

FELERMINO ALI

A Software Engineer & AI enthusiast willing to solve on world real-problem

@felerminoali@gmail.com

+258 846-689-637 / +351 910-51-7193

Rua Afredo Allen SN, Res UPorto

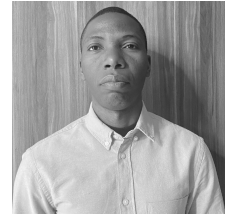
Porto, Portugal

felerminoali.github.io/

@felerminoali

felermino-ali

felerminoali



EXPERIENCE

Researcher

Laboratory of Artificial Intelligence and Computer Science (LIACC)
- Faculty of Engineering of Porto University

Sept 2021 - Present

Porto, Portugal

- Researcher and PhD student: Natural Language Processing and Machine Translation Research

Consultant - Software Engineer

Cedsif, IT

Jul 2021

Maputo, Mozambique

- Center for the Development of Finance Information Systems (CEDSIF) - CEDSIF is an institution linked to the Ministry of Economy and Finance which was set up to modernize the public finance management information systems for all state and municipal bodies
- Full stack developer for Government Municipality Management System.

Administration | Faculty Member | AI Researcher and Software Developer

Lurio University

Pemba, Mozambique

- Assistant Lecturer of the following subjects: Object Oriented Programming, Data Structure, Database Lab, Software Engineering Lab, Artificial Intelligence, Multimedia, Programming Foundation
- Developer of E-Sira (seul.unilurio.ac.mz)
- Deputy Director of Faculty of Engineering
- Research

Research Assistant

Research Assistant

SEGi, Malaysia

- Assistant Lecturer of the following subjects: Object Oriented Programming, Database Lab
- Developer of E-Sira (seul.unilurio.ac.mz)
- Research

PROJECTS

Emakhuwa Machine Translation Dataset Collection

Lacuna Fund/Meridian Institute

2 yrs

HONOR AWARDS



Valedictorian

Given at graduation ceremony by the former President of Mozambique (His Excellency Mr Armando Emilio Guebuza)



FCT PhD scholarships

shortlisted as first place at "Consórcio de escolas de engenharias" of Science and Technology Foundation - PhD scholarships

SKILLS

Hard-working

Eye for detail

Motivator & Leader

C++

Python

R

NLP

Machine Translation

Information Retrieval

Data Science

Artificial Intelligence

Statistical Analysis

Agile Management

Software Engineering

JAVA

Web Development

LANGUAGES

Portuguese



English



Emakhuwa



EDUCATION

Ph.D. in Informatics Engineering

Faculty of Engineering of the University of Porto

Sept 2021 - ongoing Porto, Portugal

Thesis title: Machine Translation for Emakhuwa, an Extremely Low-resource Bantu Language

M.Sc. in Information Technology

- Awarded a grant to create the first large-scale dataset for Machine Translation of the Emakhuwa language, Mozambique's widely spoken indigenous language.
- <https://lacunafund.org/2021-african-language-awardees/>

Google exploreCSR

Google

📅 1 yr

Promoting computing research among underprivileged students in the North of Mozambique. The project focused on boosting self-efficacy, fostering belonging, providing practical skills in Computer Sciences, and encouraging graduate studies.
<https://research.google/outreach/explore-csr/recipients/>

Musaling - Project

Personal Project

First tool for Machine Translation and spelling check for Mozambican languages
<https://felerminoali.pythonanywhere.com/en/translator/>

Models and Data Repos

Hugging Faces

- Contributor and Member of LIACC HF Repo:
<https://huggingface.co/LIACC>
- Personal Repo: <https://huggingface.co/felerminoali>

TRAINING

11th INTERNATIONAL SCHOOL ON DEEP LEARNING

<https://irdta.eu/>

📅 July 2024

📍 Maia, Portugal

- Summer School: 6-day event (15-19 July, 2024) cover multiple topics including: computer vision, neurosciences, speech recognition, language processing, human-computer interaction, drug discovery, health informatics, medical image analysis, recommender systems, advertising, fraud detection, robotics, games, finance, biotechnology, physics experiments, biometrics, communications, climate sciences, geographic information systems, signal processing, genomics, materials design, video technology, social systems, etc.

MEDCIDS - Winter School 2024

Faculty of Medicine of the University of Porto

📅 February, 2024

📍 Porto, Portugal

- Winter School: 3-day event (5-7 February 2024) Reliability and agreement studies

The 12th Lisbon Machine Learning

Instituto Superior Técnico, Instituto de Telecomunicações

📅 July, 2022

📍 Lisbon, Portugal

SEGI University

📅 Sept 2014 - June 2019 📍 Malasya, Kuala Lumpur

CONFERENCES AND WORKSHOPS

The 2024 Conference on Empirical Methods in Natural Language Processing

📅 Nov. 2024

📍 Miami, Florida, USA

- **Paper title:** Building Resources for Emakhuwa: Machine Translation and News Classification Benchmarks.

Open Language Data Initiative WMT workshop

📅 Nov. 2024

📍 Miami, Florida, USA

- **Paper title:** Expanding FLORES+ Benchmark for More Low-Resource Settings: Portuguese-Emakhuwa Machine Translation Evaluation.

First Workshop on NLP for Indigenous Languages of Lusophone Countries

📅 2024

📍 Galicia

- **Paper title:** Network-based Approach for Stopwords Detection.

Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)

📅 May, 2024

📍 Torino, Italia

- **Paper title:** Detecting Loanwords in Emakhuwa: An Extremely Low-Resource Bantu Language Exhibiting Significant Borrowing from Portuguese.

