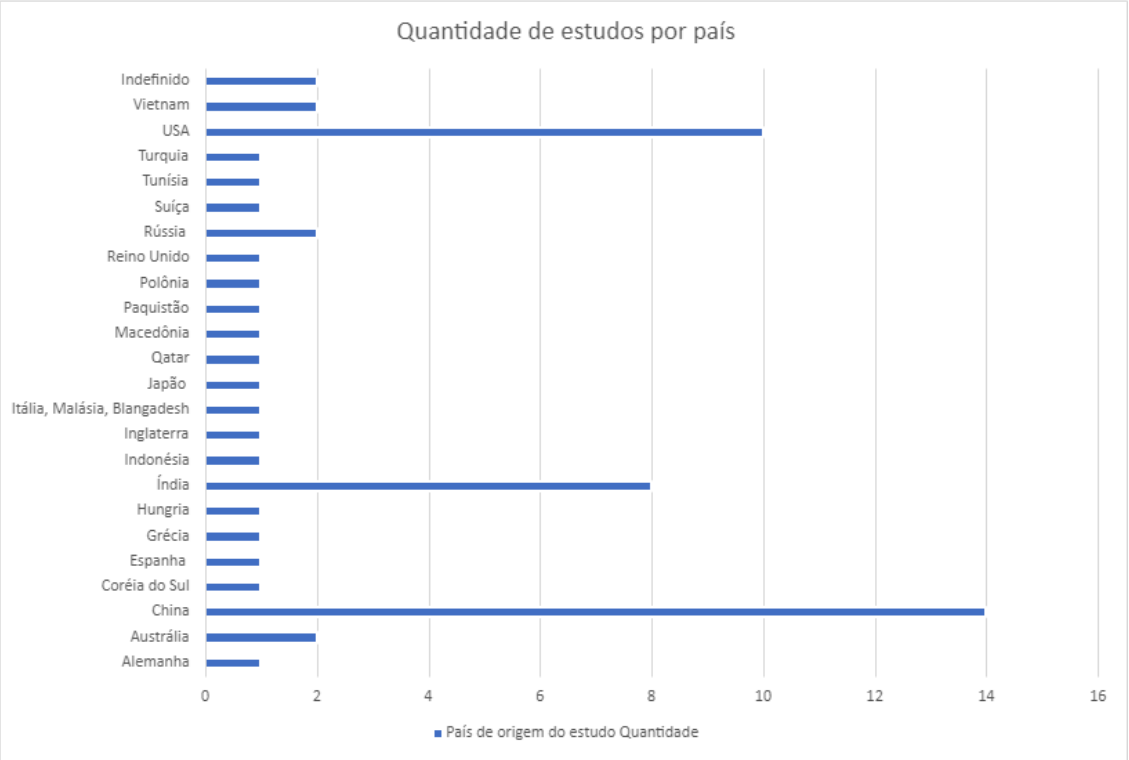


Fonte: Elaborado pela autora

Gráfico 3: Técnicas de IA para reconhecimento de entidades nomeadas de fármacos utilizados nos artigos em diversos países.

A China lidera a quantidade de publicações de estudos de reconhecimento de entidade nomeada em textos biomédicos, seguida dos Estados Unidos e Índia. Nos artigos selecionados, não foi localizado estudo oriundo do Brasil para essa temática.

A distribuição dos estudos por países é ilustrada no Gráfico 4.



Fonte: Elaborado pela autora.

Gráfico 4: Quantidade de estudos por país.

Este tema vem se mantendo com a mesma relevância nos últimos anos, como pode ser analisado no Gráfico 5.

Fonte: Elaborado pela autora.

Gráfico 5: Percentual de estudos publicados por ano, entre os selecionados.

Dos estudos selecionados, serão apresentados a seguir os resultados relatados.

Zaikis et al. (2022) propuseram o uso de modelos Transformer pré-treinados, como BioBERT e SciBERT, para extrair interações medicamentosas. As tarefas abordadas foram o reconhecimento de entidades (DNER) e a classificação de interações (DDIC), utilizando o conjunto de dados DDI-2013. O SciBERT alcançou 82,4% em classificação e 98,1% em extração de entidades, destacando a importância do pré-treinamento em textos especializados.

Os artigos seguintes discutem diferentes abordagens para extração e reconhecimento de entidades nomeadas em textos biomédicos. Pandy et al. (2023) propõem uma abordagem híbrida que combina métodos baseados em regras com aprendizado profundo, utilizando tokenização, marcação e análise de chunks, seguida de aprendizado semi-supervisionado e refinamento com DNN, resultando em um aumento de 3,52% na precisão da extração de nomes de medicamentos, dosagens e frequências.

Mathu T. et al. (2023) utilizam técnicas de deep learning e sumarização de texto para reconhecer entidades nomeadas de drogas (DNER) em dados de PubMed, combinando SciBERT, LSTM bidirecional e CRF, mostrando que SciBERT obteve os melhores resultados em precisão, recall e F1-score para nomes de

drogas, suas marcas, grupos químicos e substâncias não autorizadas.

Whitton et al. (2023) empregam um pipeline de NLP com modelos de transformadores para NER e extração de relações (RE) em ensaios clínicos randomizados (RCTs), aplicando a técnica em 600 sentenças de RCTs de seis áreas de doenças, alcançando F1 scores de 0.78 para NER e 0.77 para RE, destacando a generalização para domínios de doenças não vistos durante o treinamento.

Os artigos a seguir discutem metodologias de *deep learning* para extração e reconhecimento de informações biomédicas em notas clínicas e textos gerados por usuários.

Gan et al. (2023) propõem um método para extrair disposições de medicamentos e atributos de notas clínicas utilizando o *Contextualized Medication Event Dataset* (CMED), com três componentes principais: reconhecimento de entidades nomeadas de medicamentos (NER), classificação de eventos (EC) e classificação de contexto (CC), alcançando micro-F1 scores de 0,973 para NER, 0,911 para EC e 0,909 para CC.

Rivera-Zavala et al. (2021) exploram o impacto do aprendizado de transferência em reconhecimento e normalização de entidades biomédicas em textos clínicos espanhois, demonstrando que o modelo BERT alcançou um F1-score de 88,80% na identificação e classificação de entidades no PharmaCoNER e 78,86% no CORD-19.

Sakhovskiy et al. (2023) utilizam modelos BERT para reconhecimento biomédico de entidades nomeadas e classificação de sentenças multi-rótulos em textos em inglês e russo, com o modelo EnRuDR-BERT alcançando uma pontuação macro F1 de 70%, superando um modelo BERT de domínio geral em 8,64%.

Os artigos referenciados abaixo compilam diversas abordagens de reconhecimento de entidades nomeadas (NER) em contextos médicos e científicos.

Guan et al. (2021) apresentam o modelo BERT-BiLSTM-CRF para reconhecimento de terminologia da medicina tradicional chinesa, combinando transferência de aprendizagem, modelo de linguagem pré-treinado e aprendizado de máquina clássico, superando modelos de benchmark.

Chen et al. (2023) propõem um modelo para NER em registros médicos eletrônicos chineses, utilizando BERT, LSTM bidirecional e GAT para melhorar precisão, recall e F1 score, especialmente em entidades médicas extensas como

doenças e sintomas.

Liu et al. (2021) criaram um modelo híbrido para NER em bioquímica, combinando BERT, BiLSTM, MHATT e CRF, melhorando o reconhecimento de entidades de baixa frequência em até 80% em comparação com BILSTM-CRF, com foco em nomes químicos e de drogas.

Por fim, Ge et al. (2022) comparam modelos tradicionais de NER baseados em BERT com modelos de few-shot learning (FSL) em textos médicos, mostrando que todos os modelos têm desempenho reduzido com pequenos conjuntos de dados de treinamento e que os métodos FSL são menos eficazes em textos médicos, sugerindo a necessidade de novos métodos e datasets especializados.

Naseem et al. (2021) apresentam o BioALBERT, um modelo pré-treinado para o domínio biomédico, que superou modelos de ponta em até 23,71% em precisão em diversas entidades como doenças, drogas, proteínas e espécies.

Devika et al. (2023) utilizam o BioBERT e outras técnicas de codificação para identificar entidades relacionadas à malária, alcançando F1-scores de até 97%.

Kim et al. (2022) analisam a generalização do BioBERT, destacando suas limitações em novos conceitos e sinônimos, mas mostrando melhorias com métodos de desvio estatístico.

Kolpakov et al. (2023) combinam expressões regulares e aprendizado profundo, com BioBERT superando outros métodos em precisão e F1-score na extração de compostos químicos e códigos InChI.

Jofche et al. (2022) utilizam transferência de aprendizado para rotulação de entidades em textos farmacêuticos, atingindo F1-scores de até 96,14% para organizações farmacêuticas e drogas.

Por fim, Çelikmasat et al. (2022) combinam BioBERT, biLSTM e CRF, mostrando superioridade em precisão sobre modelos estáticos e GCN em dados como doenças, drogas e genes.

Moscato et al. (2023) apresenta uma metodologia que agrega diversos datasets de entidades únicas em um único corpus multi-entidade, utilizando modelos transformadores finamente ajustados como professores para treinar um modelo estudante multi-entidade. O TaughtNet supera referências do estado da arte em precisão, recall e F1 score, sendo eficiente para dispositivos com memória limitada e

requisitos de inferência rápidos.

Tho et al. (2023) criaram um modelo BioNER utilizando PubMedBERT, complementado por camadas de BiLSTM e CRF para obter representações ocultas mais detalhadas e etiquetagem de sequência. Testado em 29 conjuntos de dados biomédicos, o modelo mostrou melhor desempenho em termos de F1 score comparado a modelos baseados apenas em PubMedBERT. As principais entidades reconhecidas incluem genes, proteínas, produtos químicos e doenças.

Os artigos a seguir discutem abordagens distintas para o reconhecimento de entidades nomeadas (NER) em textos clínicos e biomédicos.

Ravikumar et al. (2021) apresentam um modelo baseado em uma rede neural LSTM bidirecional com CRF, utilizando embeddings de palavras genéricas. A metodologia inclui a coleta, pré-processamento e divisão dos dados clínicos em conjuntos de treinamento e teste, alcançando alta precisão na classificação de entidades como doenças, diagnósticos, testes e tratamentos, facilitando o trabalho dos especialistas.

Alam et al. (2021) destacam o uso de técnicas de deep learning (DL) para mineração de texto biomédico em tarefas de NER, extração de relações (RE) e resposta a perguntas (QA). A metodologia aplica modelos de DL, como CNNs, LSTMs e BERT, em grandes bases biomédicas, demonstrando que esses modelos superam métodos tradicionais em precisão e generalização, reconhecendo entidades como genes, proteínas, produtos químicos e doenças.

Wang et al. (2021) apresentam um modelo conjunto que realiza simultaneamente o reconhecimento de entidades nomeadas (NER) e a extração de relações (RE) em textos médicos antigos chineses. A técnica utiliza uma camada de Multi-Head Attention para capturar características relacionadas a longas distâncias e um método de treinamento adversarial para melhorar a robustez do modelo. A metodologia envolve a codificação das entradas com a camada Bi-LSTM e a utilização de etiquetas BIO para identificar as entidades. Os resultados experimentais mostraram que o modelo obteve um F1-score de 82,31% no conjunto de dados TCM BIO, demonstrando a eficácia da abordagem na extração conjunta de entidades e relações em literaturas médicas antigas.

Os próximos autores discutem diferentes abordagens para reconhecimento de entidades nomeadas (NER) no domínio clínico e de uso de drogas.

Kormilitzin et al. (2021) apresentam o Med7, um modelo NER treinado em 2 milhões de registros do MIMIC-III e ajustado para tarefas de NER em sete

categorias: nomes de medicamentos, via de administração, frequência, dosagem, força, forma e duração. A técnica de pré-treinamento auto-supervisionado e ajuste fino com dados anotados manualmente resultou em um F1 score micro-médio de 0.957 em MIMIC-III, melhorando de F1=0.762 para F1=0.944 após ajuste fino no CRIS do Reino Unido.

Bajaj et al. (2022) propõem um modelo NER específico para uso de drogas, utilizando supervisão distante com a Drug Abuse Ontology (DAO) e spaCy. A metodologia combina relatórios de tendências de drogas da OSAM e o conjunto de dados OntoNotes 5.0, treinando BERT e BioClinicalBERT para reconhecer entidades genéricas e específicas, como tipos de drogas, condições de saúde e rotas de administração, resultando em um aumento de 8% no F1-score do modelo.

Vários outros autores criaram seus modelos próprios [Zhao et al., 2023; Priya et al., 2021; Wang et al., 2022; Cai et al., 2022; Ardra et al., 2023; Xiao et al., 2021; Mathu et al., 2021; Kashtriya et al., 2023; Miah et al., 2023; Nguyen et al., 2021; Lee et al., 2021; Ramachandran et al., 2021; VeerasekharReddy et al., 2023].

Inicialmente, a presente pesquisa incluiu o somatório dos contextos jurídico e biomédico, porém não obteve êxito por não apresentar nenhum resultado para área de medicamentos. Também não foram localizados, entre os estudos selecionados, tratamentos específicos para o idioma português do Brasil; além de não encontrar aplicação que gere relatório automático que conste os fármacos a serem atendidos pelo SUS oriundos de pedidos concedidos pela justiça.

Neste contexto, as lacunas do presente trabalho de pesquisa foram revisadas e definidas como:

- contribuir com uma base de petições judiciais com medicamentos anotados;
- implementar um classificador de tipo de fármaco nos pedidos de acesso à saúde pública por meio da justiça, com recorte inicial para o Tribunal de Justiça do MS;
- disponibilizar um painel de fármacos mais solicitados ao SUS por meio da justiça;
- disponibilizar um painel de fármacos a serem atendidos pelo SUS oriundos de pedidos;
- contribuir para a gestão do SUS, provocando reflexão sobre os pedidos comuns para uma possível redução da judicialização deles; e

- contribuir para a avaliação de modelos que têm como base o BERT em reconhecimento de entidades nomeadas para extração de medicamentos constantes em textos jurídicos brasileiros.

2.6. Considerações Finais

Ao concluir esta revisão de literatura, destaca-se a importância da aplicação de Inteligência Artificial (IA) e técnicas de Processamento de Linguagem Natural (NLP) no reconhecimento de entidades nomeadas (NER) em textos biomédicos e jurídicos.

Estudos mostram que modelos como BERT, LSTM e suas variantes, quando aplicados a contextos específicos, conseguem identificar com alta precisão nomes de medicamentos e outras entidades biomédicas, contribuindo para a eficiência na extração de informações relevantes em processos judiciais. A pesquisa revela um avanço significativo no uso de modelos de aprendizado profundo para melhorar a precisão e a generalização das tarefas de NER, superando métodos tradicionais baseados em regras.

A análise dos estudos e artigos selecionados revelou que os métodos mais eficazes mundialmente incluem técnicas de aprendizado profundo, como BERT, LSTM e suas variações, que demonstraram alto desempenho em reconhecimento de entidades nomeadas (NER) em textos biomédicos, superando métodos tradicionais baseados em regras. Modelos como BERT-BiLSTM-CRF, SciBERT e BioBERT foram destacados por sua alta precisão e generalização.

Para documentos em português, a pesquisa identificou uma lacuna significativa, com a localização apenas do estudo de Schneider et al. (2020) abordando o NER em textos clínicos. Isso destaca a necessidade de desenvolvimento de modelos treinados em corpora em português, especialmente adaptados à terminologia e estrutura dos textos médicos e jurídicos brasileiros. Os desafios enfrentados no reconhecimento de nomes de medicamentos incluem a complexidade da terminologia biomédica, a variabilidade na estrutura textual dos documentos e a escassez de dados anotados em português. Além disso, a adaptação dos modelos de NER para lidar com as nuances da língua portuguesa e a necessidade de rotulação manual extensiva foram identificados como obstáculos significativos.

Estudos ou aplicações específicas para a extração de informações de petições judiciais no contexto da judicialização da saúde no Brasil não foram localizados entre os artigos revisados. Isso revela uma área de pesquisa inexplorada, com potencial para desenvolver soluções que possam auxiliar na

automação e eficiência da análise de documentos judiciais relacionados à saúde pública. Portanto, há um potencial significativo para o desenvolvimento de modelos treinados em textos jurídicos brasileiros, capazes de extrair informações detalhadas sobre pleitos de medicamentos no âmbito da saúde pública, promovendo maior transparência, precisão e eficiência no sistema judiciário.

CAPÍTULO 3

MODELO DE INTELIGÊNCIA ARTIFICIAL PARA IDENTIFICAR FÁRMACOS JUDICIALIZADOS

3.1. Considerações Iniciais

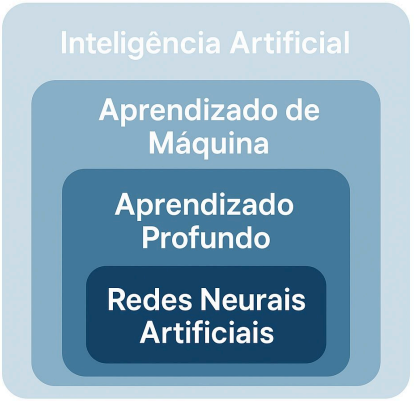
O presente trabalho tem como objetivo principal propor e validar um modelo de inteligência artificial de processamento de linguagem natural (NPL) aplicado à mineração de textos jurídicos, para identificação automática de medicamentos solicitados ao SUS em processos de judicialização da saúde no TJMS.

