

Harsh Soni

+1 (352) 328-4717 | harsh.soni@ufl.edu | linkedin.com/in/harshh06 | github.com/harshh06 | [Website](#)

Education

University of Florida, Gainesville, FL

Master of Science in Computer Science · GPA: 4.0

Jaypee University of Engineering and Technology, India

Bachelor of Technology in Computer Science and Engineering · GPA: 3.4/4.0 · Top 15%

Work Experience

Attentive.ai, Noida, India

Jun 2023 – Jun 2025

Software Development Engineer | Skills: TypeScript, React, Python, PostgreSQL, Redux, Next.js

- Reduced settings-page load time by **86%** (12s → 1.7s) by implementing React code-splitting with lazy loading, debounced zoning calculations, and IndexedDB caching for **500+ construction enterprise clients**.
- Architected a hierarchical folder system enabling **1,000+** takeoff documents per client using **PostgreSQL recursive queries** and optimistic UI updates, reducing folder navigation latency by **40%**.
- Consolidated **8 fragmented data sources** that required manual aggregation by rebuilding the CSM Dashboard with Chart.js visualizations and **RESTful API** integration, boosting team productivity by **30%**.
- Cut estimation time by **25%** by developing “Typical Takeoffs” template system with **Redux** state management and a backend caching layer, eliminating duplicate data entry for recurring project types.
- Improved client task completion speed by **60%** by building **Beam GPT**, an LLM-powered assistant using **RAG** (Pinecone vector DB + OpenAI GPT-4) for construction document Q&A, serving **500+ users**.
- Led Feature Panel redesign by migrating to Next.js **dynamic routing** with lazy-loaded modals, reducing bundle size by **35%**; mentored 3 junior engineers on component architecture patterns.

Infosys, Bangalore, India

Feb 2022 – Jun 2023

Specialist Programmer | Skills: Python, Flask, React, Java, PostgreSQL, Redis

- Built a React/Flask microservices portal to replace a fragmented toolset used by **30k Apple IS&T staff**; implemented PostgreSQL connection pooling and Redis caching, achieving **99.9% uptime** and reducing release cycles by **20%** via **Jenkins CI/CD**.
- Diagnosed concurrency failures in the **neo4j-graphql-java** query executor and contributed **5 merged PRs** fixing the bugs and enhancing GraphQL schema validation, driving a **10% adoption increase** across internal teams.
- Identified unindexed queries and missing pagination causing slow Flask API responses; resolved via composite index tuning, request pagination, and APM monitoring integration, improving response times by **45%**.

Projects

Automated Multi-Agent PR Reviewer

[GitHub](#)

FastAPI, LangGraph, Pinecone, Google Gemini, PyGithub, Python

- Architected a **multi-agent RAG pipeline** using **LangGraph** that spawns **3 parallel sub-agents** (Security, Performance, Code Quality) via `asyncio.gather()`, each autonomously querying a **Pinecone** vector index with tool calling to post a consolidated Markdown review directly on the GitHub PR thread.
- Designed **incremental vector ingestion** using per-repo **Pinecone namespaces** — re-embedding only changed files on each PR instead of the full codebase, reducing ingestion time by **~90%** on repeat reviews.
- Diagnosed and resolved **429 rate limit** failures caused by 3 concurrent **Gemini API** threads firing simultaneously; implemented randomized exponential backoff jitter, eliminating failures across **100%** of test runs.

AI-Powered Histopathology Scoring System

[GitHub](#) | [Live Demo](#)

Next.js, FastAPI, PyTorch, PostgreSQL, Docker, Hugging Face, GitHub Actions

- Built a diagnostic tool for pancreatic tissue analysis by fine-tuning **ResNet18** (transfer learning on ImageNet weights), achieving **4-metric** pathology scoring and reducing manual analysis time for researchers by estimated **40%**.
- Engineered a client-side upload queue in React with a **FIFO buffer** to handle **50MB+ TIFF batches**, preventing FastAPI server RAM exhaustion by throttling concurrent requests to 3 via **async/await**.
- Deployed containerized **FastAPI** backend to **Hugging Face Spaces** via **GitHub Actions CI/CD**, automating **Docker** builds and achieving zero-downtime deployments.

Technical Skills

- Languages:** Python, TypeScript, JavaScript (ES6+), C++, SQL, HTML/CSS
- Frameworks & Libraries:** React, Next.js, Redux, FastAPI, Flask, Node.js, PyTorch, Chart.js, Tailwind CSS
- Tools & Platforms:** Docker, Git, PostgreSQL, MongoDB, Redis, AWS (Amplify, CloudFront), GitHub Actions, Linux
- Competitive Programming:** Ranked **Top 10%** in Google Kickstart (723/7000); [Codeforces](#) Specialist; solved 300+ [LeetCode](#) problems