UlyssesLegalNER-Br: from Legislative to Legal, a comprehensive corpus of Brazilian legal documents for Named Entity Recognition

Hidelberg O. Albuquerque[1,2] (✉), Ellen Souza[1,2], Danilo C.G. Lucena[2,3], Héldon J.O. Albuquerque[1], Nádia F.F. da Silva[4], Márcio de S. Dias[5], Rafael O. Nunes[6], Adriano L.I. Oliveira[2], and André C.P.L.F. de Carvalho[7],

[1]UAST-UFRPE, Brazil [2]Cin-UFPE, Brazil [3]Zup Innovation, Brazil [4]UFG, Brazil
[5]UFCAT, Brazil [6]UFRGS, Brazil [7]ICMC-USP, Brazil

{hidelberg.albuquerque,ellen.ramos,heldon.jose}@ufrpe.br danilocglucena@gmail.com nadia.felix@ufg.br
marciodias@ufcat.edu.br ronunes@inf.ufrgs.br alio@cin.ufpe.br andre@icmc.usp.br

Abstract

The legal domain presents several challenges for Natural Language Processing (NLP), particularly due to its linguistic complexity and lack of public datasets. Named Entity Recognition (NER), a subarea of NLP, has been successfully used to extract useful knowledge from legal texts. Its widespread use is limited by the lack of legal text corpora. This paper introduces UlyssesLegalNER-Br, a comprehensive corpus of Brazilian legal documents for NER, covering bills, case laws and laws, including the first NER corpus based *exclusively* on Brazilian laws. This research expand the UlyssesNER-Br corpus, previously focused only on the Brazilian legislative domain. The proposed corpus has 560 public documents annotated using a hybrid approach, organized in 9 categories and 23 fine-grained types, experimentally evaluated with the CRF, BiLSTM, and BERTimbau architectures. The corpus was experimentally evaluated regarding predictive performance, computational cost and label-level results. The best micro F1 96.18% was achieved by BERTimbau on the unified corpus, providing a strong baseline for Brazilian legal NER. At the label level, six categories and seven types presented a F1-score above 95%, while the lowest were distributed in the interval 71-82%.

1 Introduction

The analysis of documents from the legal domain presents significant challenges for Natural Language Processing (NLP), mainly due to the strong use of technical language, and specialized vocabulary (Guha et al., 2023). Legal texts vary widely across categories (e.g., bills, case law, statutes), with linguistic patterns distinct from general domains (Finatto and Macohin, 2024). Constant legal updates (e.g., revoked decisions and/or laws), further reinforce the need for scalable automation (Souza et al., 2024).

Named Entity Recognition (NER) can improve the identification and classification of semantic entities in texts (Santos and Cardoso, 2006). Legal NER has been used for identifying and classifying, e.g., legal process, roles, and case law (Castro, 2019; Fernandes et al., 2020), used in legal question-and-answer systems, precedent retrieval, legal analysis, and other applications (Bonifacio et al., 2020). However, applying legal NER in Portuguese presents multiple challenges: scarcity of annotated corpora, lack of standardization in entity schemes for comparison, and structural diversity of texts hinder the generalization of existing models.

In the last years, there has been a significant growth in legal NER studies, although Portuguese research do not offer the same number of resources as in English (Solihin et al., 2021; Souza et al., 2024). For Brazilian Portuguese, several studies developed public resources (Luz de Araujo et al., 2018; Brito et al., 2023; Guimarães et al., 2024), representing important milestones in the availability of data, and the diversity of legal entities. Despite these advances, we can observe that an imbalance still exists: most works are concentrated in a specific subdomain (e.g., court decisions, bills, specific jurisdictions), while the lack of this integration limit the ability to generalize different legal genres.

In this work, we present UlyssesLegalNER-Br, an enhanced and unified corpus developed for NER in the Brazilian legal domain. It is composed of 560 documents from three distinct subdomains: *bills*, from the federal legislative chambers; *case law decisions*, extracted from rulings of superior courts and equivalents; and *laws*. This research substantially expands the UlyssesNER-Br corpus (Albuquerque et al., 2022), previously focused only on the Brazilian legislative domain, seeking to mitigate previously identified gaps (Garcia et al., 2024; Nunes et al., 2024) through revised guidelines, annotation review, broader legal coverage, introduction of new entities, and hybrid corpus annotation.

In addition to expanding document diversity and

Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)
Vol. 1, pages 331–341

Category	Type	Description	Example
DATA (Date)	DATA	Date	01 de janeiro de 2020
EVENTO (Event)	EVENTO	Event	Pandemia de Covid-19
FUNDAMENTO (Law foundation)	FUNDlei	Legal norm	Lei no. 11.340/2006
	FUNDapelido	Legal norm nickname	Lei Maria da Penha
	FUNDprojetodelei	Bill	PEC 187/2016
	FUNDsolicitacaotrabalho	Legislative consultation	Solicitação no. 43/2019
JURISPRUDENCIA (Case law)	**JURISprocesso**	Judicial process	Processo 000.111/2025-0
	JURISprocedimento	Case citations	ADI no. 3.767/STF
	JURISprecedente	Legal precedent	Súmula nº 80/STF
	JURISoutros	Others case law references	Decisão 123/2007-TCDF
LOCAL (Location)	LOCALconcreto	Concrete place	Cajazeiras-PB, Pernambuco
	LOCALvirtual	Virtual place	Jornal de Notícias, UOL
ORGANIZAÇÃO (Organization)	ORGpartido	Political party	PSB, PV, PDT, PT
	ORGgovernamental	Governamental organization	Câmara do Deputados, Senado
	ORGnãogovernamental	Non-governmental organization	ONU, OAB, CRM
PESSOA (Person)	PESSOAindividual	Individual	FHC, Lula da Silva
	PESSOAgrupoind	Group of individuals	Testemunhas, agravantes
	PESSOAcargo	Occupation	Deputado, Professora
	PESSOAgrupocargo	Group of occupations	Senadores, Professores
PRODUTO DE LEI (Law product)	PRODUTOsistema	System product	Sistema Único de Saúde (SUS)
	PRODUTOprograma	Program product	Programa Minha Casa, Minha Vida
	PRODUTOoutros	Others products	Fundo partidário, FIES, IPVA
VALORMONETARIO (Monetary value)	**VALORMONETARIO**	Monetary values	R$ 10,000.00; um milhão de reais

Table 1: UlyssesLegalNER-Br corpus: categories and types, extended from Albuquerque et al. (2022)

semantic coverage, the main contributions from this work include: (i) the first NER corpus annotated *exclusively* with Brazilian laws; (ii) a hybrid annotation workflow combining automatic suggestions and expert curation; and (iii) a systematic evaluation of CRF, BiLSTM, and BERTimbau models, providing a strong experimental baseline for Brazilian Portuguese legal NER. These advances reinforce the empirical foundations of Brazilian legal NLP and support more accurate models across granularities and domains.

