




ARTIGO DE PESQUISA/RESEARCH PAPER

Análise comparativa do BERT e ChatGPT no reconhecimento de entidades nomeadas do domínio jurídico

Comparative Analysis of BERT and ChatGPT in the Recognition of Named Entities in the Legal Domain

Geovane Araujo   [Instituto Federal de Educação, Ciência e Tecnologia do Ceará | geovanesv.dev@gmail.com]

Raquel Silveira   [Instituto Federal de Educação, Ciência e Tecnologia do Ceará | raquel_silveira@ifce.edu.br]

 Instituto Federal de Educação, Ciência e Tecnologia do Ceará, CE-187, s/n - Estádio, Tianguá - CE, 62320-000, Brasil.

Resumo. Este trabalho apresenta uma análise comparativa de dois modelos de linguagem pré-treinados, BERT e ChatGPT, para o reconhecimento de entidades nomeadas em textos jurídicos. O objetivo é avaliar o desempenho desses modelos em uma tarefa essencial para a extração de informações relevantes e específicas de documentos legais não estruturados. O modelo BERT é treinado com o dataset LENER-BR, que contém textos jurídicos anotados com as seguintes categorias de entidades: Jurisprudência, Legislação, Local, Organização, Pessoa e Tempo. O modelo ChatGPT é utilizado sem treinamento adicional, aproveitando o seu conhecimento prévio de linguagem natural. As métricas utilizadas para a avaliação são acurácia, precisão, cobertura e F1-score. Os resultados mostram que o modelo BERT supera o ChatGPT no NER no dataset LENER-BR em todas as métricas. O trabalho contribui para a comunidade de PLN ao fornecer uma análise dos modelos e suas limitações, bem como sugestões para trabalhos futuros.

Abstract. This work presents a comparative analysis of two pre-trained language models, BERT and ChatGPT, for named entity recognition in legal texts. The objective is to evaluate the performance of these models in an essential task for extracting relevant and specific information from unstructured legal documents. The BERT model is trained with the LENER-BR dataset, which contains legal texts annotated with the following categories of entities: Jurisprudence, Legislation, Location, Organization, Person and Time. The ChatGPT model is used without additional training, taking advantage of its prior knowledge of natural language. The metrics used for evaluation are accuracy, precision, recall and F1-score. The results show that the BERT model outperforms the ChatGPT model in NER on the LENER-BR dataset in all metrics. The work contributes to the NLP community by providing an analysis of the models and their limitations, as well as suggestions for future work.

Palavras-chave: Inteligência Artificial. NER. Bert. Modelo GPT. ChatGPT

Keywords: Artificial intelligence. NER. Bert. GPT model. ChatGPT

Recebido/Received: 06 Dezembro 2023 • Aceito/Accepted: 19 March 2025 • Publicado/Published: 30 May 2025

1 Introdução

Em meio à grande quantidade de dados textuais, identificar automaticamente partes do texto, tais como nomes de pessoas, organizações, locais e datas, é crucial para a eficiência e precisão em várias aplicações. O Reconhecimento de Entidades Nomeadas (do inglês, *Named Recognition Entity*, NER) é uma técnica essencial no campo do Processamento de Linguagem Natural (PLN), projetada para identificar e classificar entidades específicas em textos, como nomes de pessoas, organizações, locais, datas e outras informações específicas, em textos não estruturados [Ratinov and Roth, 2009].

A aplicação do NER reside na capacidade de extrair informações relevantes e específicas de grandes volumes de texto de maneira automatizada. Isso é crucial para aprimorar a compreensão do conteúdo textual e facilitar uma variedade de tarefas, incluindo a indexação eficiente de documentos, a personalização de assistentes virtuais, a análise de sentimentos, a busca de informações e a automação de processos que requerem a identificação e categorização de entidades específicas. O NER desempenha um papel fundamental em tarefas que envolvem o manuseio de dados não estruturados,

proporcionando uma base para uma análise mais precisa e informada em diversas áreas, incluindo aplicações no domínio jurídico, médico, dentre outros [Che *et al.*, 2013] [Ratinov and Roth, 2009].

