

# Gabriel de Jesus

## CURRICULUM VITAE

[gabriel.jesus@inesctec.pt](mailto:gabriel.jesus@inesctec.pt) | [gabrieldejesus.labadain.com](http://gabrieldejesus.labadain.com) | Dili, Timor-Leste  
[Google Scholar](#) | [LinkedIn](#) | [GitHub](#)

## PROFILE

I am a Research Scientist affiliated with INESC TEC (Porto, Portugal), based in Timor-Leste, and the creator of **Labadain**, an integrated ecosystem of algorithms, datasets, evaluation resources, search technologies, and conversational AI applications for the Tetun language. I have pioneered research in Information Retrieval (IR) and Natural Language Processing (NLP) for Tetun and hold a PhD in Informatics Engineering from the Faculty of Engineering of the University of Porto, Portugal. I am also an invited lecturer at the Institute of Business (IOB).

My research focuses on text retrieval, neural IR, retrieval-augmented generation (RAG), conversational AI, and the development of reproducible language resources for Tetun. In addition to my academic work, I bring more than a decade of professional experience in technology development and management, digital governance, and enterprise information systems across public- and private-sector organizations.

## RESEARCH INTERESTS

Information retrieval \* natural language processing \* Tetun \* low-resource languages \* neural retrieval \* conversational search \* retrieval-augmented generation \* large language models \* agentic AI

## RESEARCH EXPERIENCE

### Affiliated Researcher

2021–present

INESC TEC, Porto, Portugal [www.inesctec.pt](http://www.inesctec.pt)

Research on information retrieval, NLP, and AI, including language resources, datasets, and conversational systems.

### Founder and Research Scientist

2025–present

Labadain [www.labadain.com](http://www.labadain.com)

Research and development of datasets, algorithms, software, search tools, and AI applications for the Tetun language.

## EDUCATION

### PhD in Informatics Engineering

Sep 2021–Sep 2025

Faculty of Engineering, University of Porto, Portugal [repositorio-aberto.up.pt/handle/10216/169208](https://repositorio-aberto.up.pt/handle/10216/169208)

Thesis: Text Information Retrieval in Tetun. The research established datasets, algorithms, tools, test collections, and retrieval baselines for Tetun.

### Master in Computer Science

Sep 2011–Jul 2013

Faculty of Science, University of Porto, Portugal [hdl.handle.net/10216/103895](https://hdl.handle.net/10216/103895)

Dissertation: Relatórios e Análise de Tendências de Rede Social Desportiva Playnify.

### Bachelor in Informatics Engineering

Sep 2005–Jul 2008

Fundação das Universidades Portuguesas and Universidade Nacional Timor Lorosa'e, Timor-Leste

Final project: Software para a gestão dos combatentes da libertação nacional.

## SELECTED PEER-REVIEWED PUBLICATIONS

de Jesus, G., & Nunes, S. (2026). *Labadain Chat: A Conversational Agent for the Tetun Language*. In Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM. [doi:10.1145/3805712.3808372](https://doi.org/10.1145/3805712.3808372)

de Jesus, G., Singh, S. A. K., Nunes, S., & Yates, A. (2025). *Zero-Shot and Hybrid Strategies for Tetun Ad-Hoc Text Retrieval*. In Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval, 264-274. ACM. [doi:10.1145/3731120.3744593](https://doi.org/10.1145/3731120.3744593)

- de Jesus, G., & Nunes, S. (2025). *Insights into LLM-Based Conversational Search: A Study of Tetun-Speaking Users' Search Behavior*. In Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval, 297-306. ACM. [doi:10.1145/3731120.3744596](https://doi.org/10.1145/3731120.3744596)
- de Jesus, G., & Nunes, S. (2024). *Exploring Large Language Models for Relevance Judgments in Tetun*. In Proceedings of the First Workshop on Large Language Models for Evaluation in Information Retrieval, CEUR Workshop Proceedings 3752, 19-30. [Paper](#)
- de Jesus, G., & Nunes, S. S. (2024). *Data Collection Pipeline for Low-Resource Languages: A Case Study on Constructing a Tetun Text Corpus*. In Proceedings of LREC-COLING 2024, 4368-4380. ELRA and ICCL. [ACL Anthology](#)
- de Jesus, G., & Nunes, S. (2024). *Labadain-30k+: A Monolingual Tetun Document-Level Audited Dataset*. In Proceedings of the 3rd Annual Meeting of SIGUL at LREC-COLING 2024, 177-188. ELRA and ICCL. [ACL Anthology](#)

