

Hasan Mohd Hussain

Machine Learning Engineer | Generative AI, LLMs, RAG & Agentic AI
Mumbai | 7400341331 | hasan98209@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Machine Learning Engineer with a **Master's in Computer Engineering**, specializing in **Generative AI, LLMs, RAG** and **agentic AI** systems. Experienced in building, evaluating and deploying ML applications with **Python, FastAPI, Docker** and **Azure OpenAI**.

EDUCATION

The University of Texas at Dallas

Master of Science (MS) in Computer Engineering | GPA 3.74 / 4.0

Relevant Coursework - Machine Learning, NLP, Data Structures & Algorithms, Database Design, Computer Architecture, Embedded Systems

Richardson, TX

Aug 2024 – Aug 2026

University of Mumbai

Bachelor of Engineering (B.E.) in Computer Engineering | GPA 7.88 / 10.0

Mumbai, India

Aug 2017 – Jul 2021

SKILLS

AI / ML: Machine Learning, Deep Learning, NLP, Computer Vision, CNNs, Transformers, Supervised Learning, Fine-Tuning, Hyperparameter Tuning, Model Evaluation

Generative AI & LLMs: LLMs, Generative AI, RAG, Agentic AI, AI Agents, Multi-Agent Systems, Prompt Engineering, Embeddings, Semantic Search, LLM Evaluation, LLM-as-a-Judge, Knowledge Graphs

LLM Frameworks & APIs: LangChain, LangGraph, OpenAI API, Azure OpenAI, Gemini API, Hugging Face, Vector Databases

ML Libraries: PyTorch, Scikit-learn, Hugging Face Transformers, NumPy, Pandas, Matplotlib, Seaborn

MLOps & Deployment: Model Deployment, Model Serving, FastAPI, Flask, REST APIs, Docker, Kubernetes, CI/CD, RabbitMQ

Cloud Platforms: AWS, GCP, Azure

Databases: PostgreSQL, MySQL, MongoDB, SQLite, ChromaDB, RDF4J, SPARQL

Programming Languages: Python, SQL, C++, C#, JavaScript

Tools & Practices: Git, GitHub, GitLab, Linux, Shell Scripting, pytest, Playwright, Claude Code, Data Structures & Algorithms, OOP, Agile/Scrum

WORK EXPERIENCE

Mphasis

Machine Learning Engineer

New York, United States

May 2025 – Aug 2025, Jun 2026 – Aug 2026

- Developed an autonomous **LLM-based QA agent (Python, Flask, Playwright)** that browser-tests a production mortgage-income dashboard end-to-end and grades results against ground-truth keys via a **multi-provider LLM layer** (Claude, GPT-4o, Gemini, self-hosted Qwen).
- Applied **LLM-as-a-judge evaluation** to replace manual QA, cutting per-case test time by **95%** (minutes to seconds).
- Built a **synthetic data generator** for **5 financial document types** (W-2, paystub, VOE, Schedule C, P&L) with computed answer keys, scaling **evaluation coverage to 200 cases per run** with zero manual authoring.
- Built a **RAG system** ingesting **PDFs** into **ChromaDB** with a **FastAPI** dashboard and **Azure OpenAI LLM**, improving retrieval by **30%**.
- Optimized **Python backend logic** for full-stack AI apps (**FastAPI, LangChain, AWS**), improving **API response performance by 20%**.

Dot9 Games

Game Developer Intern

Mumbai, India

Apr 2024 – Aug 2024

- Built gameplay systems in **Unity** and **C#** from game design documents, shipping mechanics that raised **user engagement by 25%**.
- Collaborated with developers, artists and designers in **Agile/Scrum** sprints, using **Git** for version control and iterative delivery.

Shree Infotech

Software Engineer

Mumbai, India

Apr 2023 – Mar 2024

- Developed **Python/Flask backend services** and internal tools that automated manual team workflows and reduced operational overhead.
- Containerized applications with **Docker** to standardize development and deployment environments across the team.

Nilee Games

Game Developer Intern

Mumbai, India

Nov 2022 – Mar 2023

- Built a **2D physics-based driving game** in **Unity/C#** within **two weeks**, contributing to **500K downloads** in its first month.
- Increased **user engagement by 30%** by tuning gameplay mechanics: terrain design, vehicle upgrades and a progression system.

TECHNICAL PROJECTS

RAG Document QA System with Gemini Integration

- Implemented a **Retrieval-Augmented Generation (RAG)** pipeline with **LangChain, ChromaDB** and **Hugging Face embeddings** for semantic search, document chunking and PDF text retrieval.
- Integrated the **Google Gemini LLM** to generate context-aware answers, improving query accuracy by **35%**.

LangGraph Agentic Nutrition System

- Built a **multi-agent** nutrition AI system (**LangGraph, GPT-4o-mini**) with **ReAct agents** and a critic-revise loop using score-based routing.
- Streamed live Think-Act-Observe traces per agent via **Flask SSE**, generating a 7-day meal plan, supplement protocol and safety-audited report.

Deep Learning Image Classification (MLP vs CNN) – PyTorch

- Developed **MLP** and **CNN** architectures in **PyTorch** on **MNIST** and **CIFAR-10** using **SGD/Adam, BatchNorm** and **Dropout**.
- Ran **30+ hyperparameter experiments** (learning rate, batch size, regularization): **99%+ accuracy** on **MNIST**, **78%** on **CIFAR-10**.

CERTIFICATIONS

Anthropic: Claude Code 101, Claude 101, Introduction to Claude Cowork

DeepLearning.AI: Supervised Machine Learning: Regression and Classification, AI For Everyone

IBM: Machine Learning with Python, Building AI Agents and Agentic Workflows | **Udemy:** The Complete SQL Bootcamp