

BIPRAYAN CHOUDHURI

biprayanc@gmail.com | [linkedin.com/in/biprayanchoudhuri](https://www.linkedin.com/in/biprayanchoudhuri) | github.com/CodeNeuron58

PROFESSIONAL SUMMARY

AI/ML Engineer specializing in agentic systems, LangGraph-based multi-agent architectures, and production-grade RAG pipelines. Proven ability to build end-to-end AI products — from local LLM inference to full-stack deployment. Currently pursuing B.S. in AI and Data Science at IIT Guwahati.

TECHNICAL SKILLS

Languages: Python, SQL

ML / AI Frameworks: PyTorch, TensorFlow, scikit-learn, LangChain, LangGraph, Hugging Face Transformers, OpenAI API

NLP & LLMs: RAG Systems, Transformer Models, Fine-tuning, BERT, FAISS, Vector Databases, Prompt Engineering

Tools & Platforms: FastAPI, Streamlit, Ollama, Git, MLOps, Pydantic, PixiJS

TECHNICAL EXPERIENCE

AI Companion Agent

Feb 2026 – Present

Personal Project | Python · LangGraph · LangChain · FastAPI · PixiJS · Local LLMs

- Architected a zero-cloud-cost real-time AI companion running entirely on CPU using free-tier APIs and small local models, eliminating all infrastructure expenses.
- Engineered a low-latency response pipeline with persistent memory, emotional context tracking, and personality state management via LangGraph, solving consistency challenges common in production AI systems.
- Delivered a full-stack AI product by integrating an animated 2D avatar frontend (PixiJS + HTML/CSS) with a FastAPI backend, enabling a seamless end-to-end interactive experience.
- Optimized model selection, API routing, and inference strategy to run on commodity hardware owned by most users — demonstrating constraint-driven engineering at scale.

RxGuard: Agentic Medication Safety Copilot

Jan 2026 – Feb 2026

Personal Project | Python · LangGraph · Ollama (MedGemma) · FAISS · Pydantic · Streamlit

- Reduced cloud API costs to \$0 by architecting a 100% local, air-gapped clinical AI agent over a FAISS vector store, directly addressing LLM hallucination risks in medication safety workflows.
- Engineered a Safety Critic node that autonomously evaluates reasoning outputs, enforces citations, and triggers re-evaluation loops — cutting hallucination rate in complex multi-drug queries.
- Enforced deterministic state transitions via strict Pydantic schema validation and built a Streamlit UI to visualize real-time agent thought traces and execution logs, improving debuggability.

LangGraph Blog Agent

Jan 2026 – Feb 2026

Personal Project | Python · LangGraph · LangChain · Groq (Llama-3.3-70B) · Tavily API · Streamlit

- Reduced content generation time by 60% by designing a multi-agent blog system with a fan-out state graph enabling parallel section drafting versus sequential execution.
- Integrated Tavily real-time web research API; orchestrator node generates structured content plans with strict goals and word-count targets, improving output consistency.

Video-RAG-Analyst

Jul 2025 – Sep 2025

Personal Project | Python · LangChain · Cohere LLM · FAISS · HuggingFace · Streamlit

- Built an end-to-end RAG pipeline to extract, summarize, and semantically query YouTube transcripts using HuggingFace sentence-transformer embeddings and FAISS vector retrieval.
- Deployed an interactive Streamlit app supporting real-time Q&A with automated source citation, enabling verifiable video insights at zero per-query cost.

Sentiment Classification Model

Jan 2025 – Mar 2025

Personal Project | scikit-learn · LightGBM · TF-IDF

- Improved classification accuracy from 65% (Random Forest baseline) to 86% on a noisy real-world dataset by applying TF-IDF vectorization and LightGBM — a 32% relative improvement.

EDUCATION

Indian Institute of Technology (IIT) Guwahati

Expected Sep 2028

B.S. in Artificial Intelligence and Data Science

Guwahati, India

Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing, Probability & Statistics, Database Systems

CERTIFICATIONS

Applied Data Science with Python Specialization

Coursera – University of Michigan

AI Engineer Certificate

Udemy