The remainder of this article is organized as follows: Sec. 2 reviews related work. Sec. 3 details the annotation process. Sec. 4 presents the corpus. Sec. 5 explains the evaluation setup. Sec. 6 discusses the results. Sec. 7 concludes the paper and outlines future directions.

2 Related Work

NER has been applied to legal texts in several languages (Solihin et al., 2021; Souza et al., 2024), but annotated resources in Portuguese are still limited due to the cost of manual curation. Existing corpora for Brazilian Portuguese generally focus on specific subdomains. LeNER-Br (Luz de Araujo et al., 2018) was the first initiative, comprising 70 legal documents from case laws and laws, and introducing two legal entities. Using BiLSTM-CRF with GloVe embeddings, it achieved an overall F1-score[1] of 92.5%, and became a standard benchmark (Zanuz and Rigo, 2022; Garcia et al., 2024). Some works focus on administrative or judicial genres: DOU-Corpus (470 docs., F1: 44.5%, Alles (2018)), Labor Court (1,3k cases, F1: 93.8%, Castro (2019)), KauaneJunior (3k cases, F1: 94.7%, Fernandes et al. (2020)), and DrugSeizures-Br (6,2k petitions, F1: 89.3%, Bonifacio et al. (2020)). Recent efforts emphasize fine-grained or nested annotation, including STF-corpus (24 nested types, F1: 90%, Correia et al. (2022)), CDJUR-BR (21 types, F1: 67%, Brito et al. (2023)) and PersoSEG (12 types, F1: 75.6%, Guimarães et al. (2024)).

Among these initiatives, UlyssesNER-Br (Albuquerque et al., 2022) is the first corpus to address Brazilian legislative texts with legislative entities. It consists of 950 annotated documents from public bills (PL-corpus) and private legislative consultations (ST-corpus) from the Brazilian Chamber of Deputies, achieving F1 $\approx$ 81% with CRF model. Since then, the corpus has supported several studies: incorporation into the C-Corpus for analysis of legislative comments (Costa et al., 2022); benchmark of legal documents in BiLSTM and BERT hybrid pipelines (Albuquerque et al., 2023); deduplication experiments with RoBERTaLexPT model (Garcia et al., 2024); contamination filtering and self-learning approaches (Nunes et al., 2024; Nunes, 2025); and active learning strategies for annotation efficiency (Gouveia et al., 2025). This work expands UlyssesNER-Br corpus in volume, subdomains, and coverage, creating a unified and multi-generic resource designed to support the robust NER evaluation in heterogeneous legal texts.

[1]Unless otherwise stated, all F1 scores are micro F1, as commonly used in NER evaluation.

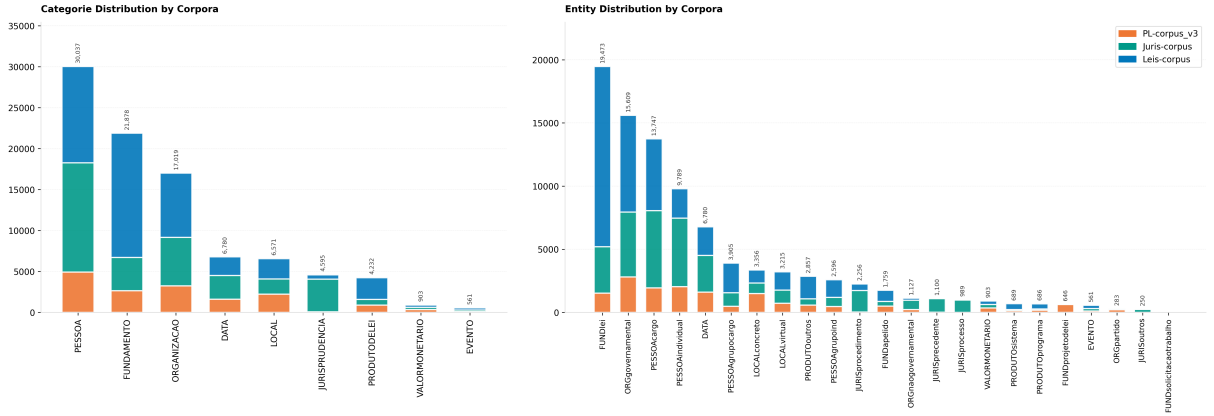

Figure 1: Distribution of categories and entities

(a) PL-corpus_v3 (b) Juris-corpus (c) Leis-corpus

Figure 2: Annotation workflow in INCEpTION: curated entities in color, automatic suggestions in gray

3 Method

The UlyssesNER-Br contains seven semantic classes or categories, five of which were based on HAREM (Santos and Cardoso, 2006), and two specifics to the legislative domain. These categories were also refined into 18 entity types. In this research, two new semantic categories and five types were included, to ensure comprehensive coverage of the main legal genres: case law (*jurisprudência*), and monetary value (*valor monetário*). The *case law* seeks to identify references to legal processes or courts decisions. In turn, references to monetary values are usual in legal documents, such as budget bills and court rulings. In addition to expanding the scope of the corpus, the inclusion of these entities also aligns with the Brazilian legal NER works (Luz de Araujo et al., 2018; Guimarães et al., 2024).

Our annotation scheme retains all original entity types from UlyssesNER-Br to ensure consistency. Nested or overlapping mentions were not annotated. Even though it is not present in our corpus, we keep the entity type *FUNDsolicitacaotrabalho* to maintain full compatibility with the ST-corpus, where it appears exclusively. Table 1 lists the nine categories and 23 entity types, including the newly added labels. Figure 1 shows their mention distribution across the three subcorpora, whereas Table 4 reports aggregate mentions and document coverage in the unified corpus. The category layer reduces sparsity caused by rare entity types, while the fine-grained labels preserve the underlying long-tail frequency distribution.

3.1 Annotation Process

The original annotation process was carried out by three groups across three phases. In the present work, given the resource constraints, we adopted an interactive, machine-assisted annotation workflow using the recommendation system of the INCEpTION tool (Klie et al., 2018)[2], illustrated in Figure 2. All suggestions were manually validated by an expert human. As annotation was performed by a single expert, inter-annotator agreement (IAA) was not assessed. Annotation quality checks included script-based detection of label inconsistencies and boundary mismatches, manual expert review, and revision of the previously annotated corpus version used to support the recommendation workflow.

