





Neloy Barman

Agentic AI Engineer • Automation Specialist


 Neloy-Barman  portfolio  neloy-barman  neloycareer018@gmail.com

 +880-1708-708167  Dhaka, Bangladesh

PROFESSIONAL EXPERIENCE

Chatbot Developer

Jul 2024 – present

Gravitas AI 

Bavdhan, Pune, India (Remote)

- Architected **LangGraph-based multi-agent workflow** replacing legacy **AWS Lex/ Lambda** system.
- Improved **maintainability** and **control flow** across the entire conversational AI pipeline.
- Evaluated **6 latency reduction approaches** (keyword matching, semantic search, hybrid methods), selecting **few-shot** prompted intent-based navigation as the optimal strategy.
- Integrated **Groq inference**, reducing response latency by **~65%** (6s → 1.5–2s)
- Implemented **audio response caching strategy**, achieving an additional **0.5s latency reduction**.
- Improved **real-time voice UX**, validated across **100+** test runs.
- Designed and deployed **conversational AI systems** integrating **Meta Developer Platform**.
- Led **architecture decisions** involving **latency tradeoffs, token cost analysis, and scalability**.

Freelance AI Data Trainer

Dec 2025 – Present

Outlier 


Oakland, California, USA (Remote)

- Benchmarked **4 LLMs** across reasoning, accuracy, and instruction-following for model selection.
- Curated **450+ multilingual** audio samples, correcting **STT** noise to enhance **ASR** model training.
- Generated **computer vision datasets** by segmenting interactive UI components.

EDUCATION

Bachelor of Science in Computer Science and Engineering

2019 – 2023 | Dhaka, Bangladesh

Ahsanullah University of Science & Technology 

CGPA: 3.542

PROJECTS

Automated Email Query Resolution System

n8n, qdrant, mxbai-embed, llama-3.2, lm studio, google drive api, gmail api, docker

- Automated incoming queries via **semantic search** and **LLM pipeline**, cutting manual response time
- Extracted FAQ data to generate embeddings, storing them in **Qdrant DB** for **semantic retrieval**
- Deployed a **Dockerized** workflow integrating **n8n** and **Gmail API** to generate replies with **llama-3.2**

AI-Powered Customer Feedback Automation

n8n, qwen2-vl, lm studio, postgresql, discord api, tally, docker

- Architected a customer feedback processing pipeline with intelligent support routing, **reducing manual triage**
- Engineered a multimodal analysis pipeline using **Qwen2-VL** via LM Studio
- Deployed a **Dockerized microservices** architecture with PostgreSQL to route real-time alerts

Affiliate Submission Hub: Form to Sheets Automation

n8n, google sheets api, google drive api, slack api, tally, docker

- Streamlined affiliate lead capture and smart data routing, **reducing manual data entry** effort
- Connected webhooks to the **Google Sheets API** to auto-create and populate specific spreadsheets

- Containerized n8n workflow orchestration via Docker, broadcasting **real-time Slack** alerts.

An Interpretable Bengali Fish Recognizer

fastai, resnet-50, densenet-121, vgg-19, grad-cam, hugging face

- Achieved **81.379% accuracy** on image classification by fine-tuning a ResNet-50 architecture
- Applied **Grad-CAM** to interpret convolutional feature maps and validate anatomical region targeting
- Deployed the optimized computer vision pipeline as an interactive web app via **Hugging Face** Spaces

Scientific Paper Keywords Categorization

python, selenium, bert, distilbert, roberta, onnx, hugging face, render

- Achieved an **87.45% reduction in model size** (to 105.3 MB) with only a 2.8% F1-score drop by compressing a fine-tuned BERT text classifier via ONNX
- Extracted and pre-processed a custom training dataset of **36,398 IEEE research abstracts**
- Deployed the highly efficient natural language processing inference pipeline on **Render**

Daraz 11.11 Top Selling Product Data Analysis

python, selenium, pandas, tableau, kaggle

- Extracted e-commerce data across 119 subcategories, publishing a dataset of **12,907 products**
- Implemented an **EDA** pipeline to compute break-even points and categorical discount distributions
- Deployed an interactive Business Intelligence dashboard via **Tableau** to visualize pricing trends

Artwork Caption Generator

python, selenium, pytorch, transformers, swin transformer, gpt-2, streamlit, hugging face

- Extracted and preprocessed a custom dataset of **47,115 artwork-title pairs** using **Selenium**
- Fine-tuned an **image-to-text** model combining a **Swin** Transformer encoder and **GPT-2 decoder**
- Built an interactive inference app with **Streamlit** and **Hugging Face** Spaces

An XAI based OSCC Detection System

python, tensorflow, keras, sklearn, grad-cam, lime

- Achieved **97.84% accuracy** on **oral cancer classification** using **DenseNet-121**
- Applied **Grad-CAM** and **LIME** to make CNN predictions interpretable
- Benchmarked **4 training strategies** to identify the **optimal approach**

SKILLS

Agentic AI & LLM Orchestration: *LangChain, LangGraph, LangSmith, RAG, Tool Calling, Claude Computer Use* |

AI Models & Infrastructure: *OpenAI, Groq, Hugging Face, Ollama, LM Studio* | **Voice AI:** *LiveKit, Deepgram, Speechify, AssemblyAI, ElevenLabs* | **Vector DBs & Databases:** *Pinecone, Qdrant, FAISS, PostgreSQL, Redis, AWS DynamoDB* |

Automation & Platforms: *n8n, CI/CD, Slack, Discord, Telegram, WhatsApp* | **Languages & Tools:** *Python, JavaScript, Git/GitHub, Streamlit* | **Backend & Cloud:** *FastAPI, AWS (Lambda, SES, S3), Docker*

CERTIFICATES

Claude Code in Action (Anthropic) 