



IA para o Domínio Legal Desafios e oportunidades

Profa. Dra. Nádia Félix

NÁDIA FÉLIX



- **Graduação em Ciência da Computação - UFG**
- **Mestre em Ciência da Computação - UFU**
- **Doutorado em Ciência da Computação ICMC-USP.**
- **Pós-doutorado - Tema: IA no contexto legislativo**
 - **Câmara Legislativa/ICMC-USP - André Ponce**
- **Professora/Pesquisadora UFG/CEIA**
- **Bolsista de Produtividade em Pesquisa FAPEG/CNPq - Nível C**
- **"Corredora nos finais de semana."**

Desafios da Justiça Brasileira



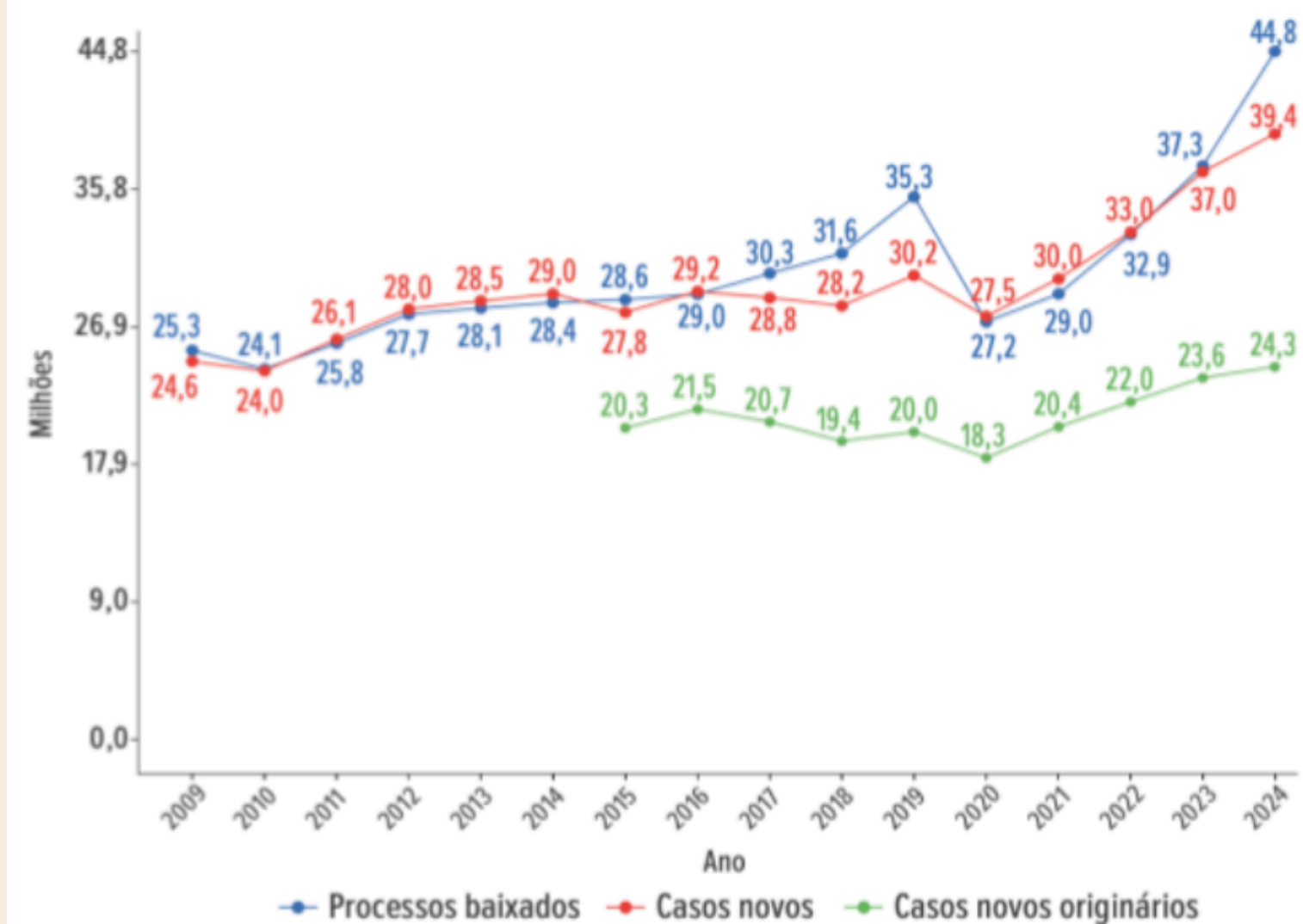
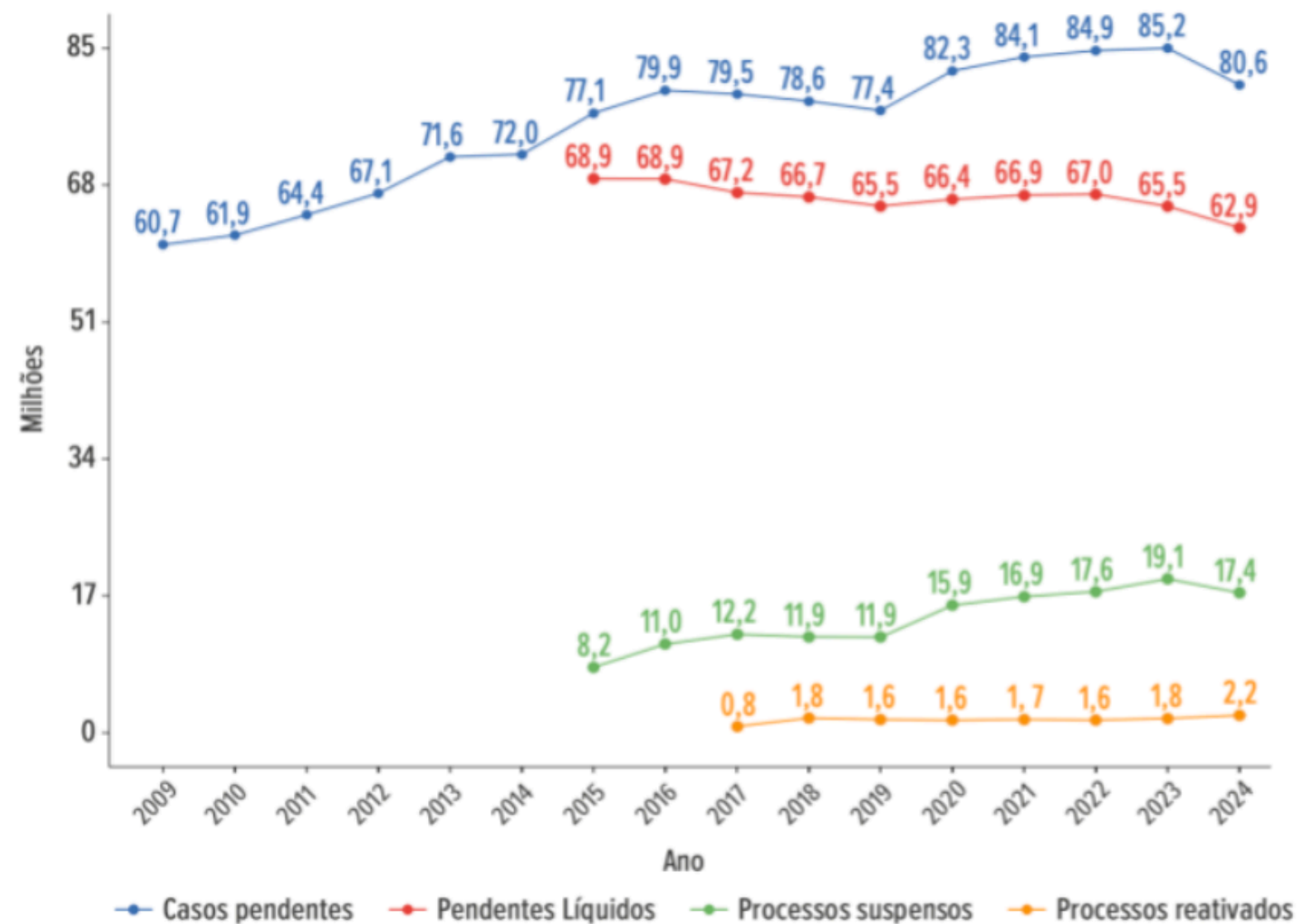