Based on the ambiguities identified in the original corpus (Garcia et al., 2024; Nunes et al., 2024), we undertook the following steps: (i) revision and updating of the annotation scheme, including the new entities; (ii) revision of prior annotations and correction of inconsistencies (see Table 2); (iii) expansion of the bill corpus; and (iv) inclusion of

[2] https://inception-project.github.io/

	Documents		Mentions					
Granularity	**Reviewed**	**With corrections**	**Corrected**	**Added**	**Removed**	**Relabelled**	**Boundary-adjusted**	**Boundary+ label changed**
Category-level	149	149	1,607	1,249	85	69	178	26
Type-level	149	149	1,666	1,249	85	128	162	42

Table 2: Summary of original PL-corpus (v1) revisions at the category and type levels

			Sentences		Tokens		Named Entities (T/C)[1]	
Corpus	**Period**	**Docs**	**Total**	**Filtered**	**Total**	**Filtered**	**Mentions**	**Density (%)**
UlyssesLegalNER-Br	**1940-2025**	**560**	**87,184**	**56,698**	**2,189,406**	**1,641,303**	**92,576 (T/C)**	**17.47 (T/C)**
(A) *PL-corpus_v3* (Bills)	2011-2025	300	26,053	7,971	345,291	246,917	16,130 (T/C)	18.38 (T/C)
(B) v2 (2024)	2011-2019	149	2,528	2,348	110,900	85,326	3,006 (T) / 3,537 (C)	10.51 (T) / 10.61 (C)
(C) v1 (2022)	2011-2019	150	9,526	3,070	138,741	103,244	3,763 (T/C)	10.18 (T/C)
(D) *Juris-corpus* (Case law)	2015-2025	160	20,545	19,469	600,412	492,437	33,245 (T/C)	21.13 (T/C)
(E) *Leis-corpus* (Laws)	1940-2025	100	40,586	29,258	1,243,703	901,949	43,201 (T/C)	15.46 (T/C)

Table 3: Corpus-level statistics. [1]"T/C": granularity in types and categories

the new subcorpora. In the end, all comparable previously annotated documents were systematically reviewed, and the new ones were annotated. Documents were randomly sampled while ensuring topical and domain diversity.

4 The UlyssesLegalNER-Br corpus

In this section, we present UlyssesLegalNER-Br, an extended and enhanced version of the UlyssesNER-Br corpus (Albuquerque et al., 2022). We describe the structural properties of the corpus (Table 3) and provide a performance evaluation of state of the art (SOTA) models for NER tasks. In total, 560 legal documents were annotated, comprising about 87k sentences and 2M tokens. The corpus includes three types of Brazilian legal documents: bills (PL-corpus_v3), case law (Juris-corpus), and laws (Leis-corpus). Only public documents were included from the Tesemõ corpus (Siqueira et al., 2025)[3], or directly from official portals. A total of 92,6k entities were annotated (overall density: 17.47%), reflecting the high informational load characteristic of legal texts. The subcorpora show comparable entity density, indicating consistent annotation.

Brazilian legal documents, especially legislative and judicial texts, frequently repeat prior references, leading to many single-token sentences (e.g., "." or ","). Following the original methodology, we estimated the number of tokens per sentence (n) both for the unified corpus and for each subcorpus. Sentences outside the range of $n \in (1, 100]$ were classified as outliers and excluded only from the statistical analysis. After filtering, 56,7k sentences and 1,64M tokens were retained for the statistics, while the full corpus remained intact for all experiments. The descriptive statistics are reported with 95% confidence intervals (CI). The estimated average number of sentences per document in the corpus is 155.69 (CI: [121.87; 189.50]). The estimated average number of filtered tokens per sentence is 28.95 (CI: [28.76; 29.14]). These results suggest structural uniformity, supporting fair comparisons and minimizing bias from textual variation.

Categories/Types	**Mentions**	**Docs**
PESSOA	**30,037**	**559**
PESSOAcargo	13,747	559
PESSOAindividual	9,789	559
PESSOAgrupocargo	3,905	352
PESSOAgrupoind	2,596	282
FUNDAMENTO	**21,878**	**544**
FUNDlei	19,473	526
FUNDapelido	1,759	323
FUNDprojetodelei	646	152
FUNDsolicitacaotrabalho	0	0
ORGANIZACAO	**17,019**	**560**
ORGgovernamental	15,609	560
ORGnaogovernamental	1,127	185
ORGpartido	283	94
DATA	**6,780**	**559**
LOCAL	**6,571**	**485**
LOCALconcreto	3,356	454
LOCALvirtual	3,215	361
JURISPRUDENCIA	**4,595**	**222**
JURISprocedimento	2,256	171
JURISprecedente	1,100	119
JURISprocesso	989	114
JURISoutros	250	56
PRODUTODELEI	**4,232**	**287**
PRODUTOoutros	2,857	222
PRODUTOsistema	689	112
PRODUTOprograma	686	98
VALORMONETARIO	**903**	**104**
EVENTO	**561**	**95**

Table 4: UlyssesLegalNER-Br entity distribution

[3] https://github.com/ulysses-camara/ulysses-tesemo

4.1 PL-corpus_v3

The original PL-corpus (PL-corpus_v1) was constructed from 150 legislative documents collected from the Brazilian Chamber of Deputies between 2011 and 2019, covering bills from various areas. For PL-corpus_v1, Garcia et al. (2024) and Nunes et al. (2024) applied deduplication and decontamination procedures. However, only the latter released a revised version, in which one duplicate document was removed. We refer to this version as PL-corpus_v2. Although motivated by those studies, the current third version was rebuilt directly from PL-corpus_v1. Our version includes a systematic revision of the existing annotations, a expansion with bills from the Brazilian Senate[4], and the inclusion of the new entities, totaling 300 documents with broader legislative coverage and less subdomain bias. Table 2 summarizes the revision of the PL-corpus_v1, whose corrected annotations supported the expanded corpus annotation. Across the 149 directly comparable documents, corrections affected all reviewed documents in each granularity (categories and types), and increased the number of annotated mentions from 3,720 to 4,854 (+30.5%). The estimated average of sentences per document is 86.84 (CI: [72.60; 101.09]), and the estimated tokens per sentence is 30.98 (CI: [30.47, 31.49]). Table 3 (A–C) provides comparative analysis.

4.2 Juris-corpus

Case law is a fundamental source of legal language, reflecting how courts interpret and apply legislation. The hierarchical judicial system generates distinct argumentation patterns and technical vocabularies, resulting in substantial linguistic variability (Finatto and Macohin, 2024). Publicly available case law corpora are therefore strategic for legal NLP (Souza et al., 2024). Existing Brazilian NLP works typically cover specific courts or subdomains (Castro, 2019; Correia et al., 2022; Brito et al., 2023).

