



ARTIGO

INTELIGÊNCIA ARTIFICIAL E ECOSSISTEMA JURÍDICO: INICIATIVAS EM BENEFÍCIO DA SOCIEDADE

POR

Nádia Félix F. da Silva, Hidelberg O. Albuquerque e André C. P. L. F. de Carvalho
nadia.felix@ufg.br, hidelberg.albuquerque@ufrpe.br e andre@icmc.usp.br

O sistema judicial brasileiro enfrenta um gargalo operacional histórico de gerir um acervo processual que desafia a capacidade analítica humana. O Conselho Nacional de Justiça¹ (CNJ) mantém um relatório anual chamado Justiça em Números [8] no qual consolida os indicadores de litigiosidade nacional. Conforme a série histórica de casos pendentes (Figura 1.1-a), ao final de 2024, 80,6 milhões de processos continuavam sem resolução. Neste mesmo ano, o volume de processos baixados superou o ingresso de

¹ <http://cnj.jus.br>

novos casos em 7,5 milhões (Figura 1.1-b). Embora a produtividade tenha atingido 44,6 milhões de processos julgados, um aumento de 28,2% (9,8 milhões de casos) em relação a 2023, a disparidade frente ao estoque acumulado permanece crítica, evidenciando a necessidade de métodos escalonáveis de automação e análise.

Para enfrentar o desafio de atender a maior demanda por serviços, o Judiciário² institucionalizou um Programa de inovação e efetividade na realização da Justiça para todos, com a finalidade

² Incluímos aqui também o domínio Legislativo, usando a nomenclatura “Domínio Jurídico ou Legal”.

de promover o acesso à Justiça, através do desenvolvimento de projetos colaborativos de ecossistemas baseados em Inteligência Artificial (IA). Utilizando os conceitos de Justiça como um Serviço (Justice as a Service), o programa atua como um vetor de transformação digital, aproximando ainda mais esse Poder das necessidades da sociedade e ampliando o acesso à Justiça. Estas inovações buscam dar celeridade à prestação processual e reduzir despesas orçamentárias. Essa iniciativa promoveu um rol de soluções digitais cujas medidas têm sido adotadas pelo Poder Judiciário em um ritmo acelerado desde 2020 [8]. A IA, em particular o Processamento de Linguagem Natural (PLN), deixa de ser apenas uma ferramenta de automação para atuar estrategicamente como suporte à decisão e à equidade³.

³ <https://www.cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/>

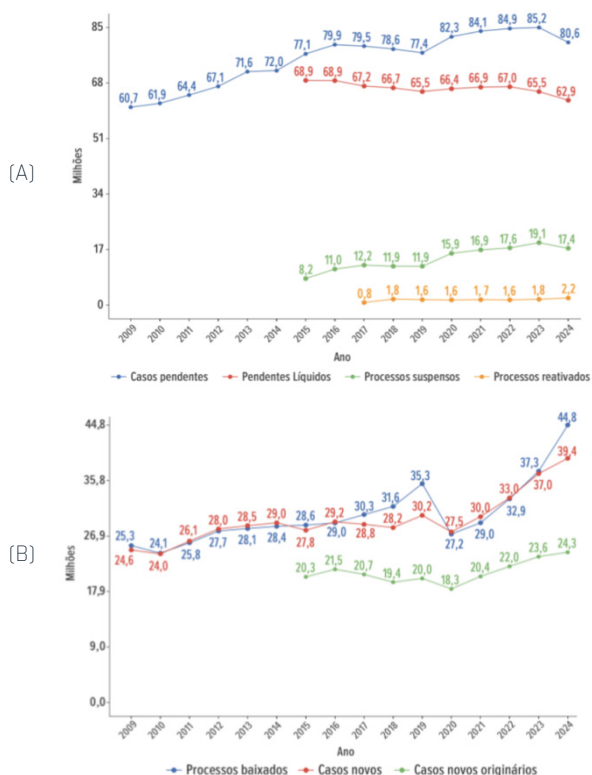


FIG. 01 | SÉRIES HISTÓRICAS DO PODER JUDICIÁRIO: (A) CASOS PENDENTES E (B) FLUXOS DE PROCESSOS NOVOS E BAIXADOS. FONTE: RELATÓRIO JUSTIÇA EM NÚMEROS