Foram testados o modelo BioBERTpt, subcategoria de modelo de linguagem específicos de domínio da área biomédica, o Clinicalnerpt-pharmacologic, que utiliza aprendizado supervisionado para reconhecimento de entidades nomeadas (NER) de dados biomédicos. Também foi testado o modelo Longformer-base-4096, com *fine tuning* para reconhecimento de entidades nomeadas. Entre os dois, teve melhor desempenho o BioBERTpt para essa tarefa, o que será apresentado de forma mais detalhada neste capítulo.

Ambos utilizam uma arquitetura de redes neurais transformers (arquitetura BERT) e *Fine-Tuning* em tarefas específicas de reconhecimento de entidades nomeadas (NER).

As redes neurais artificiais (RNAs) são modelos matemáticos inspirados no funcionamento do cérebro humano, com neurônios artificiais organizados em camadas (de entrada, ocultas e de saída), com capacidade computacional adquirida por meio da aprendizagem, por meio de exemplos, e generalização [Rezende, 2005].

As RNAs são a estrutura-base do Aprendizado Profundo (*Deep Learning*), subcampo do aprendizado de máquina que utiliza redes neurais artificiais com múltiplas camadas (*Deep Neural Networks*) para identificar e modelar padrões complexos em grandes volumes de dados. O *Deep Learning*, por sua vez, integra o Aprendizado de Máquina (*Machine Learning*), área que desenvolve algoritmos capazes de aprender a partir de dados, sem programação explícita para a tarefa. Todo esse conjunto se insere no campo mais amplo da Inteligência Artificial (IA), que se dedica à criação de sistemas computacionais capazes de executar funções tradicionalmente humanas, como raciocínio lógico, tomada de decisão e aprendizado contínuo [Rezende, 2005]. A Figura 4 ilustra essa cadeia de tecnologias.



Fonte: Elaborado pela autora com o uso de IA generativa.

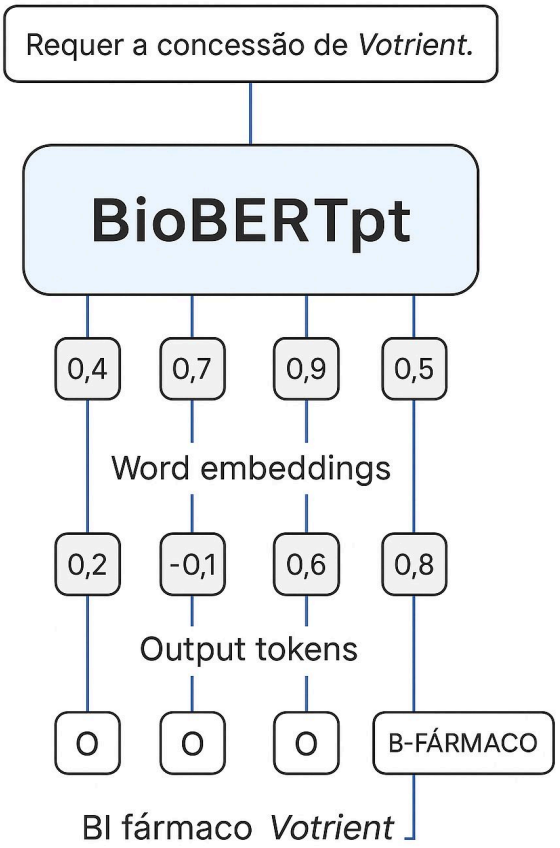
Figura 4: Enquadramento das Redes Neurais na Inteligência Artificial.

Exemplificando o funcionamento de uma RNA, na Figura 5, com o modelo BioBERTpt, ela recebe uma frase, transforma as palavras em vetores numéricos (word embeddings), processa numa rede neural, e devolve rótulos indicando onde há entidades farmacológicas, com o esquema BIO (*Begin, Inside, Ouside*), onde o B indica o início da entidade, o I a continuação dela e o O o não pertencimento a ela.

A abordagem proposta teve os seguintes objetivos específicos:

1- Construir um dataset anonimizado e estruturado a partir das petições iniciais ajuizadas no período de 01/01/2022 a 31/12/2024 de processos judiciais da área da saúde no TJMS, respeitando princípios éticos e legais de privacidade e proteção de dados;

2 - Aplicar os modelos de aprendizado de máquina, como BioBERTpt e Longformer-base-4096, para extrair os medicamentos solicitados ao SUS, utilizando o processo de mineração de textos para analisar petições iniciais ajuizadas no período de 01/01/2022 a 28/05/2025, nas varas de juizados especiais da comarca de Campo Grande, com o assunto de código 12484 – Direito da Saúde – Pública – Fornecimento de Medicamentos e seus correlacionados (12480, 12481, 12492, 12493, 12494, 12495 e 12496), estabelecido pela tabela processual unificada nacional do Conselho Nacional de Justiça;



Fonte: Elaborado pela autora com o uso de IA generativa.

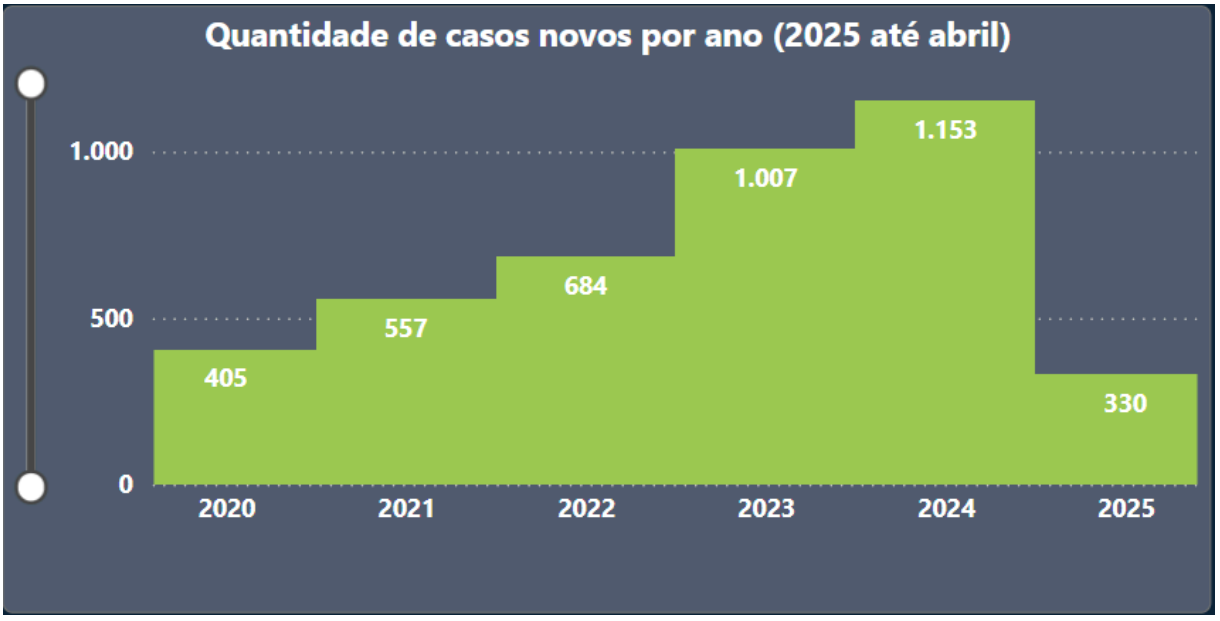
Figura 5: Exemplo do funcionamento de uma RNA, com o modelo BioBERTpt.

3 - Avaliar e comparar a performance dos modelos propostos, adotando métricas robustas como precisão, recall e F1-score, promovendo decisões baseadas em evidências;

4 - Implementar um painel interativo intitulado MED-SUS-MS, Figura 6, que ilustra dados extraídos dos modelos, com a relação dos fármacos vinculados aos processos judicializados, a fim de complementar, com esses dados, as informações no Painel Nacional do Direito à Saúde, do CNJ, onde consta:

a) total de processos ajuizados (distribuídos) na linha do tempo;

Esse dado possibilita observar a quantidade de casos novos por ano da saúde pública. Uma informação importante para prever o crescimento da demanda. No filtro utilizado, no exemplo do Gráfico 6, registra que entre janeiro/2020 e abril/2025 a judicialização da saúde pública em Campo Grande teve um aumento percentual de 184,69%.



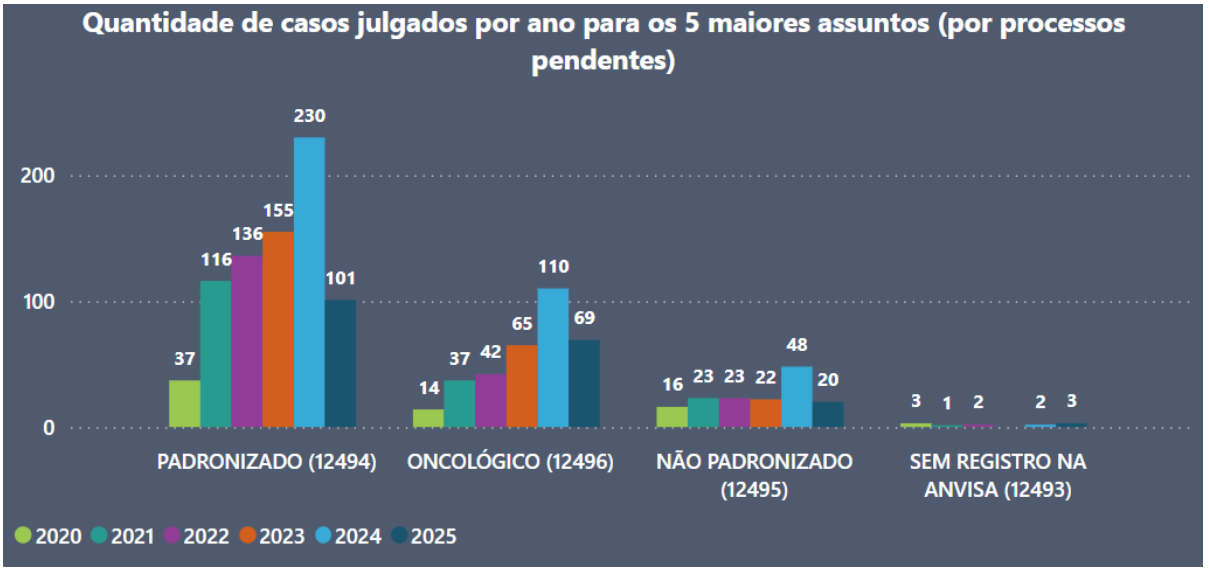
Fonte: Painel de Estatísticas Processuais de Direito à Saúde do CNJ, mai. 2025.
Gráfico 6: Quantidade de casos novos por ano da saúde pública, em Campo Grande entre janeiro/2020 e abril/2025.

b) total de julgados (sentenças proferidas nos processos) por tipo de medicamento;

É a partir do julgamento que o pedido passa a ser uma demanda real para o SUS. Atualmente, a informação existente, demonstrada no Gráfico 7, representa apenas o enquadramento dos medicamentos pedidos. Esse dado é importante para o judiciário, mas não traz nenhuma informação relevante para tomada de decisão pelo executivo.

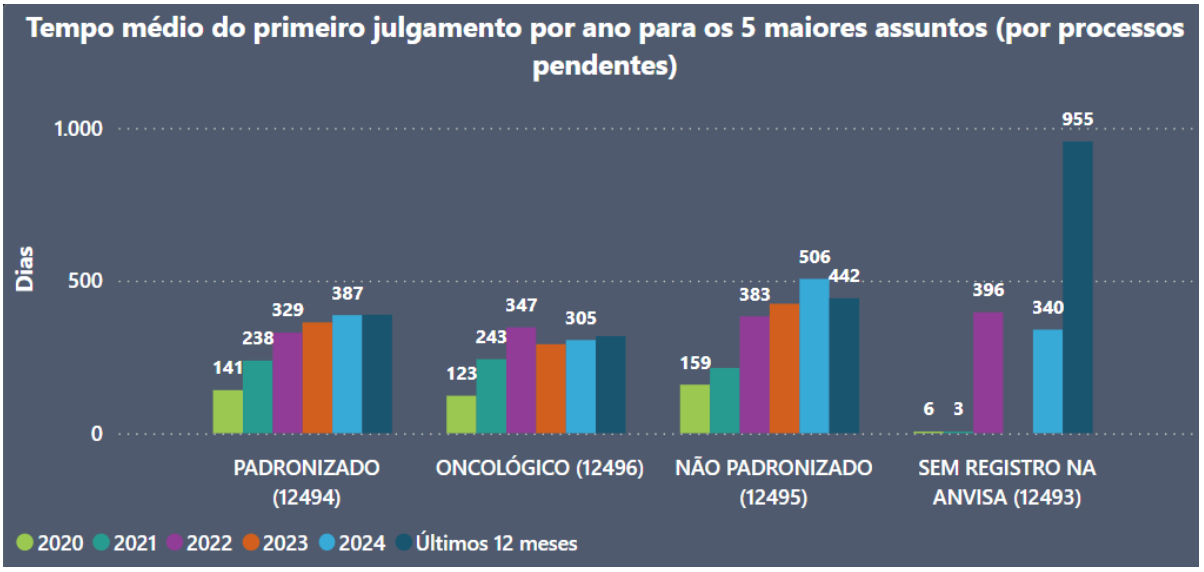
c) tempo médio de tramitação (dias até o primeiro julgamento).

O tempo médio de tramitação até o primeiro julgamento, apresentado no Gráfico 8, pode ser utilizado pelo executivo para calcular quais medicamentos serão demandados ao SUS na média de tempo apresentada. Uma informação importante para preparação de aquisição desses medicamentos, desde que sejam identificados.



Fonte: Painel de Estatísticas Processuais de Direito à Saúde do CNJ, mai. 2025.

Gráfico 7: Quantidade de casos julgados de medicamentos por ano da saúde pública, em Campo Grande entre janeiro/2020 e abril/2025.

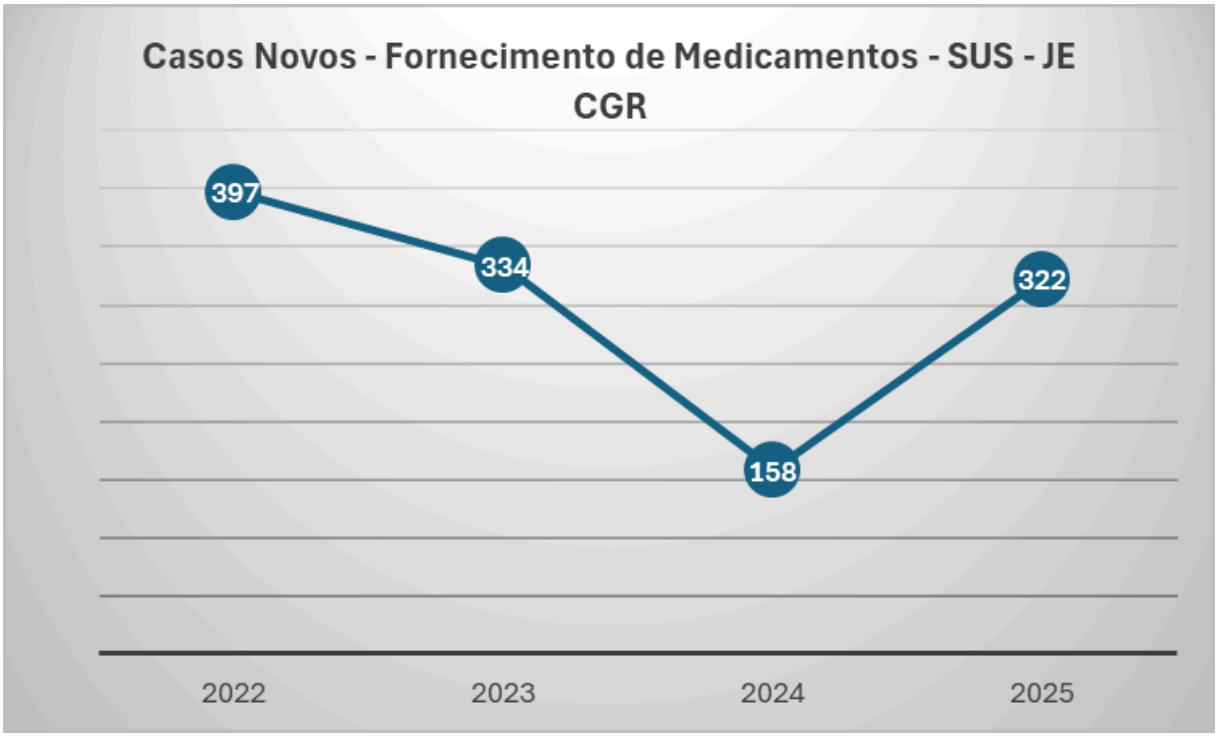


Fonte: Painel de Estatísticas Processuais de Direito à Saúde do CNJ, mai. 2025.

Gráfico 8: Tempo médio de tramitação até o primeiro julgamento, de ações de medicamentos, por ano da saúde pública, em Campo Grande entre janeiro/2020 e abril/2025.

Em resumo, a meta principal deste trabalho era a identificação dos medicamentos solicitados nas petições judiciais. Por se tratar de dados não estruturados, foi utilizado modelo de inteligência artificial para processamento da linguagem natural capaz de recuperar informações de texto.