The Juris-corpus contributes a multi-jurisdictional and multi-instance case law corpus, integrating 120 decisions from all five Brazilian superior courts (STF, STJ, STM, TSE, and TST), and with 25 rulings from appellate courts (TRFs and TJs). We also included 15 decisions from the Federal Court of Accounts (TCU), which, although outside the Judiciary, produce technically rulings comparable in structure to judicial decisions. This comprehensive coverage of different legal documents enables us to capture the terminological and stylistic heterogeneity of Brazilian jurisprudence, thereby contributing to the development of legal NLP in Portuguese. In total, 160 documents were annotated (Table 3, D), with the estimated metrics: average of 128.41 sentences per document (CI: [109.62; 147.20]) and 25.29 tokens per sentence (CI: [24.96; 25.62]).

4.3 Leis-corpus

Robust legal NLP systems must process the entire cycle of legal genre: formulation (bills), standardization (laws), and application (case law) (Corazza et al., 2024). Although, most of legal NER studies focus mainly on case law. Laws, however, exhibit distinct linguistic attributes, as technical terminology, frequent cross-references, and structured formatting (Sjansen, 2022), making them a crucial for the legal NLP. Although some studies include laws (Váradi et al., 2020; Siqueira et al., 2025), to the best of our knowledge, there is no NER corpus composed mainly of laws in Portuguese or English, leaving a major gap in legal NLP.

Aiming to fill the identified gap and provide greater robustness to Portuguese legal NLP, we present the Leis-corpus, the first NER corpus composed *exclusively* of Brazilian laws[5]. The corpus contains 100 laws, including the current Brazilian Federal Constitution (1988), three State Constitutions (Paraíba, Pernambuco, and Goiás), seven statutes (e.g., Estatuto da Magistratura), four legal codes (e.g., Código Penal), and several federal laws and decrees, highlighting this corpus from previous works. The selected laws cover diverse thematic areas, including administrative law, social policies, criminal, and digital rights. This multi-record approach enables examining how NLP models generalize/specialize across legal text types, a central question in legal systems (Costa et al., 2022). This thematic diversity reinforces the broad coverage of the Brazilian legal system, from historical (1940s–1970s) to contemporary legislation (2020–2025).

The Leis-corpus contains 405.86 sentences per law (CI: [231.54; 580.18]). The considerable range of this interval reflects the heterogeneity of the selected laws, including short documents such as decrees (12-30 sentences) and extensive examples such as constitutions and codes (thousands of sentences). The average lexical density is 30.83 tokens

[4] https://www25.senado.leg.br/

[5] To ensure legal coherence, only compiled versions of the current legislation were used, from https://www4.planalto.gov.br/legislacao

Model	Key Hyperparameters
CRF	L-BFGS; c1=c2=0.1; max iterations=100 all_possible_transitions=True
BiLSTM-CRF +GloVe	GloVe NILC 300d; 2-layer BiLSTM (hidden size per direction: 128, 256); dropout=0.5; batch=32; epochs=75 RMSProp (lr=0.001); early stopping (patience=10)
BERTimbau$_{base}$	maxlen=512; batch=20; epochs=20; warmup=10% AdamW (lr=2e-5, wd=0.01); grad clip=1.0 label_all_tokens=True
BERTimbau$_{base}$ -CRF	Same as BERTimbau$_{base}$ + CRF decoding layer

Table 5: Essential hyperparameters used, based on Albuquerque et al. (2023), with minor adaptations

per sentence (CI: [30.57; 31.09]). It is worth highlighting that this corpus has the fewest documents, but the largest number of sentences, tokens, and annotated entities, as shown in Table 3 (E).

5 Evaluation

To evaluate our proposal, we used three widely adopted architectures in a structured experimental environment. Each subcorpus was evaluated individually to identify the model that offered the most suitable result. This model was then applied to the unified corpus. The following sections describe the training data, setup, hardware, hyperparameters, metrics, and preliminary studies.

5.1 Data

All documents were used without text preprocessing, or additional normalization. With the exception of the PL-corpus_v1 (already provided in IOB2 format), the remaining documents were downloaded in original format and converted to plain text, followed by manual inspection to correct occasional conversion errors. After annotation, the documents were exported in IOB2 format

5.2 Experimental Setup

To evaluate the corpus, we conducted all experiments using 5-fold cross-validation at the sentence level, with sentences as instances rather than documents. This protocol was adopted to maintain consistency with UlyssesNER-Br, improve data utilization, and obtain more robust performance estimates across folds while preserving reproducibility and comparability between models. Deterministic randomized splits were generated with `KFold` (`shuffle=True`, `random_state=42`), and the same folds were reused across all models to ensure consistent comparisons. No stratification by entity type was applied. All sources of randomness were controlled via fixed seeds (`global seed=1999`; `fold generation random state=42`).

We selected four representative models commonly used in legal NER: Conditional Random Fields (CRF) (Lafferty et al., 2001), BiLSTM-CRF with GloVe embeddings (Lample et al., 2016; Pennington et al., 2014), and two variants of the BERT architecture (Devlin et al., 2019): BERTimbau$_{base}$ (Souza et al., 2020) and BERTimbau$_{base}$ with a CRF decoding layer. Table 5 details the key hyperparameters used. All experiments were conducted on Google Colab Pro+[6]. Different hardware setups were used both to reduce computational cost and because lighter models do not require high-end GPUs, unlike Transformers-based models.

The **CRF** baseline is trained at the sentence level using the CoNLL-formatted sequences, implemented with `sklearn-crfsuite`[7]. We follow the UlyssesNER-Br feature templates, using a fixed context window of ± 2 tokens and combining lexical/orthographic cues, POS tags (when available), and a small set of legal-domain patterns. Training was performed on an Intel Xeon CPU (2.20 GHz).

The **BiLSTM-CRF+GloVe** ensemble model adopts the standard neural sequence-labeling architecture, with a CRF layer with `pytorch-crf`[8], and 300-dim Brazilian Portuguese NILC GloVe embeddings[9]. The GloVe are fine-tuned and concatenated with character-level CNN features before the BiLSTM. Dropout is applied after the word+character concatenation and after each BiLSTM layer, and CRF decoding uses masking to ignore padded positions. Hyperparameters were selected based on a prior comparative evaluation for Brazilian legal documents, and the experiments were performed on a NVIDIA T4 GPU.