O levantamento anual do CNJ revelou um aumento significativo de projetos de IA no Poder Judiciário em 2023⁴ e 2024⁵. Em comparação com 2023, 63% dos tribunais participaram no desenvolvimento de soluções baseadas em IA, em um crescimento de 27% (de 140 para 178 soluções mapeadas), incluindo modelos de IA Generativa, na busca contínua pela celeridade processual. Em 2018, uma parceria com a Universidade de Brasília e o Supremo Tribunal Federal (STF) resultou no projeto “Victor” [11], que aplica IA na classificação de processos e identificação de temas recorrentes. Em 2022, o STF lançou a ferramenta “RAFA 2030”, para apoiar a classificação de processos de acordo com a Agenda 2030 da Organização das Nações Unidas (ONU). Em 2023, buscando automatizar a identificação e o agrupamento de processos por assunto na Corte, foi lançada a ferramenta de IA “VitorIA”⁶. Em 2026, o CNJ lançou o “Promptus”⁷, um repositório nacional de *prompts* de IA padronizados e integrados à Plataforma Digital do Poder Judiciário (PDPJ), visando aumentar a eficiência operacional e a segurança jurídica em todo o Judiciário brasileiro.

O domínio jurídico ou legal abrange textos diversos, como legislação e jurisprudência [17]. A natureza destes documentos impõem desafios às aplicações de PLN, pois esses documentos são tipi-

⁴ <https://www.cnj.jus.br/programa-justica-4-0-divulga-resultados-de-pesquisa-sobre-ia-no-judiciario-brasileiro/>

⁵ www.cnj.jus.br/wp-content/uploads/2025/09/relatorio-entregas-do-programa-justica-4-0-versao-interativa.pdf

⁶ <https://portal.stf.jus.br>

⁷ <https://www.cnj.jus.br/tecnologia-da-informacao-e-comunicacao/justica-4-0/conheca-o-conecta/promptus/>

camente extensos, ruidosos e repletos de jargão técnico [12]. Existem iniciativas em IA e PLN na comunidade científica para estes desafios, tais como Recuperação de Informação Legal [7,20], Sumarização de Texto [13], Previsão de Julgamentos [6], Segmentação Semântica [16] e Reconhecimento de Entidades Nomeadas [2, 4,5,14].

Em 2019, a Câmara dos Deputados lançou o “Ulysses”, um conjunto de iniciativas de IA para ampliar a transparência e o suporte legislativo por meio de análises complexas [3]⁸. Esses novos algoritmos permitiram a otimização de tarefas como a busca por documentos similares

[1,21], análise do posicionamento de cidadãos sobre projetos de lei [18, 15], reconhecimento de entidades nomeadas para recuperação de informação [1, 9], além da construção de modelos de linguagem como o RoBERTaLexPT [10] e a criação do “Ulysses Tesemō” [19], um grande corpus composto por mais de 3,5 milhões de arquivos, com jurisprudência, leis e trabalhos acadêmicos.

É importante salientar que a tecnologia deve ser eticamente orientada, que as decisões automatizadas precisam respeitar os direitos fundamentais e a inovação deve promover equidade, inclusão e justiça social, e não apenas eficiência técnica.

⁸ <https://github.com/ulysses-camara>

Referências

1. Albuquerque et al. (2022). UlyssesNER-Br: A Corpus of Brazilian Legislative Documents for Named Entity Recognition. Proceedings of the 14th International Conference on Computational Processing of Portuguese (PROPOR 2022).
2. Alles, V. (2018). Construção de um corpus para extrair entidades nomeadas do diário oficial da união utilizando aprendizado supervisionado. Master's thesis UNB.
3. Almeida, P. G. R. (2021). Uma jornada para um Parlamento inteligente: Câmara dos Deputados do Brasil. Red Información, 24.
4. Brito et al. (2023). CDJUR-BR - Uma Coleção Dourada do Judiciário Brasileiro com Entidades Nomeadas Refinadas. In Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana.
5. Castro, P. (2018). Deep learning for named entity recognition in legal domain. Master's thesis, Universidade Federal de Goiás.
6. Chalkidis, I., Androutsopoulos, I., and Aletas, N. (2019). Neural Legal Judgment Prediction in English. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
7. Chalkidis et al. (2021). Regulatory Compliance through Doc2Doc Information Retrieval: A case study in EU/UK legislation where text similarity has limitations. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics.
8. Justiça em números 2025. <https://www.cnj.jus.br/wp-content/uploads/2025/11/justica-em-numeros-2025.pdf>. Último acesso: 12/01/2026.
9. Costa et al. (2022). Expanding UlyssesNER-Br Named Entity Recognition Corpus with Informal User-Generated Text. Proceedings of the 21st EPIA Conference on Artificial Intelligence (EPIA 2022).
10. Garcia et al. (2024). RoBERTaLexPT: A Legal RoBERTa Model pretrained with deduplication for Portuguese. In Proceedings of the 16th International Conference on Computational Processing of Portuguese (PROPOR 2024).
11. Hartmann Peixoto, F. (2020). Projeto Victor: relato do desenvolvimento da Inteligência Artificial na Repercussão Geral do Supremo Tribunal Federal. Revista Brasileira de Inteligência Artificial e Direito - RBIAD, 1(1):1–22.
12. Kapoor et al. (2022). HLDC: Hindi legal documents corpus. In Findings of the Association for Computational Linguistics: ACL

2022.