Destaca-se que, como escopo inicial, para fins de treinamento e piloto da aplicação, foi realizado um recorte para todas a petições do TJMS, ajuizadas no período de 01/01/2022 a 24/05/2025, nas varas do Juizado Especial da comarca de Campo Grande, de solicitação de medicamentos na Saúde Pública, ou seja, onde a parte requerida é o poder público (SUS), totalizando 1.211 processos, conforme distribuição demonstrada no Gráfico 9. Não foram analisadas as petições de processos contra planos de saúde, identificadas pelo assunto Saúde Complementar.



Fonte: Elaborado pela autora com dados extraídos do sistema de processo eletrônico (SAJ) do TJMS.

Gráfico 9: Casos novos (processos distribuídos) nas varas de Juizados Especiais de Campo Grande referentes a fornecimento de medicamentos pelo SUS, de 01/01/2022 a 24/05/2025.

A título de comparação, analisando o cenário nacional as demandas de saúde pública somaram 379.741 processos novos, frente a um total de 39.347.752, representando menos de 1%.

Todavia, apesar de representar um número baixo de ações dessa natureza se comparado ao cenário geral de processos, o crescimento desse tipo de ações tem causado preocupação ao poder público.

De acordo com a lista de alto risco da administração pública federal de 2024, do Tribunal de Contas da União [TCU, 2024], o acesso e a sustentabilidade do SUS é um tema de alto risco devido à tendência de elevação dos gastos, o que tem acontecido por causa da crescente demanda por serviços médicos e hospitalares público, impulsionados por fatores como o envelhecimento da população e pelo aumento de doenças crônicas.

Segundo o relatório, os custos federais com saúde pública foram de 182,9 bilhões em 2023 e podem chegar a 219,5 bilhões em 2030.

Um dos riscos apontados pelo relatório é o crescimento da judicialização da saúde, que tem como efeito o aumento de gastos desassociados da programação orçamentária e financeira da área de saúde. Num cenário onde 70% da população brasileira depende do SUS, o aumento dos gastos pode implicar em diminuição do nível de assistência.

Assim, a aplicação de tecnologias para entendimento das demandas deve contribuir para políticas públicas de contenção de elevação dos gastos com saúde pública no Brasil.

3.2. Metodologia

A seguir são apresentados os procedimentos metodológicos adotados para aplicar técnicas de IA para extração dos nomes dos fármacos em petições iniciais de pleitos ao SUS, os quais foram extraídos de dados totalmente não estruturados, seguinte os seguintes passos: extração das petições e preparação dos dados, escolha do algoritmo e critérios de avaliação de desempenho.

I- Dataset:

O *Dataset* é um conjunto de dados que fornecido pelo TJMS, extraídos do sistema de processo eletrônico SAJ, com 1.504 petições iniciais utilizadas para treinamento e 1.258 petições iniciais utilizadas para teste, oriundas de processos judiciais das Varas dos Juizados Especiais da comarca de Campo Grande, selecionadas a partir do assunto 12484 - Fornecimento de medicamentos da saúde pública e seus subcódigos (12496 - Oncológico, 12492 - Registrado na ANVISA, 12495 - Não

padronizado, 12494 - Padronizado e 12493 - Sem registro na ANVISA).

II- Pré-processamento

Para o pré-processamentos das petições foram realizados os seguintes procedimentos:

1- Ocerização ou Reconhecimento Ótico de Caracteres (OCR)

Essa técnica foi aplicada às petições iniciais, armazenadas no banco de dados do sistema de processo eletrônico do TJMS, no formato PDF, para convertê-las em textos digitais editáveis. Este procedimento é necessário para o processamento do texto na utilização da técnica de reconhecimento de entidade nomeada, usada nesta pesquisa. Para ocerização foi utilizado o software aberto Tesserat OCR, com customizações para o TJMS.

2- Anonimização dos Dados Pessoais das 1504 petições usadas no treinamento

Esse procedimento foi necessário para garantir a conformidade com a Lei Geral de Proteção de Dados, Lei Federal nº 13.709 de 14/08/2018, que estabelece no seu Art. 5º, que dados referentes à saúde são sensíveis, sendo imprescindível o tratamento adequado, neste caso, a anonimização.

A primeira entrega desta pesquisa foi o desenvolvimento de uma aplicação, utilizando spaCy, uma biblioteca de software de código aberto para processamento de linguagem natural, para anonimização de petições iniciais. Ela foi publicada em maio/2024, na plataforma SINAPSES, que é a plataforma nacional do Conselho Nacional de Justiça (CNJ) para armazenamento, treinamento supervisionado, controle de versionamento, distribuição e auditoria dos modelos de IA, além de estabelecer os parâmetros de implementação e funcionamento dessas mesmas IAs. O anonimizador reconhece as entidades nomeadas de dados pessoais e as substitui por tags. (https://www.opantaneiro.com.br/geral/tjms-publica-primeiro-modelo-de-inteligencia-artificial-na-plataforma/214619/)

3- Anotação manual dos dados

A fim de construir uma base de dados anotada para ser utilizada no desenvolvimento da aplicação de reconhecimento dos medicamentos em textos de petições, foram selecionadas 1504 petições iniciais, ocerizadas e anonimizadas, provenientes de processos judiciais referentes à solicitação de fornecimento de medicamentos pelo poder público.

A anotação manual foi conduzida pela própria autora desta pesquisa, com o auxílio de mais dois anotadores. Cada petição foi analisada individualmente, sendo realizada a leitura integral do conteúdo com o objetivo de identificar e marcar os medicamentos solicitados, na seção de pedidos da petição, utilizando o Label Studio, uma ferramenta de rotulagem de dados.

Foram anotados, utilizando a label “MED”, apenas medicamentos, sendo ignorados os insumos, tratamentos e outros pedidos. Foram anotados 3.291 fármacos, sendo que destes, 1776 eram distintos. As anotações levaram em conta a posologia, considerando o nome genérico e comercial como um único fármaco, pois de fato é.

A escolha por uma abordagem manual foi mediante a constatação de que as petições não são padronizadas, cada autor escreve à sua própria maneira e forma de organização do texto. Também a forma de menção dos medicamentos são distintas, onde algumas vezes aparecem com seus nomes comerciais, outras de seus princípios ativos.

Na ampliação do projeto, essa base utilizada para treinamento pode e deve ser ampliada, com petições de outros tribunais, de outras regiões do país, de maneira a evitar o risco de *overfitting*. Esse fenômeno ocorre quando o modelo se ajusta demais aos padrões da base de treinamento, neste caso, às petições regionais, muitas vezes padronizadas pela Defensoria Pública do Estado de MS, e depois tem dificuldades de generalizar para outras. Ou seja, o modelo acerta muito para as petições do MS, mas pode não reconhecer medicamentos em petições de outros estados, ou de outros advogados.

Sobretudo por que as tarefas de NER são sensíveis ao domínio e contexto. Apesar da utilização de modelos capazes de reconhecer contexto, baseado em arquitetura de *transformers*, ele pode memorizar padrões específicos com base no corpus limitado, com risco de não aprender generalizações úteis.

A utilização de um corpus limitado, também oferece o risco já previsto na Resolução CNJ 615/2025, de enviesar o modelo, com padrões de linguagem, estrutura e tipos de ação baseados apenas nas petições ajuizadas no TJMS, limitando também a generalização para utilização nos demais tribunais brasileiros.

4- Tokenização

É a segmentação de palavras por meio da quebra da sequência de caracteres

em um texto, localizando o início e fim de cada palavra, que a partir disso passa a ser chamada de token, [Barbosa (2017) 2017]. Isso, para permitir a remoção de tokens que não são alfabéticos, remoção das stopwords e remoção da acentuação.

5- Lowercasing

É a transformação de todo o texto em letras minúsculas.

6- Pré-processamento baseado em regras

Na primeira execução do modelo de reconhecimento de entidade nomeada, alguns medicamentos foram rotulados incorretamente no esquema BIO (Begin,, Inside, Ouside), usado para segmentação e rotulagem de entidades. Muitos fármacos foram apresentados como “O” (O de *Outside*, onde o token não pertence a nenhuma entidade de interesse), quando deveriam apresentar o tipo *B-PharmacologicSubstance* (B de *Begin*, que marca o início de uma entidade do tipo substância farmacológica) ou *I-PharmacologicSubstance* (I de *Inside*, que indica que está dentro da mesma entidade farmacológica iniciada anteriormente, ou seja, tokens subsequentes da mesma substância). Para correção disso, foi implementada regras linguísticas e semânticas baseadas num dicionário especializado contendo a lista de fármacos da ANVISA.

7- Undersampling

Tem a função de reduzir a quantidade de exemplos da classe majoritária, neste caso, a O - Ouside, que são os “não-medicamentos”. Com essa função foram removidos, por exemplo, cabeçalhos e rodapés.

III- Aplicação do modelo de reconhecimento de entidade nomeada

O modelo escolhido para aplicação neste trabalho tem como base o BERT, que, de acordo com o resultado da revisão de literatura, tem sido utilizado com mais sucesso na tarefa de reconhecimento de fármacos em processamento de linguagem natural (NPL).

De acordo com o [Deep Learning Book], BERT vem da sigla em inglês para *Bidirectional Encoder Representations for Transformers,* traduzido para o português como representações codificadoras bidirecionais de transformadores, significando que aprende as relações contextuais entre as palavras de um texto com base no que está tanto à sua esquerda quanto à sua direita. É um algoritmo de aprendizado profundo, de código aberto, que foi desenvolvido pelos cientistas da *Google IA*

Language em 2018, que utiliza a técnica computacional de redes neurais, para processamento de linguagem natural, que é um ramo da inteligência artificial.

Na revisão de literatura não foi localizado modelo pré-treinado para recuperação de entidades nomeadas em textos jurídicos na língua portuguesa, nem em outros idiomas compatíveis. Assim, foi utilizado o modelo clinicalnerpt-pharmacologic, modelo de Recuperação de Entidade Nomeada (NER), pré-treinado para a reconhecimento e recuperação de fármacos, o qual faz parte do projeto BioBERTpt, da PUCPR.

Esse modelo foi treinado para o corpus SemClinBr, com 1000 anotações clínicas, rotulado com 65.117 entidades, a partir de textos clínicos na língua portuguesa [Oliveira et all., 2022].

Também foi testado o modelo Longformer-base-4096, com *fine tuning* para reconhecimento de entidades nomeadas.

3.3. Resultados

O resultado foi medido quanto ao desempenho dos algoritmos, utilizando testes estatísticos de referência, como acurácia e F1-score. E também foi apresentada a análise das extrações feitas pelos modelos em comparação direta, por amostragem, com a consulta processual realizada no sistema de processo eletrônico do TJMS.

Rezende 2005] informa que existem muitos testes estatísticos para medir a diferença entre algoritmos. Essas métricas são definidas por [Mariano 2021] da seguinte forma:

> Acurácia: considerada uma das métricas mais importantes, avalia o percentual de acertos. Acurácia = Total de acertos / total de itens;
>
> F1-score: métrica que mostra a média harmônica entre a precisão e a revocação: F1-score = 2 * (precisão * sensibilidade) / (precisão + sensibilidade);
>
> Precisão: avalia a quantidade de verdadeiros positivos sobre a soma de todos os positivos: Precisão = Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Positivos);
>
> Revocação: avalia a quantidade de verdadeiros positivos sobre a soma dos verdadeiros positivos e falsos negativos: Precisão = Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Negativos);
>
> Sensibilidade ou Recall: avalia a capacidade de detecção com sucesso dos resultados classificados como positivos: Sensibilidade = Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Negativos);
>
> AUC: Área abaixo da curva ROC (traduzido em curva característica de

> operação do receptor plotada num gráfico para avaliar classificador binário) que indica a probabilidade de duas previsões serem corretamente ranqueadas, onde é obtido valor entre 0 e 1, em que quanto maior, melhor a capacidade do modelo em separar classes.

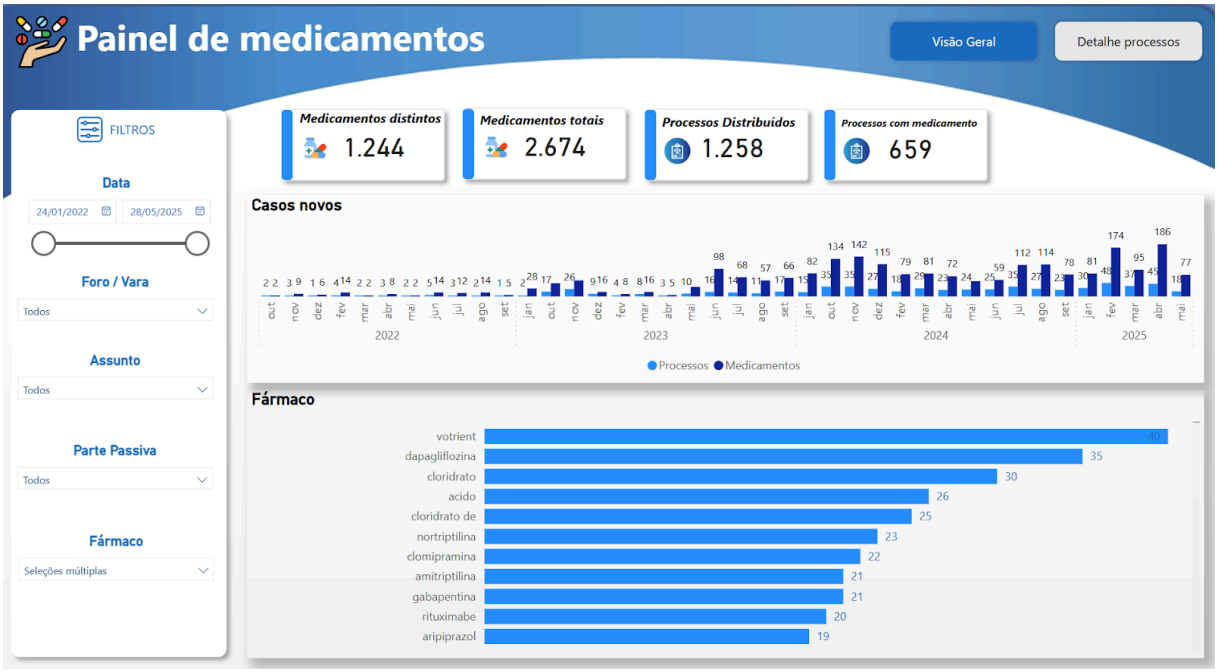
Para o modelo BioBERTpt, a precisão foi de 0.71, o F1 Score de 0.60 e a acurácia de 0.32. O modelo Longformer-base-4096 apresentou a precisão de 0.34, o F1 Score de 0.50 e a acurácia de 0.98. É importante esclarecer que existe uma razão para esses resultados estarem com esses valores, e que a análise das extrações devem ser levadas em consideração para conclusão do melhor modelo entre aqueles testados.

Aplicado em 1.258 petições judiciais, o BioBertpt - Clinicalnerpt-pharmacologic apresentou resultados relevantes, sendo capaz de reconhecer 2.674 entidades nomeadas de fármacos, mesmo em textos com características distintas daquelas para o qual foi inicialmente treinado, conforme exemplos nas Figuras 6 e 7.

A partir desse reconhecimento, foi possível apresentar os resultados no painel MED-SUS-MS, onde o total de medicamentos reconhecidos foi de 2.674, sendo 1.244 distintos, totalizando o valor de causa de R$ 7,7 milhões, no período de 24/01/2022 a 28/05/2025.

Importante realçar que os valores encontrados são com base em processos que tramitam no juizado da fazenda pública, cuja competência se refere a ações com valor da causa até 60 salários mínimos.

Se aplicada a metodologia aos processos que tramitam nas varas da fazenda pública, da justiça comum, cujo valor da causa é superior a 60 salários mínimos, o valor apontado para judicialização será muito superior.



Fonte: Elaborado pela autora com informações extraídas do modelo BioBERTpt - Clinicalnerpt-pharmacologic, com petições e dados entre 22/0/2022 e 28/05/2025, das varas dos juizados especiais de Campo Grande - MS.

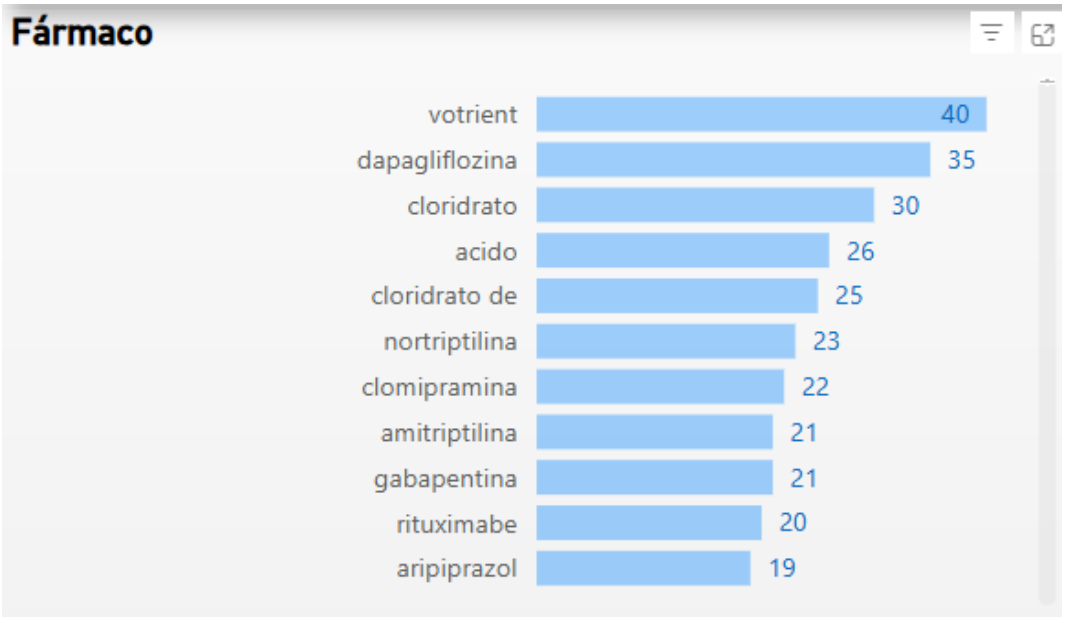
Figura 6: Painel de medicamentos MED-SUS-MS.