A pesquisa visa a análise comparativa entre BERT e ChatGPT (nomeamos neste trabalho de NER-ChatGPT), permitindo avaliar o desempenho desses modelos no reconhecimento de entidades nomeadas no contexto jurídico, utilizando o dataset LENER-BR [Luz de Araujo *et al.*, 2018]. Os modelos de classificação serão avaliados em relação às métricas de acurácia, precisão, cobertura e F1-score. A aplicação em domínio jurídico é realizada para identificar o quanto os modelos conseguem se adequar às especificidades de um determinado tipo de linguagem. Ao final, espera-se fornecer *insights* valiosos para a comunidade de PLN e contribuir para o desenvolvimento de ferramentas mais eficazes na extração de entidades nomeadas em textos legais e preencher a lacuna da comparação dos algoritmos explorados no domínio jurídico e utilizando o dataset Lener-BR.

2 Fundamentação Teórica

Nesta seção serão apresentados os principais conceitos relacionados à temática deste trabalho, que envolvem o processamento de linguagem natural e algoritmos de aprendizagem de máquina, assim como a tarefa de reconhecimento de entidades nomeadas.

2.1 Aprendizado de máquina

Aprendizado de máquina é uma área da inteligência artificial que se concentra em ensinar computadores a aprender com experiências passadas, tornando-os melhores em tarefas específicas ao longo do tempo. Ao invés de programar regras fixas, como é feito em muitos softwares tradicionais, o aprendizado de máquina permite que as máquinas melhorem automaticamente ao analisar dados e identificar padrões [Faceli *et al.*, 2021].

2.1.1 BERT

O BERT, do inglês *Bidirectional Encoder Representations from Transformers*, é um modelo de linguagem estruturado em redes neurais, desenvolvido para processamento de texto que se destaca por seu desempenho em uma variedade de tarefas de processamento de linguagem natural. As entradas do BERT são sequências de tokens de texto, que podem ser palavras, subpalavras ou caracteres e sua saída é a classificação de um texto, com base em classes informadas no treinamento. Em comparação com a arquitetura original do transformer, o modelo BERT-BASE possui 12 camadas no codificador, enquanto o modelo BERT-LARGE possui 24 camadas. [Devlin *et al.*, 2019].

2.1.2 Modelo GPT

O modelo usado na análise deste trabalho, é treinado com uma ampla combinação de conjuntos de dados que totalizam 400 bilhões de tokens e composto por um impressionante total de 175 bilhões de parâmetros, demonstrando a capacidade de aprendizado em poucas etapas, conforme relatado por [Brown *et al.*, 2020], em que o modelo aprende por meio de múltiplos exemplos da tarefa de PNL ou Geração de Linguagem Natural (do inglês, *Natural Language Generation*, NLG), denominados *prompts*.

2.1.3 ChatGPT

O chatGPT, do inglês *Chat Generative Pre-Trained Transformer*, é um modelo de linguagem baseado em *deep learning* desenvolvido em 2019 pela OpenAI. Essa ferramenta utiliza uma base de conhecimento atualizada para decodificar palavras e fornecer respostas textuais em conversas, perguntas sobre diversos assuntos, conceitos, letras de músicas, poemas e informações Sarrion [2023].

2.2 Named Entity Recognition (NER)

O reconhecimento de entidades nomeadas pode ser definido como, um ramo do processamento de linguagem natural que realiza a tarefa de extração e classificação de entidades mencionadas em um determinado texto que está escrito em linguagem natural. Essas entidades são representadas por categorias pré-definidas que representam conceitos ou objetos de um determinado domínio, tais como pessoas, lugares, organizações, normas e autoridades [Môro, 2018].

2.2.1 Aplicações do NER

A técnica de reconhecer entidades pode ser aplicada em diversas áreas, tais como análise automática de textos [Sousa and Mello, 2022], notificações de eventos adversos e queixas técnicas de dispositivos médicos Garcia [2021], detecção automática de tópicos de aplicativos [Gutterres, 2022], busca semântica em documentos da área biomédica [Lara and Lobo, 2021], modelagem e análise de base de conhecimento para chatbots [Xavier, 2022].