Fig. 1. Série histórica dos casos pendentes no Poder Judiciário. Dados obtidos através do relatório do Justiça em Números*.

Fig. 2. Série histórica dos casos novos e processos baixados (finalizados) no Poder Judiciário. Dados obtidos através do relatório do Justiça em Números*.

* <https://www.cnj.jus.br/wp-content/uploads/2025/11/justica-em-numeros-2025.pdf>.

Digitalização: A grande maioria desses textos é eletrônica. Em 2023, 90,6% dos processos em tramitação eram digitais e 99,6% dos casos novos ingressaram eletronicamente.

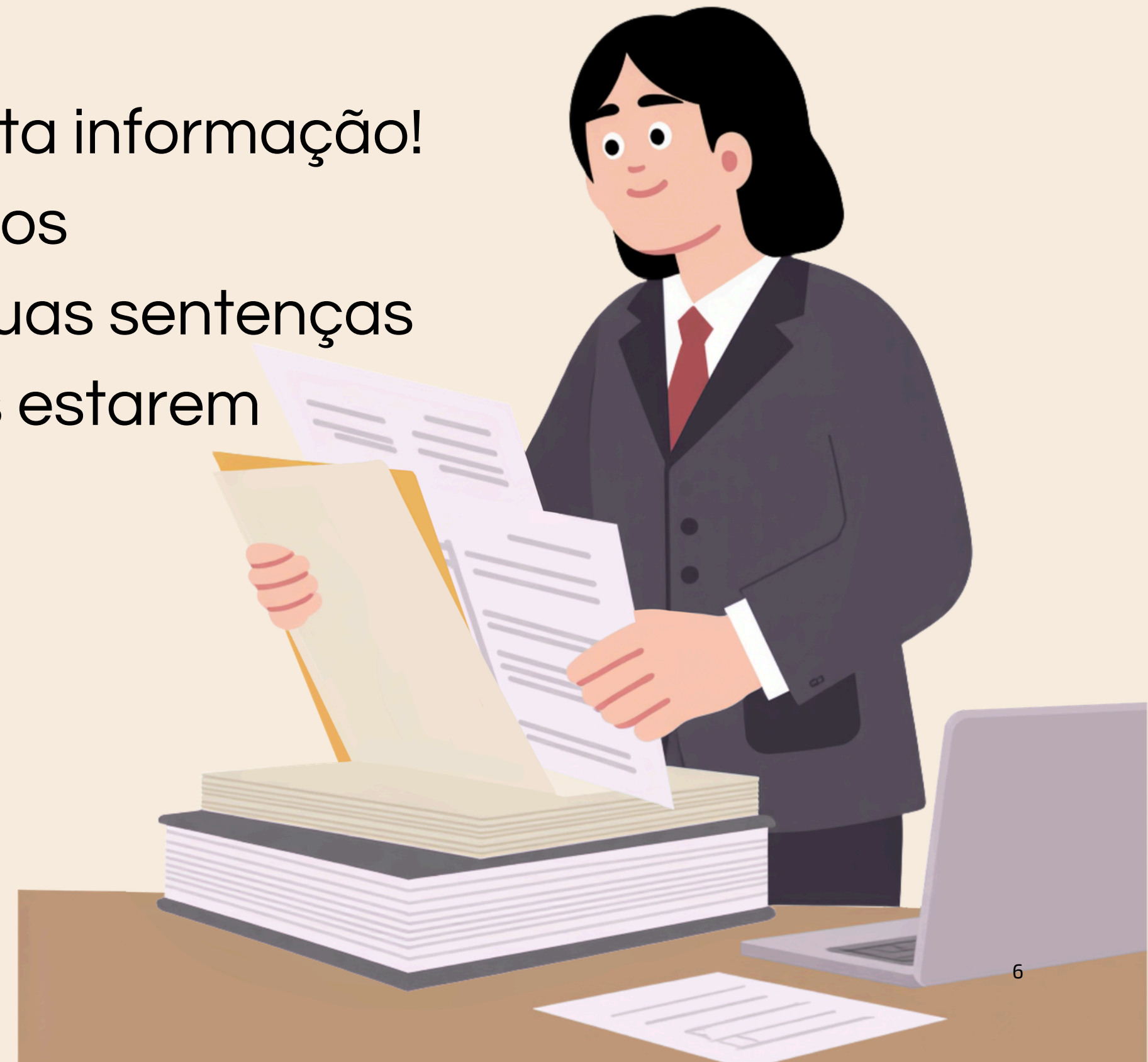
O programa Justiça 4.0, focado na digitalização e uso de inteligência artificial (IA) no Poder Judiciário brasileiro, tem gerado um volume massivo de documentos e automações para lidar com um estoque de cerca de 80 milhões de processos pendentes.