Com base nos medicamentos encontrados e valores atribuídos poderá ser direcionada à política pública de saúde. Entre os 5 medicamentos mais demandados no período, e que são identificáveis, apareceram o Votrient em 40 processos, o Dapagliflozina em 35, o Nortriptilina em 23, o Clomipramina em 22 e o Amitriptilina em 21 processos.

Não é possível identificar o valor individual de cada medicamento, uma vez que isso, em geral, não consta nas petições. E também porque nelas, é comum constar a relação de vários fármacos na seção do pedido, além do uso contínuo que pode ser considerado no valor da causa. Mas é possível identificar que ações que contém esses medicamentos somam R$ 327.432,23 para o Votrient, R$ 99.272,64 para o Dapagliflozina, R$ 198.902,76 para o Nortriptilina, R$ 220.366,58 para o Clomipramina e R$ 192.434,98 para o Amitriptilina, para o mesmo período.

Exemplificando o impacto da judicialização, o Votrient 400 mg com 30 comprimidos, que em junho de 2025 custa R$ 6.798,00 para o consumidor, no mês de novembro de 2023 custava R$ 8.130,00 para compra emergencial para o SUS, conforme aponta do DIOGRANDE n. 7.275 na página 26. Aplicando a atualização de medicamentos (IPCA) de 2024 e 2025, o valor a ser considerado é R$8.821,24, ou

seja, 29,76% mais caro para o SUS. Em valores, significaria a economia de aproximadamente R$105 mil no período apenas com esse medicamento, ou R$1,2 milhões no mesmo período, em apenas um município.



Elaborado pela autora com informações extraídas do modelo BioBERTpt - Clinicalnerpt-pharmacologic, com petições e dados entre 22/0/2022 e 28/05/2025, das varas dos juizados especiais de Campo Grande - MS.

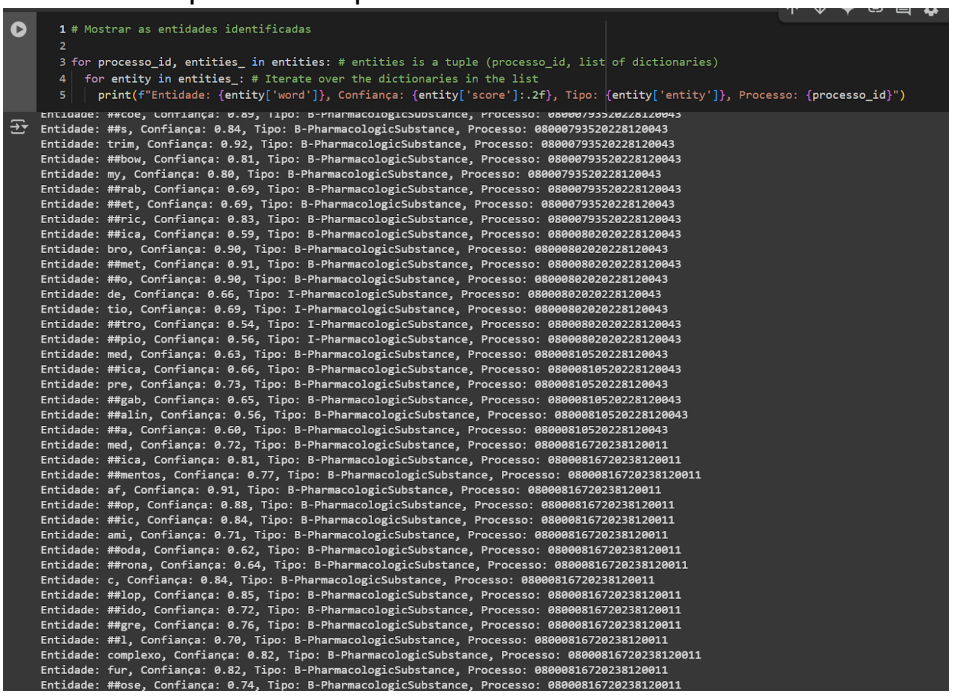
Figura 7: Painel de medicamentos MED-SUS-MS, com ampliação dos medicamentos listados.

Um dado interessante extraído é o valor das ações que contém Rituximabe, que juntas somam R$730.686,36 mesmo tendo sido encontrado em apenas 20 processos. O Prednisona, constante em 15 processos que somam R$646.311,14 de valor de causa, e o Canabidiol, constante em 10 processos, somam R$173.411,60, também integram os valores mais altos apesar do volume menor de pedidos.

Esses 8 medicamentos mencionados fazem parte de uma demanda de 2,8 milhões de reais aos SUS, apenas em Campo Grande, em pouco mais de três anos, sendo responsáveis por um terço do valor total demandado. Para chegar neste resultado, foi aplicado o modelo BioBERTpt que extraiu as entidades farmacológicas em formato de *tokens*, conforme Figura 9.

Fonte: Elaborado pela autora com informações extraídas do modelo BioBERTpt - Clinicalnerpt-pharmacologic, com petições e dados entre 22/0/2022 e 28/05/2025, das varas dos juizados especiais de Campo Grande - MS.

Figura 8: Painel de medicamentos MED-SUS-MS, com total de valor da causa dos processos listados que contém pedidos de medicamentos.



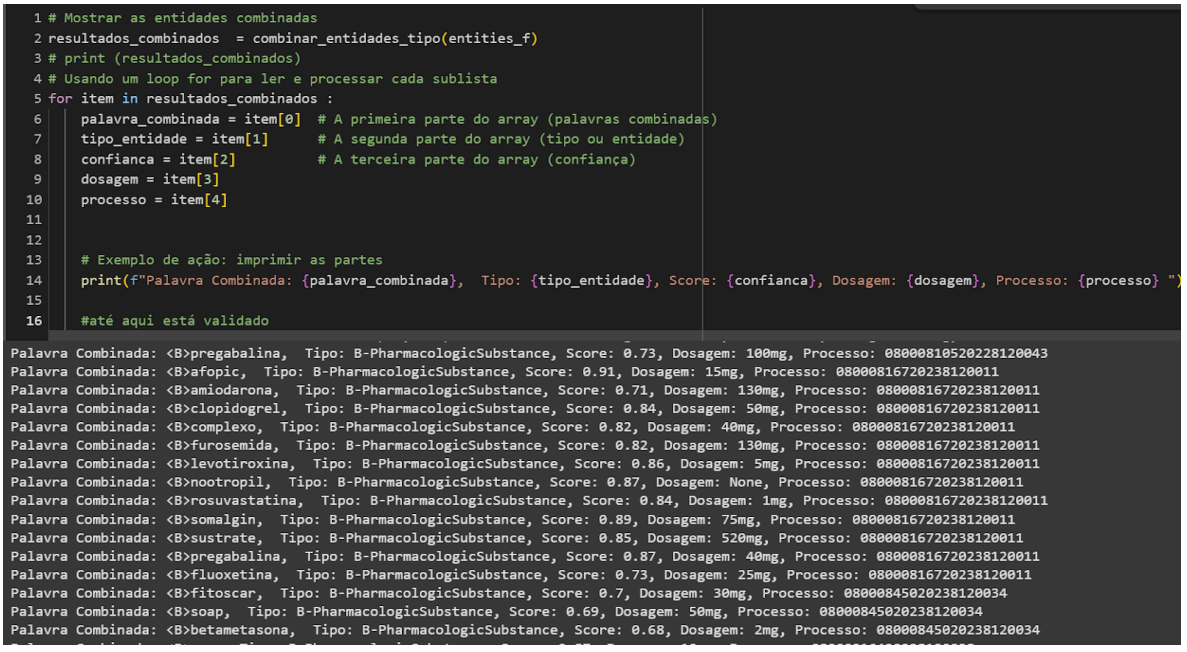
```
# Mostrar as entidades identificadas

for processo_id, entities_ in entities: # entities is a tuple (processo_id, list of dictionaries)
  for entity in entities_: # Iterate over the dictionaries in the list
    print(f"Entidade: {entity['word']}, Confiança: {entity['score']:.2f}, Tipo: {entity['entity']}, Processo: {processo_id}")
```

```
Entidade: ##s, Confiança: 0.84, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: trim, Confiança: 0.92, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: ##bow, Confiança: 0.81, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: my, Confiança: 0.80, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: ##rab, Confiança: 0.69, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: ##et, Confiança: 0.69, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: ##ric, Confiança: 0.83, Tipo: B-PharmacologicSubstance, Processo: 08000793520228120043
Entidade: ##ica, Confiança: 0.59, Tipo: B-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: bro, Confiança: 0.90, Tipo: B-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: ##met, Confiança: 0.91, Tipo: B-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: ##o, Confiança: 0.90, Tipo: B-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: de, Confiança: 0.66, Tipo: I-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: tio, Confiança: 0.69, Tipo: I-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: ##tro, Confiança: 0.54, Tipo: I-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: ##pio, Confiança: 0.56, Tipo: I-PharmacologicSubstance, Processo: 08000802020228120043
Entidade: med, Confiança: 0.63, Tipo: B-PharmacologicSubstance, Processo: 08000810520228120043
Entidade: ##ica, Confiança: 0.66, Tipo: B-PharmacologicSubstance, Processo: 08000810520228120043
Entidade: pre, Confiança: 0.73, Tipo: B-PharmacologicSubstance, Processo: 08000810520228120043
Entidade: ##gab, Confiança: 0.65, Tipo: B-PharmacologicSubstance, Processo: 08000810520228120043
Entidade: ##alin, Confiança: 0.56, Tipo: B-PharmacologicSubstance, Processo: 08000810520228120043
Entidade: ##a, Confiança: 0.60, Tipo: B-PharmacologicSubstance, Processo: 08000810520228120043
Entidade: med, Confiança: 0.72, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##ica, Confiança: 0.81, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##mentos, Confiança: 0.77, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: af, Confiança: 0.91, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##op, Confiança: 0.88, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##ic, Confiança: 0.84, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ami, Confiança: 0.71, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##oda, Confiança: 0.62, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##rona, Confiança: 0.64, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: c, Confiança: 0.84, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##lop, Confiança: 0.85, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##ido, Confiança: 0.72, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##gre, Confiança: 0.76, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##l, Confiança: 0.70, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: complexo, Confiança: 0.82, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: fur, Confiança: 0.82, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
Entidade: ##ose, Confiança: 0.74, Tipo: B-PharmacologicSubstance, Processo: 08000816720238120011
```

Fonte: Resultado da execução do modelo BioBERTpt - Clinicalnerpt-pharmacologic pela autora.

Figura 9: Entidades PharmacologicSubstance identificadas pelo modelo em tokens.

O resultado foi combinado, conforme Figura 10, para gerar o nome identificável do medicamento, e a partir disso, foram feitas comparações analíticas entre o resultado apresentado pelo modelo e o processo judicial. No exemplo utilizado, para o processo nº 0800081-67.2023.8.12.0011, o modelo apresentou os seguintes medicamentos: Afopic, Amiodarona, Clopidogrel, Complexo, Furosemida, Levotiroxina, Nootropil, Rosuvastatina, Somalgin, Sustrate, Pregabalina e Fluoxetina, exatamente os doze medicamentos solicitados na petição, Figura 11, demonstrando a capacidade de reconhecimento de fármacos pelo modelo, mesmo em textos diferentes daqueles para o qual foi inicialmente treinado (corpus biomédicos).

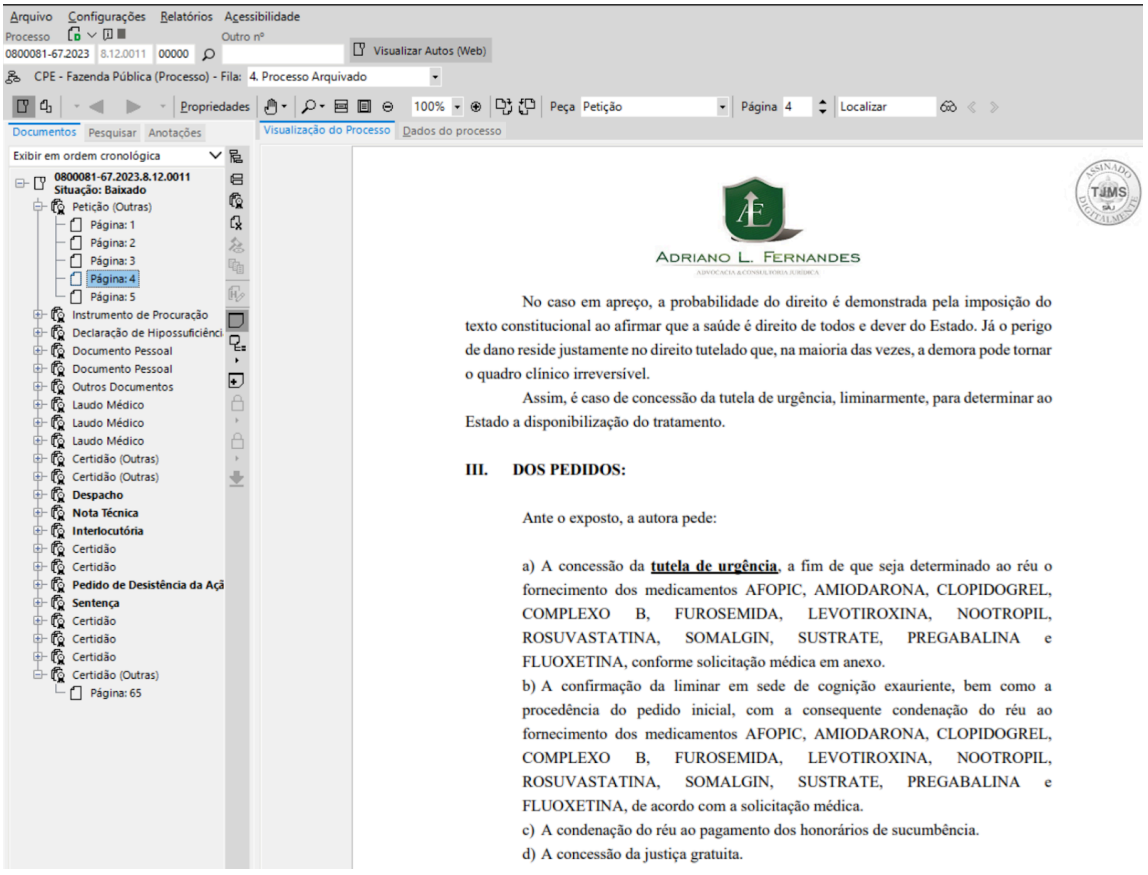


```
# Mostrar as entidades combinadas
resultados_combinados  = combinar_entidades_tipo(entities_f)
# print (resultados_combinados)
# Usando um loop for para ler e processar cada sublista
for item in resultados_combinados :
    palavra_combinada = item[0]  # A primeira parte do array (palavras combinadas)
    tipo_entidade = item[1]      # A segunda parte do array (tipo ou entidade)
    confianca = item[2]          # A terceira parte do array (confiança)
    dosagem = item[3]
    processo = item[4]


    # Exemplo de ação: imprimir as partes
    print(f"Palavra Combinada: {palavra_combinada},  Tipo: {tipo_entidade}, Score: {confianca}, Dosagem: {dosagem}, Processo: {processo} "

    #até aqui está validado
```

```
Palavra Combinada: <B>pregabalina,  Tipo: B-PharmacologicSubstance, Score: 0.73, Dosagem: 100mg, Processo: 08000810520228120043
Palavra Combinada: <B>afopic,  Tipo: B-PharmacologicSubstance, Score: 0.91, Dosagem: 15mg, Processo: 0800081672023812 0011
Palavra Combinada: <B>amiodarona,  Tipo: B-PharmacologicSubstance, Score: 0.71, Dosagem: 130mg, Processo: 08000816720238120011
Palavra Combinada: <B>clopidogrel,  Tipo: B-PharmacologicSubstance, Score: 0.84, Dosagem: 50mg, Processo: 08000816720238120011
Palavra Combinada: <B>complexo,  Tipo: B-PharmacologicSubstance, Score: 0.82, Dosagem: 40mg, Processo: 0800081672023812 0011
Palavra Combinada: <B>furosemida,  Tipo: B-PharmacologicSubstance, Score: 0.82, Dosagem: 130mg, Processo: 08000816720238120011
Palavra Combinada: <B>levotiroxina,  Tipo: B-PharmacologicSubstance, Score: 0.86, Dosagem: 5mg, Processo: 08000816720238120011
Palavra Combinada: <B>nootropil,  Tipo: B-PharmacologicSubstance, Score: 0.87, Dosagem: None, Processo: 0800081672023812 0011
Palavra Combinada: <B>rosuvastatina,  Tipo: B-PharmacologicSubstance, Score: 0.84, Dosagem: 1mg, Processo: 08000816720238120011
Palavra Combinada: <B>somalgin,  Tipo: B-PharmacologicSubstance, Score: 0.89, Dosagem: 75mg, Processo: 0800081672023812 0011
Palavra Combinada: <B>sustrate,  Tipo: B-PharmacologicSubstance, Score: 0.85, Dosagem: 520mg, Processo: 0800081672023812 0011
Palavra Combinada: <B>pregabalina,  Tipo: B-PharmacologicSubstance, Score: 0.87, Dosagem: 40mg, Processo: 08000816720238120011
Palavra Combinada: <B>fluoxetina,  Tipo: B-PharmacologicSubstance, Score: 0.73, Dosagem: 25mg, Processo: 0800081672023812 0011
Palavra Combinada: <B>fitoscar,  Tipo: B-PharmacologicSubstance, Score: 0.7, Dosagem: 30mg, Processo: 0800084502023812 0034
Palavra Combinada: <B>soap,  Tipo: B-PharmacologicSubstance, Score: 0.69, Dosagem: 50mg, Processo: 0800084502023812 0034
Palavra Combinada: <B>betametasona,  Tipo: B-PharmacologicSubstance, Score: 0.68, Dosagem: 2mg, Processo: 08000845020238120034
```

Fonte: Resultado da execução do modelo BioBERTpt - Clinicalnerpt-pharmacologic pela autora, com destaque para o processo 0800081-67.2023.8.12.0011.

