

Rennan Cordeiro Lima

Mobile: +55-31-98348-6510

Email: rennancel@gmail.com

Homepage: www.linkedin.com/in/rennancel

Tech Lead and Data Scientist with experience in machine learning, information retrieval, and large-scale data systems across industry and academia. Published researcher in IR/NLP with experience leading teams and building production-grade search and predictive modeling systems.

EDUCATION

- Federal University of Minas Gerais (UFMG)** Belo Horizonte, MG, Brazil
Phd Candidate in Computer Science. Area: Natural Language Processing. March 2026 - Present
- Federal University of Minas Gerais (UFMG)** Belo Horizonte, MG, Brazil
M.Sc. in Computer Science. Area: Information Retrieval. Average Grade: 98/100 March 2021 - December 2023
- Federal University of Minas Gerais (UFMG)** Belo Horizonte, MG, Brazil
B.Sc. in Computer Science - Honored as one of the best students during Bachelors degree March 2016 - March 2021

WORK EXPERIENCE

- Kunumi** Remote
Tech Lead Jan 2025 - Present
 - Led a 5-10 person data science team developing short-term delinquency prediction models using Python, Spark, Databricks, SQL, LightGBM, and MLflow on large-scale customer data to anticipate payment delays.
- Data Scientist* May 2024 - Dec 2025
 - Developed a national-scale machine learning income prediction model using Python, Spark, Databricks, SQL, LightGBM, and MLflow, engineering features and training models on large-scale bank data, including transactions and salary information, to support more profitable business decisions.
- Public Prosecutor's Office of Minas Gerais (MPMG) project at DCC/UFMG** Belo Horizonte, Brazil
Software Engineer March 2021 - Present
 - Developed a governmental data search engine using information retrieval tools, including Transformers and Elasticsearch. The corpus contains more than 1 million documents.
 - Managed data processing, indexing, and search model training, as well as collaborated on the web interface for the search engine using Django. Two teams of multiple prosecutors currently use this search engine.
- Software Engineer Tech Lead* March 2021 - August 2023
 - Led a development team with almost 10 people as Tech Lead for a crawler configuration software, which included mentoring team members. Worked with stakeholders to define the project scope and tasks.
 - Utilized expertise in web scraping to architect solutions for complex data extraction needs. The software allowed the configuration of more than 5000 crawlers to target websites.
- Brazilian Institute of Geography and Statistics (IBGE) project at DCC/UFMG** Belo Horizonte, Brazil
Data Scientist October 2020 - March 2021
 - Created Python scripts with Scrapy, Selenium, Google Maps API to collect shopping places' sites and extract their products' data, which collected more than 100 places' websites across Brazil
 - Developed a price retrieval system with the data for IBGE, using more than 10,000 products from the crawled websites.
- Studio Sol** Belo Horizonte, Brazil
Data Scientist January 2021 - March 2021
 - Maintained a music recommender system and executed data analysis. Collaborated on deploying microservice architecture system based on Docker and Kubernetes for 3 million users.
- Data Scientist Intern* August 2019 - January 2021
 - Implemented music recommender systems and reports involving data analysis and modeling using Jupyter, SciPy, and Sklearn, with data from more than 1 million songs and 100,000 artists.

ACADEMIC EXPERIENCE

- Probabilistic Machine Learning Group (PML) at Aalto University** Espoo, Finland
Research Assistant Summer Intern May 2023 - July 2023
 - Supervised by Prof. Samuel Kaski on text summarization of electronic health records in a 3-month summer internship.
 - Proposed an LLM-based factual review pipeline for electronic health record summarization, reducing hallucinations by 50% and improving summary quality by nearly 5% over baseline metrics.
- Laboratory for Treating Information (LATIN) at DCC/UFMG** Belo Horizonte, Brazil
Undergraduate Student Researcher January 2018 - August 2019
 - Created a crawler, including a CAPTCHA-solving algorithm, to extract data from the Brazilian national researchers' profile repository platform (Lattes), which collected more than 100,000 profiles.
 - Implemented a name disambiguation algorithm utilizing Apache Spark to match more than 100,000 profiles.
- Computer Science Department (DCC) at UFMG** Belo Horizonte, Brazil
Teaching Assistant August 2017 - March 2018
 - Conducted classes for the "Computer Programming" discipline of fundamental programming concepts, engaging more than 30 students.

SKILLS SUMMARY

- **Platforms and Programming:** Python, C, SQL, Bash, Web
- **Tools and Frameworks:** PyTorch, Docker, Git, Elasticsearch, HuggingFace, Django, Jupyter, Transformers, Sklearn
- **Languages:** Portuguese (Native), English (Fluent), Spanish (Beginner)

PUBLICATIONS

- **On extractive summarization for profile-centric neural expert search in academia:** **Rennan C. Lima** and Rodrygo L. T. Santos. ACM SIGIR International Information Retrieval Conference, Spain, 2022.
- **On the presence of abusive language in mis/disinformation:** Breno Matos, **Rennan C. Lima**. et al. SocInfo International Conference, UK, 2022. Springer.
- **Intellectual dark web, alt-lite and alt-right: Are they really that different? A multi-perspective analysis of the textual content produced by contrarians:** Breno Matos, **Rennan C. Lima**, J. M. Almeida, M. A. Gonçalves, and Rodrygo L. T. Santos. Social Network Analysis and Mining, 14(1):32, 2024.
- **I've heard this before: Initial results on Tiktok's impact on the re-popularization of songs:** B. Matos, F. Galuppo, **R. Cordeiro**, and F. Figueiredo. arXiv preprint arXiv:2411.01239, 2024.
- **On representation learning-based methods for effective, efficient, and scalable code retrieval:** C. França, **Rennan C. Lima**, C. Andrade, W. Cunha, P. O. S. V. de Melo, B. Ribeiro-Neto, et al. Neurocomputing, 600:128172, 2024.