[Lara and Lobo, 2021] desenvolveram um buscador semântico para recuperar documentos em um *corpus* específico para Covid-19, nomeado como CORD-19. Para isso, foi utilizada uma representação simplificada do conhecimento e a biblioteca Python spaCy para recuperar entidades nomeadas nos documentos.

[Xavier, 2022] fez um estudo de caso para auxiliar o setor de recursos humanos na automatização do suporte às dúvidas mais comuns de funcionários, utilizando-se do NER para a modelagem e análise da base de conhecimento para um chatbot.

3 Trabalhos Relacionados

Nesta seção, são apresentados alguns trabalhos que estão relacionados à temática deste trabalho, ou seja, que aplicam o NER no domínio jurídico, bem como usam modelos de linguagem pré-treinados para tarefas de PLN.

O trabalho de [Zhong *et al.*, 2023] explora a capacidade de compreensão do modelo de linguagem ChatGPT em relação ao modelo BERT na tarefa de reconhecimento de entidades nomeadas (NER). O estudo conclui que o ChatGPT pode ser útil em outras tarefas de linguagem natural, como geração de texto, tradução automática e respostas a perguntas, mas o BERT é mais adequado para a tarefa de NER.

Silva and Lopes [2023], propôs um modelo de classificação para o NER em português, baseado no BERT. O objetivo do trabalho foi avaliar o desempenho do BERT para o NER em português, comparando-o com outros modelos baseados em aprendizado de máquina, como *Conditional Random Field* (CRF), *Support Vector Machine* (SVM) e *Multi-Layer Perceptron* (MLP). O método utilizado treinou o BERT com diferentes conjuntos de dados anotados para o NER em português, como o HAREM, o CoNLL-2002 e o WikiNER, e testar o modelo com um conjunto de dados inédito, o Corpus Jurídico. Os resultados obtidos foram os seguintes: o BERT superou os outros modelos em todos os conjuntos de dados, alcançando uma F1-score média de 0,87 no Corpus Jurídico.

[Silveira *et al.*, 2023] apresentam o LegalBert-pt, um modelo de linguagem pré-treinado para o domínio jurídico em português brasileiro. O modelo foi pré-treinado em um corpus diverso de textos jurídicos do Conselho Nacional de Justiça do Brasil e é de código aberto e personalizável para tarefas jurídicas específicas. O trabalho descreve o processo de desenvolvimento do modelo, incluindo a escolha da arquitetura do modelo, a seleção dos datasets LENER-BR e CDJUR para o treinamento e a avaliação do modelo em várias tarefas jurídicas, como classificação e reconhecimento de entidades nomeadas.

Em conclusão, nesta seção, foram revisados os princi-

país trabalhos relacionados ao NER em português, utilizando modelos de classificação baseados em BERT ou em GPT-3, ou comparando esses modelos com outros. O diferencial deste trabalho em relação aos demais se concentra na proposta de uma comparação entre o modelo BERT treinado com o conjunto de dados de domínio jurídico (LENER-BR) e um modelo do ChatGPT, utilizando a versão GPT-3, para o NER em português aplicado ao domínio jurídico.

4 Metodologia

Considerando que este trabalho visa analisar e comparar o desempenho de diferentes modelos de classificação para o reconhecimento de entidades nomeadas em textos jurídicos, esta seção descreve os procedimentos utilizados para o NER pelos diferentes algoritmos, definindo-se o conjunto de dados, assim como os algoritmos de aprendizagem de máquina utilizado nos experimentos.

Ambos os modelos desenvolvidos para o NER foram avaliados usando métricas *precision*, *recall*, F1-score e acurácia.