Figura 10: Exemplo 1: Entidades PharmacologicSubstance identificadas pelo modelo combinadas.



Fonte: Visualização do processo 0800081-67.2023.8.12.0011 no Sistema SAJ do TJMS.

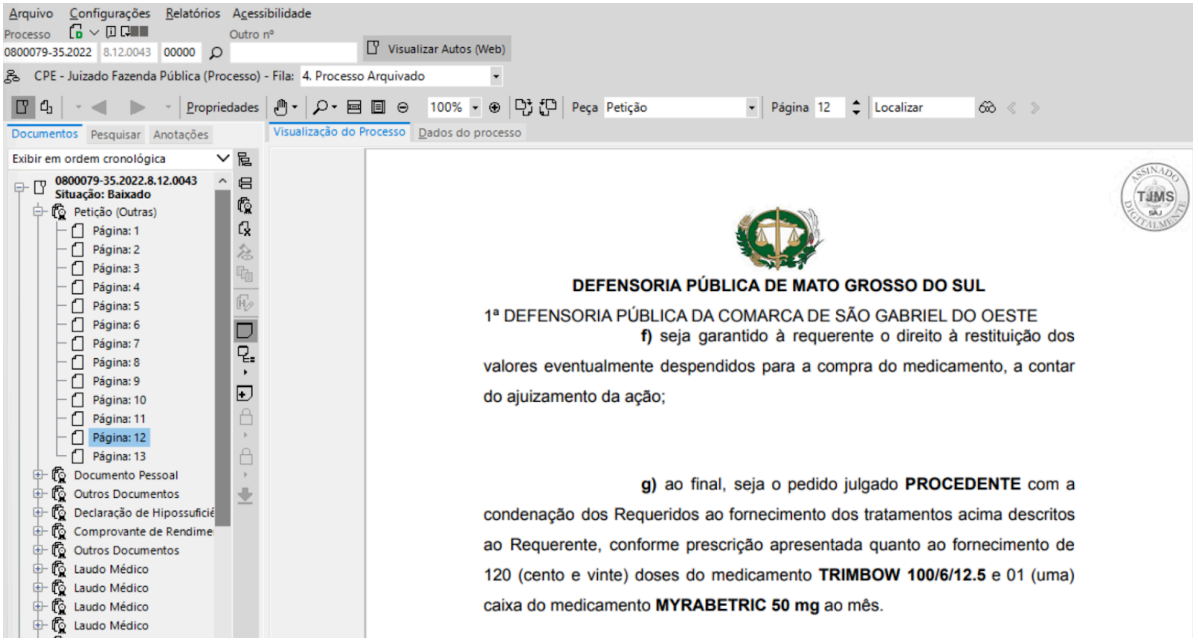
Figura 11: Exemplo 1: Petição inicial do processo 0800081-67.2023.8.12.0011.

Outros exemplos também reconhecidos pelo modelo Trimbow e Myrabetric, que contém os mesmos dois da petição do processo nº 080079-35.2022.8.12.0043, Figuras 12 e 13.

```
Palavra Combinada: <B>trimbow,  Tipo: B-PharmacologicSubstance, Score: 0.92, Dosagem: 60mg, Processo: 08000793520228120043
Palavra Combinada: <B>myrabetric,  Tipo: B-PharmacologicSubstance, Score: 0.8, Dosagem: 400g, Processo: 08000793520228120043
```

Fonte: Resultado da execução do modelo BioBERTpt - Clinicalnerpt-pharmacologic pela autora, com destaque para o processo 080079-35.2022.8.12.0043.

Figura 12: Exemplo 2: Entidades *PharmacologicSubstance* identificadas pelo modelo combinadas.



Fonte: Visualização do processo 080079-35.2022.8.12.0043 no Sistema SAJ do TJMS.

Figura 13: Exemplo 2: Petição inicial do processo 080079-35.2022.8.12.0043.

E também o Prebictal, Cymbi, Duloxetina, Glifager, Metformina, Qtern, Sexagliptina e Degagliflozina para o nº 0800136-46.2022.8.12.0110. Neste, existe uma observação sobre o comportamento do modelo, onde ele reconhece de forma individualizada os medicamentos com nomes genéricos e comerciais, tratando-os como distintos, como é o caso do Cymbi (Duloxetina), Figuras 14 e 15.

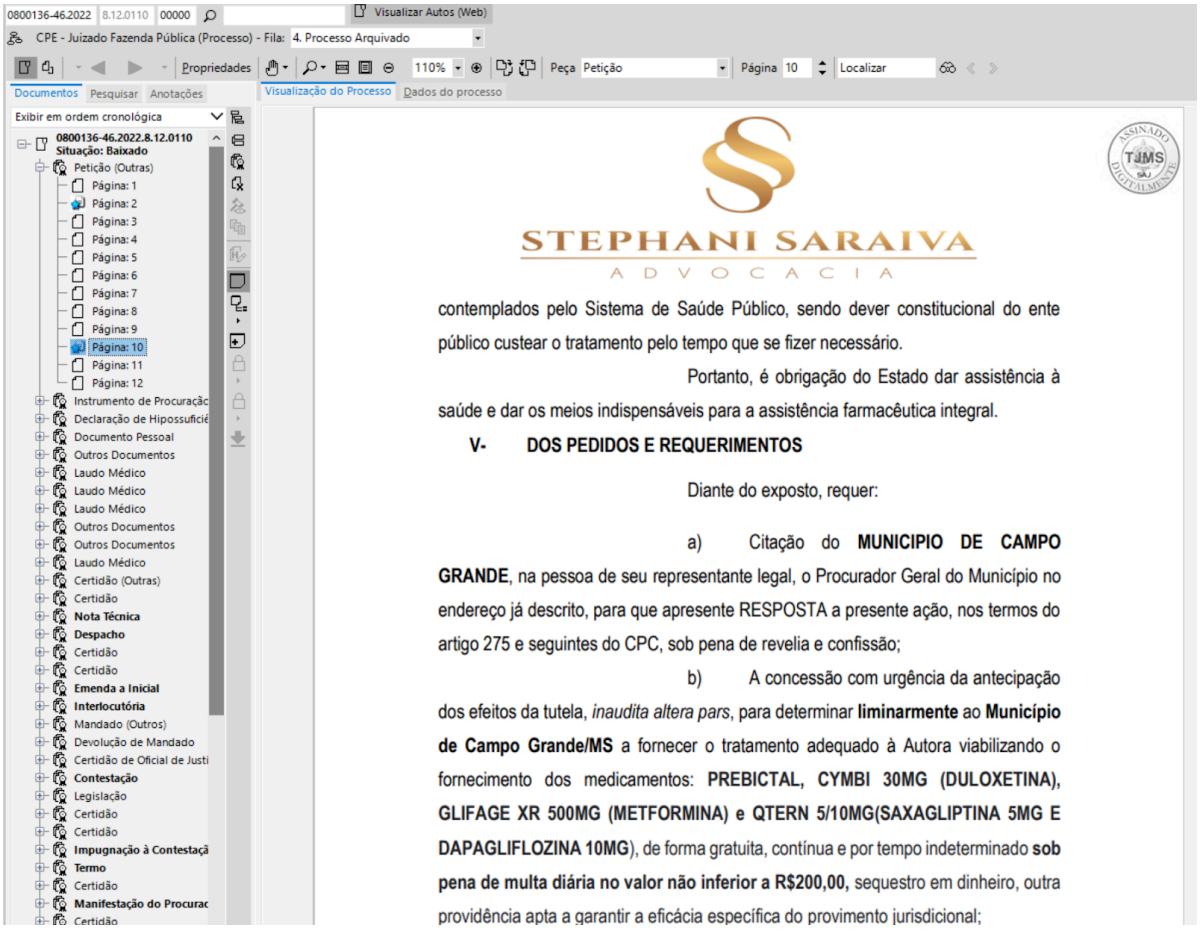
```
Palavra Combinada: <B>prebictal,  Tipo: B-PharmacologicSubstance, Score: 0.91, Dosagem: 100 mg, Processo: 08001364620228120110
Palavra Combinada: <B>cymbi,  Tipo: B-PharmacologicSubstance, Score: 0.94, Dosagem: None, Processo: 0800136462022812011O
Palavra Combinada: <B>duloxetina,  Tipo: B-PharmacologicSubstance, Score: 0.93, Dosagem: 10mg, Processo: 08001364620228120110
Palavra Combinada: <B>glifager,  Tipo: B-PharmacologicSubstance, Score: 0.88, Dosagem: 1mg, Processo: 0800136462022812011O
Palavra Combinada: <B>metformina,  Tipo: B-PharmacologicSubstance, Score: 0.93, Dosagem: 15 mg, Processo: 08001364620228120110
Palavra Combinada: <B>qtern,  Tipo: B-PharmacologicSubstance, Score: 0.91, Dosagem: 5 ml, Processo: 0800136462022812011O
Palavra Combinada: <B>saxagliptina,  Tipo: B-PharmacologicSubstance, Score: 0.9, Dosagem: 025mcg, Processo: 08001364620228120110
Palavra Combinada: <B>dapagliflozina,  Tipo: B-PharmacologicSubstance, Score: 0.93, Dosagem: None, Processo: 08001364620228120110
```

Fonte: Resultado da execução do modelo BioBERTpt - Clinicalnerpt-pharmacologic pela autora, com destaque para o processo 0800136-46.2022.8.12.0110.

Figura 14: Exemplo 3: Entidades PharmacologicSubstance identificadas pelo modelo combinadas.

Apesar do seu potencial, o desempenho foi insatisfatório perante as métricas de F1-Score e Acurácia, sendo 0.60 e 0.32, para os textos jurídicos. Isso pode ser

explicado a partir de algumas considerações, expostas a seguir.



Fonte: Visualização do processo 0800136-46.2022.8.12.0110 no Sistema SAJ do TJMS.

Figura 15: Exemplo 3: Visualização da petição inicial do processo 0800136-46.2022.8.12.0110.

A primeira delas é a diferença entre o que foi anotado e o resultado apresentado pelo modelo. As anotações foram feitas, de forma manual, utilizando os fármacos da seção “Dos pedidos”, que nem sempre está identificada desta forma, mas que, normalmente, está presente na última ou penúltima página da petição judicial. Essas anotações consideraram a posologia como parte do nome do medicamento. Também foi considerado nomes entre parênteses como sendo um medicamento só.

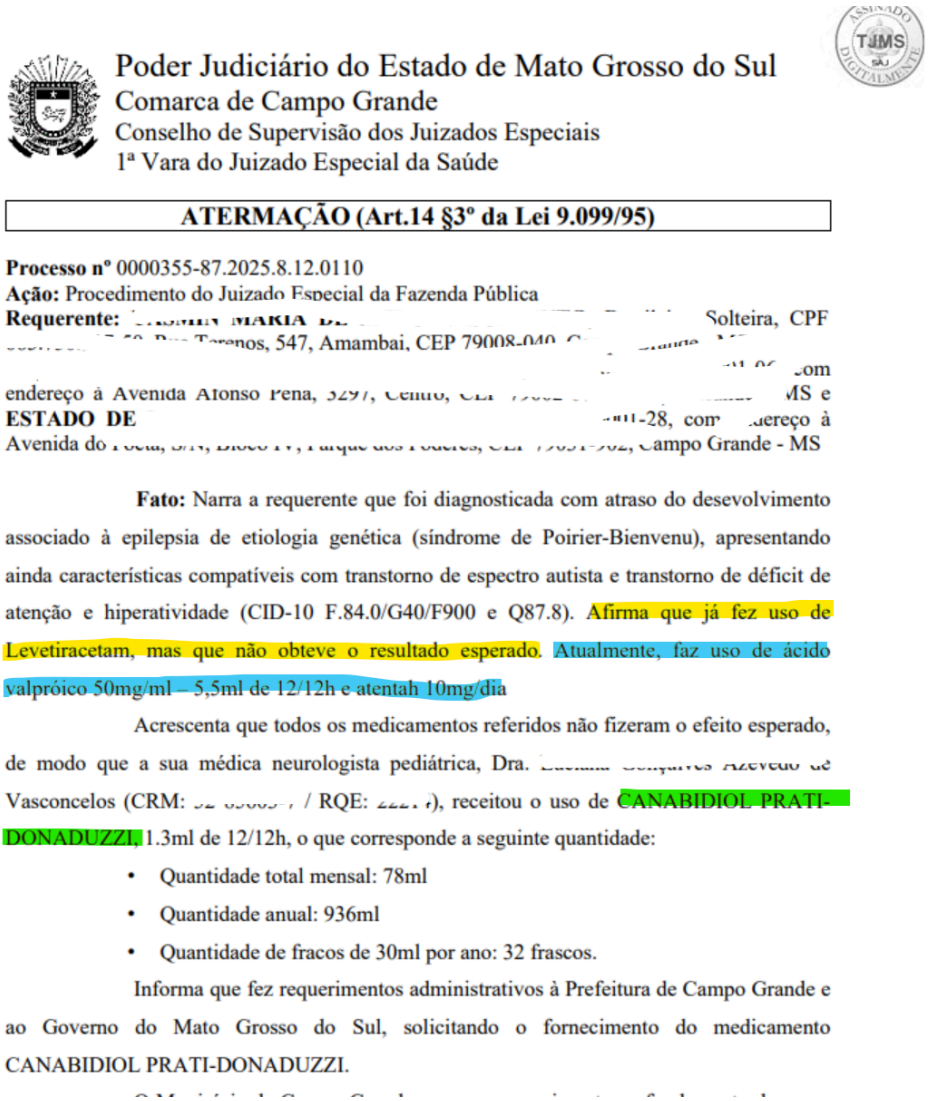
O modelo utilizado tende a reconhecer todos os medicamentos do texto, e é comum que, na justificativa do pedido, sejam informados fármacos testados no paciente e que não tiveram resultado, por exigência do Tema Repetitivo nº 106 do Superior Tribunal Justiça. Porém, estes não foram anotados por não fazerem parte

do escopo da informação pretendida, que não são aqueles solicitados. Mas, isso compromete a medição do resultado como falso positivo, como no exemplo da Figura 16.

Os nomes entre parênteses ora referem-se à nomenclatura genérica, ora comercial. E, para o modelo, são entidades distintas, como demonstrado nas Figuras 14 e 15, também comprometendo a medição do resultado na comparação entre anotações e descobertas.

Outro ponto é a posologia, pois, uma vez que o modelo não apresenta e a anotação sim, é lido como falso positivo, apesar de não ser.

Outra consideração que deve ser feita é a diferença de Domínio Textual, uma vez que o modelo foi pré-treinado para o corpus clínico, e está sendo aplicado no domínio jurídico. A disposição e linguagem desses textos em nada se comparam entre si. São estruturas sintáticas diferentes, não havendo, nas petições, o contexto clínico, além de terminologias médicas e jurídicas específicas de cada área. Para finalizar, as petições possuem quantidades de tokens muito mais elevadas (texto maior) do que um receituário médico, por exemplo.