For BERT-based evaluation, we used the **BERTimbau$_{base}$** checkpoint (`neuralmind/bert-base-portuguese-cased`)[10], adopting a previously used hyperparameter configuration for legal NER, with the number of epochs selected based on preliminary experiments. Training and inference are performed over the sentence instances. Each sentence is tokenized and padded/truncated up to a maximum sequence length of 512 tokens (no sliding-window or block-based document segmentation is used). The **BERTimbau+CRF** variant adds a

[6] https://colab.google/
[7] https://sklearn-crfsuite.readthedocs.io/
[8] https://pytorch-crf.readthedocs.io/
[9] https://huggingface.co/collections/nilc-nlp/nilc-embeddings
[10] https://github.com/neuralmind-ai/portuguese-bert

Corpus (Sentences / Tokens)	Model	Training Time (h)	Training Cost (rel.)	Accuracy	Micro Precision	Micro Recall	Micro F1	Weighted F1	Macro F1	F1-cost
PL-corpus_v3 (Types) (26,251 / 345,737)	CRF	0.51	0.51	97.63 ± 0.49	91.64 ± 0.90	85.67 ± 1.96	88.55 ± 1.42	88.14 ± 1.55	74.49 ± 5.36	145.82
	BiLSTM-CRF+GloVe	4.09	16.36	98.12 ± 0.23	88.62 ± 1.42	91.02 ± 1.01	89.80 ± 1.15	89.72 ± 1.13	79.98 ± 5.16	4.89
	BERTimbau$_{base}$	12.58	251.60	**98.69 ± 0.12**	**93.33 ± 0.74**	**94.55 ± 0.30**	**93.94 ± 0.60**	**93.74 ± 0.57**	**85.64 ± 3.15**	0.34
	BERTimbau$_{base}$+CRF	24.61	492.25	98.15 ± 0.28	92.15 ± 1.45	94.58 ± 0.31	93.35 ± 0.53	93.35 ± 0.32	85.53 ± 3.08	0.17
Juris-corpus (Types) (20,663 / 600,564)	CRF	0.92	0.92	98.46 ± 0.04	94.21 ± 0.23	91.92 ± 0.25	93.05 ± 0.23	93.01 ± 0.23	89.02 ± 1.17	96.73
	BiLSTM-CRF+GloVe	3.95	15.80	98.73 ± 0.13	93.57 ± 0.62	94.67 ± 0.33	94.12 ± 0.46	94.08 ± 0.45	89.70 ± 1.62	5.67
	BERTimbau$_{base}$	9.38	187.60	**98.98 ± 0.08**	**95.71 ± 0.49**	**96.56 ± 0.29**	**96.13 ± 0.28**	**96.12 ± 0.27**	**92.71 ± 0.84**	0.49
	BERTimbau$_{base}$+CRF	23.31	466.19	98.70 ± 0.09	95.07 ± 1.43	96.36 ± 0.41	95.71 ± 0.57	95.82 ± 0.56	91.89 ± 1.46	0.20
Leis-corpus (Types) (40,586 / 1,243,703)	CRF	1.69	1.69	99.01 ± 0.07	95.16 ± 0.21	92.28 ± 0.60	93.70 ± 0.37	93.54 ± 0.38	82.25 ± 1.66	48.54
	BiLSTM-CRF+GloVe	7.27	29.08	98.63 ± 0.21	92.13 ± 1.19	91.26 ± 2.97	91.69 ± 0.99	91.11 ± 1.56	82.56 ± 4.52	2.84
	BERTimbau$_{base}$	18.29	365.77	**99.32 ± 0.08**	**95.49 ± 0.40**	**96.49 ± 0.37**	**95.99 ± 0.37**	**95.99 ± 0.38**	**88.87 ± 3.13**	0.24
	BERTimbau$_{base}$+CRF	29.98	599.72	98.55 ± 0.15	95.31 ± 0.23	96.15 ± 0.23	95.73 ± 0.58	93.55 ± 0.77	88.73 ± 2.78	0.15
Ulysses* (Types) Ulysses* (Categories) (87,500 / 2,190,004)	BERTimbau$_{base}$	35.23 33.09	704.53 661.82	**99.14 ± 0.04 99.23 ± 0.08**	**95.42 ± 0.08 95.67 ± 0.21**	**96.27 ± 0.20 96.70 ± 0.41**	**95.84 ± 0.13 96.18 ± 0.24**	**95.87 ± 0.12 96.16 ± 0.23**	**93.21 ± 0.05 94.95 ± 0.78**	0.13 0.14

Table 6: Model performance and relative training cost across corpora (mean ± std over 5-fold CV). Relative training cost: training time × hardware factor, based on Strubell et al. (2019): CRF (CPU) = 1x, BiLSTM (GPU T4) = 4x, and Transformer (GPU A100) = 20x. F1-cost: macro-F1 per unit cost. Ulysses*: full corpus. Best results in bold.

CRF decoding layer (with `pytorch-crf`[8]), on top of BERTimbau to capture tag-transition dependencies. All hyperparameters of BERTimbau were kept identical for comparability. All BERT-based models experiments used a NVIDIA A100 GPU.

The performance of the models was evaluated using standard NER metrics: overall Accuracy, micro Precision, micro Recall, and F1-score (micro, weighted, and macro), with the `seqeval library`[11]. For comparable experimental settings, statistical analyses were performed using the Wilcoxon Signed-Rank test ($p < 0.05$). In addition, we assessed computational efficiency examining training cost vs. predictive performance, based in the relative efficiency ratios by Strubell et al. (2019). For each corpora, we report the training time, the relative compute cost, and the macro-F1 per unit cost (F1-cost), used as an efficiency indicator[12].

5.3 Preliminary Experiments

This section reports preliminary sensitivity experiments (e.g., folds, epochs, model size) to confirm data integrity and evaluate alternative training configurations under computational constraints. Increasing the number of CRF folds (5→10) did not produce statistically significant gains (Wilcoxon: [*PL_v3*, p=0.0625]; [*Leis*, p=0.0625]; [*Juris*, p=0.1875]), despite small increases in macro F1 (+5.3pp, +3.5pp, +1.5pp). For the BERTimbau$_{base}$, extending training from 55→75 epochs (Albuquerque et al., 2023) also showed no improvement, with performance stabilizing around the 20th epoch (p=0.8125). Both adjustments added a substantial computational cost, and exploratory runs with BERTimbau$_{large}$ offered only marginal benefits at a considerably higher cost. After validating the pipeline, we fixed the final hyperparameters based on the preliminary experiments and the computational constraints of this work, and then conducted a full evaluation on the subcorpora to identify the most suitable model for the unified corpus. Model selection considered both mean micro F1 and standard deviation across folds, balancing predictive performance and stability.

6 Results and Discussion

Table 6 summarizes the results of evaluation, revealing a clear and consistent overall pattern. BERTimbau systematically outperforms in every corpus and metrics. Its performance higher in absolute terms and stability, with low standard deviation (std) regardless of text genre or corpus size. In *PL_v3*, BERTimbau reaches a micro F1 of 93.9%. A similar advantage appears in *Juris* (96.1%), reinforcing the model's ability to capture the variability of legal entities. In *Leis*, the largest subcorpus, BERTimbau reaches 95.9%, surpassing the previous one (+2-4pp). Regarding BERTimbau+CRF, the addition of a CRF decoding layer does not yield consistent gains. Across the subcorpora, its results remain statistically similar to the vanilla BERTimbau, but slightly inferior, while incurring substantially higher computational cost. Based in our results, the additional CRF layer did not justify its complexity.