The image shows a screenshot of the CNJ (Conselho Nacional de Justiça) website. The browser address bar shows the URL: cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/. The page header features the CNJ logo and the text 'CONSELHO NACIONAL DE JUSTIÇA'. A search bar with the placeholder text 'Pesquisar...' is located on the right side of the header. Below the header is a navigation menu with the following items: 'O CNJ', 'Gestão da Justiça', 'Programas e Ações', 'Publicações e Pesquisas', 'Sistemas e Serviços', and 'Comunicação e Eventos'. The main content area has a breadcrumb trail: 'Início » Tecnologia da Informação e Comunicação » Justiça 4.0'. The title 'Justiça 4.0' is prominently displayed. On the left side, there is a sidebar menu under the heading 'Institucional' with the following links: 'Sobre o Programa', 'Painel de Monitoramento à Integração', and 'Painel de Monitoramento de'. The main content area features a large graphic with the 'JUSTIÇA 4.0' logo and an illustration of a diverse group of people.

Tomadores de decisão

- Precisam estar familiarizados com muita informação!
- Precisam conhecer as leis, regulamentos
- Precisam rever casos semelhantes e suas sentenças
- Precisam formular argumentos e estes estarem consistentes com os precedentes

AI em Ação!



TEXTOS JURÍDICOS

"Juridiquês"

- Textos Legislativos
- Corpus/Corpora específicos para o português brasileiro e o domínio jurídico.