Poder Judiciário do Estado de Mato Grosso do Sul
Comarca de Campo Grande
Conselho de Supervisão dos Juizados Especiais
1ª Vara do Juizado Especial da Saúde

ATERMAÇÃO (Art.14 §3º da Lei 9.099/95)

Processo nº 0000355-87.2025.8.12.0110
Ação: Procedimento do Juizado Especial da Fazenda Pública
Requerente: [redacted] Solteira, CPF [redacted] Terenos, 547, Amambai, CEP 79008-040 [redacted] com endereço à Avenida Afonso Pena, 3297, Centro, [redacted] MS e **ESTADO DE** [redacted]-28, com [redacted]ereço à Avenida do [redacted], Campo Grande - MS

Fato: Narra a requerente que foi diagnosticada com atraso do desevolvimento associado à epilepsia de etiologia genética (síndrome de Poirier-Bienvenu), apresentando ainda características compatíveis com transtorno de espectro autista e transtorno de déficit de atenção e hiperatividade (CID-10 F.84.0/G40/F900 e Q87.8). Afirma que já fez uso de Levetiracetam, mas que não obteve o resultado esperado. Atualmente, faz uso de ácido valpróico 50mg/ml – 5,5ml de 12/12h e atentah 10mg/dia

Acrescenta que todos os medicamentos referidos não fizeram o efeito esperado, de modo que a sua médica neurologista pediátrica, Dra. [redacted] Vasconcelos (CRM: [redacted] / RQE: [redacted]), receitou o uso de CANABIDIOL PRATI-DONADUZZI, 1.3ml de 12/12h, o que corresponde a seguinte quantidade:

- Quantidade total mensal: 78ml
- Quantidade anual: 936ml
- Quantidade de fracos de 30ml por ano: 32 frascos.

Informa que fez requerimentos administrativos à Prefeitura de Campo Grande e ao Governo do Mato Grosso do Sul, solicitando o fornecimento do medicamento CANABIDIOL PRATI-DONADUZZI.

Fonte: Visualização da petição inicial do processo 0000355-87.2025.8.12.0110 no sistema SAJ do TJMS

Figura 16: Exemplo de petição com menção de medicamentos em outras seções distintas do pedido.

O modelo também apresentou ruídos durante a etapa de extração, identificando termos genéricos como “medicamentos”, “medicamento”, “insumo”, “produto”, como entidades relevantes, o que compromete parcialmente os resultados.

Por fim, foi realizado um teste com o modelo Longformer-base-4096, que não possuía *fine tuning* específico para reconhecimento de fármaco, mas que foi treinado com um conjunto de dados anotados com 900 documentos. Foi feita uma distribuição das classes, com aplicação de pesos por classe devido ao número muito maior da classe O - Ouside, mas, apesar da métrica estatística de acurácia superior

ao BioBertpt - Clinicalnerpt-phamacologic, os fármacos extraídos possuem muito mais ruídos, conforme Figura 17. Razões pelas quais deixou de ser considerado

word	Contagem de word	Contagem de processo
Insulina	43	18
o	43	34
Novo	42	40
Bairro	39	38
fraldas	39	21
Nilton	37	29
ins	30	28
dieta	20	16
insumos	15	8
fórmula	13	5
Hiram	13	10
ml	13	12
de	12	10
fármaco	12	8
as	11	9
Helkis Clark Ghizzi	11	7
Campo	10	10
insulinas	9	5
da	8	5
DAS	7	7
Lorraine	7	1
•	6	4
Almeida	6	5
noventa	6	6
.	5	5
6 (	5	5
a	5	5
Amarildo	5	4
Av	5	5
fralda	5	4
frascos	5	5
I)	5	5
Jardim	5	4
LANTUS	5	2
setenta	5	5
Total	**1938**	**323**

Fonte: Resultado compilado da lista oriunda da execução do modelo Longformer-base-4096 pela autora

Figura 17: Lista de “medicamentos” identificados pelo modelo Longformer-base-4096.

3.4. Considerações Finais

A proposta de aplicação de modelo para reconhecimento automatizado de nomes de medicamentos em petições judiciais é tecnicamente viável, embora haja grandes desafios práticos. A revisão de literatura demonstrou que, em que pese a existência de muitos estudos sobre reconhecimento de entidade nomeada no domínio Biomédico, para extração de medicamentos, há uma lacuna na aplicação em textos do domínio jurídico, sobretudo na língua portuguesa.

O desenvolvimento da solução inicialmente proposta, que era a geração de estatísticas com base na extração de medicamentos solicitados nas petições judiciais, enfrentou dificuldades pelas particularidades do domínio jurídico, como por exemplo, a presença dos nomes de medicamentos em várias partes das petições,

não somente na seção do pedido em si, o que gera inconsistência no resultado apresentado. Além disso, as petições não são padronizadas, dificultando a criação de regras acessórias para atingir a finalidade.

Com isso, o modelo apresentou falsos positivos, como a identificação da palavra “medicamentos” e derivadas, “insumos”, “farmacos/’, “0, 9 %” (que refere-se a soro fisiológico), em várias petições, como no caso dos processos n. 0800792-24.2022.8.12.0006, 0800784-69.2023.8.12.0052, 0800794-14.2021.8.12.0043, 0800784-98.2023.8.12.0010, 0800790-05.2023.8.12.0011 e 0800798-25.2023.8.12.0029.

Casos como no processo n. 0800015-77.2022.8.12.0058, onde o pedido é por Pazopanibe (votrient) ou Sunitinibe, também foram apresentados como três medicamentos distintos sendo solicitados, mesmo não sendo o caso. Situações onde os medicamentos foram listados apenas para cumprimento dos requisitos do Tema 106 do STJ, como comprovação da ineficácia de outros fármacos, a exemplo do processo n. 0800028-63.2023.8.12.0051, que apresenta Metformina e Varfarina na lista do que já foi utilizado sem sucesso, mas solicita Xarelto 20 mg (01 comprimido ao dia) e Glifage 1000 mg também contribuem para falsos positivos listando todos os quatro como solicitados, por exemplo.

Nesta avaliação dos processos, há aquelas petições que listam medicamentos como exemplos de que não há protocolos unificados de tratamento, como ocorreu no processo n. 0810673-96.2025.8.12.0110, que listou vários, mas requer apenas o Cloridrato de Paroxetina, e o modelo também apresentou como vários medicamentos requisitados, não considerando o contexto.

Também há falsos negativos, como o caso do processo n. 0812790-94.2024.8.12.0110, que não listou medicamentos, porém há solicitação do fármaco HEPA-MERZ – 0,6 G/G.

Não obstante esses obstáculos, a criação de uma base anotada manualmente, de 1504 petições, com identificação de 3291 medicamentos, sendo 1776 distintos, representa um avanço relevante para as próximas pesquisas nessa área, pois é um dos primeiros datasets com as características da junção do domínio jurídico e do domínio biomédico, na língua portuguesa.

A visualização dos dados extraídos no painel MED-SUS-MS, ainda que em fase embrionária, tem o potencial de se tornar uma importante ferramenta de apoio à tomada de decisões acerca da judicialização da saúde. Apesar de, ainda necessitar de aprimoramento para representar os quantitativos solicitados por medicamento no

judiciário sul-mato-grossense, a partir deste painel é possível saber quais aparecem com mais frequência nas petições judiciais, apontando tendências.

CAPÍTULO 4
CONCLUSÃO

4.1. Resultados

No presente trabalho foi apresentado o desenvolvimento de uma abordagem computacional baseada em técnicas de mineração de texto e aprendizado de máquina, com foco na identificação de medicamentos em processos judiciais relacionados à saúde, no âmbito do TJMS. A aplicação do modelo de IA demonstrou viabilidade na extração automatizada de entidades nomeadas em textos jurídicos, com potencial para apoiar a formulação de políticas públicas e contribuir para maior eficiência na gestão da judicialização da saúde.

Os resultados do trabalho reafirmam o compromisso com a inovação responsável, o uso estratégico de dados e a melhoria contínua da administração pública, abrindo caminho para diversas linhas de pesquisa e desenvolvimento, tanto na área técnico-computacional quanto no contexto jurídico-institucional, especialmente considerando o avanço das normativas e ferramentas do Conselho Nacional de Justiça (CNJ). O presente trabalho representa não apenas inovação tecnológica e uma aproximação entre academia e judiciário, mas sim um avanço efetivo na democratização do acesso à justiça e na governança pública baseada em evidências.

A partir deste trabalho é possível destacar que a aplicação de modelos de IA para extrair nomes de medicamentos em petições judiciais, ainda que com limitações, é uma das estratégicas mais ágeis e eficientes mediante o grande volume de dados nos tribunais. Uma das principais contribuições deste trabalho foi a

criação de um *dataset* de 1.504 petições judiciais com os medicamentos anotados, com potencial ineditismo no contexto da língua portuguesa.

Embora o desempenho dos modelos testados não tenham sido totalmente satisfatórios, os resultados são relevantes para futuras melhorias, como o *fine-tuning* dos modelos existentes e a superação do desafio da delimitação da extração dos medicamentos a apenas a seção dos pedidos da petição, uma vez que são esses os verdadeiramente solicitados. Não podem ser considerados na análise, os medicamentos que estão em outras partes da petição.

O reconhecimento de medicamentos em petições judiciais no Brasil precisa ser perseguido pela ciência pela importância do impacto positivo no custo financeiro ao SUS. Este sistema é exclusivo do Brasil, e garante o acesso universal e gratuito à saúde para mais de duzentos milhões de brasileiros, sendo a judicialização uma contribuidora para o desequilíbrio financeiro do sistema. Por isso a necessidade de buscar a desjudicialização.

O painel implementado neste trabalho, MED-SUS-MS, representa uma inovação na visualização dos dados de judicialização da saúde, com indicação dos medicamentos mais demandados. A partir destes dados de judicialização, o trabalho propiciou o estreitamento da parceria entre o judiciário e o executivo sul-mato-grossense, com vistas a criar políticas de minimização da judicialização da saúde para acesso aos medicamentos que são de direito aos usuários do SUS, beneficiando a sociedade e reduzindo desperdícios de tempo e recursos público com judicializações que podem ser evitadas, tornando as cidades mais inteligentes.

Sanadas as limitações apresentadas, o produto da pesquisa tem potencial de ser expandido para nível nacional, uma vez que a judicialização da saúde ocorre em todo o país e segue padrões semelhantes.

Por fim, os resultados deste trabalho reforçam o compromisso da Ciência, Tecnologia e Inovação com os Objetivos de Desenvolvimento Sustentável (ODS), especialmente o ODS 3 (Saúde e Bem-Estar) e o ODS 16 (Paz, Justiça e Instituições Eficazes).

4.2. Trabalhos Futuros

Diante dos resultados obtidos neste trabalho, diversas possibilidades de continuidade e aprofundamento da pesquisa são possíveis, organizadas nas seguintes esteiras do conhecimento:

1- Alinhamento com a Agenda Estratégica do CNJ e Justiça 4.0. O CNJ vem

estimulando a modernização do Judiciário com base no uso responsável de IA, e a área de judicialização da saúde deveria ser um tema estratégico. O programa Justiça 4.0, do qual o SINAPSES é parte integrante, visa reduzir a morosidade e promover maior efetividade nos serviços jurisdicionais, especialmente em áreas sensíveis como o direito à saúde. O desenvolvimento e expansão desta pesquisa poderão contribuir diretamente para esse ecossistema, sendo replicável em outros tribunais com adaptações locais. Assim, a proposta pode ser ampliada para uma rede nacional de cooperação entre tribunais e universidades, com foco na disseminação de soluções compartilhadas por meio da plataforma SINAPSES, do Conselho Nacional de Justiça (CNJ), consolidando o modelo de busca de medicamentos judicializados, como apresentado neste trabalho como referência replicável em todo o país.

2- Desenvolver um painel de inteligência que permita monitoramento em tempo real da judicialização da saúde no Brasil, com filtros por região, município, unidade de saúde, grupo populacional ou medicamento, alimentando diretamente gestores da saúde e tribunais. Tais dados extraídos deverão ser utilizados para apoiar a formulação de políticas públicas preventivas, contribuindo para a desjudicialização e para o fortalecimento da equidade no acesso ao SUS.

3- Expandir o escopo da abordagem proposta e dos modelos para reconhecer outras entidades além de medicamentos, patologias, exames, tratamentos, insumos, internações e outros pleitos de acesso à saúde pública, utilizando técnicas de Named Entity Recognition (NER) adaptadas ao domínio biomédico e jurídico, possibilitando uma análise mais completa das demandas judiciais e, consequentemente, redução do impacto à saúde financeira do SUS.

4- Avaliar e comparar o desempenho de vários modelo e arquiteturas baseadas em transformers, como BERTimbau, SciBERT e BERT-BiLSTM-CRF, com os modelos utilizados e treinados em textos jurídicos brasileiros a fim de avaliar acurácia e interpretabilidade ou explicabilidade, que é capacidade de entender e explicar como um modelo de IA chega a uma determinada decisão ou previsão.

5- Construir um corpus anotado jurídico-sanitário em língua portuguesa, com dados anonimizados, conforme os parâmetros da Lei Geral de Proteção de Dados (LGPD) e diretrizes da Resolução CNJ nº 332/2020, que sirva como base aberta para pesquisas futuras em Processamento de Linguagem Natural (PLN) aplicado ao direito e à saúde pública.

6- Integrar o painel MED-SUS-MS ao sistema de tramitação processual do TJMS (por exemplo, e-SAJ), viabilizando a extração automática em tempo real e a geração

de painéis gerenciais para magistrados e gestores da saúde,

7- Utilizar os dados extraídos para subsidiar decisões estratégicas da Secretaria Estadual de Saúde do Estado de Mato Grosso do Sul, como a previsão de demanda, a negociação com laboratórios e o aprimoramento dos protocolos clínicos.

8- Publicar relatórios periódicos de judicialização por região, medicamento ou grupo populacional, contribuindo para uma gestão pública baseada em evidências.

9- Realizar parceria com as Universidades para, em termos de impacto social e de governança, em aplicar técnicas de IA para avaliar e pesquisar os efeitos da judicialização no orçamento público, identificar padrões de desigualdade no acesso judicial à saúde, e subsidiar políticas públicas com base em evidências.

10- Adoção de modelos Legal-BERT ou camadas jurídicas customizadas. Testar modelos de linguagem mais alinhados ao domínio jurídico, como o Legal-BERT ou variantes do tipo LegalBertPT, desenvolvido com corpus do sistema judiciário brasileiro. A proposta consiste em avaliar se modelos especializados em textos jurídicos oferecem melhor desempenho na tarefa de reconhecimento de entidades farmacológicas em petições, sobretudo por estarem adaptados ao vocabulário, estilo e construções sintáticas típicas do Direito. Alternativamente, pode-se explorar uma abordagem de domain adaptation, combinando embeddings biomédicos (como os do BioBERT) com embeddings jurídicos gerados a partir de corpus oriundos dos tribunais brasileiros, visando capturar tanto o contexto técnico da saúde quanto a estrutura linguística legal das petições judiciais.

11- Utilização de modelo multitarefa (*multi-task learning*). Testar modelos multitarefas capazes de, simultaneamente, realizar a classificação de entidades nomeadas e a extração de contexto semântico. Nesse cenário, o modelo não apenas reconheceria o nome do medicamento, mas também inferiria, por exemplo, se está sendo solicitado, deferido, negado ou apenas mencionado no texto. Essa abordagem permitiria agregar maior valor informacional ao processo de extração, contribuindo com análises mais precisas para fins de gestão e decisão judicial no contexto da judicialização da saúde.

12- *Prompt-based Learning* com LLMs: com os avanços dos modelos de linguagem de larga escala (LLMs), outra linha de investigação relevante seria a utilização de *prompt-based learning* com modelos de código aberto, como LLaMA, Mistral ou DeepSeek. Essa estratégia visa explorar a capacidade desses modelos em compreender instruções em linguagem natural para realizar a tarefa de identificação de entidades, sem necessidade de re-treinamento supervisionado tradicional. O uso

de prompts bem estruturados pode permitir generalização para novos domínios com menos dependência de grandes volumes de dados anotados, tornando o processo mais escalável.

13- Aplicação de técnicas de *Distant Supervision* ou *Weak Supervision* para anotações. Diante das limitações e custos associados à anotação manual de corpora, uma alternativa a ser considerada é a adoção de técnicas de *distant supervision* ou *weak supervision*. Essas abordagens utilizam fontes externas de conhecimento, como listas oficiais de medicamentos da ANVISA ou da RENAME. Apesar de menos precisas que anotações humanas, essas técnicas permitem construir grandes conjuntos de dados pseudo-rotulados, viabilizando o treinamento de modelos robustos com menor investimento em recursos humanos especializados.

14- Exploração de modelos com extração baseada em RAG (Retrieval-Augmented Generation). Experimentar arquiteturas do tipo RAG (Retrieval-Augmented Generation), que combinam mecanismos de recuperação de informação com geração textual. Nessa abordagem, ao identificar uma possível entidade, o modelo consulta uma base externa, como por exemplo, uma lista oficial de medicamentos, antes de confirmar a classificação. Essa estratégia tem o potencial de aumentar a precisão do reconhecimento, especialmente em casos de grafias variantes, medicamentos novos ou contextos ambíguos, promovendo maior confiança nos resultados.

15- Incorporação de auditabilidade por meio de LLMs. Por fim, em compatibilidade com a Resolução CNJ 615/2025, sugere-se investigar formas de incorporar auditabilidade nas decisões dos modelos, especialmente com apoio de LLMs. Isso significa não apenas extrair entidades ou classificar contextos, mas também justificar, de forma interpretável, o motivo da classificação de determinado trecho como um medicamento. A possibilidade de gerar explicações auditáveis sobre as decisões do modelo é crucial para ampliar a transparência, a confiabilidade e a adoção institucional de soluções baseadas em inteligência artificial, especialmente no âmbito do Poder Judiciário.