4.1 Dataset utilizado para o NER em textos jurídicos

Para o reconhecimento de entidades nomeadas em textos jurídicos, utilizou-se o conjunto de dados LENER-BR [Luz de Araujo *et al.*, 2018], composto por um total de 70 documentos legais provenientes de diversos tribunais brasileiros, como o Supremo Tribunal Federal, Superior Tribunal de Justiça, Tribunal de Justiça de Minas Gerais e Tribunal de Contas da União, além de documentos legislativos, como a Lei Maria da Penha [Luz de Araujo *et al.*, 2018].

Originalmente, os documentos foram submetidos a um processo de pré-processamento, que incluiu a divisão dos documentos em frases e a divisão das frases em tokens, a fim de preparar os tokens para a identificação de entidades nomeadas. Os documentos foram divididos aleatoriamente para as etapas de treinamento, validação e teste, sendo 50 documentos separados para conjunto de treinamento, 10 documentos para conjunto de validação e 10 documentos para conjunto de teste [Luz de Araujo *et al.*, 2018], mostrado na **Tabela 1**.

Tabela 1. Contagem de sentenças, tokens e documentos no LENER-BR.

Dataset	Documentos	Sentenças	Tokens
Treino	50	7.827	299.277
Validação	10	1.176	41.166
Teste	10	1.389	47.630

O dataset possui as seguintes entidades: Organização, Pessoa, Tempo, Local, Legislação e Jurisprudência, na **Tabela 2** apresenta a quantidade de entidades em cada dataset Luz de Araujo *et al.* [2018].

Foi empregado o esquema de marcação *Inside-Outside-Beginning* (IOB), onde “B-” é usado para tokens que iniciam entidades nomeadas, “I-” para tokens dentro de entidades nomeadas e “O” para aqueles tokens que não pertencem a nenhuma entidade Luz de Araujo *et al.* [2018].

Tabela 2. Contagem de palavras de entidade nomeada para cada conjunto.

Entidade	Dataset Treino	Dataset Validação	Dataset Teste
Jurisprudência	3.967	743	660
Legislação	13.039	2.609	2.669
Local	1.417	244	132
Organização	6.671	1.608	1.367
Pessoa	4.612	894	735
Tempo	2.343	543	260

4.2 Modelos para NER

Foram implementados dois algoritmos para o NER: BERT e ChatGPT, utilizando as bibliotecas pertinentes e seguindo as arquiteturas e configurações descritas nos trabalhos originais. A seguir são apresentadas as configurações de cada um dos modelos.

- **Bidirecional Encoder Representations from Transformers (BERT):** O modelo BERT foi implementado no cenário do dataset LENER-BR, que visa o reconhecimento de entidades nomeadas no âmbito jurídico, seguindo várias etapas fundamentais. Nesse processo, foi utilizado o modelo bert-base-cased-pt-lenerbr, 10 épocas e tamanho de lotes de 4 amostras. Como otimizador utilizou-se o AdamW com uma taxa de aprendizado de $1e-4$.
- **NER-ChatGPT:** Para a implementação do NER utilizando o ChatGPT, utilizou-se o modelo text-davinci-003 (modelo pré-treinado do ChatGPT), capaz de processar 4.097 tokens por requisição. Foi utilizado um prompt específico (promptfy) para que o ChatGPT atuasse no NER Silva [2020], com um grupo de entidades específicas mostrado a seguir: “Você é um sistema de reconhecimento de entidade nomeada (NER) altamente inteligente e preciso. Você recebe um texto como entrada e sua tarefa é reconhecer e extrair tipos específicos de entidades nomeadas nesse texto e classificá-las de acordo com um conjunto de tipos de entidades. As entidades predefinidas são: PESSOA, LOCAL, ORGANIZAÇÃO, TEMPO, LEGISLAÇÃO, JURISPRUDÊNCIA. Seu formato de saída é apenas [(‘T’: tipo de entidade predefinida, ‘E’: entidade no texto de entrada)]. A saída do ChatGPT foi submetida a um tratamento de dados para que fique com o mesmo padrão do *corpus* LENER-BR, ou seja, no formato IOB.

