Yves Gaetan Nana Teukam

Researcher in Artificial Intelligence & Machine Learning

+41 767283121 | yves.g.nana@gmail.com | https://yvesnana.github.io/ | Italian | Work permit B | Zurich

Pre-Doctoral research scientist with 4 years of experience developing AI/ML methods for accelerating scientific discovery and open sourcing. Ph.D. at IBM Research Zürich and Eindhoven University of Technology (expected March 2025). Expert in language modelling for biological applications, specializing in protein engineering and enzyme design.

PROFESSIONAL EXPERIENCE

IBM Research - Zürich, Switzerland **01/2022 – Present**

Pre-Doctoral Researcher

- Developed MTL4AD, a repo containing techniques for fine-tuning large language models (LLMs) on multi-task datasets, using parameter-efficient methods to integrate cross-domain knowledge.
- Developed Enzeptional, end-to-end ML pipeline for enzyme optimization, combining LLMs with genetic algorithms.
- Developed LM-ABC (Language Model Agent for BioCatalysis), a chatbot framework that streamlines and automates bioinformatics workflows.
- Developed RXNAAMapper, a state-of-the-art transformer-based tool for enzymatic binding sites prediction, achieving a 38% improvement in accuracy and a 30% reduction in false positives over baseline methods.
- Contributed to the GT4SD library for training and fine-tuning generative models to accelerate scientific discovery.
- Developed molecular dynamics simulation framework for validating AI-generated protein designs.

IBM Research - Zürich, Switzerland **02/2021 – 07/2021**

Research Intern

- Developed a synthesis planning approach integrating biocatalysis with transformer-based learning.
- Implemented transfer learning techniques using OpenNMT for reaction prediction.
- Conducted analysis of attention mechanism analysis in transformer architectures for model interpretability.

StemAway – California, USA **05/2020 – 09/2020**

Bioinformatics Project Lead

- Led international group of 30 students through all stages of gene expression analysis.
- Implemented automated QC pipeline reducing analysis time by 50% using Bioconductor.

Sequentia Biotech - Barcelona, Spain **04/2019 – 07/2019**

Research Intern

- Conducted microbiome analysis using bioinformatics tools and sequence alignment tools.
- Performed enterotype classification and diabetes prediction from microbiome data.

SKILLS

Machine Learning & AI:

- Model Architecture: Transformers, Large Language Models, Multi-Modal Architectures, Generative Models, Agent-Based LMs.
- Advanced Methods: Zero and Few-Shot Learning, Self-Supervised Learning, Unsupervised Learning, Model Compression (Quantization, Pruning).
- Model Efficiency: Distributed Training (DeepSpeed), Model Optimization, MLOps, LoRA Fine Tuning.

Domain Expertise:

- Computational Biology: Protein Optimization Design, Molecular Dynamics, Structure Prediction.
- Chemistry: Biocatalysis, Drug Design, Green Chemistry, Reaction Prediction.
- Research: Open-Source Contribution, Scientific Publication, Conferences Presentations.

Core Technologies:

- ML Frameworks: Hugging Face, PyTorch Lightning.
- Experiment Tracking: MLflow, Weights & Biases.
- Infrastructure: CI/CD, Docker, Cloud Object Storage (S3).

Languages:

- English, French, Italian (Native)
- Spanish (Excellent)

EDUCATION

Ph.D. in Biomedical Engineering
IBM Research Zürich & Eindhoven University of Technology – Zürich, Switzerland & Eindhoven, Netherlands
Thesis: "Leveraging Large Language Models for Enzyme Design, Functional Modelling, and Optimization in Biocatalysis"
Supervisors: Dr. Francesca Grisoni (TU/e) and Dr. Matteo Manica (IBM Research)
01/2022 to 03/2025 (expected graduation)

Master of Science in Data Science, grade: 109/110
University of Rome La Sapienza – Roma, Italy
09/2019 to 10/2021

Bachelor of Science in Bioinformatics, grade: 105/110
University of Rome La Sapienza – Roma, Italy
09/2016 to 06/2019

AWARDS

- 1st IEEE Open Software Service Awards as part of the GT4SD team. 2022.
- Sandmeyer Award of the Swiss Chemical Society as part of the RXN for Chemistry project team. 2022.

PUBLICATIONS

- **Teukam, Yves Gaetan Nana**, et al. "Language models can identify enzymatic binding sites in protein sequences." Computational and Structural Biotechnology Journal 23 (2024): 1929-1937.
- **Teukam, Yves Gaetan Nana**, et al. "Integrating Genetic Algorithms and Language Models for Enhanced Enzyme Design." (2024), (Soon in Briefings in Bioinformatics).
- **Teukam, Yves Gaetan Nana**, et al. "A language model assistant for biocatalysis." (2024), ChemRxiv, (Preprint).
- Manica, Matteo, Jannis Born, Joris Cadow, Dimitrios Christofidellis, Ashish Dave, Dean Clarke, **Yves Gaetan Nana Teukam** et al. "Accelerating material design with the generative toolkit for scientific discovery." npj Computational Materials 9, no. 1 (2023): 69.
- Probst, Daniel, Matteo Manica, **Yves Gaetan Nana Teukam**, Alessandro Castrogiovanni, Federico Paratore, and Teodoro Laino. "Biocatalysed synthesis planning using data-driven learning." Nature communications 13, no. 1 (2022): 964.