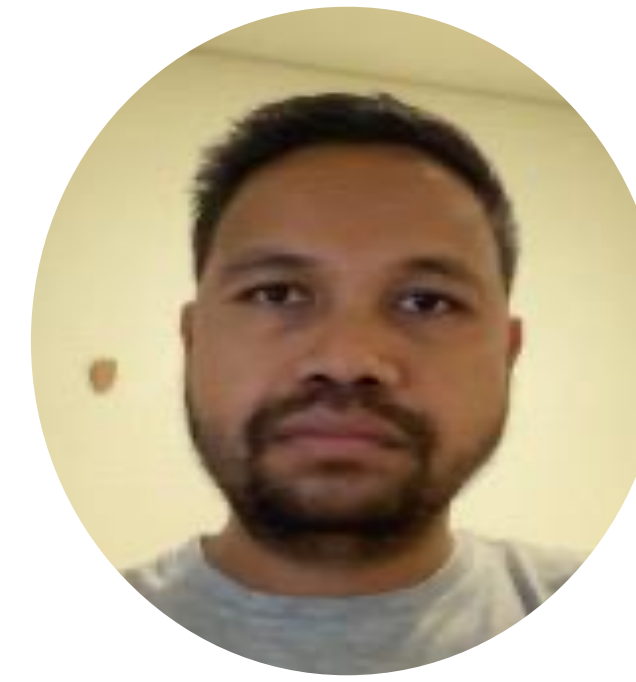




RESEARCH INTEREST



Information Retrieval (IR)



Natural Language Processing (NLP)



AI for Health and Legal Domains



Conversational Search



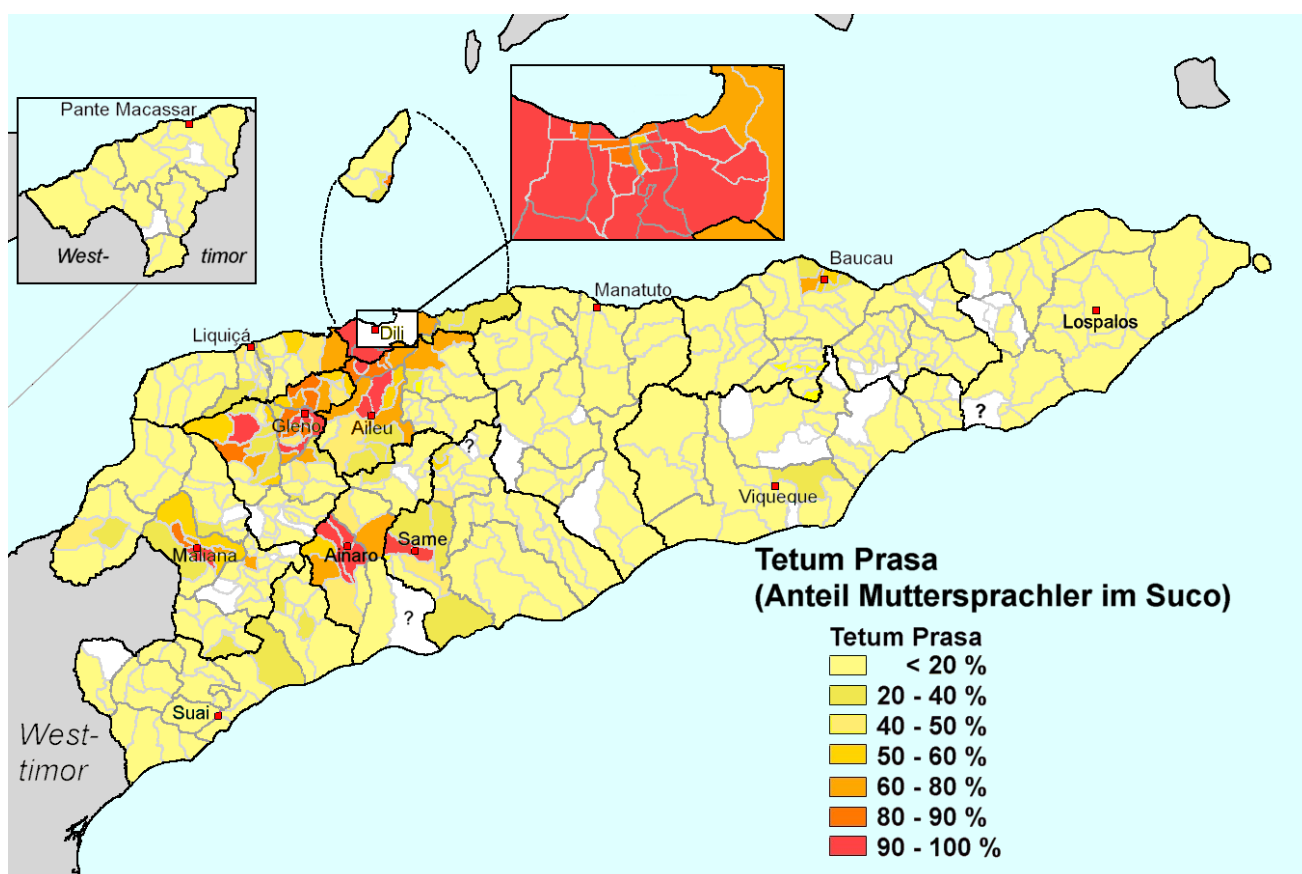
Retrieval-Augmented Generation (RAG)

RESEARCH VISION

Advancing digital inclusion through language, culture, and innovation by developing AI and information access for under-represented languages, particularly Tetun.

