

Abhinav Khanna

✉ abhinavkhanna05@gmail.com ☎ (216) 235-1517 🌐 <https://www.linkedin.com/in/abhinavk005/>

🐙 <https://github.com/abhinav1005>

EDUCATION

Case Western Reserve University, Bachelor of Science in Computer Science
Davis United World College Scholar, Veale Snyder Fellow 2022 **GPA: 3.8/4.0**
CodePath Advanced Technical Interview Prep Course, Summer 2023 Cohort

August 2020 – May 2025
Cleveland, Ohio

PROFESSIONAL EXPERIENCE

MRI Software, Software Engineer II (AI Team)

July 2025 – July 2026
Solon, Ohio

- Architected and shipped an **MCP server** to **production**, exposing live Power BI data as **LLM-accessible tools** — enabling AI agents to run filtered DAX queries across 3 datasets.
- Built a **LangGraph pipeline** running 6 **LLM agents** to generate customer-facing insights from large-scale business datasets, reducing insight turnaround time to 4 minutes.
- Designed a **CrewAI multi-agent** system that automates discrepancy detection in accounting workflows, cutting manual review time by **8 hours per month**.
- **Engineered backend APIs** that stream ledger/transaction data from multiple internal APIs and store it in CSV format for the CrewAI reconciliation system.
- Led a complete frontend redesign in **TypeScript/React**, shipping redesigned views to **production**.
- Utilized **AI coding tools** including **Copilot** and **Claude Code** for code refactoring, feature implementation, and design documentation.

MRI Software, Software Engineer Intern (AI Team)

May 2024 –
December 2024
Solon, Ohio

- Built a lease data extraction pipeline using **LangChain**, **Azure**, and **OpenAI LLMs**, experimenting with search, embeddings, and relevance ranking extraction; engineered **600+ Pydantic fields**, improving extraction accuracy by **30%**.
- Designed and optimized **REST APIs**, with **concurrent processing** to handle large documents, cutting **extraction runtime by 70%**.

RESEARCH EXPERIENCE

Case Western Reserve University, Machine Learning Research Assistant

December 2022 –
May 2024
Cleveland, United States

- Built a framework to generate **SYCL** implementations of convolutional algorithms from pseudocode using the **OpenAI API**.
- Contributed to a dynamic mapping framework for convolutional algorithms to optimize utilization across different layers of **Convolutional Neural Networks (CNNs)**, extracting details from CNN structures using PyTorch, ONNX, and CNN inference.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, Java, C/C++

AI / Agents: LangChain, LangGraph, CrewAI, MCP, OpenAI & Azure OpenAI APIs, RAG, Embeddings

Backend / Infra: FastAPI, REST APIs, Docker, Kubernetes, PostgreSQL, Okta Auth, Git, CI/CD

Frontend: React, TypeScript, HTML/CSS | **Tools:** Claude Code, GitHub Copilot, Azure, Figma

PROGRAMMING PROJECTS

KnowledgeOS 📄

Built a local-first AI-powered knowledge base using **FastAPI**, **LangGraph**, and **ChromaDB** to enable natural-language search over personal notes and documents; designed the multi-step RAG pipeline, and implemented real-time file indexing via a watchdog watcher with SSE.

Notes Repo Project 📄

Built a static knowledge management web app that converts Markdown notes into a searchable, tag-based website with automated metadata generation via **CI/CD**. Implemented dynamic UI with **React and TypeScript**, automated content processing using **Python + GitHub Actions**, and **deployed** 📄 on Vercel.

Web Server Analysis (Python) 📄

Developed Python scripts to collect, analyze, and visualize data for web server performance analysis, focusing on metrics like Round Trip Times (RTT), Time to First Byte (TTFB), and Packet Loss across 35 websites.