

Jaldert François

PhD Candidate | Applied Machine Learning, Data Science and Database Development

✉ jaldert.francois@kuleuven.be

📞 +32 470 59 12 07

📍 Leuven, Belgium

🔗 <https://www.jaldertfrancois.org/>

linkedin.com/in/jaldert-françois

Profile

I am a Bioinformatician (M.Sc.) and PhD researcher with a broad background in applied machine learning, deep learning development, and statistical analysis in life sciences. I have hands-on experience building end-to-end ML pipelines, developing scientific software tools and databases, with computational protein engineering as my core research topic.

SKILLS

- Applied Machine Learning & Deep Learning
- Statistical Analysis & Data Science
- Bioinformatics
- Python, R, SQL, Bash, HPC computing, MLOps & DevOps
- Database Design and Data Mining

AWARDS

FWO - SB Research Grant,
Research Foundation - Flanders
01/10/2022

Honours College Program,
University of Antwerp
01/09/2017

LANGUAGES

English

Native or Bilingual Proficiency

Dutch

Native or Bilingual Proficiency

WORK EXPERIENCE

PhD Candidate, FWO - KU Leuven

10/2021 – 11/2026 | Leuven

Machine Learning & Deep Learning

- Fine-tuned language models (ESM-2, ProtTrans) for enzyme EC classification using PEFT techniques (LoRA, full fine-tuning) with DeepSpeed CPU offloading, trained on datasets of 500,000+ sequences
- Developed an above SOTA enzyme optimal temperature prediction model with novel resampling strategies for imbalanced regression — outperforming current methods
- Achieved state-of-the-art pH optimum prediction by implementing population based training (PBT) for neural architecture search (NAS)
- Built and benchmarked predictive models across enzyme property tasks: optimal temperature, temperature range, pH optimum, EC classification, and solubility. Achieving near SOTA or better performance.

Platform & Software Engineering

- Architected and deployed a full stack enzyme annotation platform: React.js web application (dashboard + search interface) and CLI frontend, backed by a Dockerised and Singularity-containerised ML serving pipeline
- Designed a multi source relational database from scratch integrating UniProt, BRENDA, Rhea and other public resources at thousands of sequences scale, with a custom built search interface on top

Teaching

- Prepared and delivered lectures, slides and exercises for Bioinformatics (3rd year BSc Bioscience Engineering, ~60 students) and Introduction to Object-Oriented Programming (MSc Artificial Intelligence & MSc Bioinformatics, 20–50 students), KU Leuven

Research Internship (Honours College), University of Antwerp
08/2018 – 09/2018 | Antwerp

- Investigated novel mutations in Charcot-Marie-Tooth disease; conducted DNA/RNA extraction and molecular analysis

EDUCATION

Bioinformatics (M.Sc), KU Leuven

2019 – 2021

- Cum Laude
- Electives: (1) Neural Computing (2) Univariate Data Analysis and Modelling (3) Gene Technology
- Thesis: "Exploring mutational patterns of evolution experiments: Meta-analyses of microbial data from the CAMEL database"

Biochemistry and Biotechnology (B.Sc), University of Antwerp

2016 – 2019

- Summa Cum Laude
- Honours College Program

PROJECTS

FWO Strategic Basic Research, "Optimising the Enzyme Discovery Pathway through Omics Integration and AI"

10/2022 – 10/2026

- Competitive national research grant (FWO-SB) focused on advancing AI-driven enzyme discovery through novel ML approaches for classification, reaction parameter prediction and substrate affinity modelling
- Delivered above SOTA predictive models for temperature stability and pH optimum prediction, and a fine-tuned protein language model pipeline for large-scale enzyme classification (*see Work Experience for technical detail*)

PhD Society, Board Member - KU Leuven

2025 – 2026

Organise academic and social events connecting PhD researchers and postdoctoral fellows across KU Leuven; contribute to building an interdisciplinary research community

Enzymares,

<https://www.ugent.be/marine/en/partnerships/enzymares>

10/2021 – 10/2025

- Delivered an end-to-end enzyme selection platform for industry use: a unified multi-source database at thousands of sequences scale and a full-stack annotation toolbox (web app + CLI + containerised ML backend) (*see Work Experience for technical detail*)