TETUN AT A GLANCE

Tetun is an official language of Timor-Leste, spoken by approximately 79% of its population (**Census 2015**).



900k+ speakers



National building and cultural identity



Under-represented in global AI and search ecosystem



Our work helps bridge the digital language gaps



Technology becomes inclusive when it speaks the language of its users.

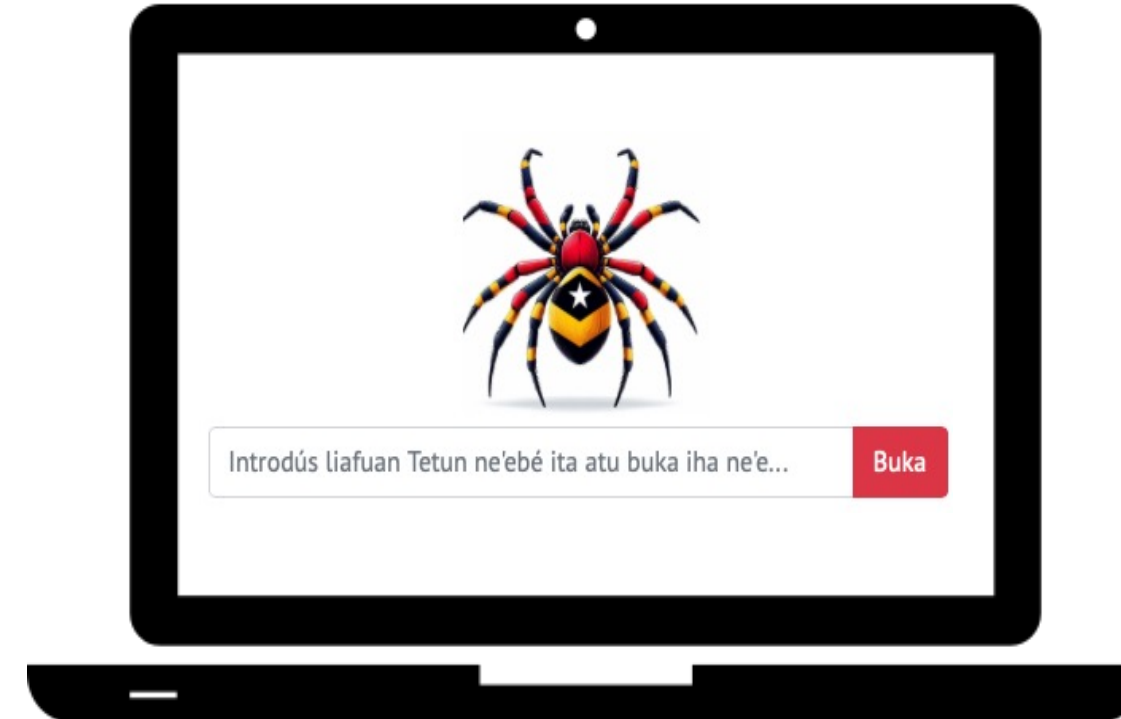
WORK AND CONTRIBUTIONS

THE LABADAIN ECOSYSTEM



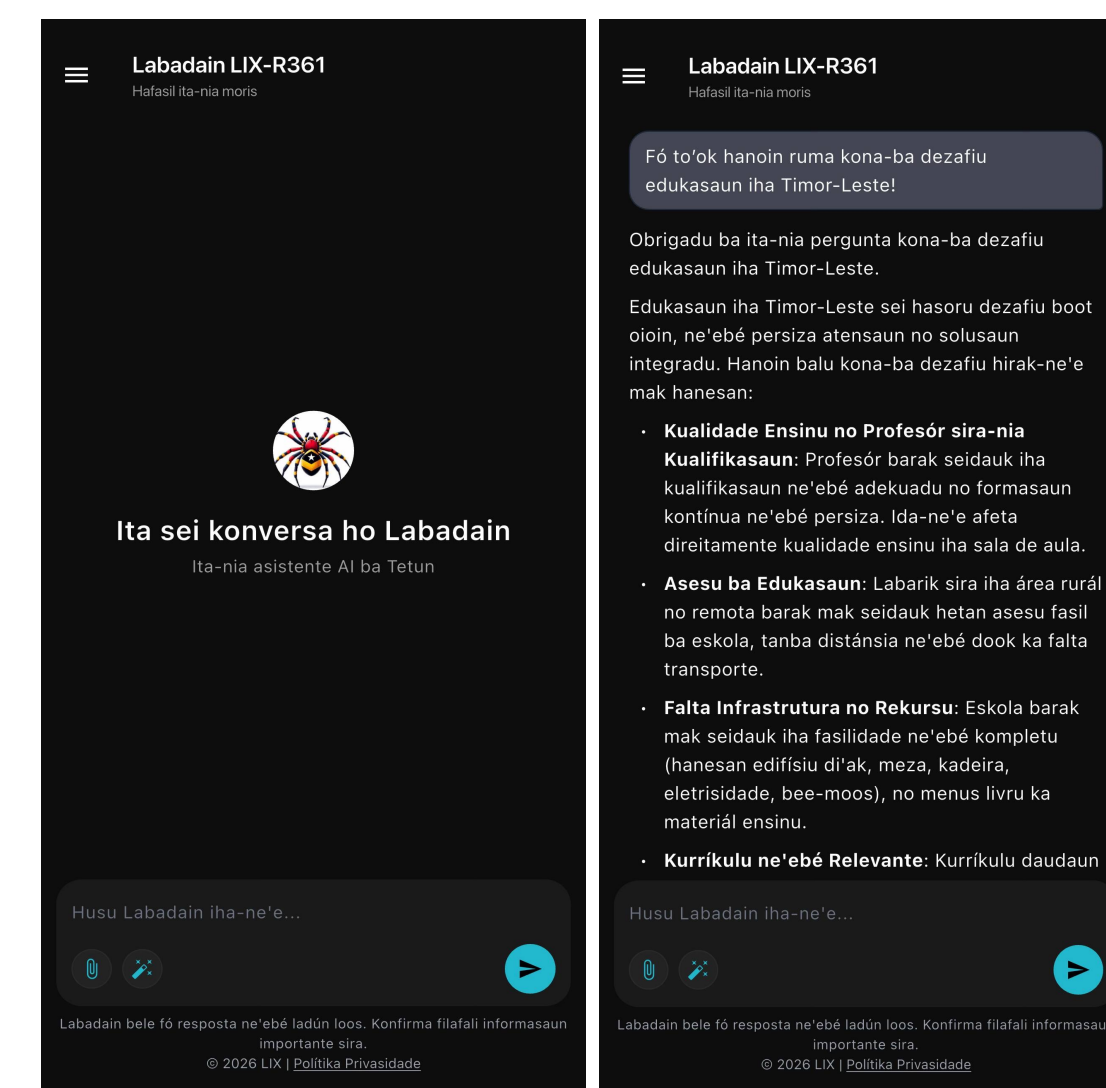
Labadain Search

The first search engine platform enabling information discovery in Tetun.



Labadain Chat

An AI conversational assistant designed specifically for Tetun-speaking communities.



Open Datasets & Resources

- Labadain-30k+ dataset
- Labadain-Avaliadór
- Labadain-Stopwords
- LabadainLog-17k+
- Labadain-ZSRun
- LabadainLog-60

Available on
INESC TEC repository



Algorithms, Toos & Software

- Labadain Stemmer
- Labadain Crawler
- Tetun Tokenizer
- Tetun LID

Available on
GITHUB



SELECTED PUBLICATIONS



Labadain Chat: A Conversational Agent for the Tetun Language

Adaptation of existing LLMs to Tetun using language-specific prompting strategies.



Zero-shot and Hybrid Strategies for Tetun Ad-Hoc Text Retrieval