4.2.1 Avaliação Experimental

As avaliações de sistemas de NER são geralmente baseadas na comparação das saídas dos sistemas com textos anotados por especialistas Mõro [2018]. Algumas métricas utilizadas no reconhecimento de entidades nomeadas são descritas a seguir.

4.2.1.1 Acurácia

A acurácia mede a proporção de todas as entidades nomeadas identificadas corretamente pelo sistema em relação ao número total de entidades nomeadas. A acurácia é útil quando se deseja avaliar o desempenho geral de um sistema de NER, incluindo a identificação correta e incorreta de entidades Strubell *et al.* [2018].

4.2.1.2 Precisão

A precisão (do inglês, *Precision*) mede a proporção de entidades nomeadas identificadas corretamente pelo sistema em relação ao número total de entidades nomeadas identificadas pelo sistema. Em outras palavras, é a capacidade do sistema de evitar falsos positivos Yan *et al.* [2019]

4.2.1.3 Cobertura ou Revocação

A cobertura (ou revocação) (do inglês, *Recall*) mede a proporção de entidades nomeadas identificadas corretamente pelo sistema em relação ao número total de entidades nomeadas presentes no texto. É a capacidade do sistema de evitar falsos negativos Wu [2019].

4.2.1.4 F1-score

F1-score é a média harmônica das métricas *Precision* e *Recall* demonstradas anteriormente. Essa métrica combina o *Precision* e o *Recall* para trazer um número único que indique a qualidade geral do modelo. O F1-score é especialmente útil para conjuntos de dados que possuem classes desproporcionais, pois trabalha bem, mesmo quando as classes são desbalanceadas Chaves [2021].

5 Resultados

Nesta seção será apresentada uma análise do desempenho dos algoritmos BERT e NER-ChatGPT na tarefa de reconhecimento de entidades nomeadas, utilizando o conjunto de dados LENER-BR. Esta seção fornece uma avaliação dos resultados obtidos, destacando a eficácia de cada algoritmo na identificação de entidades em textos do domínio jurídico em português, demonstrando insights valiosos para a compreensão das diferenças entre as abordagens de BERT e NER-ChatGPT nessa tarefa específica.

A **Tabela 3** apresenta os resultados das métricas de avaliação dos algoritmos BERT e NER-ChatGPT para o NER. O algoritmo BERT obteve resultados superiores em todas as métricas avaliadas, com precisão de 0.881, cobertura de 0.909, F1-score de 0.895 e acurácia de 0.975, enquanto o algoritmo NER-ChatGPT obteve precisão de 0.239, cobertura de 0.462, F1-score de 0.315 e acurácia de 0.860.

Tabela 3. Resultados dos algoritmos NER-ChatGPT e BERT para a tarefa de NER, no dataset LENER-BR.

Algoritmo	Precisão	Cobertura	F1-score
NER-ChatGPT	0.239	0.462	0.315
BERT	0.881	0.909	0.895

Os resultados mostrados na **Tabela 3** indicam que o algoritmo BERT é mais adequado para a tarefa de NER em comparação com o NER-ChatGPT, considerando os experimentos realizados com o dataset LENER-BR. A precisão do BERT é aproximadamente quatro vezes maior do que a do NER-ChatGPT, enquanto a cobertura do BERT é duas vezes maior do que a do NER-ChatGPT. O F1-score do BERT é aproximadamente três vezes maior do que o do NER-ChatGPT. Esses resultados sugerem que o BERT é mais eficaz na identificação de entidades em textos do domínio jurídico em português do que o NER-ChatGPT.

A **Tabela 4** apresenta os resultados dos algoritmos NER-ChatGPT e BERT para cada uma das entidades, permitindo uma análise mais aprofundada do desempenho de cada algoritmo nas diferentes categorias de entidades.

