Alexander P. Kranias

U.S. Citizen | 727-457-4433 | alexander.kranias@gatech.edu
alexkranias.com | linkedin.com/in/alexanderkranias | github.com/alexkranias

Education

Georgia Institute of Technology | Atlanta, GA *August 2022 – Present*
Bachelor of Science in Computer Science, GPA 4.00 Expected Graduation, May 2026

Concentrations: Intelligence (AI/ML), Systems and Architecture
Organizations: GT Create-X Incubator Program, GT Start Up Exchange, Georgia Tech Varsity Rowing, GT Student Tour Guides
Coursework: Data Structures**,** Algorithms**,** Computer Architecture, Systems and Networks, Intro to AI, Robotics, Linear Algebra

Skills

Languages: Python, C, Java, CUDA, C++, React, HTML/CSS, JavaScript, Node.js
Tools/Frameworks: Microsoft Azure**,** Docker, Flask**,** PostgresSQL, SQL, ROS2, MATLAB/Simulink, git, Jupyter, Android Studio
OS/Hardware/Networking: Linux, SSH, TCP/UDP protocols, Sensors (IMUs, Potentiometers, Encoders), Raspberry Pi, Arduino

Experience

Embedded Software Engineer Intern *May 2024– August 2024*
Vermeer Corporation Pella, Iowa

- On Automation to Autonomy team creating path planning/controls for mobile agricultural robot products; using Linux Ubuntu.
- Developing ROS2 (C++) occupancy grid functionality and implementing behavior trees in local and global path planning models.
- Simulating motion control systems in Simulink for design option comparisons.

Software Engineer Intern *May 2023 – July 2023*
Raymond James Financial St. Petersburg, Florida

- Developed a full-stack LLM copilot platform using HTML/CSS, JavaScript, Python, and git on the RJ Innovation Team.
- Improved email response time by 80+%; saving 200,000+hrs/week among our 23,000+ employees pending full integration.
- Created a scalable pipeline with Python, Regex, Azure OpenAI API, Outlook API, and test CRM environment for automated data extraction, training, and deployment of personalized fine-tuned GPT-3 models for financial advisors and investment bankers.
- Established relationships with 20+ stakeholders; onboarded team to continue project; developed roadmap with the COO.

Research

HPArch Lab (High Performance Architecture Lab) | C, C++, CUDA, Linux, Docker *August 2023 – Present*

- Formulated cache replacement algorithm for Row Hammer mitigation that de-prioritizes cache blocks originating from hot rows.
- Resulted in up to 200+% improvement to IPC among single and multicore systems for 12+ traces.
- Built cache block memory access trackers in a remote Linux environment integrated into open-source ChampSim repo (C++).
- Developed memory paging and multithreaded round-robin process scheduling projects in a Linux docker environment.
- Assisted with creation of GPU Hardware/Software course; designed Tiled Matrix Multiplication project in CUDA run on cluster.

DuckAI | Python, Regex, OCR, Microsoft Azure *October 2022 – October 2023*

- Built DuckTrack: Accurate Computer Activity Tracking (tool to build multimodal computer agent datasets, 100+ downloads).
- Created ARB: Advanced Reasoning Benchmark for Large Language Models (accepted to MATH-AI Workshop at NeurIPS ’23).

Extracurriculars

Co-Founder, Sideline (GT Create-X Incubator) | Python Flask, LangChain, React, Pinecone, PyTorch *July 2023 – Present*

- AI search for sport recruiting; demoed to 17 colleges so far: 7 D1, 3 D2, 5 D3, and 2 NAIA; met with NBA VP of Strategy.
- Developed multimodal vector search using LangChain, Pinecone VectorDB, Postgres, and Flask CRUD requests.
- Designed vector embeddings to enable “player similarity” search, and “video clip” search using action-timestamp annotations.
- Creating multiagent CV-RL simulation for determining optimal next action in sports using PyTorch, CUDA, and OpenAI gym.

Projects

3dReal (Stanford TreeHacks 2024 Winner) | Swift, CUDA, Python, Firebase, instant-ngp *January 2024 - Present*

- Developed in 48 hours a computer vision iOS app and extension to NVIDIA instant-ngp repo using Firebase, Swift, and CUDA.
- Synchronously generates Neural Radiance Fields (NeRFs) using the cameras of multiple iOS devices.
- Creating CUDA extensions for live 3D video reconstruction via efficient GPU memory access and online model weight fine-tuning.

Bite (HackMIT 2023 Winner - $2K prize) | Raspberry Pi, Python, Flask, Node.js, React, IMU, Computer Vision *September 2023*

- Built in 24 hours an intelligent wearable watch that uses computer vision to track your daily food and nutritional intake.
- Developed band-pass filter to process our IMU data and designed a robust state machine for gesture recognition.
- Created image processing pipeline with REST API and Python Flask server to our deep learning classification/estimation model.