The CRF model remains competitive in micro F1. But, its overall performance lags behind neural models, and macro F1 shows a bias towards frequent classes, with poor handling of rare but semantically important entities for legal context. BiLSTM-CRF reduces this difference, but BERTimbau still outperform both by a substantial margin.

[11] https://pypi.org/project/seqeval/

[12] The factors used are approximate rather than exact energy measurements, but they provide a consistent basis for comparing model families under a unified cost framework.

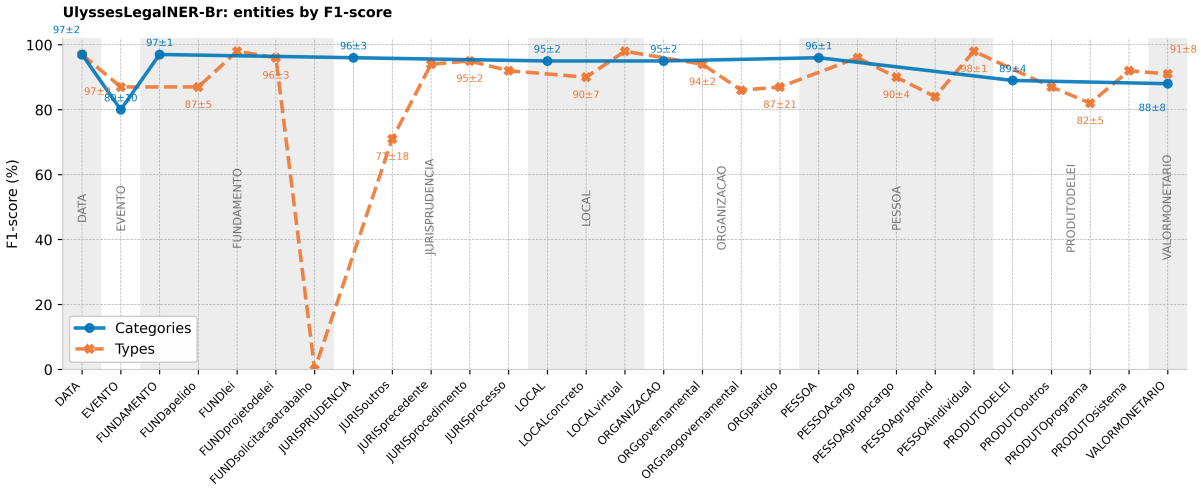

Figure 3: UlyssesLegalNER-Br, BERTimbau results: chances of success by categories and types

Corpus	Granularity	Model	Best micro-F1
PL-corpus_v1 (138k tokens)	Types	CRF [1]	76.30
		Albertina-PT-BR$_{xlarge}$ [2]	86.62
	Categories	BiLSTM-CRF+GloVe [3]	76.94
		RoBERTaLexPT$_{base}$ [2]	88.56
PL-corpus_v2 (110k tokens)	Types	WE. GloVe [4]	79.14
		Albertina 100M [4]	86.86
	Categories	WE. GloVe [4]	80.75
		BERTimbau$_{large}$ [4]	88.09
PL-corpus_v3 (345k tokens)	Types	CRF	88.55
		BiLSTM-CRF	89.80
		BERTimbau$_{base}$	93.94

Table 7: Historical overview of the PL-corpus versions. Only the best micro-F1 scores are shown. Results reported in: [1]Albuquerque et al. (2022), [2]Garcia et al. (2024), [3]Albuquerque et al. (2023), [4]Nunes (2025)

In the across-corpora comparison, BERTimbau achieves the highest F1-scores on *Juris*. This result may be attributed to corpus-specific characteristics, particularly the more argumentative and structured nature of judicial texts and their high entity density (21.13%). In contrast, the slightly lower stability and performance observed in *Leis* may be associated with its strong class imbalance. *PL_v3* shows the lowest scores and the highest variance, which may be linked to greater structural heterogeneity.

About computational cost, CRF are the most efficient, achieving the highest F1-cost values[13], but with a substantially lower macro-F1. BiLSTM-CRF offers moderate gains, but reduced efficiency due to longer training time. Transformer models are considerably more expensive (Strubell et al., 2019), but consistently exhibit the best predictive performance. Among them, BERTimbau provides the best results, while BERTimbau-CRF is approximately twice as slow and half as efficient. Considering the predictive result, BERTimbau was the chosen option.

The unified corpus was evaluated with BERTimbau at two annotation granularities: categories and types. As shown in Table 6, both settings exhibit very similar behavior across all metrics, with close mean values and low variances. The category-level model achieves a slightly higher micro F1 (96.18±0.24% vs. 95.84±0.13%), which is consistent with the coarser label space. Because these settings correspond to different prediction granularities, we avoid inferential statistical comparison and report only a descriptive comparison of the aggregate results. From an operational perspective, the category-level setting also shows a slightly higher reported F1-cost ratio (0.143 vs. 0.132).

Comparing the overall results, the unified category-based model shows competitive performance across all domains: it outperforms *PL_v3* (+2pp), improves on *Leis* (+0.19pp), and is practically tied with *Juris* (+0.05pp). It also demonstrates greater stability (cv: 0.24%) than the individual models (cv: 0.28-0.6%), indicating that cross-domain training does not reduce consistency. For the unified corpus, the category and type-based settings both show strong performance, with only small descriptive differences in score and variation across folds.

Figure 3 shows that categories with well-defined lexical structures (e.g., DATA, FUNDAMENTO) achieve stable micro F1 of ≥96%, while broad or sparse classes (e.g., LOCAL, ORGANIZACAO, EVENTO) remain ≤88% with higher variance. At the type level, again well-defined entities (e.g., FUNDlei, LOCALvirtual, PESSOAindi-

[13]A higher F1 cost indicates a lower computational cost for a given performance.