A análise dos resultados da **Tabela 4** revela diferenças significativas no desempenho dos algoritmos NER-ChatGPT e BERT o NER. O algoritmo BERT apresentou um desempenho geral superior ao NER-ChatGPT. O NER-ChatGPT demonstrou imprecisão e confusão na identificação das entidades. Nas entidades JURISPRUDÊNCIA e LEGISLAÇÃO, a imprecisão na identificação das entidades pelo NER-ChatGPT é maior. Essas entidades são conceitos relacionados especificamente ao campo legal. Nas entidades ORGANIZAÇÃO, PESSOA e TEMPO há uma imprecisão na identificação das entidades, no entanto por serem entidades de domínio genérico, a imprecisão na identificação dessas entidades é menor. Destaca-se ainda que, o NER-ChatGPT não conseguiu identificar 5.245 entidades, marcando-as como 'O'. Diante desse contexto, o modelo BERT destaca-se como mais adequado para a tarefa de NER no domínio jurídico.

Esses resultados podem ser justificados devido o modelo de linguagem GPT-3, que é a base do NER-ChatGPT, ser treinado para tarefas de linguagem natural em geral, enquanto que o BERT é um modelo de linguagem que se baseia em uma arquitetura de codificador bidirecional, usando uma técnica de máscara de palavras para prever palavras faltantes em uma sentença, permitindo uma compreensão contextual bidirecional das palavras, e foi treinado especificamente para a tarefa de reconhecimento de entidades nomeadas.

6 Conclusão e Trabalhos futuros

Neste trabalho realizou-se uma análise comparativa de dois modelos de linguagem pré-treinados, BERT e ChatGPT, para o reconhecimento de entidades nomeadas em textos jurídicos. A tarefa de NER é essencial para a extração de informações relevantes e específicas de documentos legais não estruturados, que apresentam uma linguagem altamente especializada e complexa.

Os modelos foram testados com o dataset LENER-BR, que contém textos jurídicos anotados com as categorias de entidades: Jurisprudência, Legislação, Local, Organização, Pessoa e Tempo. As métricas utilizadas para a avaliação foram: acurácia, precisão, cobertura e F1-score. Os resultados mostraram que o modelo o BERT superou o ChatGPT em todas as métricas. A intuição é que o modelo BERT superou os resultados por ter um processo de fine-tuning específico para a tarefa de NER, utilizando no treinamento documentos do dataset LENER-BR, portanto, essa etapa possibilita a aprendizagem de características específicas para o reconhecimento das entidades do domínio jurídico. Enquanto que o ChatGPT, originalmente, se baseia no seu conhecimento prévio genérico da linguagem natural, o que dificulta reconhecimento das entidades específicas.

Para futuras pesquisas, sugerimos explorar o potencial do modelo GPT-4, uma versão avançada do modelo de linguagem GPT, para obter respostas mais precisas, superando as capacidades do GPT-3. Além disso, recomenda-se o uso de diferentes prompts para melhorar o reconhecimento de entidades em textos pelo ChatGPT. A avaliação do desempenho

Tabela 4. Resultados dos algoritmos NER-ChatGPT e BERT para cada entidade do dataset LENER-BR.

Tipo de entidade	Métrica	NER-ChatGPT	BERT	Diferença
JURISPRUDÊNCIA	Precisão	0.0942	0.8188	+0.7246
	Cobertura	0.1405	0.8987	+0.7582
	F1-score	0.1128	0.8569	+0.7441
LEGISLAÇÃO	Precisão	0.1222	0.9484	+0.8262
	Cobertura	0.3122	0.9229	+0.6107
	F1-score	0.1756	0.9355	+0.7599
LOCAL	Precisão	0.1640	0.5934	+0.4294
	Cobertura	0.6596	0.7941	+0.1345
	F1-score	0.2627	0.6792	+0.4165
ORGANIZAÇÃO	Precisão	0.3893	0.8601	+0.4708
	Cobertura	0.4631	0.8754	+0.4123
	F1-score	0.4230	0.8677	+0.4447
PESSOA	Precisão	0.3580	0.9029	+0.5449
	Cobertura	0.7468	0.9743	+0.2275
	F1-score	0.4840	0.9372	+0.4532
TEMPO	Precisão	0.2804	0.9937	+0.7133
	Cobertura	0.6719	0.9224	+0.2505
	F1-score	0.3957	0.9567	+0.5610

nho do algoritmo NER-ChatGPT utilizando outros conjuntos de dados como CDJUR-BR Mauricio *et al.* [2023] e Ullisses Albuquerque *et al.* [2022].