REFERÊNCIAS

[Aguiar et al. 2022] Aguiar, A., Silveira, R., Furtado, V., Pinheiro, V., and Neto, J. (2022). Using topic modeling in classification of brazilian lawsuits. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13208 LNAI:233–242. cited By 0; Conference of 15th International Conference on the Computational Processing of Portuguese, PROPOR 2022; Conference Date: 21 March 2022 Through 23 March 2022; Conference Code:275209.

[Akbik et al., 2019] Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S. and Vollgraf, R. 2019. FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), pages 54–59, Minneapolis, Minnesota. Association for Computational Linguistics.

[Akira et al. 2014] Akira, V., Gonçalves, B., da Costa, B. R., and de Carvalho Silva, J. E. (2014). Modelos de predição estruturada em part-of-speech tagging para português do brasil. https://www.researchgate.net/publication/273141061_Modelos_de_Predicao_Estruturada_em_Part-of-Speech_Tagging_para_Portugues_do_Brasil. [Online: acesso em 24-novembro-2022].

[Alam et al., 2021] Alam, T., & Schmeier, S. (2021). Deep Learning in Biomedical Text Mining: Contributions and Challenges. In M. Househ et al. (Eds.), Multiple Perspectives on Artificial Intelligence in Healthcare (Lecture Notes in Bioengineering). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-67303-1_14. Acesso em 18 mar. 2024.

[Ali et al., 2022] ALI, Sajid; MASOOD, Khalid; RIAZ, Anas; SAUD, Amna. Named Entity Recognition using Deep Learning: A Review. In: 2022 International Conference on Business Analytics for Technology and Security (ICBATS), 2022, Lahore, Paquistão. Proceedings [...]. IEEE, 2022. DOI: 10.1109/ICBATS54253.2022.9759051. Disponível em: https://doi.org/10.1109/ICBATS54253.2022.9759051. Acesso em: 31 mar. 2024.

[ANVISA 2022] ANVISA (2022). Lista de medicamentos de referência da Anvisa. https://www.gov.br/anvisa/pt-br/setorregulado/regularizacao/medicamentos/medicamentos-de-referencia/lista-de-medicamentos-de-referencia. [Online: acesso em 24- novembro-2022].

[Ardra et al., 2023] ARDRA, K. R.; ANOOP, V. S.; PANTA, Prashanth. OralMedNER: A Named Entity Recognition System for Oral Medicine and Radiology. In: 2023 9th International Conference on Smart Computing and Communications (ICSCC), 2023, Thiruvananthapuram, Índia. Proceedings [...]. IEEE, 2023. DOI: 10.1109/ICSCC59169.2023.10334994. Disponível em: https://doi.org/10.1109/ICSCC59169.2023.10334994. Acesso em: 31 mar. 2024.

[Asghari et al., 2022] ASGHARI, M.; SIERRA-SOSA, D.; ELMAGHRABY, A. S. BINER: A low-cost biomedical named entity recognition. Information Sciences, v. 602, p. 184-200, 2022. DOI: 10.1016/j.ins.2022.04.037. Disponível em: https://www.sciencedirect.com/science/article/pii/S0020025522003838. Acesso em: 18 fev. 2024.

[Bajaj et al., 2022] BAJAJ, Goonmeet; KURSUNCU, Ugur; GAUR, Manas; LOKALA, Usha; HYDER, Ayaz; PARTHASARATHY, Srinivasan; SHETH, Amit. Knowledge-Driven Drug-Use Named Entity Recognition with Distant Supervision. Studies in Health Technology and Informatics, v. 290, p. 140-144, 2022. DOI: 10.3233/SHTI220048. Disponível em: https://scholarcommons.sc.edu/aii_fac_pub/545. Acesso em: 17 mar. 2024.

[Barbosa 2017] Barbosa (2017). [a study of cross-validation and bootstrap for accuracy estimation and model selection]. https://www.facom.ufu.br/˜wendelmelo/terceiros/tutorial_nltk.pdf. [Online: acesso em 19- junho-2022].

[Brasil 1988] Brasil (1988). Constituição da República Federativa do Brasil. https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. [Online: acesso em 22-novembro-2022].

[Cai et al., 2022] CAI, Xiaoya; GUO, En; ZHUANG, Xuqiang; YU, Hui; XU, Weizhi. MTBC-BioNER: Multi-task Learning Using BioBERT and CharCNN for Biomedical Named Entity Recognition. In: 2022 12th International Conference on Information Technology in Medicine and Education (ITME), 2022, JiNan, China. Proceedings [...]. IEEE, 2022. DOI: 10.1109/ITME56794.2022.00078. Disponível em: https://ieeexplore-ieee-org.ez51.periodicos.capes.gov.br/document/10086255/ . Acesso em: 17 mar. 2024.

[Carline 2014] Carline, Angélica (2014). Judicialização da Saúde Pública e Privada. Livraria do Advogado.

[Çelikmasat et al., 2022] ÇELIKMASAT, Gökberk; AKTÜRK, Muhammed Enes; ERTUNÇ, Yunus Emre; ISSIFU, Abdul Majeed; GANIZ, Murat Can. Biomedical Named Entity Recognition Using Transformers with biLSTM + CRF and Graph Convolutional Neural Networks. In: 2022 International Conference on Innovations in Intelligent Systems and Applications (INISTA), 2022, Istanbul, Turquia.

Proceedings [...]. IEEE, 2022. DOI: 10.1109/INISTA55318.2022.9894270. Disponível em: https://ieeexplore.ieee.org/document/9894270. Acesso em: 18 fev. 2024.

[Chen et al., 2022] CHEN, Aokun; YU, Zehao; YANG, Xi; GUO, Yi; BIAN, Jiang; WU, Yonghui. Contextualized medication information extraction using Transformer-based deep learning architectures. Journal of Biomedical Informatics, 2022. DOI: 10.1016/j.jbi.2022.104119. Disponível em: https://doi.org/10.1016/j.jbi.2022.104119. Acesso em: 17 mar. 2024.

[Chen et al. 2023]. CHEN, Xueliang; LIU, Tianming; TANG, Gongzheng; LIU, Yunjiang. Named Entity Recognition Based on Boundary Enhanced for Chinese Electronic Medical Records. In: 2023 12th International Conference of Information and Communication Technology (ICTech), Beijing, China. Proceedings [...]. IEEE, 2023. DOI: 10.1109/ICTech58362.2023.00025. Disponível em: https://doi.org/10.1109/ICTech58362.2023.00025. Acesso em: 17 mar. 2024.

[CNJ 2020] CNJ (2020). Resolução nº 332 de 21/08/2020. https://atos.cnj.jus.br/atos/detalhar/3429. [Online: acesso em 24-novembro-2022].

[CNJ 2024a] CNJ (2024a). Justiça em número 2024. https://www.cnj.jus.br/wp-content/uploads/2025/04/justica-em-numeros-2024.pdf [Online: acesso em 21-abril-2025].

[CNJ 2025] CNJ, C. N. d. J. (2025). Painel de estatísticas processuais de direito da saúde. https://justica-em-numeros.cnj.jus.br/painel-saude/ [Online: acesso em 19-abril-2025].

[Curdin et al., 2022] Marxer, C., Rölke, H., Alfieri, A., & Halatsch, M.-E. (2022). A Comparison of Automated Named Entity Recognition Tools Applied to Clinical Text. 7th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). DOI: 10.1109/ICIIIBMS55689.2022.9971532.

[das Nações Unidas ONU, 2015] das Nações Unidas ONU, O. (2015). Agenda 2030: Objetivos do desenvolvimento sustentável. https://atos.cnj.jus.br/atos/detalhar/3429. [Online: acesso em 24-novembro-2022].

[de Justiça 2007] de Justiça, C. N. (2007). Resolução nº 46 de 18/12/2007. https:// atos.cnj.jus.br/atos/detalhar/167. [Online: acesso em 22-novembro-2022].

[de Justiça 2022] de Justiça, C. N. (2022). Repositório de modelos de ia em produção. https://git.cnj.jus.br/ia/docs/-/wikis/Modelos-de-IA-em-Produ%C3%A7%C3%A3o. [Online: acesso em 23- novembro-2022].

[Devika et al., 2023] DEVIKA, N.; ANOOP, V. S.; THEKKINIATH, Jose. Biomedical Named Entity Recognition from Malaria Literature using BioBERT. In: 2023 9th International Conference on Smart Computing and Communications (ICSCC), 22-24 maio 2023, Thiruvananthapuram, Índia. Proceedings [...]. IEEE, 2023. p. 239-244. DOI: 10.1109/ICSCC59169.2023.10335049. Disponível em: https://doi.org/10.1109/ICSCC59169.2023.10335049. Acesso em: 18 fev. 2024.

[Devlin et al. 2018] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805.

[Fudholi et al., 2022] FUDHOLI, D.H.; NAYOAN, R.A.N.; HIDAYATULLAH, A.F.; ARIANTO, D.B. A Hybrid CNN-BiLSTM Model for Drug Named Entity Recognition. Journal of Engineering Science and Technology, v. 17, n. 1, p. 730-744, fev. 2022. ISSN 1823-4690. Disponível em: https://jestec.taylors.edu.my/ e https://www.scopus.com/record/display.uri?eid=2-s2.0-85124628532&origin=inward&txGid=3c478c3ade032fe166cc9369ed0fd4c5. Acesso em: 18 fev. 2024.

[Gan et al., 2023] GAN, Qiwei et al. A deep learning approach for medication disposition and corresponding attributes extraction. Journal of Biomedical Informatics, v. 143, 2023, p. 104391. Disponível em: https://doi.org/10.1016/j.jbi.2023.104391. Acesso em: 12 fev. 2024.

[Ge et al., 2022] GE, Yao; GUO, Yuting; YANG, Yuan-Chi; AL-GARADI, Mohammed Ali; SARKER, Abeed. A comparison of few-shot and traditional named entity recognition models for medical text. In: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), 2022, Atlanta, GA. Proceedings [...]. IEEE, 2022. DOI: 10.1109/ICHI54592.2022.00024. Disponível em: https://doi.org/10.1109/ICHI54592.2022.00024. Acesso em: 12 fev. 2024.

[Guan et al., 2021] GUAN, Y.; LI, H.; XU, W. A Traditional Chinese Medicine Terminology Recognition Model Based on Deep Learning: A TCM Terminology Recognition Model. In: 6th International Conference on Big Data and Computing (ICBDC 2021), 22-24 maio 2021, Virtual, Online, China. ACM International Conference Proceeding Series. New York: Association for Computing Machinery, 2021. p. 15-20. ISBN 978-145038980-8. DOI: 10.1145/3469968.3469971. Disponível em: https://doi.org/10.1145/3469968.3469971. Acesso em: 31 mar. 2024.

[He et al., 2014] He L, Yang Z, Lin H, Li Y. Drug name recognition in biomedical texts: a machine-learning-based method. Drug Discov Today. 2014 May;19(5):610-7. doi: 10.1016/j.drudis.2013.10.006. Epub 2013 Oct 16. PMID: 24140287.

[Jacob et al., 2020] Jacob Turton, David Vinson, Robert Elliott Smith. (2020). Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings. arXiv preprint arXiv:2012.15353. https://doi.org/10.48550/arXiv.2012.15353.

[Jin et al., 2022] JIN, Ran; HOU, Tengda; YU, Tongrui; LUO, Min; HU, Haoliang. A Multitask Deep Learning Framework for DNER. Computational Intelligence and Neuroscience, v. 2022, artigo ID 3321296, 10 páginas. DOI: 10.1155/2022/3321296. Disponível em: https://doi.org/10.1155/2022/3321296. Acesso em: 14 fev. 2024.

[Jofche et al., 2022] JOFCE, Nasi; MISHEV, Kostadin; STOJANOV, Riste; JOVANOVIK, Milos; ZDRAVEVSKI, Eftim; TRAJANOV, Dimitar. Named Entity Recognition and Knowledge Extraction from Pharmaceutical Texts using Transfer Learning. Procedia Computer Science, v. 203, p. 721-726, 2022. DOI: 10.1016/j.procs.2022.07.107. Disponível em: https://doi.org/10.1016/j.procs.2022.07.107. Acesso em: 17 mar. 2024.

[Kashtriya et al., 2023] KASHTRIYA, Poonam; SINGH, Pardeep; BANSAL, Parul. A Review on Clinical Named Entity Recognition. In: 2023 International Technology Conference on Emerging Technologies for Sustainable Development (OTCON), 2023, Hamirpur, Índia. Proceedings [...]. IEEE, 2023. DOI: 10.1109/OTCON56053.2023.10113977. Disponível em: https://ieeexplore.ieee.org/document/10113977. Acesso em: 14 fev. 2024.

[Kim et al., 2022] KIM, Hyunjae; KANG, Jaewoo. How Do Your Biomedical Named Entity Recognition Models Generalize to Novel Entities? IEEE Access, v. 10, p. 31513-31523, 2022. DOI: 10.1109/ACCESS.2022.3157854. Disponível em: https://doi.org/10.1109/ACCESS.2022.3157854. Acesso em: 17 mar. 2024.

[Kohli et al. 2021] Kohli, G., Kaur, P., and Bedi, J. (2021). Automatic detection of rhetorical role labels using ernie2.0 and roberta. *CEUR Workshop Proceedings*, 3159:581–588. cited By 0; Conference of Working Notes of FIRE - 13th Forum for Information Retrieval Evaluation, FIRE-WN 2021; Conference Date: 13 December 2021 Through 17 December 2021; Conference Code:180530.

[Kolpakov et al., 2023] KOLPAKOV, Nikolay A.; MOLODCHENKOV, Alexey I.; LUKIN, Anton V. Methods of extracting biomedical information from patents and scientific publications (on the example of chemical compounds). Discrete and Continuous Models and Applied Computational Science, v. 31, n. 1, p. 64-74, 2023. DOI: 10.22363/2658-4670-2023-31-1-64-74. Disponível em: http://journals.rudn.ru/miph. Acesso em: 17 mar. 2024.

[Kormilitzin et al., 2021] KORMILITZIN, Andrey; VACI, Nemanja; LIU, Qiang; NEVADO-HOLGADO, Alejo. Med7: a transferable clinical natural language processing model for electronic health records. arXiv preprint arXiv:2003.01271,

2020. Disponível em: https://arxiv.org/abs/2003.01271. Acesso em: 17 mar. 2024.

[Krammes 2010] Krammes, A. G. (2010). *Workflow em processos Judiciais Eletrônicos*. LTr, 1ª edition.

[Lee et al., 2019] LEE, Lung-Hao; LU, Yi. Multiple Embeddings Enhanced Multi-Graph Neural Networks for Chinese Healthcare Named Entity Recognition. IEEE Journal of Biomedical and Health Informatics, v. 25, n. 7, p. 2801-2810, jul. 2021. DOI: 10.1109/JBHI.2020.3048700. Disponível em: https://ieeexplore-ieee-org.ez51.periodicos.capes.gov.br/document/9312396/. Acesso em: 17 mar. 2024.

[Lee et al., 2021] L. -H. Lee and Y. Lu, "Multiple Embeddings Enhanced Multi-Graph Neural Networks for Chinese Healthcare Named Entity Recognition," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 7, pp. 2801-2810, July 2021, doi: 10.1109/JBHI.2020.3048700.

[Lei 8080/90] Lei Nº 8.080, de 19 de setembro de 1990. https://www.planalto.gov.br/ccivil_03/leis/l8080.htm [Online: acesso em 08 - junho - 2023].

[Lewis et al., 2019] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. arXiv preprint arXiv:1910.13461. DOI: 10.48550/arXiv.1910.13461.

[Li et al., 2020] Jing, Li, Sun, Aixin, Han, Jiamglei, and Li, Chenliang (2020). A Survey on Deep Learning for Named Entity Recognition. Disponível na url https://arxiv.org/abs/1812.09449. [Online: acesso em 30 - maio - 2024].

[Liu et al., 2021] LIU, Jian et al. A hybrid deep-learning approach for complex biochemical named entity recognition. Journal of Chemical Information and Modeling, v. 62, n. 10, p. 2531-2543, 2023. DOI: 10.1021/acs.jcim.2c01119. Disponível em: https://doi.org/10.1021/acs.jcim.2c01119. Acesso em: 12 fev. 2024.

[Liu et al. 2023] Liu, J., Gao, L., Guo, S., Ding, R., Huang, X., Ye, L., Meng, Q., Nazari, A., & Thiruvady, D. (2023). A hybrid deep-learning approach for complex biochemical named entity recognition. *Journal of Cheminformatics, 7*(S1). DOI: 10.1007/s13321-020-00473-w.

[López Úbeda et al., 2019] López Úbeda, P., Díaz Galiano, MC, Urena Lopez, LA, & Martin, M. (2019, novembro). Usando o Snomed para reconhecer e indexar menções químicas e de medicamentos. Em K. Jin-Dong, N. Claire, B. Robert, & D. Louise (Eds.), Anais do 5º Workshop sobre Tarefas Compartilhadas

Abertas de BioNLP (pp. 115–120). doi:10.18653/v1/D19-5718

[Lucena, 2021] Lucena, Sidrak Braz (2021). Judicialização da saúde: ferramenta de gestão de demandas judiciais com vistas à redução dos custos assistenciais: um exemplo aplicado numa operadora de saúde de autogestão. https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/30437/Trabalho%20Aplicado%20-%20Vers%c3%a3o%20final%2003.pdf?sequence=1&isAllowed=y [Online: acesso em 14 - junho - 2023].