Work	Corpus	Domain	Docs.	Tokens	Ann.	Ent./Gr.	Models	Best micro-F1
Luz de Araujo et al. (2018)	LeNER-Br	Case law+Law	70	318k	M	6U	BiLSTM, CRF	92.53
Alles (2018)	DOU-corpus	Gazette	470	597k	H	10U	OpenNLP	44.50
Castro (2019)	Labor Court	Case law	1,3k	9M	H	10U	ELMo	93.81
Collovini et al. (2019)	Police dataset	Police	30	37k	M	1U	BiLSTM, CRF, ELMo	90.95
Bonifacio et al. (2020)	DrugSeizures-Br	Police	6,2k	6,4k	A	6C, 25T	BERT	89.39
Fernandes et al. (2020)	KauaneJunior	Appellate	3k	221k	A	6U	BiLSTM, BIGRU, CRF	94.79
Albuquerque et al. (2022)	UlyssesNER-Br	Legislative	950	216k	M	7C, 18T	HMM, CRF, BiLSTM	81.04
Correia et al. (2022)	STF-corpus	Case law	594	1,7M	M	4C, 24T	BiLSTM, CRF	90.00
Costa et al. (2022)	C-Corpus	Leg. Comments	1,1k	178k	M	7C, 18T	BERT	73.90
Zanuz and Rigo (2022)	LeNER-Br	Case law+Law	70	318k	No	6U	BERT	91.14
Albuquerque et al. (2023)	LeNER-Br, PL-corpus	Mixed	220	456k	No	6U, 7C	CRF, BiLSTM, BERT	88.27
Brito et al. (2023)	CDJUR-BR	Case law	1,2k	NI	M	6C, 21T	BERT	67.00
Guimarães et al. (2024)	PersoSEG	Gazette	127	NI	M	11U	CNN, CRF, LSTM	75.65
Nunes et al. (2024)	PL-corpus	Legislative	149	110k	No	7C	BERT LMs	86.70
Nunes (2025)	LeNER-Br	Case law+Law	220	318k	No	6U	CRF, BERT LMs	95.38
This work	UlyssesLegalNER-Br	Mixed	560	2,1M	H	9C, 23T	CRF, BiLSTM, BERT	96.18

Table 8: Contextual overview of reported results in Brazilian legal NER studies. Ann.: type of annotation (A/H/M/No=Automatic, Hybrid, Manual, No annotation effort). Ent./Gr.: Entities and Granularity (U/C/T=Unique, Categories, Types). BERT: all BERT-based variants. LMs: language models. Mixed: Legislative+Case law+Law.

vidual) reach ≈ 98%, whereas ambiguous or low-frequency types (e.g., JURISoutros, ORGpartido) exhibit high variability (std ≈18-20%). Overall, semantic ambiguity and data sparsity remain key challenges for fine-grained legal NER.

To provide a historical overview, Table 7 summarizes the best micro F1 results previously reported for different versions of the PL-corpus. Although encoder-based models consistently appear among the strongest approaches in earlier versions, and our results for *PL_v3* are also strong in the type-level setting, these values are not directly comparable, as they were obtained on different corpus versions using different model architectures and evaluation settings. Table 8 situates our results within the broader Brazilian legal NER research. As the reported studies differ in corpus composition, label granularity, and evaluation settings, the values are shown for contextual reference only. Our unified model achieves 96.18% micro F1 and provides a strong baseline on the proposed corpus.

7 Conclusion and future works

We present UlyssesLegalNER-Br, a unified corpus of Brazilian legal texts for NER, composed of 560 public documents (bills, case law, and laws). The corpus comprises 2,1M tokens in nine entity categories and 23 refined types, constituting the first large-scale resource that harmonizes multiple legal genres under a single annotation scheme. Experiments with CRF, BiLSTM, and BERT-based models demonstrate that the BERTimbau$_{base}$, trained across all domains achieves 96.18% micro F1. Category-level results are slightly higher than refined-type results. The BERTimbau and BERTimbau-CRF models perform similarly, but the former is 1.9x more cost-efficient. Overall, these results provide a strong and computationally efficient baseline for Brazilian legal NER on the proposed corpus and unified cross-domain setting.

Some limitations in our work. The corpus does not cover all legal genres, and our evaluation focused on BERTimbau, leaving other Brazilian language models (e.g., Albertina, RoBERTaLexPT) outside the experiments. Annotation was performed by a single expert; therefore inter-annotator agreement (IAA) was not computed, and annotation quality was not assessed through a multi-annotator setting. Resource constraints also limited category-level evaluation in all subcorpora. Ongoing work includes annotation with LLM-assisted and integration into legal information retrieval systems. Future works involves expanding to additional genres, evaluating more recent Brazilian language models, and conducting multi-annotator studies with systematic annotation error analysis, as well as exploring cross-domain transfer. The Ulysses project repository provides access to the PL-corpus versions (v1 and v2), all resources generated in this work (including raw text and annotated data at both the type and category levels of granularity), and more in-depth analyses. [14]

Acknowledgments

This research was conducted in the context of Ulysses Project. Ellen Souza is supported by FACEPE. Nádia Félix is supported by FAPEG and CNPq. Rafael O. Nunes is supported by Petrobrás and CNPq. André C.P.L.F. de Carvalho and Adri-

[14] https://github.com/ulysses-camara/ulysses-ner-br

ano .L.I. Oliveira are supported by CNPq. We thank the Institute of Artificial Intelligence (IAIA) and the funding agencies for their support. Generative AI tools (ChatGPT 5.4 and Claude 4.6 Sonnet) were used for code optimization and linguistic refinement; all outputs were validated by the authors.

References

Hidelberg O. Albuquerque, Rosimeire Costa, Gabriel Silvestre, Ellen Souza, Nádia F.F. da Silva, Douglas Vitório, Gyovana Moriyama, Lucas Martins, Luiza Soezima, Augusto Nunes, Felipe Siqueira, João P. Tarrega, Joao V. Beinotti, Marcio Dias, Matheus Silva, Miguel Gardini, Vinicius Silva, André C.P.L.F. de Carvalho, and Adriano L.I. Oliveira. 2022. UlyssesNER-Br: A corpus of Brazilian Legislative Documents for Named Entity Recognition. In *15th Computational Processing of the Portuguese Language (PROPOR 2022)*.

Hidelberg O. Albuquerque, Ellen Souza, Adriano L. I. Oliveira, David Macêdo, Cleber Zanchettin, Douglas Vitório, Nádia F. F. da Silva, and André C.P.L.F. de Carvalho. 2023. On the Assessment of Deep Learning Models for Named Entity Recognition of Brazilian Legal Documents. In *22st EPIA Conference on Artificial Intelligence (EPIA 2023)*.

V.J. Alles. 2018. Construção de um corpus para extrair entidades nomeadas do diário oficial da união utilizando aprendizado supervisionado. Master's thesis, Master's thesis, Universidade Federal de Brasília.

Luiz H. Bonifacio, Paulo A. Vilela, Gustavo R. Lobato, and Eraldo R. Fernandes. 2020. A Study on the Impact of Intradomain Finetuning of Deep Language Models for Legal Named Entity Recognition in Portuguese. In *9th Brazilian Conference on Intelligent Systems (BRACIS 2020)*.

Maurício Brito, Vládia Pinheiro, Vasco Furtado, João M. Neto, Francisco Bomfim, André da Costa, and Raquel Silveira. 2023. CDJUR-BR - Uma Coleção Dourada do Judiciário Brasileiro com Entidades Nomeadas Refinadas. In *14th Symposium in Information and Human Language Technology (STIL 2023)*.

Pedro V.Q. de Castro. 2019. Aprendizagem profunda para reconhecimento de entidades nomeadas em domínio jurídico. Master's thesis, Federal University of Goiás.