## SELECTED RESEARCH OUTPUTS AND PLATFORMS

- Labadain Chat** AI assistant designed to support conversation and access to information in Tetun. [Link](#)
- Labadain Search** Search engine developed specifically for Tetun-language information retrieval. [Link](#)
- Labadain Crawler** Open-source web data-collection pipeline for low-resource languages, with Tetun language identification and processing components. [Link](#)
- Labadain-30k+** Document-level audited Tetun corpus containing 33,550 documents. [Link](#)
- Labadain-Avaliadór** Tetun test collection for ad-hoc information retrieval with topics, documents, and graded relevance judgments. [Link](#)
- LabadainLog-17k+** Search Logs from Tetun-Speaking Users Across Chat, Web, and News Platforms. [Link](#)
- Labadain-ZSRuns** Sparse and Zero-Shot Dense Retrieval Runs with LLM-Generated Summaries for Tetun Ad-Hoc Text Retrieval. [Link](#)
- Labadain-Stemmer** Open-source Tetun stemming algorithms with light, moderate, and heavy variants. [Link](#)

## RESEARCH GRANTS AND SCHOLARSHIPS

- Doctoral Research Grant** 2021–2025  
*Fundação para a Ciência e a Tecnologia, Portugal* [doi.org/10.54499/SFRH/BD/151437/2021](https://doi.org/10.54499/SFRH/BD/151437/2021)  
 Project: Pesquisa e recomendação computacional de conteúdo noticioso. Grant number: SFRH/BD/151437/2021.
- Erasmus Mundus ACP Scholarship** 2011–2013  
*European Union*  
 Scholarship for Master studies in Computer Science at the University of Porto, Portugal.

## TEACHING EXPERIENCE

- Lecturer in AI Systems** Apr 2026–present  
*Faculty of ICT, Institute of Business (IOB), Timor-Leste*  
 Undergraduate course in Informatics Engineering.
- Lecturer in Object-Oriented Programming** Feb–Jun 2016  
*Faculty of Computer Engineering, Instituto Profissional de Canossa (IPDC), Timor-Leste*  
 Undergraduate course in Computer Engineering.
- Lecturer in Data Mining and Data Warehouse** Feb–Jun 2015  
*Faculty of Computer Engineering, Instituto Profissional de Canossa (IPDC), Timor-Leste*  
 Undergraduate course in Computer Engineering.
- Lecturer in Information and Communication Technology** Sep–Dec 2013  
*Faculty of Law, Universidade Nacional Timor Lorosa'e (UNTL), Timor-Leste*  
 Undergraduate introductory ICT course.
- Lecturer in Operating Systems** Feb–Jun 2009  
*Fundação das Universidades Portuguesas (FUP) and Universidade Nacional Timor Lorosa'e (UNTL), Timor-Leste*  
 Undergraduate course in Informatics Engineering.

## STUDENT SUPERVISIONS

### Co-supervisor of Master's Thesis

Sep 2024–Jul 2025

*Faculty of Engineering, University of Porto, Portugal*

Semantic Search in Sports Search Engine. The project integrated semantic retrieval into a sports search engine.

### Supervisor of Undergraduate Research Project

Sep 2024–Jul 2025

*Escola Superior de Tecnologia e Gestão, Felgueiras, Portugal*

Cross-Lingual Information Retrieval in Tetun. The work led to a peer-reviewed publication on cross-lingual Tetun retrieval ([doi:10.1007/978-3-032-05179-0\\_20](https://doi.org/10.1007/978-3-032-05179-0_20)).

