

Nikil Kumar Kandala

Orlando, FL | nkandala3@gatech.edu | 407-929-0900 | linkedin.com/in/nikilkandala | github.com/Nikil456 | nikilkandala.netlify.app

EDUCATION

Georgia Institute of Technology

Atlanta, GA

Bachelor of Science in Computer Science

August 2022 - Expected December 2026

- GPA:** 3.65/4.0, Dean's List, Faculty Honors
- Concentrations:** Artificial Intelligence and Information Internetworks
- Coursework:** Data Structures and Algorithms, Design and Analysis of Algorithms, Computer Systems and Networks, Computer Organization and Architecture, Machine Learning, Artificial Intelligence, Linear Algebra, OOP, Intro to Databases

Georgia Institute of Technology

Atlanta, GA

Master of Science in Machine Learning

August 2025 - Expected May 2027

SKILLS

Languages: Python, C/C++, Java, SQL, JavaScript, Typescript, HTML/CSS

Frameworks/Libraries: PyTorch, TensorFlow, NumPy, Pandas, CUDA, Kubernetes, React.js, Node.js, Playwright

Tools: Git, AWS, Jupyter Notebook, Docker, REST API, Conda, NoSQL, MongoDB, Figma, Firebase, Word, PowerPoint, Excel

PROFESSIONAL EXPERIENCE

Frontline Insurance

Lake Mary, FL

Software Development Engineer in Test Intern

June 2024 – August 2024

- Automated over 150 test cases for an insurance claim center application using TypeScript and Playwright
- Refactored 62 Selenium test scripts into Playwright, improving test execution by 12% and long-term maintainability
- Implemented API-based automation testing which reduced dependency on UI and manual testing by 40%

3D Hologroup Startup

Orlando, FL

Software Engineer Intern

January 2021 – May 2022

- Built a SQL speech db system for CommunicatAR, enabling those with speech disabilities to vocalize through an AR interface
- Developed backend logic to manage context-aware word menus, reducing sentence creation time by about 30s
- Captured patterns 15k+ user interactions using PostgreSQL to generate training data for speech prediction models

PROJECTS

GPU Kernel Benchmarking | Python, CUDA, Triton, Nsight Compute

June 2025 – Present

- Built custom matrix multiplication kernels in CUDA/Triton to achieve up to 80% of cuBLAS GEMM performance
- Profiled memory access with Nsight revealing 30% lower stall rates in Triton auto tiling compared to custom CUDA
- Benchmarked and visualized performance across the kernels comparing throughput, memory access, and launch occupancy

ChatGPT Caching Solution | Python, FAISS, RAG, Vector Databases, Prompt Engineering

June 2025 – Present

- Engineered a caching solution for LLM prompts using sentence-transformer embeddings and FAISS
- Integrated a RAG pipeline to reduce duplicate API calls by 42% saving about \$48 per 1K calls in token usage
- Logged and visualized API call reduction statistics across caching tiers to evaluate real-time cache hit rate of about 63%
- Tuned vector similarity thresholds and prompt chunking strategies to minimize hallucinations during retrieval

ML Ad Ranking | Python, PyTorch, CI/CD, Kubernetes, Docker, CUDA

June 2025 – Present

- Built an end-to-end ml ad ranking system that improved click through rates by 9.4% compared to other models
- Orchestrated a CI/CD workflow automating model retraining/deployment and containerized w/Docker
- Optimized inference throughput by 22% using NVIDIA CUDA and Kubernetes GPU resource allocation
- Reduced end-to-end ETL time by 40% on a 10M-row dataset by optimizing preprocessing pipeline with Apache Arrow

Extropian Sports Analytics – Junior Design Capstone | Python, React, Node.js, Firebase

August 2024 – May 2025

- Placed 2nd Overall at Georgia Tech's CS Junior Design Capstone Expo Spring 2025
- Worked w/startup client Extropian to create data analytics/visualizations for golf wearable tech with expected 3000 users
- Architected the development of a mobile application for Android/iOS with seamless pairing with biometric devices
- Conducted heuristic testing for early minimum marketable features using Figma with 20 testers
- Demonstrated end to end ownership, including scoping with client, architectural design, and feature delivery

TEACHING

Undergraduate Teaching Assistant

Atlanta, GA

CS 4400 Intro to Databases

January 2025 – May 2025

LEADERSHIP AND INVOLVEMENTS

Secretary - Phi Kappa Tau Fraternity, Georgia Tech

January 2025 - Present

Athlete – Georgia Tech Rowing Club

August 2022 – Present