Yonas Atinafu

+1 226-581-5923 | yma9130@nyu.edu | github.com/yonas650 | linkedin.com/in/yonas-atinafu | yonasatinafu.com

Education

University of Waterloo *Sep 2025 – Aug 2027*
Master of Mathematics in Computer Science *Waterloo, Canada*

New York University *Aug 2021 – May 2025*
Bachelor of Science in Computer Science and Mathematics *Abu Dhabi, UAE*

Relevant Coursework: Advanced Topics in AI and Machine Learning, Applied Machine Learning, Machine Learning with Graphs (Stanford University), Natural Language Processing, Computer Networks, Software Engineering, Operating Systems, Computer Systems Organization, Data Science, Algorithms, Data Structures, Linear Algebra, Calculus, Multivariable Calculus, Probability and Statistics, Applied Internet Technology

Technical Skills

Languages: Java, Python, C/C++, SQL (Postgres), JavaScript, HTML/CSS
Frameworks & Libraries: React, Node.js, MongoDB, Flask, pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch.

Experience

Undergraduate Research Assistant May 2024 – July 2025
@AIM Lab, NYU Abu Dhabi *Abu Dhabi, UAE*

- Built EEG-to-graph pipelines across **4** datasets and automated **LOSO** for **28** subjects — reduced training and evaluation time by **45%** and increased throughput **3**×
- Benchmarked GCN, GAT, and GraphSAGE for link prediction and graph classification — standardized entropy features and PLV/correlation edges to cut run-to-run variance and training overhead by **40%**

AI/ML Engineer September – October 2024
@Dismantly *Remote (UAE)*

- Prototyped a 4-stage CV pipeline (preprocessing, detection, postprocessing, inference API) — CPU→GPU migration lowered **p50** inference latency by **60%** and improved throughput by **4**× in bench tests

AI Peer Mentor May 2024 – June 2024
@Design Lab: AI, NYUAD *Abu Dhabi, UAE*

- Mentored a team on ideation, data prep, and evaluation — the team won the **Sustainable Project Award** against 7 other teams

Full-Stack Web Developer May 2023 – September 2023
Project KUWA (Start-up) *Remote (Michigan)*

- Delivered MERN features end-to-end (MongoDB schemas, Express REST routes, React UI, auth/validation) — introduced smoke tests that reduced regression bugs during handoff by **30%**

Publications

PixLift: Accelerating Web Browsing via AI Upscaling — ACM COMPASS 2025 (Toronto)

- Created a browser extension that performs on-device ML upscaling — reduced data usage and decreased average page load time by **7s** via client-side upscaling of low-resolution assets

Projects

Decaf-bytecode *Python/Compilers/VM*

- Built a small language with a handwritten lexer, Pratt parser, AST with source spans, and a semantic resolver
- Implemented a stack-based VM with globals and call frames, plus a CLI, bytecode JSON, and trace mode

Production RAG Pipeline *Python/FastAPI/FAISS/Flan-T5*

- Built a local RAG service with PDF ingestion, chunking, **FAISS** retrieval, and **FastAPI** endpoints
- Returned answers with inline citations and a web UI, with configurable top_k and CPU/MPS/CUDA support

Smart-Nutrition-Assistant *Node/Express/MongoDB/GPT-4 Vision*

- Built a meal-logging app with Passport auth, sessions, and Multer photo uploads
- Used **GPT-4 Vision** and **MongoDB/Mongoose** with Chart.js trend dashboards

LoRA_SFT *Python/PEFT/TRL/HF Transformers*

- Fine-tuned SmolLM-135M with LoRA using PEFT and TRL, reducing perplexity from **1.455** to **1.238**
- Compared with a **GPT-2** baseline (**6.447**) and exported merged weights with an inference script