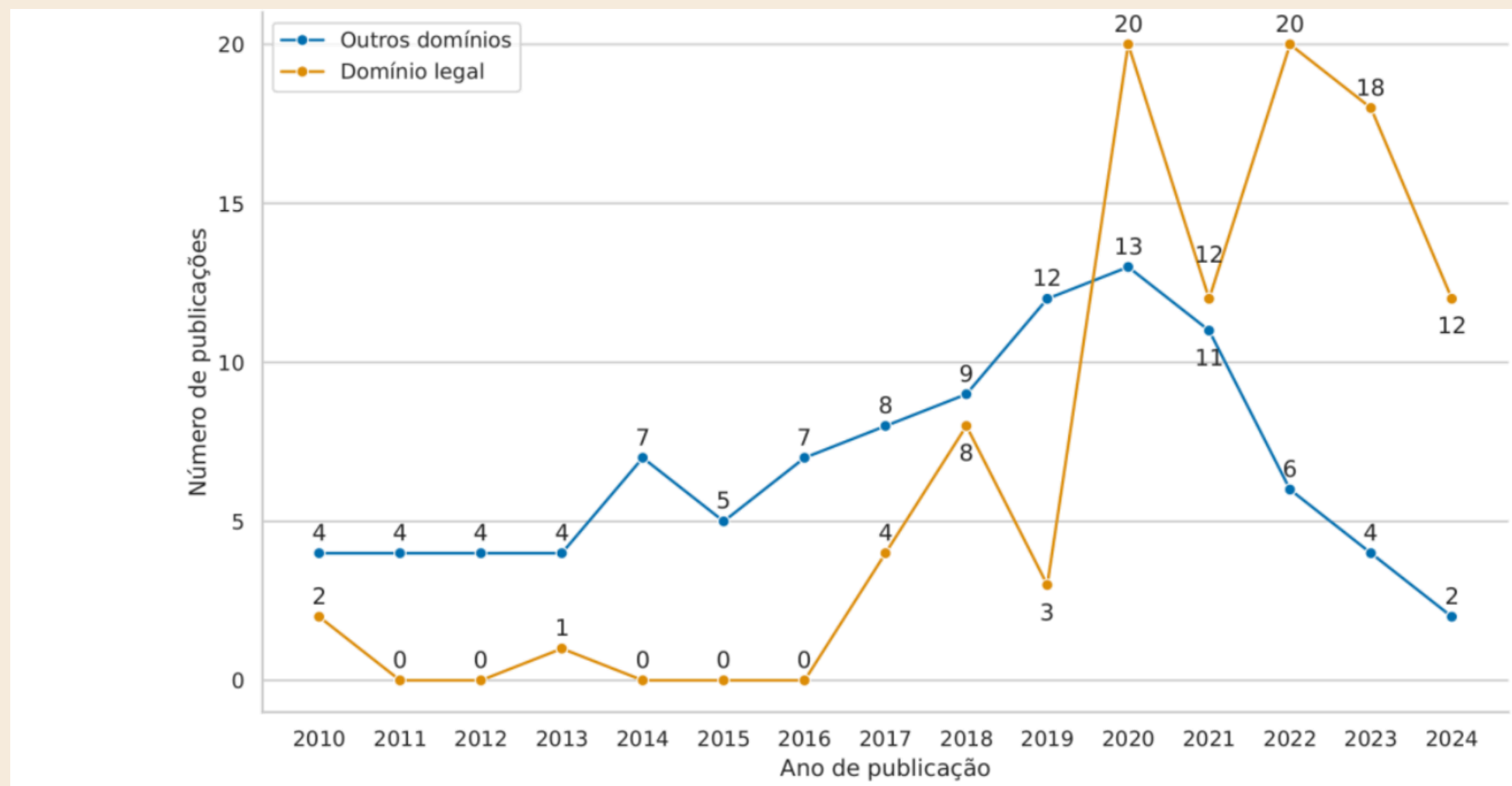


PROJETO ULYSSES



- Brazilian Portuguese Legislative Documents
 - <https://github.com/ulysses-camara>
 - Stance Detection
 - Sumarização textual
 - Segmentação Brazilian legislative text
 - Reconhecimento de Entidades Nomeadas
 - Classificação de textos
 - Modelagem de tópicos
 - Agrupamento de textos

Amostra da evolução do estado da arte de Reconhecimento de Entidades Nomeadas-Legal*



*<https://brasileiraspln.com/livro-pln/>

Corpora REN-Legais em língua portuguesa com acesso público (PB) ou privado (PV)

- LeNER-Br / Araujo et al. ([2018](#)) - PB
- DOU-corpus / Alles ([2018](#)) - PB
- [Castro \(2019\) - Domínio Trabalhista - PV](#)
- DrugSeizures-Br / Bonifacio et al. ([2020](#)) - PV
- KauaneJunior / Fernandes et al. ([2020b](#)) - PV
- Fernandes et al. ([2020a](#)) - PB
- Rodríguez; Bezerra ([2020](#)) - PB
- Aposentadoria / Neto; Faleiros ([2021](#)) - PB
- Lima; Borges ([2022](#)) - PB
- [UlyssesNER-Br / Albuquerque et al. \(2022\), Costa et al. \(2022\) - PB](#)
- [UniLegalNER-Br - Albuquerque et al. \(2026\) PB](#)

CORPORA ESPECÍFICO DE DOMÍNIO

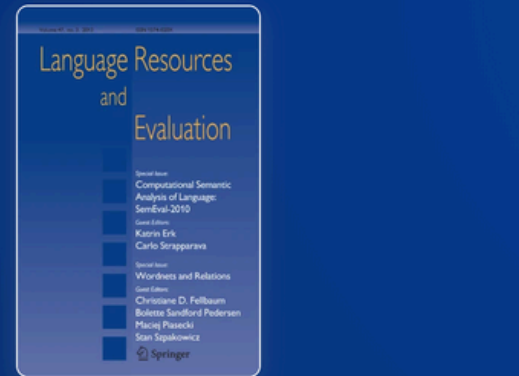
Home > Language Resources and Evaluation > Article

Ulysses Tesemõ: a new large corpus for Brazilian legal and governmental domain

Project Note | Published: 18 July 2024

Volume 59, pages 1685–1704, (2025) [Cite this article](#)

[Save article](#)



[Language Resources and Evaluation](#)

[Aims and scope](#) →

[Submit manuscript](#) →

[Access this article](#)

[Log in via an institution](#) →

Data categories

The Tesemõ Corpus is organized into 30 data categories, listed below:

ID	Source name	Files	Volume
J1	Court journals	1,014,636	12 GiB
J2	CNJ's public documents	4,317	58 MiB
L1	Legislative journals	10,014	8.2 GiB
L2	State legislation	314,461	1.6 GiB
L3	Federal legislative bills	165,448	829 MiB
L4	Speech transcriptions from Chamber of Deputies	310,057	545 MiB
L5	Federal Senate's transcriptions	979	173 MiB
L6	Chamber of Deputies committee materials	309	45 MiB
L7	State constitutions	26	11 MiB
L8	TSE's normative instructions	502	8.1 MiB
L9	Legal codes	19	6.7 MiB
L10	Federal District organic law	116	2.7 MiB
L11	Governmental standing rules	6	1.9 MiB
L12	Statutes	18	1.5 MiB

SPRINGER NATURE Link

[Login](#)

[Find a journal](#)

[Publish with us](#)

[Track your research](#)

[Search](#)

[Saved research](#) [Cart](#)

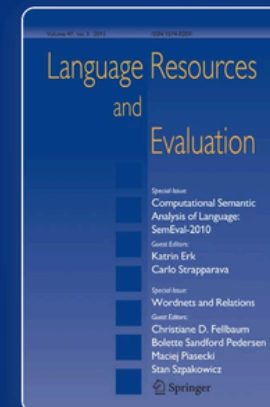
Home > Language Resources and Evaluation > Article

Building a relevance feedback corpus for legal information retrieval in the real-case scenario of the Brazilian Chamber of Deputies

Original Paper | Published: 18 August 2024

Volume 59, pages 1257–1277, (2025) [Cite this article](#)

[Save article](#)



[Language Resources and Evaluation](#)

[Aims and scope](#) →

[Submit manuscript](#) →

ulysses-camara/ulysses-tesemo



Ulysses Tesemõ corpus: Brazilian corpus for judicial, legislative, and governmental textual data, scraped from more than 150 sources.