Declarações complementares

Agradecimentos

Agradece-se à orientadora Raquel Silveira por sua inestimável dedicação e pelo compartilhamento de seu vasto conhecimento em Inteligência Artificial. Sua orientação foi fundamental para a condução e o aprimoramento deste estudo, proporcionando insights valiosos e contribuindo significativamente para a qualidade e profundidade da pesquisa. Seu comprometimento e suporte foram essenciais em cada etapa do trabalho, e por isso, expressa-se sincera gratidão.

Contribuições dos autores

Geovane Araujo foi responsável pela escrita do manuscrito, revisão bibliográfica e implementação dos algoritmos para comparação e análise dos resultados, além de conduzir a análise dos dados obtidos. Raquel Silveira supervisionou a pesquisa, fornecendo direcionamento teórico e metodológico, além de contribuir com trabalhos anteriores que serviram de embasamento para este estudo. Ambos os autores participaram da implementação dos algoritmos e da validação dos resultados.

Conflitos de interesse

Os autores declaram que não têm nenhum conflito de interesses.

Disponibilidade de dados e materiais

Os conjuntos de dados (e/ou softwares) gerados e/ou analisados durante o estudo atual estão disponíveis em: <https://github.com/geovanesv/NER-BERT-VS-ChatGPT>

Referências

Albuquerque, H. O., Costa, R., Silvestre, G., Souza, E., da Silva, N. F. F., Vitória, D., Moriyama, G., Martins, L., Soezima, L., Nunes, A., Siqueira, F., Tarrega, J. P., Beinnotti, J. V., Dias, M., Silva, M., Gardini, M., Silva, V., de Carvalho, A. C. P. L. F., and Oliveira, A. L. I. (2022).

UlyssesNER-Br: A Corpus of Brazilian Legislative Documents for Named Entity Recognition. In *Computational Processing of the Portuguese Language*, pages 3–14, Cham. Springer. DOI: 10.1007/978-3-030-98305-5_1.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language Models are Few-Shot Learners. Disponível em: <https://doi.org/10.48550/arXiv.2005.14165>.

Chaves, L. S. (2021). *Utilizando um modelo transformer no processo de identificação de entidades nomeadas em textos criminais*. Trabalho de Conclusão de Curso Graduação em Ciência da Computação, Universidade Federal do Ceará, Campus de Quixada, Quixada. Disponível em: <https://repositorio.ufc.br/handle/riufc/61224>.

Che, W., Wang, M., Manning, C. D., and Liu, T. (2013). Named entity recognition with bilingual constraints. In Vanderwende, L., Daumé III, H., and Kirchhoff, K., editors, *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 52–62, Atlanta, Georgia. Association for Computational Linguistics. Disponível em: <https://aclanthology.org/N13-1006/>.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Disponível em: <https://arxiv.org/abs/1810.04805>.

Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A. d., and de Carvalho, A. C. P. d. L. F. (2021). *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. LTC. Disponível em: <https://repositorio.usp.br/>

