



Hasnain Ali

Senior AI Engineer

6+ years shipping software, 5+ in AI, starting with RLHF training data at Turing. I build AI voice agents and document AI, and the Python backends they run on.

Remote from Pakistan (UTC+5). Open to relocation. US roles need visa sponsorship.

codingwithhasnain@gmail.com

hasnainali.vercel.app

LinkedIn: <https://www.linkedin.com/in/hasnainali4/> GitHub: <https://github.com/HasnainAli47>

Experience

Jul 2026 – now

Independent contractor,
remote

Freelance AI Engineer

Building Mentanima, a legal copilot that cites its source for every answer, and Automation Studio, four agent workflow demos you can run in your browser.

Nov 2024 – Jun 2026

Full-time

Senior Software Engineer at Devsinc

Led eight AI engineers across legal, healthcare and real-estate products, including retrieval across about five million documents.

Senior AI Lead Engineer on Smart Advocate: led its document AI, from OCR, classification and structured extraction to knowledge-graph retrieval, with a 40% accuracy gain over standard RAG.

Mar 2022 – Nov 2024

Full-time

Associate Software Engineer at Amrood Labs

Django APIs on PostgreSQL and AWS for fleet and employee management software.

Contract analysis on the OpenAI API that cut manual review by about 60%.

Earlier

Jan 2022 – Aug 2023

Part-time

Django Developer at Arbisoft

Apr 2021 – Feb 2022

Full-time, remote

Software Engineer, POD lead at Turing

Led an annotation team producing RLHF training data for a frontier AI lab.

Mar 2020 – Mar 2021

Founder and lead engineer at 2ndPlace

Built a Django and React travel platform: 3,500+ active users, 15,000+ listings.

Skills

Agents: LangGraph, Tool calling, Human-in-the-loop, MCP, Voice AI, Twilio, ElevenLabs

Models and tuning: OpenAI, Anthropic, Gemini, LoRA fine-tuning

Retrieval: RAG, KAG, Neo4j, pgvector

Backend: Python, FastAPI, Django, PostgreSQL, GCP, AWS

Selected systems

Speculo.ai Senior AI Communication Engineer

AI that calls, texts and emails real-estate leads, then hands the qualified ones to the client's human agents. It contributed to \$2M+ in annual revenue.

My part: the Python backend that times every call, text and follow-up.

117K+ emails and SMS, plus about 10K AI voice calls, in one 6-hour run.

How the pipeline works Python, LangGraph, OpenAI, Twilio, ElevenLabs, GCP Pub/Sub, pgvector

Smart Advocate Senior AI Lead Engineer, at Devsinc

Document AI for legal and medical case files, shipped into a platform used by 20,000+ legal professionals.

I led the AI work: OCR, classification and extraction, and the knowledge-graph retrieval.

40% accuracy gain over standard RAG.

Knowledge graph plus a fine-tuned model. Answers scored by a GPT-4-class judge and BLEU/ROUGE, retrieval by precision@k and recall@k.

How it was measured Python, LangGraph, RAG, KAG, Neo4j, LoRA

Mentanima In development

A legal copilot: document search backed by a knowledge graph. Every answer cites its source, or says it can't answer.

Python, RAG, KAG, Neo4j, LangGraph

Automation Studio

Four agent workflow demos you can run in your browser.

Try them Python

toon-mcp-server

An open-source MCP server on PyPI. It converts JSON and system prompts to TOON, a compact format that uses fewer tokens.

PyPI Python, MCP

Education

BS Computer Science, COMSATS University Islamabad

Sources

Speculo.ai: hasnainali.vercel.app/blog/gcp-pubsub-ai-outreach

Smart Advocate: hasnainali.vercel.app/blog/kag-vs-rag

Automation Studio: hasnain-automation-studio.vercel.app

toon-mcp-server: pypi.org/project/toon-mcp-server/