1 Contributor 0 Issues 8 Stars 1 Fork

[ulysses-camara/ulysses-tesemo: Ulysses Tesemõ corpus: Brazilian corpus for judicial, legislative, and governmental textual data, scraped...](#)

Ulysses Tesemõ corpus: Brazilian corpus for judicial, legislative, and governmental textual data, scraped from more than 150 sources. - ulysses-camara/ulysses-tesemo

[GitHub](#)

TREINAMENTO DE MODELOS

Model	LeNER	UlyNER-PL Coarse/Fine	FGV-STF Coarse	RRIP	Average (%)
BERTimbau _{base} (Souza et al., 2020)	88.34	86.39/83.83	79.34	82.34	83.78
BERTimbau _{large} (Souza et al., 2020)	88.64	87.77/84.74	79.71	83.79	84.60
Albertina-PT-BR _{base} (Rodrigues et al., 2023)	89.26	86.35/84.63	79.30	81.16	83.80
Albertina-PT-BR _{xlarge} (Rodrigues et al., 2023)	90.09	88.36/ 86.62	79.94	82.79	85.08
BERTikal _{base} (Polo et al., 2021)	83.68	79.21/75.70	77.73	81.11	79.99
JurisBERT _{base} (Viegas et al., 2022)	81.74	81.67/77.97	76.04	80.85	79.61
BERTimbauLAW _{base} (Viegas et al., 2022)	84.90	87.11/84.42	79.78	82.35	83.20
Legal-XLM-R _{base} (Niklaus et al., 2023)	87.48	83.49/83.16	79.79	82.35	83.24
Legal-XLM-R _{large} (Niklaus et al., 2023)	88.39	84.65/84.55	79.36	81.66	83.50
Legal-RoBERTa-PT _{large} (Niklaus et al., 2023)	87.96	88.32/84.83	79.57	81.98	84.02
RoBERTaTimbau _{base}	89.68	87.53/85.74	78.82	82.03	84.29
RoBERTaLegalPT _{base}	90.59	85.45/84.40	79.92	82.84	84.57
RoBERTaLexPT _{base}	90.73	88.56/86.03	80.40	83.22	85.41

Table 8: Macro F1-Score (%) for multiple models evaluated on PortuLex benchmark test splits.

PROJETO IAIA

UlyssesNER-Br

7 categorias, 17 entidades

Category	Type	Description	Example
DATA (Date)	—	Date	01 de janeiro de 2020
EVENTO (Event)	—	Event	Eleições de 2018
FUNDAMENTO (Law foundation)	FUNDlei	Legal norm	Lei no 8.666, de 21 de junho de 1993
	FUNDapelido	Legal norm nickname	Estatuto da Pessoa com Deficiência
	FUNDprojetoilei	Bill	PEC 187/2016
	FUNDSolicitacaotrabalho	Legislative consultation	Solicitação de Trabalho n° 3543/2019
LOCAL (Location)	LOCALconcreto	Concrete place	Niterói-RJ
	LOCALvirtual	Virtual place	Jornal de Notícias
ORGANIZAÇÃO (Organization)	ORGpartido	Political party	PSB
	ORGgovernamental	Governmental organization	Câmara do Deputados
	ORGnãogovernamental	Non-governmental organization	Conselho Reg. de Medicina (CRM)
PESSOA (Person)	PESSOAindividual	Individual	Jorge Sampaio
	PESSOAgрупoid	Group of individuals	Família Setúbal
	PESSOAcargo	Occupation	Deputado
	PESSOAgрупocargo	Group of occupations	Parlamentares
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



PROJETO IAIA

PL-corpus

PESSOAindividual Afonso Florence) Altera a **FUNDlei** Lei nº 6.088, de 16 de julho de 1974, que dispõe sobre a criação de uma **ORGgovernamental** Companhia de Desenvolvimento do Vale do São Francisco - Codevasf, localizada na **LOCALconcreto** Bacia do Rio Paraguaçu entre suas áreas de atuação, nos termos que especifica as providências.

C-corpus

LOCALconcreto No **Brasil**, **PESSOAgрупocargo** super-ricos e banqueiros não pagam imposto. Classe média e pobres pagam impostos na tributação de seus salários e especialmente sobre produtos alimentícios da cesta básica. Lamentavelmente, o **ORGpartido** PT traiu os trabalhadores e se aliou aos **PESSOAgрупocargo** banqueiros. Por que acham que **PESSOAindividual** Joseph Safra, as famílias **PESSOAgрупoind** Setúbal, **PESSOAgрупoind** Vilella e **PESSOAgрупoind** Moreira Salles não pagam altos impostos? Por qual **LOCALconcreto** razão os dividendos ainda não são tributados no **Brasil**? Cadê o imposto sobre

ST-corpus

PESSOAcargo O parlamentar solicita a gentileza em elaborar consulta técnica alterando o **FUNDapelido** Código de Defesa do Consumidor (CDC), no sentido de permitir que a loja que efetuar a venda de um produto, e o mesmo após dentro do prazo de garantia, a custear o reparo ou troca, e o

