Babar Ali

AI Engineer | Generative & Agentic AI | Machine Learning | Forward Deployed AI Engineering

LinkedIn | +92-342-1015978 | Portfolio | maharbabar79@gmail.com | GitHub

Professional Summary

AI Engineer specializing in **Generative AI, Large Language Models and Agentic AI systems**, with a track record of taking AI from concept to production. Designed, built and deployed **six end-to-end AI systems**: a tool-calling customer-support agent that handles refunds, a governed text-to-SQL analytics assistant, a permission-aware RAG assistant, an LLM document-extraction pipeline, a Model Context Protocol (MCP) security gateway for AI agents, and a real-time IoT platform, backed by **1,000+ automated tests** and frozen held-out evaluations. Specialized in **reliable, measurable and safe AI**: deterministic guardrails around the model, human approval for high-risk actions, prompt-injection defenses and tamper-evident audit trails. Currently delivering AI-powered automation for clients as a remote AI Engineer. Certified by **IBM, DeepLearning.AI, Stanford Online, Google, Duke University and Vanderbilt University**.

Technical Skills

Programming & Development:

- **Languages & Frameworks:** Python, SQL, FastAPI, Pydantic, Scikit-Learn, TensorFlow, Keras, PyTorch, Hugging Face Transformers, OpenCV, Streamlit, React.
- **Tools & Libraries:** Pandas, NumPy, Matplotlib, Seaborn, SQLAlchemy, Alembic, sqlglot, REST APIs, Server-Sent Events, Docker, Docker Compose, Git, GitHub Actions, CI/CD Pipelines, pytest, mypy, OpenTelemetry.
- **Database Systems:** PostgreSQL, Microsoft SQL Server, Vector Databases (pgvector with HNSW), Neon Serverless Postgres.

Data Science & AI:

- **Core Skills:** Machine Learning, Deep Learning, Generative AI, LLMs (Gemini, GPT-OSS), Natural Language Processing (NLP), Computer Vision, Statistical Analysis, Regression Analysis, Prompt Engineering, Model Evaluation.
- **Agentic AI & LLM Engineering:** AI Agents, LangGraph, LangChain, Model Context Protocol (MCP), Tool Calling, Retrieval-Augmented Generation (RAG), Text-to-SQL, Structured Outputs, Human-in-the-Loop Workflows, n8n.
- **Evaluation & AI Safety:** Held-out Evaluation, Retrieval Metrics (Recall@k, MRR), Ablation Studies, Prompt-Injection Defense, Guardrails, Access Control in Retrieval, Audit Logging.
- **Applications:** AI-Driven Process Automation, Document Intelligence, Conversational AI, Business Intelligence (Power BI), Data Visualization, IoT Analytics.

Cloud & MLOps:

- **Platforms:** AWS (SageMaker, Lambda, Bedrock, EC2), Streamlit Cloud, Heroku.
- **Practices:** MLOps, LLMOps, AIOps, Cloud Machine Learning, MLflow, Orchestration Frameworks, Observability and Tracing.

Management & Operations:

- Forward Deployed Engineering, Requirements-to-Delivery, AI Solution Design, Technical Documentation, Stakeholder Communication, Knowledge Sharing.

Communication & Collaboration:

- Strong in technical writing, effective communication, stakeholder engagement and team collaboration, fostering a productive and inclusive environment.

Languages:

- Fluent in English and Urdu.

Experience

AI Engineer
SADA DIGITAL TECHNOLOGY

Doha, Qatar (Remote) 06/2026 -- Currently

- **AI-Powered Automation:** Build AI-powered automation and data-driven solutions integrated into custom client software, aligning technology with business objectives to enable smarter decisions and scalable growth.
- **End-to-End AI Delivery:** Take AI features from business requirement to production: scoping the use case, designing the data flow, and integrating machine learning models and LLM APIs into existing applications.
- **Business Alignment:** Translate client objectives into practical AI capabilities that fit real workflows and support data-driven decision-making as the business scales.

Data Scientist
VAKIAL TECH

Multan, Pakistan 02/2026 -- Currently

- **Machine Learning Solutions:** Build and deploy machine learning models for data analysis and business intelligence.
- **Business Intelligence:** Develop interactive Power BI dashboards and data visualizations that turn data into insights for stakeholders.
- **AI Automation Pipelines:** Design and implement AI automation pipelines that streamline recurring data workflows.

Campus Brand Ambassador
DATACRUMBS

MNS University of Agriculture, Multan 06/2025 -- 09/2025

- **Community Growth:** Promoted DataCrumbs' AI and Data Science live sessions and research projects across academic and tech communities, connecting students with industry learning resources.

