

Raaif Yousuf

raaifirfanyousuf@gmail.com Nashville, Tennessee GitHub (<https://github.com/Raaif-Yousuf>) LinkedIn (<https://www.linkedin.com/in/raaif1>)

raaif.dev (<https://raaif.dev>)

EDUCATION

Vanderbilt University, Nashville, Tennessee

EXPECTED MAY 2027

BACHELOR OF SCIENCE, COMPUTER SCIENCE, GPA 3.71

- Honors: Dean's List, three semesters. Student of the Year, Behavioral and Social Science, Tarrant County College, 2024.
- Coursework: Algorithms, Software Engineering, Operating Systems, Computer Architecture, Systems for AI, Social Network Analysis with graph neural networks, Programming Languages, Linear Algebra, Rapid Prototyping, Data Structures, Technology Strategy.

PROFESSIONAL EXPERIENCE

CLAIR Analytics

JANUARY 2026 TO PRESENT

FOUNDER AND SOLE DEVELOPER

POLARS · DUCKDB · PYTORCH · XGBOOST · ONNX RUNTIME ·
REACT · TAURI · FASTAPI

- Shipped a local-first desktop analytics app in React and Tauri over a FastAPI backend that lets researchers ask questions of their spreadsheets in plain English, computed entirely on the machine, so data covered by IRB or FERPA never leaves it. In use by a Vanderbilt principal investigator and by a non-technical business owner.
- Routed analytical questions through exact DuckDB SQL aggregates rather than the model, and badge a result as verified only when a deterministic query produced it, which removes a whole class of numeric hallucination.
- Built an end-to-end AutoML and training pipeline: a time-budgeted XGBoost hyperparameter search and PyTorch Lightning training, with leakage detection, ONNX export, and global explanations surfaced in the interface.
- Wrote a hardware-aware inference backend that detects the GPU architecture across NVIDIA CUDA, DirectML for AMD and Intel Arc, Apple Metal, and CPU, then reconciles predictions against bound ONNX providers, so a session cannot silently fall back to CPU without saying so.
- Measured link discovery across a 2,950 file, 21.3 GB public corpus: precision 0.910 and recall 0.763 over 48 labelled groups and 86 labelled pairs.

DNA-Entropy, Meers Lab, Vanderbilt University Medical Center

MAY 2026 TO PRESENT

INDEPENDENT RESEARCH PROJECT

PYTHON · EVO 2 · GOOGLE CLOUD · GENBANK · IGV
· GENEIOUS

- Built a command-line tool and cloud-orchestration pipeline that computes per-position Shannon entropy for a DNA sequence in a single forward pass through the Evo 2 genomic language model, replacing the lab's previous alignment-based conservation workflow.
- Automated the whole Google Cloud GPU virtual machine lifecycle, provisioning, health check and teardown, so a run costs only the minutes of GPU time it uses, and exports bedGraph, wig, GFF3 and TSV tracks that open directly in IGV, Geneious, SnapGene and Benchling.
- The GPU is rented inside the researcher's own cloud project, so no server of mine is in the path and the bill and the sequences stay theirs. Used independently by lab researchers on their own sequences.

Abako Technologies, remote

SOFTWARE ENGINEER

OCTOBER 2024 TO PRESENT

TYPESCRIPT · TAURI · RUST · NODE.JS · VITEST ·
ELECTRON · FLASK

- Built a pixel-accurate Windows, macOS and web emulator of the AB-991EX scientific calculator: a dependency-free TypeScript core whose parser and rasteriser drive a 192 by 63 dot framebuffer across the calculator's 12 calculation modes, on decimal.js precision matched to the hardware.
- Pinned the rendering behind 244 golden screen files, inside a suite of 3,560 tests, so any change that moves an LCD pixel has to ship the picture it now produces.
- Built the desktop telemetry backend: batched analytics, error ingest, per-IP and per-install rate limiting, and the auto-update feed.
- Automated a national math competition end to end in a Python and Electron app: Excel-driven 60-question exam generation with a synced answer key, percentile grading, and an individualised report card and certificate for every student, covered by 51 tests.

Engineering Identity and Engineering Education, University of Texas at Dallas, Richardson, Texas

AUGUST 2024 TO JUNE 2025

LEAD RESEARCHER

PYTHON · SCIKIT-LEARN · XGBOOST · KERAS · MATLAB

- Led a research team investigating engineering identity in STEM students across data collection, classification and predictive modelling, and built the internal Python tooling, data-cleaning scripts and onboarding material the team depended on.
- Rebuilt both original pipelines, a MATLAB wisdom-of-the-crowd ensemble and a Keras neural network, and scored them on the same cross-validation folds as a set of new models, so the original results could be audited rather than cited. Under repeated five-by-five cross-validation the original network reaches an R-squared of 0.281 plus or minus 0.027, against 0.417 plus or minus 0.040 for the best new model, a stacked ensemble.
- Found the study's largest effect by removing it: seven survey items that closely proxy the target score were doing most of the work, and taking them out drops the best model from 0.417 to 0.294 plus or minus 0.030. Predicting the end-of-semester score from start-of-semester answers alone reaches only 0.115 plus or minus 0.051, which is the honest version of the task.
- Automated key data-pipeline segments in Python, including keyword and synonym detection over open-ended definitions and error detection across the concept-map corpus, and restructured the team's optimised Excel workflows.

PUBLICATIONS AND PERSONAL PROJECTS

"Using an Artificial Neural Network to Predict Engineering Identity Shifts in STEM Students." Manuscript in preparation.

Commercial-grade syringe pump

FUSION 360 · 3D PRINTING · EMBEDDED SYSTEMS · C++ ·
ARDUINO · ACCELSTEPPER

Built a commercial-grade syringe pump from scratch using Fusion 360, 3D printing and custom hardware wiring, and wrote and deployed the embedded C++ firmware for precise flow control, an I2C LCD display and the hardware safety interlocks.

TECHNICAL SKILLS

MAIN LANGUAGES

Python for automation and data pipelines, TypeScript with Node.js and Vitest, C++ with the STL, RAII, smart pointers and GTest.

OTHER LANGUAGES

Assembly for MIPS and RISC-V, SQL, HTML, CSS.

TOOLS AND WORKFLOW

Linux, Git, GitHub Actions, Docker, uv, Postman, AWS, Google Cloud Platform.

LIBRARIES AND FRAMEWORKS

React, Tauri, Electron, FastAPI, DuckDB, Polars, TensorFlow, scikit-learn, NumPy, Pandas, Matplotlib.

The figures on this page are checked against the repositories and result files they came from, not carried over from an older draft of this résumé. The Engineering Identity numbers in particular are the audited replication values I produced when I rebuilt the original pipelines myself, not the ones in the initial write-up.