

Pedro Henrique Luz de Araujo

Email : pedrohluzaraujo@gmail.com

<https://github.com/peluz/>

<https://scholar.google.com/citations?user=wqaJLc0AAAAJ>

WORK EXPERIENCE

- **Research Group Data Mining and Machine Learning** University of Vienna
Research Staff. Dec 2021 – Ongoing
- **KnEDLe - Knowledge Extraction from Documents of LEgal content** University of Brasília
Master's student associated with project. Jan 2020 – Sep 2021
- **VICTOR Project: an AI system for Brazil's supreme court** University of Brasília
Master's student associated with project. Mar 2019 – May 2019
- **R&D Centre for Public Sector Transformation (NEXT)** University of Brasília
Research intern Feb 2018 – Dec 2018

ACADEMIC EDUCATION

- **Doctoral Programme in Computer Science** University of Vienna
Topic: Aligning and Personalizing NLP Models Mar 2022 – Ongoing
- **Master of Computer Science** University of Brasília
Area of concentration: Artificial Intelligence. GPA: 4.85/5. Mar 2019 – August 2021
- **Bachelor of Computer Engineering** University of Brasília
Completed 104 credits (158 ECTS credits). GPA: 4.81/5. Mar 2017 – Mar 2019
- **Bachelor of Law** University of Brasília
Degree obtained on 25 nov. 2016. GPA: 4.53/5. Mar 2012 – Nov 2016

PUBLICATIONS

- **Pedro H. Luz de Araujo**, Benjamin Roth. *Helpful assistant or fruitful facilitator? Investigating how personas affect language model behavior.* PLoS One 20(6): e0325664 (2025). Available at <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0325664>.
- Yuxi Xia, **Pedro H. Luz de Araujo**, Klim Zaporozhets, Benjamin Roth. *Influences on LLM Calibration: A Study of Response Agreement, Loss Functions, and Prompt Styles.* ACL 2025. Preprint available at <https://arxiv.org/abs/2501.03991>.
- Benjamin Roth, **Pedro H. Luz de Araujo**, Yuxi Xia, Saskia Kaltenbrunner, Christoph Korab. *Specification Overfitting in Artificial Intelligence..* Artificial Intelligence Review 58, 35 (2025). Available at <https://link.springer.com/article/10.1007/s10462-024-11040-6>.
- **Pedro H. Luz de Araujo**, Benjamin Roth. *Functionality learning through specification instructions.* Findings of the Association for Computational Linguistics: EMNLP 2024. Available at <https://aclanthology.org/2024.findings-emnlp.642/>.
- Andreas Stephan, Lukas Miklautz, Collin Leiber, **Pedro H. Luz De Araujo**, Dominik Répás, Claudia Plant and Benjamin Roth *Text-Guided Alternative Image Clustering.* In Proceedings of the 9th Workshop on Representation Learning for NLP (RepL4NLP-2024). Available at <https://aclanthology.org/2024.repl4nlp-1.13/>.
- **Pedro H. Luz de Araujo**, Benjamin Roth. *Cross-functional Analysis of Generalization in Behavioral Learning.* Transactions of the Association for Computational Linguistics (2023) 11: 1066–1081. Available at <https://direct.mit.edu/tacl/article/doi/10.1162/tacl.a.00590/117216/Cross-functional-Analysis-of-Generalization-in>.

- **Pedro H. Luz de Araujo**, Ana Paula G. S. de Almeida, Fabricio Ataidees Braz, Nilton Correia da Silva, Flavio de Barros Vidal and Teófilo E. de Campos. *Sequence-aware multimodal page classification of Brazilian legal documents*. International Journal on Document Analysis and Recognition (IJDAR) volume 26, pages 33–49 (2023). Available at <https://link.springer.com/article/10.1007/s10032-022-00406-7>.
- **Pedro H. Luz de Araujo**, Benjamin Roth. *Checking HateCheck: a cross-functional analysis of behaviour-aware learning for hate speech detection*. Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP at ACL 2022. Available at <https://aclanthology.org/2022.nlppower-1.8/>.
- **Pedro H. Luz de Araujo**, Teófilo E. de Campos. *Topic Modelling Brazilian Supreme Court Lawsuits*. International Conference on Legal Knowledge and Information Systems (JURIX 2020). Available at <http://ebooks.iospress.nl/volumearticle/56168>.
- **Pedro H. Luz de Araujo**, Teófilo E. de Campos, Fabricio Ataidees Braz, Nilton Correia da Silva *VICTOR: a dataset for Brazilian legal documents classification*. Language Resources and Evaluation Conference (LREC 2020). Available at <https://www.aclweb.org/anthology/2020.lrec-1.181.pdf>.
- **Pedro H. Luz de Araujo**, Teófilo E. de Campos, Marcelo M. Silva de Sousa. *Inferring the source of official texts: can SVM beat ULMFiT?*. International Conference on the Computational Processing of Portuguese (PROPOR 2020). Preprint available at https://cic.unb.br/~teodecampos/KnEDLe/propor2020/luz_de_araujo_et_al_propor2020.pdf.
- **Pedro H. Luz de Araujo**, Teófilo E. de Campos, Renato R. R. de Oliveira, Matheus Stauffer, Samuel Couto, Paulo Bermejo. *LeNER-Br: A Dataset For Named Entity Recognition in Brazilian Legal Text*. International Conference on the Computational Processing of Portuguese (PROPOR 2018). Preprint available at https://cic.unb.br/~teodecampos/LeNER-Br/luz_et_al_propor2018.pdf.

