

Yosif Soliman | Machine Learning Engineer

☎ +20 109 609 5283 • ✉ yosifoli981@gmail.com • 🌐 y0sif.github.io • in y0sif • 🔄 y0sif
Languages: Arabic (Native), English (Fluent)

Summary

Applied ML engineer with published research on Kolmogorov-Arnold Networks (two IEEE articles) and hands-on experience fine-tuning open models with disciplined, reproducible evaluation. Owns the full research loop: hypothesis, data pipeline, custom kernels and quantization, benchmarking (SPRT, paired eval, ablations), and validation inside real systems.

Research & Publications

- **Trainable Activation Functions for Real-Time Chess Engine Evaluation** (KANUE, IEEE-format article)
 - Integrated **KAN** B-spline activations into an NNUE trainer with custom fused **CUDA** kernels (**1.71x** training speedup) and **int8** quantized lookup-table inference; validated a **+169–226 Elo** gain over a matched MLP baseline via **SPRT** (~2556 Elo vs Stockfish 18).
- **KANs vs Traditional Neural Architectures in Chess Applications** (B.Sc. thesis + IEEE-format article)
 - Multi-domain study: a chess engine (HCE/NNUE/KAN, **2400–2600 Elo**), computer-vision piece classification (KAN **+1.81%**: 97.86% vs 96.05%), and behavioral cheat detection.

Experience

Bond

AI Engineer

– Built a **55+** scenario real-backend agent evaluation framework with rubric-based **LLM judging** across list-building, enrichment, and copy pipelines, enabling **model-per-sub-agent** selection and prompt/tool fixes that eliminated copy fabrication and ICP-filter leakage.

Remote, USA

Dec 2025 – Present

OpenQQuantify

AI Software Engineering Intern

– Built **multi-agent pipelines** for code generation and code-to-assembly compilation (RISC-V, QEMU) using Python and the OpenAI API.

Remote

Jan 2025 – Jun 2025

Highlighted Projects

- **GutWise – Fine-Tuned Gemma Health Assistant** [Model]
 - **QLoRA** fine-tune of Gemma 4 E4B with a grounded clinical data pipeline (medical-fact validators + LLM judge) and a **3-seed paired evaluation**; surpassed the base model while preserving **4/4** red-flag safety handling across all runs.
- **Arcwright – Domain Fine-Tune for Rust Frameworks** [Model]
 - Built an instruction-dataset pipeline (~4.7k rows via tree-sitter) and fine-tuned Gemma for Leptos/Axum/Rig; ranked **#1** on a custom **RustWebBench-15** eval, outscoring larger models and Claude Haiku.

Technical Skills

Languages: Python, Rust, C/C++, SQL

ML & Fine-tuning: PyTorch, Unsloth, PEFT, TRL, QLoRA, Transformers, Hugging Face, Scikit-learn, TensorFlow/Keras

Research & Eval: Benchmarking, SPRT, ablations, LLM-as-judge, custom architectures (KAN/B-splines), CUDA, quantization

LLM/Agents: LangChain, LangGraph, OpenAI, Anthropic Claude, RAG, MCP, Langfuse, LangSmith

Tools: Weights & Biases, Docker, Git, GCP, OpenSearch, ChromaDB

Education

Ain-Shams University

B.Sc. in Computer Science, **CGPA: 3.48/4.0**

Cairo, Egypt

Oct 2021 – Jul 2025

Certifications: DataCamp AI Engineer for Developers (Associate); Kaggle Computer Vision; Kaggle Intro to Deep Learning