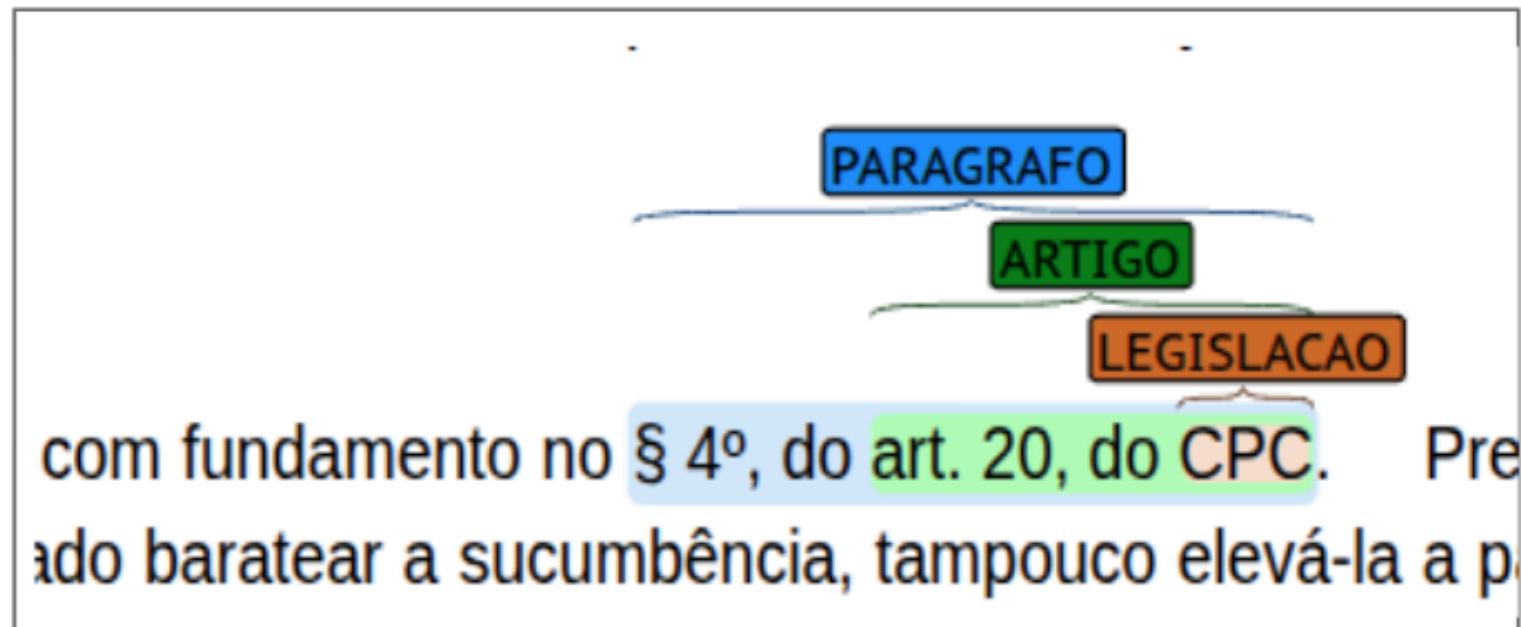


Entidades Aninhadas, Sobrepostas e Não-Contíguas

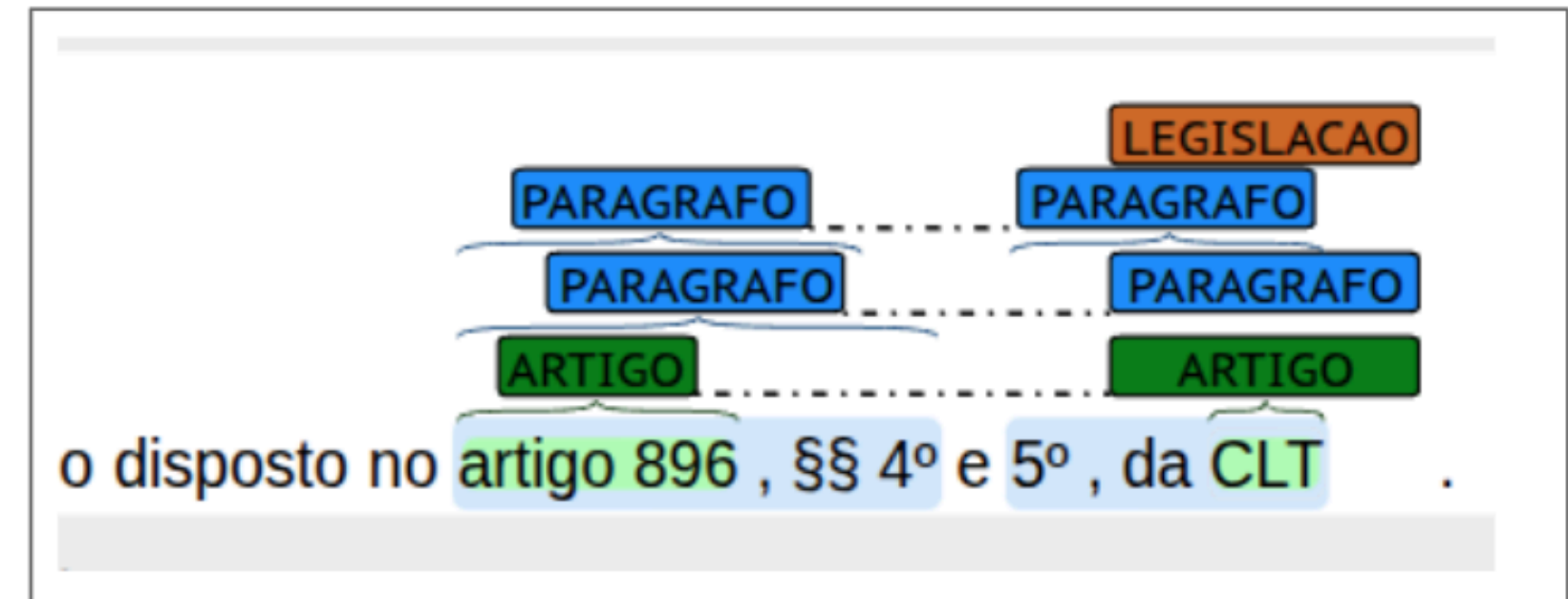
- Nested (Aninhadas):
 - Frase: "O juiz da 2ª Vara Criminal do Tribunal de Justiça de São Paulo proferiu a sentença."
 - Entidades: "Tribunal de Justiça de São Paulo" (organização) está aninhado dentro de "2ª Vara Criminal do Tribunal de Justiça de São Paulo" (jurisdição).
- Overlapped (Sobrepostas):
 - Frase: "O Ministério Público Federal e Estadual realizaram a investigação."
 - Entidades: "Ministério Público Federal" e "Ministério Público Estadual" (ambas organizações) sobrepõem-se na expressão "Ministério Público".
- Non-Contiguous (Não-Contíguas):
 - Frase: "O presidente e os ministros do Supremo Tribunal Federal se reuniram."
 - Entidades: "presidente" e "ministros" (papéis) estão relacionados à entidade "Supremo Tribunal Federal" (organização), mas não estão adjacentes.

