

Kanhaiya Madaswar

kanhaiya@gatech.edu · (404) 593-5424 · Atlanta, GA

linkedin.com/in/kanhaiya38

github.com/kanhaiya38

EDUCATION

- **Georgia Institute of Technology** Atlanta, Georgia
Master of Science in Computer Science Aug 2025 – May 2027
 - Coursework: Systems for Machine Learning, Machine Learning, Computer Networks
- **College of Engineering, Pune (COEP)** Pune, India
Bachelor of Technology in Computer Engineering Aug 2018 – May 2022
 - Coursework: Probability and Statistics, Operating Systems, Distributed Systems, Object-Oriented Programming

SKILLS

- **Languages:** Python, TypeScript, JavaScript, C, C++, SQL, Bash, HTML, CSS
- **Frameworks and Libraries:** Node.js, Express, Flask, Django, React, React Native, LangChain, NumPy, PyTorch, Jest, Vite
- **Databases:** MongoDB, PostgreSQL, MySQL, Redis, Pinecone
- **Tools and Platforms:** vLLM, WebSockets, Kafka, Kubernetes, Docker, GCP, AWS, Bitbucket Pipelines, Linux, Git, Vim

EXPERIENCE

- **UPTIQ** Pune, India
Software Development Engineer – AI Workbench Jun 2022 – May 2025
 - Built a **Retrieval-Augmented Generation (RAG)** framework integrating 10+ data sources and 15+ document types, processing documents over 10 GB in size through incremental loading and semantic chunking for higher retrieval accuracy.
 - Architected a web-based **voice agent framework** with real-time interruption detection, supporting 1,000+ concurrent connections by using WebSockets and Redis for distributed session management across multiple nodes.
 - Created a **drag-and-drop workflow builder** using React Flow, implementing custom node types, edge validation, and real-time state synchronization to enable users to design AI agent orchestration workflows.
 - Developed **20+ responsive, accessible UI components** (forms, dialogs, autocomplete) by extending shadcn/ui, adopted across 3 projects to reduce component development time and ensure consistent UX.
 - **Deployed LLMs** such as Llama 3.1 and NVIDIA Nemotron using vLLM on GCP, optimizing both cost and performance.
 - **Fine-tuned Llama 3.1 and GPT-4o** to improve accuracy by 40% for financial domain-specific tasks.
- *Software Development Engineer – Advisor Lending Platform*
 - Created a robust **synchronization system** for database documents using Kafka and MongoDB, ensuring data consistency across 100+ enterprise tenants and providing real-time observability through a dashboard.
 - **Improved MongoDB performance** by 30% through optimized indexing using the ESR rule and periodic data archival. Reduced latency from 5.2s to 0.98s by refining complex aggregation pipelines and implementing cursor-based pagination.
 - Engineered a feature-rich **React table** component with debounced search, sorting, advanced filtering, pagination, and Excel data export, used across 60% of platform interfaces, reducing development overhead.
 - Built a configurable **dashboard** that dynamically generates custom responsive layouts and **visualizations** (line, bar, pie charts) from JSON specifications using Chart.js, reducing dashboard development time by 60%.
 - Developed a **Document Management System** with granular role-based access control (**RBAC**) and audit logging, improving security and ensuring compliance with SOC 2 standards.
- **Nuance Communications (Microsoft)** Pune, India
Software Engineer Intern Jun 2021 – Aug 2021
 - Developed Azure functions for log analysis and automated **report generation**, enhancing operational efficiency.
 - Increased unit test coverage from 60% to 97% using Jest and **fixed critical bugs**, improving overall code quality.

PROJECTS

- **XV6 Kernel Threads** | Operating Systems, C Feb 2021 – Apr 2021
 - Implemented kernel-level threading within the **xv6 OS** and added a user-level threading library with one-to-one mapping.
 - Introduced **spinlocks** as synchronization primitives to ensure thread safety and support concurrency.
- **Torrent-dl: BitTorrent Engine** | Object-Oriented Programming, Python Oct 2020 – Nov 2020
 - Developed a BEP-3 compliant asynchronous BitTorrent engine in Python using the **asyncio** library to handle concurrent connections and enable faster download speeds.

LEADERSHIP AND AWARDS

- **Software Development Section, COEP:** Led a team of 7 members to design and develop an Induction Portal using MERN stack, simplifying the induction process for university clubs during the COVID-19 pandemic.
- **Runner-Up, CodeQuest 2020:** Secured second place out of 600 participants in a university-wide timed coding contest.