TEACHING EXPERIENCE

- **Foundations of Computer Vision** University of Brasília
Teaching assistant. Prof: Teófilo E. de Campos Mar 2019 – Jul 2019

GRANTS AND AWARDS

- OpenAI’s Researcher Access Program. 2024.
- EuroHPC Development Access. 2024.
- OpenAI’s Researcher Access Program. 2023.
- EuroHPC Development Access. 2023.

SCHOLARSHIPS

- **KnEDLe - Knowledge Extraction from Documents of LLegal content.** 2020
Research scholarship.
- **VICTOR Project: an AI system for Brazil’s supreme court** 2019
Research scholarship.
- **Department of Computer Science, University of Brasília Master’s Degree Scholarship** 2019
Awarded by CAPES.

PARTICIPATION IN EXAMINATION BOARDS

- Matheus Schmitz Oliveira. *Extração de Informação sobre Drogas em Conteúdos Textuais de Redes Sociais utilizando técnicas de Reconhecimento de Entidades Nomeadas* (Drug Information Extraction in Social Media Texts using Named Entity Recognition Techniques). Bachelor Thesis. University of Brasília, Brazil, 2019.

ORGANISING COMMITTEES

- 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025). Student Volunteer Co-Chair.
- 20th Conference on Natural Language Processing (KONVENS 2024). Proceedings Chair.

ADDITIONAL COURSES

- **Reinforcement Learning Specialisation.**
University of Alberta. Offered through Coursera. 82 hours. 2020
- **Philosophy and the Sciences: Introduction to the Philosophy of Cognitive Sciences.**
University of Edinburgh. Offered through Coursera. 10 hours. 2020
- **Bayesian Statistics: Mixture Models.**
University of California, Santa Cruz. Offered through Coursera. 22 hours. 2020
- **Bayesian Statistics: Techniques and Models.**
University of California, Santa Cruz. Offered through Coursera. 30 hours. 2020
- **Bayesian Statistics: From Concept to Data Analysis.**
University of California, Santa Cruz. Offered through Coursera. 12 hours. 2020
- **Natural Language Processing.**
State University Higher School of Economics. Offered through Coursera. 33 hours. 2018
- **Build a Modern Computer from First Principles: Nand to Tetris Part I & II.**
Hebrew University Of Jerusalem. Offered through Coursera. 58 hours. 2018
- **Functional Programming Principles in Scala.**
Ecole Polytechnique Fédérale de Lausanne. Offered through Coursera. 23 hours. 2018
- **Deep Learning Specialisation.**
Deeplearning.ai. Offered through Coursera. 80 hours. 2017 – 2018
- **Machine Learning.**
Stanford University. Offered through Coursera. 55 hours. 2017

VOLUNTEER WORK

- **Student Volunteer at ACL 2022** Dublin, Ireland
Volunteer May 2022
- **University of Brasília's Computer Science Junior Enterprise (CJR)** Brasília, Brazil
Volunteer Apr 2017 – Mar 2018

LANGUAGES

- **Portuguese:** Native speaker.
- **English:** Fluent: TOEFL ITP, 2018. Total score: 677. TOEFL IBT, 2019. Total Score: 112
- **German:** B1.

OTHER WORK EXPERIENCE

- **Brazil's Supreme Court** Brasília, Brazil
Legal intern Nov 2015 – Sep 2016
- **Federal District and the Territories Public Ministry** Brasília, Brazil
Legal intern Mar 2014 – Mar 2015