

# Ahad Ahmad Khan

+91 9828895893 | ahadkhan5547@gmail.com | Ahad Khan | Ahad Khan

## EXPERIENCE

### Capgemini

Dec. 2024 – Present

#### Generative AI Engineer

India

- Designed and deployed **agentic AI workflows** using AG2, enabling conversational orchestration and execution agents
- Built a multi-role AI chatbot framework with coaching, reasoning, and task execution agents for enterprise use cases
- Developed backend services using **FastAPI** with asynchronous patterns, REST APIs, and scalable microservices
- Implemented industry-standard **Retrieval-Augmented Generation (RAG)** via Azure AI Search and Azure Blob Storage
- Applied vector search and embeddings for semantic retrieval and context grounding in AI responses
- Performed structured prompt engineering to optimize model relevance, reduce hallucination, and improve output quality
- Worked with cloud platforms (Azure), containerization (Docker), version control (Git), and CI/CD awareness
- Collaborated with cross-functional teams to translate product problems into AI-first solutions for business impact

## PROJECTS

### AI Gym & Activity Memory System

- Built an AI-powered conversational system to log gym workouts and daily activities using natural language input
- Implemented historical memory retrieval to answer queries such as “What did I train last week?”
- Used Gemini Flash for intent extraction and embedding-based semantic search for recall
- Designed REST APIs using FastAPI for structured storage and retrieval of activity data

### Document Intelligence & RAG System

- Developed a Retrieval-Augmented Generation (RAG) system for querying enterprise documents
- Implemented document ingestion, chunking, and embedding pipelines for semantic search
- Integrated Azure AI Search with vector embeddings to ground LLM responses in source documents
- Improved answer accuracy and reduced hallucination by enforcing document-based context retrieval

### AI Knowledge Assistant with Tool Calling

- Built an AI assistant capable of reasoning and invoking external APIs to complete user tasks
- Enabled dynamic tool selection and execution using MCP-style context passing
- Developed backend services with FastAPI to handle tool routing and response aggregation
- Focused on reliability, error handling, and controlled output generation for enterprise use

## EDUCATION

### GL Bajaj Institute of Technology & Management

2020 – 2024

B.Tech in Information Technology

Greater Noida, Uttar Pradesh

## SKILLS

**Languages:** Python, SQL, Java, JavaScript

**GenAI:** LLMs, RAG, prompt engineering, embeddings, vector search

**Frameworks/Tools:** FastAPI, AG2, Azure AI Search, Azure OpenAI, Docker, Git

**Cloud & Platforms:** Azure, Azure Blob Storage, REST APIs, CI/CD principles

**Concepts:** NLP, retrieval systems, agent orchestration, scalable API design