Yousif Alkhayyat

773-732-3625 | yalkhayy@gmail.com | linkedin.com/in/yalkhayyat/ | github.com/yalkhayyat

Education

University of Illinois Urbana-Champaign Expected Graduation: May 2027

Bachelor of Science, Computer Science

- Advanced Coursework: Computer Architecture, Applied Parallel Programming (CUDA), Database Systems, Artificial Intelligence, Machine Learning, Data Mining, System Programming

Experience

Nokia June 2025 – August 2025

Software Engineer Intern *Naperville, IL*

- Engineered Python tool to parse and merge 1k+ log files into a single view, including multi-threaded GUI.
- Deployed tool to 20+ engineers using Jenkins pipelines, reducing debug time by 30% per bug report.
- Implemented end-to-end system test automation for a 5G Radio Unit using Python and NETCONF/YANG.
- Optimized Radio logging and eliminated simulation environment bugs in C++ Radio Control Software.

University of Illinois March 2025 – Present

Undergraduate Researcher *Champaign, IL*

- Building distributed vector database for billion scale nearest-neighbor search in C++
- Investigated parallel algorithms to accelerate vector database search with Python, C++, CUDA on A40 GPU.
- Profiled SOTA search algorithms, analyzing runtime, throughput, and recall on 1-100M+ scale datasets.
- Presented research findings weekly, comparing performance tradeoffs across GPU search algorithms.

Milwaukee Tool June 2024 – August 2024

Software Engineer Intern *Milwaukee, WI*

- Integrated embedded systems board into Python test framework using object-oriented design.
- Authored 24 integration tests in Python to validate boot loader C++ firmware across 9 hardware variants.
- Saved 100+ engineer hours per release by implementing a CI/CD pipeline to run full system test suite.

Projects

GPT-2 Inference | *C++, CUDA, Nsight Compute*

- Developed custom CUDA kernels for full GPT-2 inference pass, such as attention and matrix multiplication.
- Increased GPU throughput by 40x by implementing KV Cache, cuBLAS, and FlashAttention.
- Analyzed & optimized kernel performance including compute/memory throughput, latency, and occupancy.

Aerohub | *Next.js, React, Javascript, Supabase, PostgreSQL*

- Launched a gaming media platform used by an online community of 30,000+ users.
- Built scalable post sharing and browsing features with file uploads and secure user access control.
- Automatically scraped, parsed, and migrated 10k+ posts into database using Python & OpenAI API.

NOVUS Flight Simulator | *Lua, Roblox Studio*

- Launched multiplayer flight simulator reaching 1.5M+ play sessions and 300k+ players.
- Built interactive frontend and secure backend with high performance client-server replication and anti-cheat.
- Grew a 3k+ member community where I received feedback, bug reports, and managed contributions.

Skills

Programming Languages: Python, C/C++, CUDA, Typescript, JavaScript, Lua, Bash, Verilog
Frameworks & Libraries: React, Node.js, Next.js, Tailwind CSS, Flask, Pytest, Pandas, PyTorch, cuPy, cuDNN
Tools & DevOps: Git, Docker, Supabase, MongoDB, Linux, Nsight Systems/Compute, Jenkins, Jira
Concepts: Backend, Frontend, Fullstack, API, Software Engineering, Artifical Intelligence, Machine Learning, Deep Learning, Neural Networks, CI/CD, Automation, High-Performance GPU Programming, Embedded Systems, Databases, Hardware-In-The-Loop (HIL) Testing, High-Performance Computing (HPC), Parallel Processing