

EDUCATION

Daffodil International University (DIU) <i>BSc in Computer Science and Engineering</i>	Savar, Ashulia 2022 – 2026
Dhaka Commerce College <i>HSC Science</i>	Dhaka, Mirpur 2019 – 2021
Monipur High School and College <i>SSC Science</i>	Dhaka, Mirpur-2 2009 – 2019

EXPERIENCE

Daffodil Computers Limited <i>AI Software Developer</i>	On-site, Dhaka, Bangladesh Feb 2026 – Aug 2026
- Engineered and deployed two AI modules within a production HR platform serving internal teams. Employee Engagement and Retention Analysis: Developed a predictive system to evaluate workforce engagement and identify at-risk employees, validated at 97% precision against historical HR outcomes. <i>Stack: Python, FastAPI, Ollama, PostgreSQL, HTML/CSS/JS.</i> AI Response Console: Engineered a RAG-based HR knowledge system using PGVector for semantic retrieval over internal documentation. Validated system reliability through an automated Python pipeline of 24,000+ test queries, achieving 99% accuracy at under 400ms average response time. <i>Stack: Python, FastAPI, Ollama, PostgreSQL, PGVector.</i>	

PROJECTS

PhotoDump.ai	GitHub
- Self-hosted, private photo-sharing platform featuring password-protected events and guest moderation. - Engineered a decoupled dual-server architecture using Docker Volumes for data persistence and a standalone GPU microservice for AI face recognition via cosine similarity, tested at 96.87% match accuracy on a 10,000+ image benchmark dataset with sub-second lookup latency. <i>Stack: Python 3.12, FastAPI, PostgreSQL 16, DeepFace, Docker Compose, JWT.</i>	
LocalPilot-ide.ai	GitHub
- Fully local, browser-based AI code editor powered by Monaco Editor and Ollama, delivering sub-200ms code-completion latency on consumer-grade GPUs. - Developed a FastAPI backend bridging local LLMs (Qwen 2.5 Coder) for real-time code completion and chat, with containerized deployment tested across 3 CPU/GPU hardware configurations. <i>Stack: Monaco Editor, Python, FastAPI, Ollama, Docker.</i>	
Fluency.ai	GitHub
- Offline voice-to-voice English conversation partner utilizing a low-latency pipeline (Moonshine STT, Mistral 7B, Kokoro TTS) via fastrtc, achieving under 1.2s end-to-end response latency fully offline. - Benchmarked on 500+ test utterances, maintaining 90%+ transcription accuracy fully offline. <i>Stack: Python 3.13, Gradio, fastrtc, Mistral 7B, Ollama, Kokoro.</i>	

PROGRAMMING SKILLS

Languages: Python, SQL, Java, C++, JavaScript, HTML, CSS
AI/ML: PyTorch, RAG, PGVector, Ollama, DeepFace, NLTK, Label Studio
Backend: FastAPI, PostgreSQL, JWT
Infrastructure & Tools: Docker, Docker Compose, Linux, Git, SSL/TLS