Yves Gaetan Nana Teukam

+41 76 728 31 21 • yves.g.nana@gmail.com • LinkedIn • EU passport • B permit • Zürich, CH

PROFESSIONAL SUMMARY

Detail-oriented Data Scientist with 4 years of experience in developing AI/ML-based tools to address complex biological problems. Currently completing a Ph.D. in Biomedical Engineering with a focus on Language Modelling for Protein Design at IBM Research Zürich and Eindhoven University of Technology, with expected graduation in January 2025. Expert in machine learning, generative modelling, NLP, Omics, and bioinformatics. Proven track record of optimizing biomolecules and enhancing model performance for biocatalysis, drug discovery, and green chemistry. Proficient in advanced machine learning frameworks and deep learning libraries, including TensorFlow, Transformers, and PyTorch, for analysing and modelling complex biological datasets. Contributed to open-source projects such as GT4SD and published in high impact journals like Nature Communications. Fluent in English, French, Italian, and Spanish. Strongly committed to driving innovative projects, fostering collaboration, and achieving team goals.

WORK HISTORY

PRE-DOCTORAL RESEARCHER | 01/2022 to Current
IBM Research - Zürich, Switzerland - Zürich, Switzerland

- Main developer of Enzeptional, an AI/ML-based computational tool for biological process modelling, leveraging pretrained large language models and evolutionary algorithms to optimize biomolecules, improving their feasibility and stability.
- Main developer of the RXNAAMapper for protein binding site predictions, leveraging transformer-based language models combined with Byte Pair Encoding for high-precision predictions. Improved accuracy of 38% and reduced the false positive rate by 30% compared to the baseline.
- Implemented a framework combining molecular dynamics simulations (GROMACS) to validate optimized protein sequences and ensure their structural integrity.
- Contributed to GT4SD, an open-source library for training and fine-tuning generative models (e.g., VAE, GAN) to accelerate scientific discovery in computational biology.

RESEARCH INTERN | 02/2021 to 07/2021
IBM Research - Zürich, Switzerland - Zürich, Switzerland

- Developed synthesis planning approach integrating biocatalysis with data-driven learning, enhancing efficiency and sustainability in chemical synthesis, leveraging transfer learning techniques.
- Achieved 49.6% top-1 accuracy in biocatalyzed forward predictions using a Transformer model, supported by OpenNMT, marking a significant advancement in computational chemistry and AI-driven drug discovery.
- Conducted in-depth attention mechanism analysis, providing insights into model interpretability and decision-making processes.
- Leveraged Python tools for data analysis and modelling, including Pandas, NumPy, SciPy, TensorFlow, Keras, and Biopython, to process and analyse large-scale biological datasets.

DATA SCIENCE AND BIOINFORMATICS PROJECT LEAD | 05/2020 to 09/2020
StemAway - California, USA - California, USA (Remote)

- Mentored 30 students from various countries and academic backgrounds through all stages of gene expression analysis, covering data collection, processing, and analysis.
- Utilized Python and R, focusing on bioinformatics tools such as DESeq2 and edgeR for comprehensive gene expression analysis.

RESEARCH INTERN | 04/2019 to 07/2019
Sequentia Biotech - Barcelona, Spain - Barcelona, Spain

- Collected microbe data from NCBI and integrated it with data from various experiments and research studies.
- Conducted human gut microbiome analysis using bioinformatics tools, including Samtools, BLAST, and Bowtie for sequence alignment and variant calling.

SKILLS

- **Programming Languages & Version Control**: Python, R, Bash, Linux, Git, GitHub
- **Data Analysis and Tools**: Statistical Analysis, Data Mining, Data Visualization, Pandas, NumPy, SciPy
- **Machine Learning & AI**:
  - **Core Skills**: Machine Learning, Deep Learning, Generative Modelling, Natural Language Processing (NLP)
  - **Specific Techniques**: Transformers, LSTMs, Attention Mechanisms, Sequence-to-Sequence Models, Few-Shot learning, Zero-Shot learning
  - **Frameworks & Libraries**: TensorFlow, Keras, PyTorch, Scikit-learn, XGBoost, SFTrainer, Hugging Face Transformers, OpenNMT, SpaCy, NLTK, Fairseq
  - **Models & Architectures**: BERT, ProTrans, ESM2, T5, XLNet
  - **Optimization & Training**: Hyperparameter Tuning, Model Compression (Quantization, Pruning)
  - **Advanced Techniques**: Transfer Learning, Reinforcement Learning, Self-Supervised Learning
- **Bioinformatics & Computational Biology Tools**: Protein Optimization, Molecular Dynamics (Gromacs), Evolutionary Algorithms, Samtools, BLAST, Bowtie, DESeq2, edgeR, Rosetta, AlphaFold
- **Development Tools & Experiment Tracking**: Jupyter Notebooks, VS Code, MLflow, Weights & Biases
- **Soft Skills**: Team Collaboration, Leadership, Mentoring, SCRUM, Agile Methodologies
- **Communication Skills**: Conferences Presentations and Scientific Divulgation
- **Languages**: Fluent in English, French, Italian, Spanish

EDUCATION

Ph.D. in Biomedical Engineering
IBM Research Zürich & Eindhoven University of Technology – Zürich, Switzerland & Eindhoven, Netherlands
01/2022 to 01/2025 (expected graduation)

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

Exchange Program Erasmus
ESCI-Universitad Pompeu Fabra – Barcelona, Spain
09/2018 to 02/2019

Bachelor of Science in Bioinformatics
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), (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.