Entidades Aninhadas, Sobrepostas e Não-Contíguas

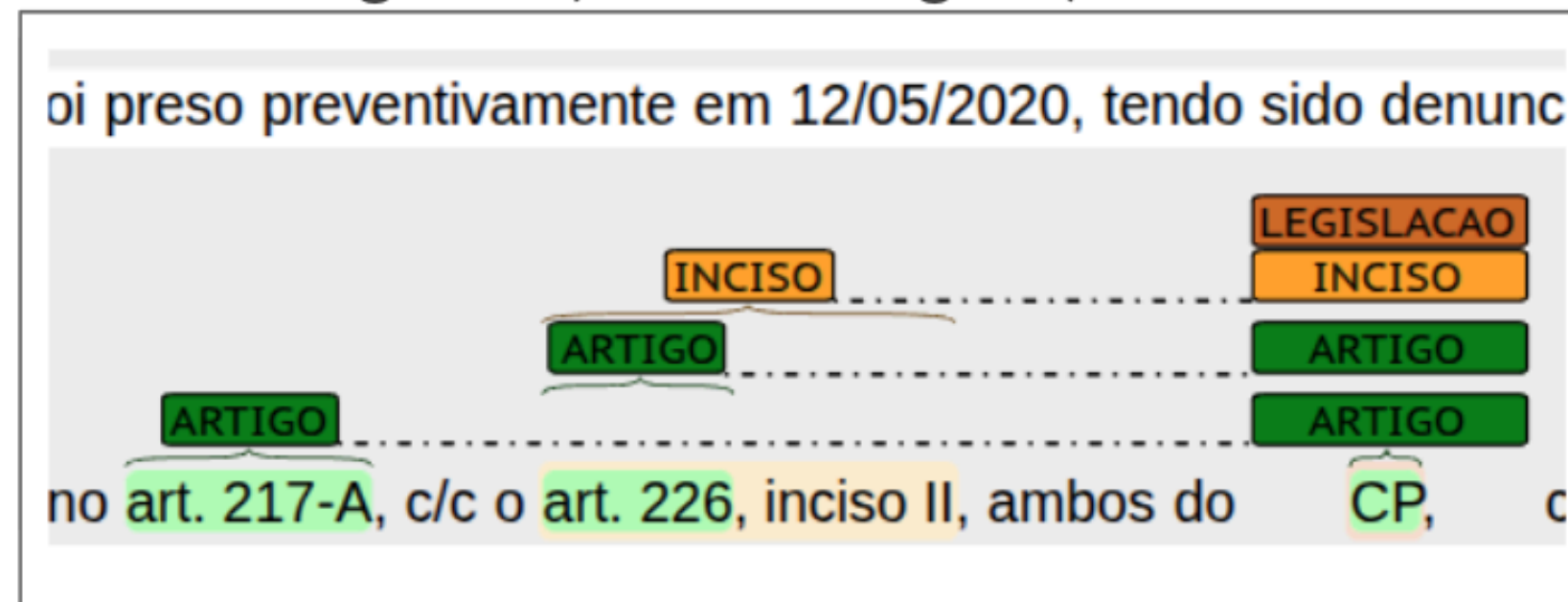
Nested (aninhadas)



Overlapped (Sobrepostas)



Non-Contiguous (Não-Contíguas)



Fase atual: Curadoria



MeLLL@UFF

313 seguidores

6 d • 🌐



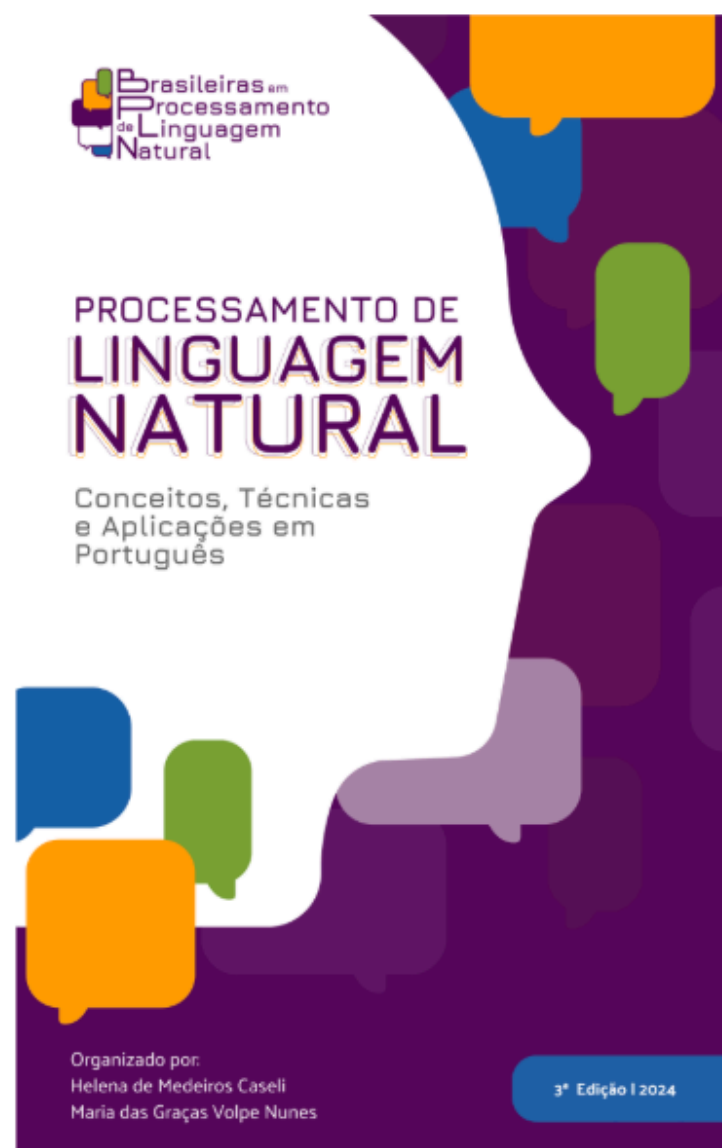
Depois do Carnaval, seguimos celebrando novas conquistas no nosso grupo 🎉.
Temos mais um trabalho aceito para o PROPOR 2026!