## TALKS AND CONFERENCE PRESENTATIONS

Labadain: Resources and Applications for Tetun Language Technology

24 Jul 2026

Invited talk, NLP Group, University of Melbourne, Melbourne, Australia

Labadain: Language Technology Resources and Tools for Tetun

03 Jun 2026

Delegation representative talk, 2nd ASEAN-NEXUS ECRs Enclave, Singapore

Labadain Chat: AI for Tetun Speakers

26 Feb 2026

Invited talk, Fundasaun Hadomi Timor Round Table, Dili, Timor-Leste

AI for Tetun: Building Timor-Leste's Inclusive Digital Future

21 Nov 2025

Invited talk, TLNOG2 Conference, Dili, Timor-Leste

Labadain: The Foundation of Tetun Language Technology

20 Nov 2025

Invited talk, DEI-FECT-UNTL National Seminar, Dili, Timor-Leste

Timor-Leste Digital Future

05 Aug 2025

Invited talk, Timor-Leste Higher Education Day, National Seminar, Dili, Timor-Leste

Labadain Crawler Pipeline for Low-Resource Languages

22 May 2024

Conference talk, LREC-COLING 2024, Torino, Italy

Labadain-30k+ Dataset Construction

20 May 2024

Conference talk, SIGUL at LREC-COLING 2024, Torino, Italy

## ACADEMIC AND SOCIAL SERVICES

**Journal Reviewer** ACM Journal on Computing and Cultural Heritage, 2026.

**Conference Reviewer** International Conference on Innovative Concepts and Theories in Information Retrieval, 2025.

**Moderator** Computing, AI, and Development session at the Science for Development Summit 2026.

**Jury Member** Served on the jury for the evaluation and selection of research grant and award applications, 2026.

## LEADERSHIP AND PROFESSIONAL EXPERIENCE

### Founder

May 2019–present

*Timor News, Timor-Leste.* Direct strategic planning, digital-platform development, editorial technology, and organizational operations.

### Human Resource Information System Manager

Jul 2019–Feb 2021

*Chemonics International, USAID HRH2030 Program, Timor-Leste.* Led feasibility analysis, requirements, customization, deployment, documentation, and training for a health-sector information system.

### Non-Executive Board Member

Dec 2018–Sep 2021

*National Communications Authority, Timor-Leste.* Contributed to governance, regulatory oversight, telecommunications policy, compliance, and sector competition.

### Database Design and Implementation Expert Consultant

Nov 2018–Mar 2019

*Food and Agriculture Organization, Timor-Leste.* Advised the Ministry of Agriculture and Fisheries on a food-security monitoring system and supported implementation and training.

**Information Technology Manager****Oct 2017–Nov 2018**

*Inovar4 Lda, Timor-Leste.* Managed network and server infrastructure, upgrades, data migration, and technical support.

**Coordinator of Administration and Information Technology****Jan 2014–Sep 2017**

*Major Project Secretariat and Infrastructure Fund, Timor-Leste.* Directed administration, HR, procurement, logistics, IT, databases, budgets, and document management.

**Software Developer and Project Manager****Aug–Dec 2013**

*Quidgest Lda, Timor-Leste.* Managed information-systems projects and supported implementation, maintenance, training, and delivery.

**Senior Database Administrator****Sep 2009–Aug 2011**

*Timor Telecom, Timor-Leste.* Administered enterprise billing and customer-management databases and supported availability, integrity, performance, and fraud analysis.

**Software Developer****Jul 2008–Aug 2009**

*Quidgest Lda, Timor-Leste.* Developed, implemented, and maintained software and databases and trained end users.

**PROFESSIONAL DEVELOPMENT****Retrieval Augmented Generation****Aug 2025**

*Coursera.* Design, implementation, and evaluation of RAG systems.

**Machine Learning Specialization****Feb 2023**

*Coursera.* Supervised and unsupervised learning, reinforcement learning, neural networks, and recommender systems.

**European Summer School in Information Retrieval****24–29 Jul 2022**

*Lisbon, Portugal.* Advanced IR models, evaluation, neural retrieval, and multilingual and multimodel retrieval.

**Lisbon Machine Learning School****18–23 Jul 2022**

*Lisbon, Portugal.* Machine learning foundations, NLP, and deep learning.

**Situational Leadership Training****Jun 2017**

*Asian Institute of Technology.* Leadership communication, goal setting, performance management, conflict resolution, and problem solving.