LITHME WG1-WG7 Joint Workshop: Bridging the gap between technology and professionals

📅 Oct. 28-30, 2023

📍 Budapest, Hungary

- **Paper title:** Data Augmentation for Low-Resource Neural Machine Translation Using Loanword Spelling Inconsistencies.

Language in the Human-Machine Era LITHME

University of Groningen, Campus Fryslân

📅 Oct. 15-16, 2022

📍 Leeuwarden, Netherlands

- Summer School: 6-day event (24-29 July 2022) covering a range of machine learning topics, from theory to practice, that are important in solving natural language processing problems arising in different application areas.

NVIDIA DLI Certificate

📅 June, 2022

- Fundamentals of Deep Learning

Microsoft Natural Language Processing Processing Event

📅 April, 2022

- Microsoft Learn Student Ambassador

ICMC 50: Python for Natural Language Processing

Institute of Mathematics Sciences and Computation

📅 2021

📍 Brazil

- Python for Natural Language Processing

Business Data Science Summer School

Schools of Economics and Business of Erasmus University Rotterdam, University of Amsterdam and Vrije Universiteit Amsterdam

📅 August 17-21, 2020

📍 Amsterdam, Netherlands

- Deep Learning

Professional Scrum Master

📅 2015

- Training on PSM I certification by Joshua Partogi

Cisco Certified Network Associate

📅 Nov 2014

- CCNA certification

PUBLICATIONS



- S. M. Saide, E. L. A. Sarmento, and F. M. D. A. Ali, "Cryptojacking malware detection in docker images using supervised machine learning," *International Conference on Intelligent and Innovative Computing Applications*, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:257780017>.
- F. D. M. A. Ali, A. Caines, and J. L. A. Malavi, "Towards a parallel corpus of portuguese and the bantu language emakhuwa of mozambique," 2021. arXiv: 2104.05753 [cs.CL].
- F. D. M. A. Ali, E. Gimo, and S. M. Saide, "A minimax agent for playing ntxuva game – the mozambican variant of mancala," pp. 1–5, 2020. DOI: 10.1109/icABCD49160.2020.9183848.

- **title:** Loanword detection in Emakhuwa language.

AfricaAI Conference

University of Groningen

📅 Oct. 15-16,2022

📍 Kigali, Rwanda

- **title:** Lacuna Fund Grantee Convening.

Doctoral Symposium, EPIA 2022

21st EPIA Conference on Artificial Intelligence

📅 Oct. 15-16,2022

📍 Lisbon, Portugal

- **title:** Machine Translation for Emakhuwa of Mozambique.

AfricaNLP Workshop

📅 2021

- **title:** Towards a parallel corpus of Portuguese and the Bantu language Emakhuwa of Mozambique

INVITED TALKS

Workshop on Translation Technology

Center of Linguistics of Porto University

📅 2022

- **title:** Machine Translation and CAT tools

IndabaX Mozambique

📅 2021

- **title:** Mozambican languages in the Natural Language Processing panorama

- B. J. Fonseca, A. Felermino D. M. A., and S. M. Saide, "A deep convolutional neural network for classifying waste containers as full or not full," pp. 54–59, 2019. DOI: 10.1109/ISC246665.2019.9071746.

Conference Proceedings

- F. D. M. A. Ali, G. de Jesus, H. L. Cardoso, S. Nunes, and R. Sousa-Silva, "Network-based approach for stopwords detection," in *Proceedings of the 16th International Conference on Computational Processing of Portuguese - Vol. 2*, P. Gamallo, D. Claro, A. Teixeira, et al., Eds., Santiago de Compostela, Galicia/Spain: Association for Computational Linguistics, Mar. 2024, pp. 55–63. [Online]. Available: <https://aclanthology.org/2024.propor-2.9>.
- F. D. M. A. Ali, H. L. Cardoso, and R. Sousa-Silva, "Building resources for emakhuwa: Machine translation and news classification benchmark," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024.
- F. D. M. A. Ali, H. L. Cardoso, and R. Sousa-Silva, "Expanding flores+ benchmark for more low-resource settings: Portuguese-emakhuwa machine translation evaluation," in *Proceedings of the Ninth Conference on Machine Translation*, Maimi, Florida, USA: The Association for Computational Linguistics, Nov. 2024. [Online]. Available: <https://www2.statmt.org/wmt24/pdf/2024.wmt-1.45.pdf>.
- F. D. M. Ali, H. Lopes Cardoso, and R. Sousa-Silva, "Building resources for emakhuwa: Machine translation and news classification benchmarks," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, November: Association for Computational Linguistics, Nov. 2024.
- F. D. M. Ali, H. Lopes Cardoso, and R. Sousa-Silva, "Detecting loanwords in emakhuwa: An extremely low-resource Bantu language exhibiting significant borrowing from Portuguese," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, Eds., Torino, Italia: ELRA and ICCL, May 2024, pp. 4750–4759. [Online]. Available: <https://aclanthology.org/2024.lrec-main.425>.
- F. M. E. Manuel, S. M. Saide, F. D. M. A. Ali, and S. Lotfi, "Ocular cataract identification using deep convolutional neural networks," in *2023 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, 2023, pp. 1–5. DOI: 10.1109/icABCD59051.2023.10220532.
- S. Muhammad, I. Abdulmumin, A. Ayele, et al., "AfriSenti: A Twitter sentiment analysis benchmark for African languages," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, and K. Bali, Eds., Singapore: Association for Computational Linguistics, Dec. 2023, pp. 13 968–13 981. DOI: 10.18653/v1/2023.emnlp-main.862.
- E. Á. Cardoso, F. D. M. A. Ali, and S. M. Saide, "Snake species identification using deep convolutional neural networks," in *2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2022, pp. 0202–0209. DOI: 10.1109/IEMCON56893.2022.9946504.