Projects

- **TicketPilot: Agentic Customer-Support AI** *(LangGraph, FastAPI, PostgreSQL, Groq)* [Live Demo | GitHub]
  Engineered a tool-calling LLM agent that resolves e-commerce support tickets end to end, including refunds, address changes and order tracking. The model only reads and proposes; a deterministic policy engine decides, and refunds above $50 pause for supervisor approval through LangGraph interrupts with a durable Postgres checkpointer. A prompt-injection guard caught **7/7 attack tickets with zero false positives** across 74 legitimate tickets, idempotent writes make double refunds impossible, and a reply checker blocks invented amounts or order numbers.
- **QueryWise: Governed Text-to-SQL Analytics** *(sqlglot, PostgreSQL, pgvector, Gemini, FastAPI)* [Live Demo | GitHub]
  Built a natural-language analytics assistant that converts business questions into verified SQL, tables, charts and summaries whose numbers are checked against the query result, over a 45k-row e-commerce warehouse. Layered safety: a SQL AST guard **blocked 57/57 malicious queries with zero false blocks**, a read-only role restricted to curated views without personal data, an EXPLAIN cost ceiling and a 5-second timeout. Generation is grounded in a semantic layer of approved metrics with embedding-based schema linking, bounded self-repair and clarifying questions for ambiguous requests.
- **LedgerLens: LLM Invoice Intelligence with Human Review** *(Gemini, FastAPI, PostgreSQL, HTMX)* [Live Demo | GitHub]
  Developed a document-intelligence pipeline that extracts invoices into validated structured data, auto-approves only clean invoices and routes anomalies to a reviewer dashboard. On a frozen 50-invoice held-out test it achieved **zero false auto-approvals**, caught 11/11 planted anomalies, reached line-item F1 of 1.00 with 100% valid structured output, at $0.0019 per invoice and 3.4 s median latency.
- **PolicyPilot: Permission-Aware RAG Assistant** *(pgvector, Gemini Embeddings, FastAPI)* [Live Demo | GitHub]
  Architected an enterprise HR-policy assistant that answers with citations, declines questions the handbook does not cover, and enforces access control inside retrieval. Held-out results: **zero permission leaks**, 97.1% citation precision, 0.906 answer accuracy and 8/8 unanswerable questions correctly declined. A retrieval ablation across full-text, vector, hybrid and reranking selected pgvector HNSW, raising Recall@1 from 0.562 to 0.906.
- **ToolGate: Security Gateway for AI Agents** *(Model Context Protocol, FastAPI, PostgreSQL)* [GitHub]
  Designed an MCP server that gives AI assistants governed access to the four systems above through a single entry point, with deny-by-default scopes per client, rate limits, a kill switch and approvals bound to the exact call. Every decision is written to a **hash-chained, tamper-evident audit log**; the policy engine is verified by 29 table-driven cases and adds only 33 ms median overhead. An evaluation regression gate fails CI whenever a prompt or model change degrades any system.
- **FreeSpot: Real-Time IoT Classroom Availability (Final Year Project)** *(ESP32, MQTT, FastAPI, PostgreSQL, React)*
  Built an end-to-end IoT platform that shows students and staff, live on their phones, whether each room is free, a class is running, a teacher is late, or a class was cancelled or moved. ESP32 nodes with privacy-preserving mmWave radar (no camera) and RFID check-in stream over MQTT to a FastAPI status engine and a React PWA with **307 ms p95 event-to-dashboard latency** (lab measurement). Safety rules live in a pure, fully tested engine (234 backend tests), and a Random Forest occupancy model is benchmarked against a calibrated threshold rule using session-wise cross-validation.

Education

Bachelor of Science in Information Technology
MNS UNIVERSITY OF AGRICULTURE
Multan, Pakistan
09/2023 -- 05/2027 (Expected)

Certifications

- **Generative AI Engineering Professional Certificate**, IBM.
- **AI Developer Professional Certificate**, IBM.
- **AI Agent Developer Specialization**, Vanderbilt University.
- **Deep Learning Specialization**, DeepLearning.AI.
- **Machine Learning Specialization**, DeepLearning.AI & Stanford Online.
- **TensorFlow Developer Professional Certificate**, DeepLearning.AI.
- **MLOps: Machine Learning Operations Specialization**, Duke University.
- **Advanced Data Analytics Professional Certificate**, Google.
- **Data Science Professional Certificate**, IBM.
- **Mathematics for Machine Learning and Data Science**, DeepLearning.AI.
- **Computer Vision Specialization**, University of Colorado Boulder.
- **AI for Medicine Specialization**, DeepLearning.AI.