

# Wahaj Uddin Ahmed

+92 3311321120 | [wahaj.uddin492@gmail.com](mailto:wahaj.uddin492@gmail.com) | [linkedin](#) | [github portfolio](#)

## PROFESSIONAL SUMMARY

AI Engineer skilled in building production-grade **digital humans**, multimodal **agentic systems**, and conversational AI. Proficient in troubleshooting Python applications, integrating **AI/ML models**, and rapidly adapting to new technologies. Experienced with **Unreal Engine 5**, **FastAPI**, scalable AI architectures. Led the development and international launch of **Tafsol Genie** at **GITEX Global 2025 (Dubai)**, with a deep interest in **embodied AI** and robotics simulations.

## TECHNICAL SKILLS

**AI/ML:** Python, LLMs, Deep Learning, Computer Vision, NLP, Generative AI, RAG, Agentic AI, Multi-Agent Systems, LoRA/QLoRA, Chain-of-Thought(CoT), MCP, Semantic Search, Prompt Engineering  
**Frameworks & Libraries:** LangChain, LangGraph, Llama, CrewAI, FastAPI, Django, TensorFlow, Keras, PyTorch, SkLearn, Transformer, Ollama, Hugging Face, OpenAI API, Anthropic API, Whisper, Firecrawl, RESTful APIs  
**Databases:** Vector DB (FAISS, ChromaDB, Qdrant, Weaviate), PostgreSQL, SQL, Neo4j Graph DB  
**Tools:** AWS, Git, Docker, Weights and Biases (WandB), CI/CD, ElevenLabs, Nvidia Omniverse  
**Soft Skills:** Problem Solving, Leadership, Cross-Team Collaboration, Communication

## EXPERIENCE

- AI Engineer** Dec 2024 – Present  
*Tafsol Technologies* Karachi, Pakistan
- Led the end-to-end development of **Tafsol Genie**, a real-time AI Digital Human platform using **Unreal Engine 5.6, Audio2Face (NVIDIA), OpenAI, Claude, Whisper, ElevenLabs, and FastAPI**.
  - Engineered **C++ WebSocket integration** with FastAPI, reducing response latency from **60s to 5s (91.7%)** and later to **under 3s**, achieving near-human response speed.
  - Optimized multimodal AI pipelines with fallback and reliability improvements, increasing uptime by **40%**.
  - Developed **AI real estate insights engine, voice-based sales/finance assistants, and an AI virtual try-on system** for apparel visualization
- Junior Developer (Remote)** Sep 2024 – Nov 2024  
*ECOM Specialists LLC* Plano, TX, USA
- Enhanced data flow automation using **IBM Sterling B2B tools**, improving reliability across client workflows.
- Software Engineer (Founding Team)(Remote)** Jan 2022 – Sep 2024  
*Techno Sphere Orbit* London, England
- Engineered a Python-based image processing engine that automated watermark application for millions of assets, reducing manual Photoshop labor by 100% and achieving near-instant batch processing
  - Integrated Meta APIs to automate social engagement monitoring for 20+ client accounts, delivering real-time performance metrics and reducing manual reporting time by 90%.
  - System Performance Optimization: Refactored backend Python logic and optimized relational database queries, resulting in a 30% improvement in application load speeds and 95% system reliability.

## PROJECT

- Negotiation Bot** | *LangChain, OpenAI, Decision-Making Agents, Guardrails, FastAPI*
- Built an AI negotiation bot that increased overall platform profitability by enabling dynamic, conversation-based pricing while maintaining guaranteed minimum margins.
- AI Powered Property Finder** | *LangGraph, SerpAPI, Firecrawl, OpenAI*
- Developed an AI assistant that helped users and agents find residential and commercial properties, reducing search time by **40% through personalized recommendations**.
- Retail AI Analytics Assistant** | *FastAPI, Deepseek LLMs, Plotly, Faiss, Embedding Model, RAG*
- Built an AI analytics assistant for retail stores that automated sales insights, generated real-time charts from natural language, and improved decision-making speed by **35%**.

## EDUCATION

**NED University of Engineering and Technology** Karachi, Pakistan  
*Bachelor of Engineering in Software Engineering* Oct. 2020 – Aug. 2024