13. Kornilova, A. and Eidelman, V. (2019). BillSum: A Corpus for Automatic Summarization of US Legislation. In Proceedings of the 2nd Workshop on New Frontiers in Summarization.
14. Luz de Araujo et al. (2018). LeNER-Br: A Dataset for Named Entity Recognition in Brazilian Legal Text. In 13th Computational Processing of the Portuguese Language (PROPOR 2018).
15. Maia et al. (2022). UlyssesSD-Br: Stance Detection in Brazilian Political Polls. Proceedings of the 21st EPIA Conference on Artificial Intelligence (EPIA 2022).
16. Malik et al. (2021). ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics , pages 4046–4062. Association for Computational Linguistics.
17. Maxwell, K. T. and Schafer, B. (2008). Concept and context in legal information retrieval. In Legal Knowledge and Information Systems, pages 63–72. IOS Press.
18. Silva et al. (2021). Evaluating topic models in portuguese political comments about bills from brazil's chamber of deputies. In Britto, A. and Valdivia Delgado, K., editors, Intelligent Systems, pages 104–120, Cham. Springer International Publishing.
19. Siqueira et al. (2024). Ulysses tesemō: a new large corpus for brazilian legal and governmental domain. Language Resources and Evaluation, pages 1–20.
20. Souza et al. (2021). An information retrieval pipeline for legislative documents from the brazilian chamber of deputies. In Legal Knowledge and Information Systems, pages 119–126. IOS Press.
21. Souza et al. (2023). Building a relevance feedback corpus for legal information retrieval in the real-case scenario of the brazilian chamber of deputies.



NÁDIA FÉLIX F. DA SILVA é professora do Instituto de Informática da Universidade Federal de Goiás (INF/UFG). É graduada em Ciência da Computação pela UFG, mestre em Ciência da Computação pela UFU e doutora em Ciências da Computação e Matemática Computacional pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP). Atua nas áreas de Inteligência Artificial com ênfase em Processamento de Linguagem Natural (PLN). É coautora do capítulo “PLN no Direito - REN: Reconhecimento de Entidades Nomeadas no Domínio Legal: um Panorama para a Língua Portuguesa” e outras publicações relevantes, refletindo sua atuação em PLN e aplicações no domínio jurídico.



HIDELBERG O. ALBUQUERQUE é professor da Unidade Acadêmica de Serra Talhada da Universidade Federal Rural de Pernambuco. É formado em Análise e Desenvolvimento de Sistemas pelo Instituto Federal da Paraíba, mestre em Informática pela Universidade Federal da Paraíba, e é doutorando em Ciência da Computação no Centro de Informática da Universidade Federal de Pernambuco. Atua nas áreas de Inteligência Artificial e Aprendizado de Máquina, com ênfase em Processamento de Linguagem Natural e Reconhecimento de Entidades Nomeadas, além de Sumarização e Recuperação de Informação Jurídica, incluindo pesquisas com Modelos de Linguagem de Grande Porte. Também é coautor do capítulo “PLN no Direito - REN: Reconhecimento de Entidades Nomeadas no Domínio Legal: um Panorama para a Língua Portuguesa”, além de outras publicações na área de PLN jurídica.



ANDRÉ CARLOS PL.F. DE CARVALHO é Professor Titular e Diretor do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP). Possui graduação e mestrado em Ciência da Computação pela Universidade Federal de Pernambuco e doutorado em Electronic Engineering pela University of Kent, além de livre-docência pela USP. Atua nas áreas de Inteligência Artificial, com ênfase em Aprendizado de Máquina, Mineração de Dados e Ciência de Dados. É coautor de livros de referência na área, incluindo “Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina”, vencedora do Prêmio Jabuti 2012, “A General Introduction to Data Analytics”, “Ciência de Dados, Fundamentos e Aplicações”, do capítulo “PLN no Direito - REN: Reconhecimento de Entidades Nomeadas no Domínio Legal: um Panorama para a Língua Portuguesa”, além de publicações relevantes.