Exploration of zero-shot dense and hybrid retrieval method for Tetun, combining with LLMs.



Insights into LLM-Based Conversational Search: A Study of Tetun-Speaking Users' Search Behavior

Analysis of real-world prompt logs from an LLM-based conversational assistant for Tetun.



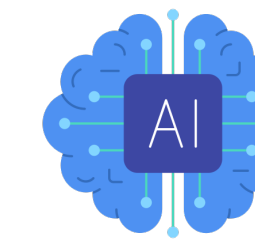
Exploring Large Language Models for Relevance Judgments in Tetun

Exploration the feasibility of automated relevance judgments for the Tetun language.



OPEN FOR COLLABORATION

I am seeking partnership in:



AI for low-resource languages



Search and conversational AI



IR and NLP datasets and benchmarking



ASEAN language technology



Digital inclusion initiative



Academia and industry collaboration



Public sector digital transformation

Hamutuk ita bele

Halo di'ak liu!

Together, we can do better!

LET'S COLLABORATE!

Inclusive AI starts with languages. By building technologies for Tetun, we empower communities, preserve our heritage, and ensure no one is left behind in the digital future.



If you share the same vision, let's build the future of inclusive AI together.

GET IN TOUCH



Website
gabrieldejesus.labadain.com



LinkedIn
gabrieljesus



Google Scholar
gabrieljesus



Email
gabriel.jesus@inesctec.pt