Sandra Collovini, Joaquim Francisco Santos Neto, Bernardo Scapini Consoli, Juliano Terra, Renata Vieira, Paulo Quaresma, Marlo Souza, Daniela Barreiro Claro, and Rafael Glauber. 2019. Iberlef 2019 portuguese named entity recognition and relation extraction tasks. In *Proceedings of IberLEF 2019, with SEPLN 2019*.

Michele Corazza, Leonardo Zilli, and Monica Palmirani. 2024. Topic Similarity of Heterogeneous Legal Sources Supporting the Legislative Process. In *10th Italian Conference on Computational Linguistics (CLiC-it 2024)*.

Fernando A. Correia, Alexandre A.A. Almeida, José L. Nunes, Kaline G. Santos, Ivar A. Hartmann, Felipe A. Silva, and Hélio Lopes. 2022. Fine-grained legal entity annotation: A case study on the Brazilian Supreme Court. *Information Processing & Management*, 59(1).

Rosimeire Costa, Hidelberg O. Albuquerque, Gabriel Silvestre, Nádia Félix F Silva, Ellen Souza, Douglas Vitório, Augusto Nunes, Felipe Siqueira, João P. Tarrega, João Vitor Beinotti, Márcio de S. Dias, Fabíola S.F. Pereira, Matheus Silva, Miguel Gardini, Vinicius Silva, André C.P.L.F. de Carvalho, and Adriano L.I. Oliveira. 2022. Expanding UlyssesNER-Br named entity recognition corpus with informal user-generated text. In *21st EPIA Conference on Artificial Intelligence (EPIA 2022)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*.

William P.D. Fernandes, Luiz J.S. Silva, Isabella Z. Frajhof, Guilherme F.C.F. de Almeida, Carlos N. Konder, Rafael B. Nasser, Gustavo R. de Carvalho, Simone D.J. Barbosa, and Hélio C.V. Lopes. 2020. Appellate court modifications extraction for Portuguese. *Artificial Intelligence and Law*, 28(3):327–360.

Maria J.B. Finatto and Aline Macohin. 2024. PLN no Direito: Perspectivas e desafios com textos jurídicos e legais. In H.M. Caseli and M.G.V. Nunes, editors, *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*, 3 edition, book chapter 30. BPLN.

Eduardo Garcia, Nádia Silva, Juliana Gomes, Hidelberg O. Albuquerque, Ellen Souza, Felipe Siqueira, Eliomar Lima, and André C.P.L.F. de Carvalho. 2024. RoBERTaLexPT: um modelo RoBERTa jurídico pré-treinado com deduplicação para língua Portuguesa. *Linguamática*, 16(2).

Rafael P. Gouveia, André C.P.L.F. de Carvalho, Ellen Souza, Hidelberg O. Albuquerque, Douglas Vitório, and Nádia F.F. da Silva. 2025. Applying Active Learning in Named Entity Recognition Corpora Expansion in Legal Domain. *Conference on Digital Government Research (DGO2025)*.

Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Aditya K, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, Dmitry Talisman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory Dickinson, Haggai Porat, Jason Hegland, and 21 others. 2023. LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, volume 36.

Gabriel M.C. Guimarães, Felipe X. B. da Silva, Lucas A. B. Macedo, Victor H. F. Lisboa, Ricardo M. Marcacini, Andrei L. Queiroz, Vinicius R. P. Borges, Thiago P. Faleiros, and Luis P. F. Garcia. 2024. Legal document segmentation and labeling through named entity recognition approaches. *Journal of Information and Data Management*, 15(1).

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In *27th International Conference on Computational Linguistics: System Demonstrations*. Association for Computational Linguistics.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *18th International Conference on Machine Learning (ICML '01)*.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural Architectures for Named Entity Recognition. In *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Pedro H. Luz de Araujo, Teófilo E. de Campos, Renato R. R. de Oliveira, Matheus Stauffer, Samuel Couto, and Paulo Bermejo. 2018. LeNER-Br: a Dataset for Named Entity Recognition in Brazilian Legal Text. In *13th Computational Processing of the Portuguese Language (PROPOR 2018)*.

Rafael O. Nunes. 2025. Data contamination in specialized named entity recognition corpora. Master's thesis, Federal University of Rio Grande do Sul.

Rafael O. Nunes, André S. Spritzer, Carla M.D.S. Freitas, and Dennis G. Balreira. 2024. Reconhecimento de Entidades Nomeadas e Vazamento de Dados em Textos Legislativos: Uma Reavaliação da Literatura. *Linguamática*, 16(2).

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of EMNLP 2014*.

Diana Santos and Nuno Cardoso. 2006. A Golden Resource for Named Entity Recognition in Portuguese. In *7th International Workshop Computational Processing of the Portuguese Language (PROPOR 2006)*.

Felipe A Siqueira, Douglas Vitório, Ellen Souza, José A.P. Santos, Hidelberg O. Albuquerque, Marcio S. Dias, Nádia F.F. Silva, André C.P.L.F. de Carvalho, Adriano L.I. Oliveira, and Carmelo Bastos-Filho. 2025. Ulysses Tesemõ: a new large corpus for Brazilian legal and governmental domain. *Language Resources and Evaluation*, 59(2).

R. Sjansen. 2022. *Named Entity Recogniton of Legislation References*. Ph.D. thesis, Tilburg University.

Firdaus Solihin, Indra Budi, Rizal Fathoni Aji, and Edmon Makarim. 2021. Advancement of information extraction use in legal documents. *International Review of Law, Computers & Technology*, 35(3).

Ellen Souza, Hidelberg O. Albuquerque, Nádia F.F. Silva, Matheus Cerqueira, André C.P.L.F. de Carvalho, and Adriano L.I. Oliveira. 2024. PLN no Direito - REN: Reconhecimento de Entidades Nomeadas no Domínio Legal: um Panorama para a Língua Portuguesa. In H.M. Caseli and M.G.V. Nunes, editors, *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*, 3 edition, book chapter 31. BPLN.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In *Proceedings of BRACIS 2020*.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of 57th Annual Meeting of the ACL*.

Tamás Váradi, Svetla Koeva, Martin Yamalov, Marko Tadić, Bálint Sass, Bartłomiej Nitoń, Maciej Ogrodniczuk, Piotr Pęzik, Verginica Barbu Mititelu, Radu Ion, Elena Irimia, Maria Mitrofan, Vasile Păiș, Dan Tufiș, Radovan Garabík, Simon Krek, Andraz Repar, Matjaž Rihtar, and Janez Brank. 2020. The MARCELL Legislative Corpus. In *12th Language Resources and Evaluation Conference*.

Luciano Zanuz and Sandro José Rigo. 2022. Fostering judiciary applications with new fine-tuned models for legal named entity recognition in portuguese. In *15th Computational Processing of the Portuguese Language (PROPOR 2022)*.