[Ma et al., 2022] MA, Yuekun; LIU, Yun; ZHANG, Dezheng; ZHANG, Jiye; LIU, He; XIE, Yonghong. A Multigranularity Text Driven Named Entity Recognition CGAN Model for Traditional Chinese Medicine Literatures. Computational Intelligence and Neuroscience, v. 2022, 2022. DOI: 10.1155/2022/1495841. Disponível em: https://doi.org/10.1155/2022/1495841. Acesso em: 12 fev. 2024.

[Mathu et al., 2021] MATHU, T.; RAIMOND, Kumudha. A novel deep learning architecture for drug named entity recognition. TELKOMNIKA Telecommunication, Computing, Electronics and Control, v. 19, n. 6, p. 1884-1891, dez. 2021. DOI: 10.12928/TELKOMNIKA.v19i6.21667. Disponível em: http://journal.uad.ac.id/index.php/TELKOMNIKA. Acesso em: 14 fev. 2024.

[Mathu et al., 2023] MATHU, T.; RAIMOND, K.; DEEPAKANMANI, S. A hybrid drug named entity recognition framework for real time pubmed data using deep learning and text summarization techniques. PRZEGLĄD ELEKTROTECHNICZNY, v. 99, n. 8, p. 106-109, 2023. DOI: 10.15199/48.2023.08.18. Disponível em: https://doi.org/10.15199/48.2023.08.18 e http://pe.org.pl/articles/2023/8/18.pdf. Acesso em: 12 fev. 2024.

[Miah et al., 2023] MIAH, M. Saef Ullah; SULAIMAN, Junaida; SARWAR, Talha Bin; ISLAM, Saima Sharleen; RAHMAN, Mizanur; HAQUE, Md. Samiul. Medical Named Entity Recognition (MedNER): A deep learning model for recognizing medical entities (drug, disease) from scientific texts. In: 2023 20th International Conference on Smart Technologies (EUROCON), 2023. Proceedings [...]. IEEE, 2023. DOI: 10.1109/EUROCON56442.2023.10199075. Disponível em: https://doi.org/10.1109/EUROCON56442.2023.10199075. Acesso em: 17 mar. 2024.

[Moscato et al., 2023] MOSCATO, Vincenzo; POSTIGLIONE, Marco; SANSONE, Carlo; SPERLÍ, Giancarlo. TaughtNet: Learning Multi-Task Biomedical Named Entity Recognition From Single-Task Teachers. IEEE Journal of Biomedical and Health Informatics, v. 27, n. 5, p. 2512-2523, maio 2023. DOI: 10.1109/JBHI.2023.3244044. Disponível em: https://doi.org/10.1109/JBHI.2023.3244044. Acesso em: 31 mar. 2024.

[Naseem et al., 2021] NASEEM, Usman; KHUSHI, Matloob; REDDY, Vinay; RAJENDRAN, Sakthivel; RAZZAK, Imran; KIM, Jinman. BioALBERT: A Simple

and Effective Pre-trained Language Model for Biomedical Named Entity Recognition. In: 2021 International Joint Conference on Neural Networks (IJCNN), 2021, Shenzhen, China. Proceedings [...]. IEEE, 2021. DOI: 10.1109/IJCNN52387.2021.9533884. Disponível em: https://doi.org/10.1109/IJCNN52387.2021.9533884. Acesso em: 18 fev. 2024.

[Nguyen et al., 2021] NGUYEN, Thi-Cham; LE, Hoang-Quynh; CAN, Duy-Cat; HA, Quang-Thuy. Models Distillation with Lifelong Deep Learning for Vietnamese Biomedical Named Entity Recognition. In: 2021 13th International Conference on Knowledge and Systems Engineering (KSE), 2021, Hanoi, Vietnã. Proceedings [...]. IEEE, 2021. DOI: 10.1109/KSE53942.2021.9648790. Disponível em: https://ieeexplore.ieee.org/document/9648790. Acesso em: 14 fev. 2024.

[Noh et al., 2021] NOH, Jiho; KAVULURU, Ramakanth. Joint Learning for Biomedical NER and Entity Normalization: Encoding Schemes, Counterfactual Examples, and Zero-Shot Evaluation. ACM BCB. 2021. Disponível em: https://doi.org/10.1145/3459930.3469533. Acesso em: 18 mar. 2024.

[Nojoo Kambar et al., 2022] Nojoo Kambar, M. E. Z., Esmaeilzadeh, A., & Heidari, M. (2022). A Survey on Deep Learning Techniques for Joint Named Entities and Relation Extraction. Proceedings of the 7th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), IEEE, pp. 218-224. DOI: 10.1109/ICIIBMS55689.2022.9971532.

[NLP-Progress 2023] NPL-Progress (2023). Repository to track the progress in natural language processing (nlp), including the datasets and the current state-of-the-art for the most common nlp tasks. https://nlpprogress.com/english/named_entity_recognition.html. [Online: acesso em 08-junho-2024].

[Oliveira et. al., 2022] Oliveira, L.E.S.e., Peters, A.C., da Silva, A.M.P. et al. SemClinBr - a multi-institutional and multi-specialty semantically annotated corpus for Portuguese clinical NLP tasks. J Biomed Semant 13, 13 (2022). https://doi.org/10.1186/s13326-022-00269-1

[Oliveira et al., 2022b] Oliveira, Leonardo & Gomes, Anderson & Enes, Yuri & Castelo Branco, Thaíssa & Paiva, Raissa & Bolzon, Andrea & Demo, Gisela. (2022). Path and future of artificial intelligence in the field of justice: a systematic literature review and a research agenda. SN Social Sciences. 2. 10.1007/s43545-022-00482-w.

[Pandy et al., 2023] PANDY, Arpad; HARANGI, Balazs; HAJDU, Andras. Extracting Drug Names from Medical Reports. In: 2023 18th International Conference on Computer Science and Information Technologies (CSIT), 2023. Proceedings [...]. IEEE, 2023. DOI: 10.1109/CSIT61576.2023.10324071.

Disponível em: https://doi.org/10.1109/CSIT61576.2023.10324071. Acesso em: 31 mar. 2024.

[Priya et al., 2021] MATHU, T.; RAIMOND, Kumudha; JEBA PRIYA, S. An Overview of Technological Revolution in Deep Learning Architectures for Biomedical Named Entity Recognition. In: 2021 Asian Conference on Innovation in Technology (ASIANCON), 28-29 agosto 2021, Pune, Índia. Proceedings [...]. IEEE, 2021. DOI: 10.1109/ASIANCON51346.2021.9544823. Disponível em: https://ieeexplore.ieee.org/document/9544823. Acesso em: 14 fev. 2024.

[QuantPedia, 2021] QuantPedia. (2021). BERT Model – Bidirectional Encoder Representations from Transformers. https://quantpedia.com/strategies/news-sentiment-and-equity-returns-bert-ml-model/. [Online:acesso em 08-junho- 2024].

[Ramachandran et al., 2021] RAMACHANDRAN, R.; ARUTCHELVAN, K. Named entity recognition on bio-medical literature documents using hybrid based approach. Journal of Ambient Intelligence and Humanized Computing, 2021. DOI: 10.1007/s12652-021-03078-z. Disponível em: https://doi.org/10.1007/s12652-021-03078-z. Acesso em: 31 mar. 2024.

[Ravikumar et al., 2021] RAVIKUMAR, J.; KUMAR, P. Ramakanth. Machine learning model for clinical named entity recognition. International Journal of Electrical and Computer Engineering (IJECE), v. 11, n. 2, p. 1689-1696, abr. 2021. ISSN 2088-8708. DOI: 10.11591/ijece.v11i2.pp1689-1696. Disponível em: http://ijece.iaescore.com e https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097839050&doi=10.11591%2fijece.v11i2.pp1689-1696&partnerID=40&md5=efde7898eb23aef75eb5ec7d300bc29b . Acesso em: 18 mar. 2024.

[Rezende 2005] Rezende, Solange Oliveira (2005). *Sistemas Inteligentes: Fundamentos e Aplicações*. Manoele, 1ª edition.

[Rivera-Zavala et al., 2021] RIVERA-ZAVALA, Renzo M.; MARTÍNEZ, Paloma. Analyzing transfer learning impact in biomedical cross-lingual named entity recognition and normalization. BMC Bioinformatics, v. 22, Suppl 1, 2021, p. 601. Disponível em: https://doi.org/10.1186/s12859-021-04247-9. Acesso em: 17 fev. 2024.

[Rodrígues et al., 2020] Rodríguez, Marcia Marina, Bezerra, Byron Dantas (2020). Processamento de linguagem Natural para Reconhecimento de Entidades Nomeadas em Textos Jurídicos de Atos Administrativos (Portarias). Revista de Engenharia e Pesquisa Aplicada. Disponível na url https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php/acervo/buscador.html?task=detalhes&source=&id=W3043345976. [Online: acesso em 04 - junho - 2024].

[Sakhovskiy et al., 2023] SAKHOVSKIY, A.S.; TUTUBALINA, E.V. Cross-Lingual Transfer Learning in Drug-Related Information Extraction from User-Generated Texts. Programming and Computer Software, v. 49, n. 7, p. 590-595, dez. 2023. DOI: 10.1134/S036176882307006X. Disponível em: https://doi.org/10.1134/S036176882307006X. Acesso em: 17 mar. 2024.

[Schneider et al, 2020] Schneider, Elisa Terumi Rubel; de Souza, João Vitor Andrioli; Knafou, Julien; Oliveira, Lucas Emanuel Silva; Copara, Jenny; Gumiel, Yohan Bonescki; Oliveira, Lucas Ferro Antunes de; Paraiso, Emerson Cabrera; Teodoro, Douglas; Barra, Cláudia Maria Cabral Moro. BioBERTpt - A Portuguese Neural Language Model for Clinical Named Entity Recognition. Association for Computational Linguistics, 2020. Disponível em https://www.aclweb.org/anthology/2020.clinicalnlp-1.7. Acesso em 10 abr. 2025.

[Schulze 2018] Schulze, Clenio Jair, Direito à saúde e a judicialização do impossível. Coletânea direito à saúde: dilemas do fenômeno da judicialização da saúde. Brasília:2018, pág. 18, disponível em https://www.ceapetce.org.br/uploads/documentos/5e8c8f60cbd576.32070875.pdf#page=15 [Online: acesso em 05 - junho - 2023].

[Shah-Mohammadi et al., 2022] SHAH-MOHAMMADI, Fatemeh; FINKELSTEIN, Joseph. A Comparison of Automated Named Entity Recognition Tools Applied to Clinical Text. In: 2022 7th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2022, Nara, Japão. Proceedings [...]. IEEE, 2022. DOI: 10.1109/ICIIBMS55689.2022.9971532. Disponível em: https://ieeexplore.ieee.org/document/9971532. Acesso em: 12 fev. 2024.

[TCU 2024] 2ª Edição - Lista de Alto Risco da Administração Pública Federal 2024. https://sites.tcu.gov.br/listadealtorisco/sistema_unico_de_saude_acesso_e_sustentabilidade.html#:~:text=Judicializa%C3%A7%C3%A3o%20da%20sa%C3%BAde,para%20sa%C3%BAde%20e%20dep%C3%B3sitos%20judiciais. [Online: acesso em 27- maio-2025]

[Tho et al., 2023] THO, Bui Duc; GIANG, Son-Ba; NGUYEN, Minh-Tien; NGUYEN, Tri-Thanh. An Architecture for More Fine-Grained Hidden Representation in Named Entity Recognition for Biomedical Texts. In: International Conference on Advances in Information and Communication Technology (ICTA), 2023, Hanoi, Vietnã. Proceedings [...]. Springer, 2023. DOI: 10.1007/978-3-031-49529-8_13. Disponível em: https://www.researchgate.net/publication/376417141_An_Architecture_for_More_Fine-Grained_Hidden_Representation_in_Named_Entity_Recognition_for_Biomedical_Texts. Acesso em: 6 jun. 2024.

[TJMS 2022] TJMS, C. d. I. (2022). Justiça em número 2022. https://www.tjms.jus.br/storage/cms-arquivos/9be8db3bb48d868d27869ab87ff6bbb4.pdf. [Online: acesso em 22- novembro-2022].

[TPU CNJ] https://www.cnj.jus.br/sgt/consulta_publica_classes.php

[Turine 2018], Turine, Joseliza A. V. Biodiversidade e biotecnologia no Brasil: marco legal em prol da sustentabilidade competitiva. Tese (Doutorado em biotecnologia e biodiversidade da rede pro-centro-oeste). Universidade Federal de Mato Grosso do Sul, Campo Grande: 2018.

[Viegas, 2022] Viegas, Charles F. O. JurisBERT: Transformer-based model for embedding legal texts. Dissertação (Mestrado em Computação Aplicada). Universidade Federal de Mato Grosso do Sul, Campo Grande: 2022.

[VeerasekharReddy et al., 2023] VEERASEKHARREDDY, B.; GOPALA KRISHNA, J. Satya Venu; NAGARAJU THATHA, Venkata; SUNDARAM, Ajith; SWAMY BIYYAPU, Narasimha; SANDEEP, Daria. Named Entity Recognition on Medical Text by Using Deep Neural Networks. In: 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), 2023, Bangalore, Índia. Proceedings [...]. IEEE, 2023. DOI: 10.1109/GCAT59970.2023.10353439. Disponível em: https://ieeexplore.ieee.org/document/10353439. Acesso em: 31 mar. 2024.

[Wang et al., 2019] Wang, Z., Shang, J., Liu, L., Lu, L., Liu, J., & Han, J. (2019, novembro). CrossWeigh: Treinamento de Named Entity Tagger a partir de Anotações Imperfeitas. Em K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), Anais da Conferência de 2019 sobre Métodos Empíricos em Processamento de Linguagem Natural e da 9ª Conferência Conjunta Internacional sobre Processamento de Linguagem Natural (EMNLP-IJCNLP) (pp. 5154–5163). doi:10.18653/v1/D19-1519

[Wang et al., 2021] WANG, Xueting; MIAO, Fang; LIU, Huixin; ZHANG, Guoting; JIN, Libiao. Joint Extraction of Entities and Relations from Ancient Chinese Medical Literature. In: 2021 International Conference on Culture-oriented Science & Technology (ICCST), 2021, Beijing, China. Proceedings [...]. IEEE, 2021. DOI: 10.1109/ICCST53801.2021.00083. Disponível em: https://ieeexplore.ieee.org/document/9538335. Acesso em: 17 mar. 2024.

[Wang et al., 2022] WANG, Bin; TANG, Dafu; WU, Qing. Improved Pre-training and Semi-supervised Learning for Domain-specific Chinese Named Entity Recognition. In: 2022 4th International Conference on Applied Machine Learning (ICAML), 2022, Changsha, China. Proceedings [...]. IEEE, 2022. DOI: 10.1109/ICAML57167.2022.00034. Disponível em: https://ieeexplore.ieee.org/document/00034. Acesso em: 17 mar. 2024.

[Watanabe et al., 2022] WATANABE, Taiki; ICHIKAWA, Tomoya; TAMURA, Akihiro; IWAKURA, Tomoya; MA, Chunpeng; KATO, Tsuneo. Auxiliary Learning for Named Entity Recognition with Multiple Auxiliary Training Data. In: Proceedings of the BioNLP 2022 Workshop, Dublin, Ireland, 26 maio 2022. Association for Computational Linguistics, p. 130-139. Disponível em: https://aclanthology.org/2022.bionlp-1.13/. Acesso em: 18 fev. 2024.

[Whitton et al., 2023] WHITTON, Jetsun; HUNTER, Anthony. Automated tabulation of clinical trial results: A joint entity and relation extraction approach with transformer-based language representations. Journal of Biomedical Informatics, 2021. Disponível em: https://arxiv.org/abs/2112.05596. Acesso em: 18 fev. 2024.

[Xiao et al., 2021] XIAO, Y.; LIANG, S.; PENG, J.; HUANG, Z.; WANG, Y.; WANG, J. Using Contextualized Representations for Biomedical Entity Recognition. In: 2021 International Conference on Computer Engineering and Application (ICCEA), 2021, Kunming, China. Proceedings [...]. IEEE, 2021. p. 456-459. DOI: 10.1109/ICCEA53728.2021.00095. Disponível em: https://doi.org/10.1109/ICCEA53728.2021.00095. Acesso em: 31 mar. 2024.

[Zaikis et al., 2022] ZAÏKIS, Dimitrios; KOKKAS, Stylianos; VLAHAVAS, Ioannis. Transforming drug-drug interaction extraction from biomedical literature. In: 12th Hellenic Conference on Artificial Intelligence (SETN 2022), 7–9 setembro 2022, Corfu, Grécia. Anais... New York: ACM, 2022. p. 1-8. Disponível em: https://doi.org/10.1145/3549737.3549753. Acesso em: 31 mar. 2024.

[Zhao et al., 2023] ZHAO, Zongyao; QIAN, Yue; LIU, Qirui; CHEN, Jiaxu; LIU, Yueyun. A Dynamic Optimization-Based Ensemble Learning Method for Traditional Chinese Medicine Named Entity Recognition. IEEE Access, v. 11, p. 99101-99110, 2023. DOI: 10.1109/ACCESS.2023.3313608. Disponível em: https://ieeexplore.ieee.org/document/3313608. Acesso em: 12 fev. 2024.