Bhuman Pandita

AI Engineer | Applied AI | Agentic Systems | GenAI

bhumanpandita.com | bhumanpandita3@gmail.com | linkedin.com/in/bhumanpandita/ | github.com/BhumanPandita

Summary

AI Engineer with hands-on production experience building and deploying **multi-agent LLM systems, RAG pipelines, and agentic AI workflows** at scale. Proven track record of owning end-to-end AI system design, from architecture and experimentation through production deployment, stakeholder delivery, and measurable business impact. Strong foundation in **Azure OpenAI, Semantic Kernel, MCP servers, LangChain/LangGraph, FastAPI**, and async Python systems. Published researcher and ASCE award recipient.

Experience

Machine Learning Engineer (Assistant Manager), **IndiGo (InterGlobe Aviation)** Jul 2026 – Present

- Promoted to **Machine Learning Engineer (Assistant Manager)** in Jul 2026.
- Built a **fully offline cabin-crew chatbot prototype** using **quantized Qwen models from Hugging Face** for local inference, targeting on-device deployment on tablets and mobile phones; converted source documents through **OCR into Markdown** and preprocessed the text for in-flight reference queries.
- Worked on an **engineering productivity tracker** integrating **Jira and GitHub** activity to assess task progress and the relevance of code contributions to assigned work.

IAL - Digital & Data Science Program, **IndiGo (InterGlobe Aviation)** Jun 2025 – Jul 2026

- Owned end-to-end design and **production deployment of a multi-agent MCP server** (Semantic Kernel, Azure OpenAI, FastAPI, FastMCP) automating BluChip loyalty profile updates across multiple internal databases, **reduced TAT from 12 days to under 1 minute** with PII compliance and OCR-based government ID verification, directly improving operational throughput for IndiGo's loyalty team.
- Architected and delivered a **hybrid AI Loyalty Fraud Detection system** integrating MS Graph API, Azure Cosmos DB, 3 Capillary APIs, and Navitaire, 6 detection modules including **LLM-driven geographic feasibility reasoning (zero hardcoded thresholds)** and self-connecting flight protection; sole engineer on end-to-end POC delivered in 3 months with a pivot from microservices to async monolith for production reliability.
- Prototyped an **Edge AI IoT asset tracker** using a custom CNN vision model on Raspberry Pi with Go + Python backend, validating embedded AI inference for digitizing non-digital airline ground assets, presented to senior engineering leadership.
- Built an **NLP/LLM invoice categorization pipeline** using embedding retrieval and LLM classification to auto-tag MRO Non-Routine Card tasks across heterogeneous vendor invoices, supporting lessor compliance and audit automation.
- Built a **clustering-based MRO cost comparison tracker** to group comparable maintenance tasks across providers servicing the same aircraft type, identify pricing differences, and support vendor negotiations; deployed via **SharePoint** for the engineering team.

AI Engineer Intern, **Swostitech Solutions** Dec 2024 – Jan 2025

- Built and optimized a **production RAG pipeline** over 1,000+ enterprise documents, benchmarked OpenAI GPT-4, LLaMA-3, and Mistral-7B across chunking strategies and embedding models, achieving measurable gains in answer relevance and retrieval grounding.
- **Reduced end-to-end RAG inference latency by 30% (p95)** via embedding model quantization, flat-to-HNSW vector store migration, and query-layer caching, enabling higher concurrency without hardware scaling.

Data Analyst Intern, **Rigi**, Bangalore Jul 2024 – Dec 2024

- Automated end-to-end transaction reporting for 100+ creators (pandas, openpyxl, cron), **saving ∼20 finance-team hours/week (50% effort reduction)** and surfacing payout discrepancies automatically, eliminating a previously manual weekly process.
- Optimized PostgreSQL queries (window functions, CTEs, index tuning) **cutting Metabase dashboard load time by 30%**, directly improving analyst productivity and creator-economy insight delivery.

Projects

RAG Red-Teaming Pipeline | *Python, Gemini, Groq/LLaMA-3.3-70B, FastAPI, Next.js* 2025

- Designed and built a **4-stage adversarial AI evaluation framework** (attacker → executor → target RAG → evaluator) to systematically probe an HR chatbot across **13 attack categories**, prompt injection, roleplay jailbreak, encoding obfuscation, RAG context poisoning, and more.
- Implemented a **hybrid AI observability scoring system** (0.7× LLM semantic judge + 0.3× heuristic signals, CRITICAL/HIGH/MEDIUM severity tiers) with real-time SSE-streaming campaign dashboard in Next.js.

CycleGAN Image Dehazing | *PyTorch, OpenCV, Deep Learning* Mar – May 2024

- Designed and trained a **CycleGAN-based image dehazing model** from scratch, unpaired hazy-to-clear image translation using cycle consistency loss and adversarial training, achieving **40% improvement in perceptual clarity** across diverse real-world haze conditions.

CNN Sound Classification | *PyTorch, Librosa, Mel spectrograms* Jan – Mar 2024

- Benchmarked ResNet, DenseNet, MobileNetV2 and a custom 6-layer CNN on **6,000+ audio samples** across 13 classes; noise reduction preprocessing on Mel spectrograms **improved classification accuracy by 15%**.

Education

Birla Institute of Technology and Science, Pilani Pilani, India

*B.E. Civil (Hons.) with **Minor in Data Science**, CGPA 7.8/10* *Aug 2021 – May 2025*

Relevant Coursework: Generative AI, Deep Learning, Machine Learning, Foundations of Data Science, Applied Statistical Methods.

Publications & Awards

- Nanavaty, A., Sharma, R., **Pandita, B.** et al. *Integrating deep learning for VQA in Agricultural Disease Diagnostics: Case Study of Wheat Rust.* **Sci Rep** 14, 28203 (2024). doi:10.1038/s41598-024-79793-2
- Rodríguez-Narvaez, O. M., **Pandita, B.**, Goyal, O., et al. *Efficacy of biochar as a catalyst for a Fenton-like reaction.* **J. Water Process Eng.**, 70, 107014 (2025).
- Morales-Verdin, B.S., Goyal, O., **Pandita, B.**, et al. *Optimizing the upcycling of microplastics to a carbon-based adsorbent for water treatment: An integrated experimental and computational approach.* **Chem. Eng. Sci.** 325, 123409 (2026). doi:10.1016/j.ces.2026.123409
- **First Prize**, ASCE International Conference (CISSC-2025), UG Event on Challenges and Innovations for Sustainable Smart Cities, ASCE India Section.

Technical Skills

Languages: Python, SQL (PostgreSQL), R

LLM & Agentic AI: Semantic Kernel, LangChain, AutoGen, FastMCP, MCP Servers, Multi-Agent Orchestration, RAG, Prompt Engineering, Function Calling, Structured Outputs, Guardrails

Models & APIs: Azure OpenAI, OpenAI, Google Gemini, Mistral, Groq, Embedding Models

ML / DL: PyTorch, TensorFlow, CNNs, GANs, Computer Vision, NLP

BI & Visualization: Tableau, Power BI, Metabase, Postman, Bruno

Cloud & Infrastructure: Azure OpenAI, AWS (familiar), Docker, production pipeline deployment & maintenance