Zaid Yusuf

ML Research Engineer | Protein AI & Scientific ML

zaidyusuf.com | github.com/Yusufteppei | linkedin.com/in/zaid-yusuf-de |
yusufteppei11@gmail.com | Open to remote or local roles

Summary

ML research engineer with ~4 years of experience across ML engineering, platform infrastructure and full-stack development, focused on scientific ML and protein AI. I implement research architectures from first principles (diffusion over SE(3)/SO(3), AlphaFold2-style structure modeling, single-cell foundation models) and bring the engineering background to make research code reproducible and deployable. 2x IMO contestant. Certified Kubernetes Administrator.

Research Projects

ProtDiffusion: protein structure diffusion | PyTorch, geometric deep learning | github.com/Yusufteppei/protdiffusion

- Research-oriented diffusion model over protein backbone geometry, started as a from-scratch reconstruction of RFdiffusion's ideas rather than a reproduction of it.
- Represented structure as SE(3) rigid frames; implemented rotation and translation diffusion, SO(3) machinery, score-based modeling and Invariant Point Attention-inspired components.
- Worked through torsion/dihedral geometry, atom placement, noise schedules, training and sampling; investigated numerical stability and training behavior.

AlphaFold2-mini | PyTorch | github.com/Yusufteppei/alphafold2-mini

- Implemented core AlphaFold2 components from first principles: Evoformer, MSA row/column attention, outer product mean, triangle attention and multiplication, Invariant Point Attention and the structure module.

scGPT-mini: single-cell foundation model | PyTorch, transformers | github.com/Yusufteppei/scgpt-mini

- Implemented a transformer-based foundation model for single-cell gene expression: gene and expression embeddings, masked gene modeling, transformer encoder, data processing and training pipeline.

BioAgent: biology-focused tool-calling agent | LLM fine-tuning, SFT/LoRA | github.com/Yusufteppei/bio-agent

- Fine-tuned open-source LLMs with SFT/LoRA to build a tool-calling agent for biology tasks. Live demo: zaidyusuf.com/projects/bio-agent

Breast cancer prognosis from TCGA data | Python, scikit-learn, RNA-seq | github.com/Yusufteppei/breast-cancer-prognosis

- End-to-end workflow on transcriptomic and clinical data: RNA-seq processing, gene-expression feature engineering, clinical data integration, model development and evaluation.
- Caught and corrected a data leakage issue that was inflating model accuracy.

FensterSQL: relational database engine | C, Flex/Bison | github.com/Yusufteppei/fenstersql

- Built from scratch to understand systems internals: SQL parsing, page-based storage, B+ tree indexing, catalogs and disk-backed persistence.

Technical Writing

- "A working intuition for diffusion on SO(3)" (Sep 2026) and "My Top Algorithms Across Transformer Architectures" (Oct 2026), zaidyusuf.com/writing

Professional Experience

Machine Learning Engineer, Digprom (Contract, part-time) | Jun 2023 - Present

- Built and maintained machine-learning workflows and supporting data pipelines for government and university client projects at an IT consulting firm.
- Developed backend software supporting data processing and ML applications.
- Applied software engineering practices to improve reliability, reproducibility and maintainability.
- Worked with engineering and domain teams to translate requirements into technical systems.

Full-Stack Engineer, Turing (Contract) | Jun 2024 - May 2025

- Built and maintained backend systems and APIs supporting AI training projects.
- Worked with cloud infrastructure and production deployment workflows, mostly Docker-based.
- Designed reliable, data-driven application architectures.

Platform & Machine Learning Engineer, 9ijakids / Lorem Excellentiam (Contract) | Dec 2023 - Jul 2024

- Built cloud-native infrastructure supporting scalable machine-learning systems.
- Developed production ML solutions and inference workflows; automated deployment and experimentation pipelines.
- Improved reliability, reproducibility and monitoring: adding monitoring cut average incident response time to one-fifth of its previous level and made it clear which team owned each incident.

Founding Engineer, Edulethics Africa (Contract) | Jan 2023 - Dec 2023

- Led engineering strategy, software architecture and technical decisions across product development.
- Designed and implemented the platform's core architecture, serving over 45,000 users.

Technical Skills

- **Machine learning:** PyTorch, JAX, Hugging Face, transformers, diffusion models, geometric deep learning, LLM fine-tuning (SFT/LoRA), model evaluation
- **Computational biology:** protein structure prediction and generation, protein language models, single-cell transcriptomics, RNA-seq, cancer genomics, bioinformatics
- **Engineering:** Python, C, C++, TypeScript, SQL, Bash; Docker, Kubernetes, AWS, Linux, CI/CD, PostgreSQL, FastAPI, Django

Education & Honors

- Nile University of Nigeria, Bachelor's coursework in Computer Science | Sep 2017 - Jan 2020
- International Mathematical Olympiad (IMO): 2x contestant (2015, 2016)
- Pan-African Mathematics Olympiad (PAMO) 2015: Bronze medal

- Cowbell Mathematics Competition 2014: 2nd place nationally (silver)
- Certified Kubernetes Administrator (CKA)
- Languages: English; German (B2)