Yves Gaetan **Nana Teukam**

Researcher in Artificial Intelligence & Machine Learning
Switzerland (B Permit)

yves.g.nana@gmail.com | https://yvesnana.github.io/ | +41 076 728 31 21

Summary

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.

Experience

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

- Developed a tool with techniques for fine-tuning large language models (LLMs) on multi-task datasets, using parameter-efficient methods to enhance adaptability and cross-domain knowledge integration. Open-sourced as MTL4AD.
- Created an end-to-end ML pipeline for enzyme optimization, integrating LLMs with genetic algorithms to accelerate biological research and development. Open-sourced as Enzeptional.
- Built a language model agent for biocatalysis: A chatbot framework that automates bioinformatics workflows, improving efficiency and reducing manual intervention. Open-sourced as LM-ABC.
- Engineered a state-of-the-art, transformer-based tool for predicting enzymatic binding sites, achieving a 38% improvement in accuracy and a 30% reduction in false positives compared to baseline methods. Open-sourced as RXNAAMapper.
- Contributed to the GT4SD: Enhanced the library for training and fine-tuning generative models to support accelerated scientific discovery, including a framework for enzyme optimization.
- Developed a molecular dynamics simulation framework for validating AI-generated protein designs, reducing potential failure rates in laboratory testing and accelerating advancements in protein engineering projects.

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

- Developed a synthesis planning approach: Integrated biocatalysis with transformer-based learning models to optimize synthetic pathways, enhancing efficiency and accuracy in reaction planning.
- Implemented transfer learning for reaction prediction: Applied OpenNMT to enable accurate transfer learning for chemical reaction predictions, improving model adaptability across varied reaction types.
- Analysed attention mechanisms in transformer architectures: Conducted detailed analysis to enhance model interpretability, providing insights into decision-making processes and increasing transparency in reaction prediction outcomes.
- Optimized feature encoding: Transitioned from low to high-dimensional representations to capture enriched enzymatic data, improving model interpretability and prediction robustness.

StemAway *California, USA*
Bioinformatics Project Lead *05/2020 – 09/2020*

- 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*
Research Intern *04/2019 – 07/2019*

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

Education

Ph.D. in Biomedical Engineering *Zürich & Eindhoven*
IBM Research Zürich & Eindhoven University of Technology *01/2022 to 03/2025 (expected graduation)*
Thesis: “Leveraging Large Language Models for Enzyme Design, Functional Modelling, and Optimization in Biocatalysis”

M.Sc. in Data Science *Rome, Italy*
University of Rome – La Sapienza *09/2019 to 10/2021*

Exchange Program *Barcelona, Spain*
ESCI – Pompeu Fabra University *09/2018 to 04/2019*

B.Sc. in Bioinformatics *Rome, Italy*
University of Rome – La Sapienza *09/2016 to 10/2019*

Skills

- **Machine Learning & AI**: Transformers, LLMs, Multi-Modal Architectures, Generative Models, Agent Based LMs.
  skills: model architecture, self-supervised learning, zero few shot learning, attention mechanisms modelling, transfer learning.
- **Model Optimization & Efficiency**: DeepSpeed, LoRA, Quantization, Pruning, PEFT, QLoRA, 8-bit Training.
  skills: distributed training, model compression, MLOps, performance tuning, memory optimization, mixed precision training.
- **ML Frameworks & Tools**: Hugging Face, PyTorch Lightning, MLflow, Weights & Biases, Accelerate.
  skills: experiment tracking, model versioning, pipeline automation, hyperparameter optimization, parallel processing.
- **Infrastructure & Deployment**: Docker, CI/CD, Cloud Object Storage (S3), Git, FastAPI, Streamlit.
  skills: containerization, automated testing, cloud scaling, infrastructure management.
- **Computational Biology**: Protein Optimization Design, Molecular Dynamics, Sequence Analysis.
  skills: structural analysis, protein engineering, fold prediction, structure minimization and equilibration.
- **Chemistry Expertise**: Biocatalysis, Drug Design, Green Chemistry, Reaction Prediction, Retrosynthesis Analysis.
  skills: molecular modelling, synthesis optimization, reaction mechanisms, catalyst design, property prediction.
- **Research & Development**: Open-Source Contribution, Scientific Publication, Conference Presentations.
  skills: technical writing, peer review, research methodology, literature review, experimental design.

- **Languages**: English (native), French (native), Italian (native), Spanish (fluent)

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).
- M. Manica, J. Born, J. Cadow, D. Christofidellis, A. Dave, D. 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.
- D. Probst, M. Manica, **Yves Gaetan Nana Teukam**, A. Castrogiovanni, F. Paratore, T. Laino. "Biocatalysed synthesis planning using data-driven learning." Nature communications 13, no. 1 (2022): 964.

Extracurricular Activities

- Officiated football matches as referee at local and regional levels in Italy.
- Active member of the 6 AM Running Club in Zurich; regularly commute to work by running.
- Modeled for Fendi and walked at Milan Fashion Week.
- An enthusiastic traveler who recently visited Mexico to experience a total solar eclipse.