O artigo "LegalSim-PT: Building a Dataset for Legal Document Simplification in Portuguese Leveraging Linguistic Metrics", de [Arthur Scalercio](#), [Maria José Bocorny Finatto](#) e [Aline Paes](#), apresenta o LegalSim-PT, o primeiro dataset em larga escala voltado à simplificação de documentos jurídicos em português.

O trabalho aborda a simplificação em nível de documento, indo além da frase isolada e considerando aspectos como fluência e coerência textual. Para reduzir a dependência exclusiva de avaliação manual, o estudo combina estratégias de aumento de dados com métricas de legibilidade, similaridade semântica e diversidade na seleção dos pares de textos. Além disso, inclui uma análise detalhada das características do corpus, avaliações automáticas e humanas, e experimentos de fine-tuning que demonstram ganhos consistentes na tarefa de simplificação em nível de documento.

Parabéns aos autores por mais essa contribuição relevante para a simplificação textual em português e para o processamento de textos jurídicos 🙌

Capítulo de Livro



Domínios > 31 PLN no Direito - REN

31 PLN no Direito - REN

Reconhecimento de Entidades Nomeadas no Domínio Legal: um Panorama para a Língua Portuguesa

Ellen Souza
Hidelberg O. Albuquerque
Nádia F. F. Silva
Matheus Cerqueira
André C. P. L. F. de Carvalho
Adriano L. I. Oliveira

PUBLICADO EM:
20/11/2024

<https://brasileiraspln.com/>



Geração textual e os problemas....

conjur.com.br/2023-nov-12/cnj-vai-investigar-juiz-que-usou-tese-inventada-pelo-chatgpt-para-escrever-decisao

EU, ROBÔ

CNJ vai investigar juiz que usou tese inventada pelo ChatGPT para escrever decisão

12 de novembro de 2023, 14h10

Judiciário Tecnologia

O CNJ (Conselho Nacional de Justiça) vai investigar uma sentença assinada por um juiz federal da 1ª Região que foi, na verdade, feita por inteligência artificial, a partir do uso do *ChatGPT*.

O caso poderia ter passado batido não fosse o fato de a inteligência artificial ter inventado, para basear a decisão, uma jurisprudência do Superior Tribunal de Justiça que não existe. Por causa disso, o advogado derrotado percebeu a fraude e acionou a Corregedoria Regional de Justiça Federal da 1ª Região.

Caso foi descoberto porque decisão se baseou em falsa jurisprudência do STJ

uol.com.br/tilt/noticias/redacao/2023/05/28/advogado-chatgpt.htm

UOL Jogos Brasil dos Privilégios Canal UOL Colunas SAC EMAIL ENTRE ASSINE UOL

Deu ruim: advogado usou ChatGPT e chatbot inventou casos que não existem



PUBLICIDADE



Fale com a John Deere mais próxima de você.

John Deere



Decisão que absolveu estupro de menina de 12 anos tem trecho de IA

Desembargador Magid Nauef Láuar, que reverteu condenação em 1ª instância por estupro de vulnerável, manteve frase que pedia para ferramenta melhorar "exposição e fundamentação".

 G1 / Feb 24



Tribunal de Justiça de Minas Gerais

A corré A.L.S., ora segunda apelante, em razão da condição de genitora da menor M.L.S., que teria sido vítima do crime previsto no art. 217-A do Código Penal, também foi denunciada pela prática do referido delitivo na modalidade omissiva, dada a sua posição de garantidora da ofendida, à luz das disposições contidas no art. 13, § 2º, do Código Penal.

Agora melhore a exposição e fundamentação deste parágrafo:

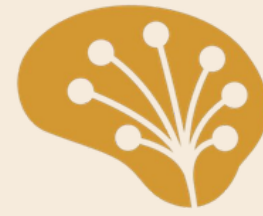
Os problemas são os humanos...



CEIA
CENTRO DE EXCELÊNCIA EM
INTELIGÊNCIA ARTIFICIAL



H.IAAC | HUB DE INTELIGÊNCIA
ARTIFICIAL E ARQUITETURAS
COGNITIVAS



FAPEG
Fundação de Amparo à Pesquisa
do Estado de Goiás



**OBRIGADA
POR SUA
atenção!**



nadia.felix@ufg.br

Referências:

- 1) <https://www.cnj.jus.br/programa-justica-4-0-divulga-resultados-de-pesquisa-sobre-ia-no-judiciario-brasileiro>
- 2) <https://portal.stf.jus.br>
- 3) <https://github.com/ulysses-camara>
- 4) <https://github.com/peluz/lener-br>
- 5) ALBUQUERQUE, H. O. et al. Named entity recognition: a survey for the portuguese language. Procesamiento del Lenguaje Natural, a2023.
- 6) ARAUJO, P. H. L. DE et al. LeNER-Br: A Dataset for Named Entity Recognition in Brazilian Legal Text. (A. Villavicencio et al., Eds.) Computational Processing of the Portuguese Language. Anais...Cham: Springer International Publishing, 2018. Disponível em: <<https://github.com/peluz/lener-br>>
- 7) CASTRO, P. Aprendizagem profunda para reconhecimento de entidades nomeadas em domínio jurídico. mathesis—[s.l.] Master's thesis, Universidade Federal de Goiás, 2019.