- item/002208293.
- Garcia, G. C. (2021). *Reconhecimento de Entidades Nomeadas na base de notificações de eventos adversos e queixas técnicas de dispositivos médicos no Brasil*. Dissertação de Mestrado em Profissional em Computação Aplicada, Universidade de Brasília, Brasília. Disponível em: <http://repositorio.unb.br/handle/10482/42718>.
- Gutterres, E. D. (2022). *Definição de modelo de reconhecimento de entidade nomeada para detecção automática de tópicos de aplicativos em português*. Trabalho de Conclusão de Curso Graduação em Ciências da Computação, Universidade Federal de Santa Catarina, Florianópolis. Disponível em: <https://repositorio.ufsc.br/handle/123456789/243293>.
- Lara, J. and Lobo, G. (2021). *Busca semântica em documentos da área biomédica relacionados com COVID-19*. Trabalho de Conclusão de Curso Graduação em Sistemas de Informação, Anima Educação, Brasil. Disponível em: <https://repositorio.animaeducacao.com.br/handle/ANIMA/20862>.
- Luz de Araujo, P. H., de Campos, T. E., de Oliveira, R. R., Stauffer, M., Couto, S., and Bermejo, P. (2018). LeNER-Br: a dataset for named entity recognition in Brazilian legal text. In *Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings 13*, pages 313–323. Springer. DOI: 10.1007/978-3-319-99722-3_32.
- Mauricio, A., Pinheiro, V., Furtado, V., Neto, J. A. M., Bomfim, F. d. C. J., da Costa, A. C. F., Silveira, R., and Araújo, N. (2023). Cdjur-br—a golden collection of legal document from brazilian justice with fine-grained named entities. *arXiv preprint arXiv:2305.18315*. Disponível em: <https://arxiv.org/abs/2305.18315>.
- Môro, D. K. (2018). *Reconhecimento de Entidades Nomeadas em Documentos de Língua Portuguesa*. Trabalho de Conclusão de Curso Graduação em Tecnologias da Informação e Comunicação, Universidade Federal de Santa Catarina, Florianópolis. Disponível em: <https://repositorio.ufsc.br/handle/123456789/191926>.
- Ratinov, L. and Roth, D. (2009). Design challenges and misconceptions in named entity recognition. In Stevenson, S. and Carreras, X., editors, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 147–155, Boulder, Colorado. Association for Computational Linguistics. Disponível em: <https://aclanthology.org/W09-1119/>.
- Sarrion, E. (2023). What Is ChatGPT? In *Proceedings of the Conference on Exploring the Power of ChatGPT*. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-9529-8_1.
- Silva, A. V. e. and Lopes, M. (2023). Um modelo híbrido para o Reconhecimento de Entidades Nomeadas em português. *Estudos Linguísticos (São Paulo. 1978)*, 51(3):1317–1335. DOI: 10.21165/el.v51i3.3271.
- Silva, P. R. (2020). Promptify: Prompt engineering for named entity recognition (ner). Disponível em: <https://dev.to/praveenr2998/promptify-prompt-engineering-for-named-entity-recognitionner-5dh8>.
- Silveira, R., Ponte, C., Almeida, V., Pinheiro, V., and Furtado, V. (2023). Legalbert-pt: A pretrained language model for the brazilian portuguese legal domain. In *Brazilian Conference on Intelligent Systems*, pages 268–282. Springer. DOI: 10.1007/978-3-031-45392-2_18.
- Sousa, E. B. G. and Mello, R. F. (2022). Aplicação de Reconhecimento de Entidades Nomeadas para Análise Automática de Textos Narrativos em Produções Textuais do Ensino Fundamental. In *Anais Estendidos do XI Congresso Brasileiro de Informática na Educação*. SBC. DOI: 10.5753/cbie_estendido.2022.226747.
- Strubell, E., Verga, P., Andor, D., Weiss, D., and McCallum, A. (2018). Linguistically-informed self-attention for semantic role labeling. Disponível em: <https://doi.org/10.48550/arXiv.1804.08199>.
- Wu, e. a. (2019). Bert for named entity recognition. *arXiv preprint*. Disponível em: <https://doi.org/10.48550/arXiv.1911.04474>.
- Xavier, N. P. (2022). Modelagem e análise de base de conhecimento para um chatbot. Disponível em: <https://repositorio.ifes.edu.br/handle/123456789/2645>.
- Yan, H., Deng, B., Li, X., and Qiu, X. (2019). Tener: Adapting transformer encoder for named entity recognition. Disponível em: <https://arxiv.org/abs/1911.04474>.
- Zhong, Q., Ding, L., Liu, J., Du, B., and Tao, D. (2023). Can chatgpt understand too? a comparative study on chatgpt and fine-tuned bert. Disponível em: <https://arxiv.org/abs/2302.10198>.