

Anubrat Bora

+916002933403 | anubratbora25@gmail.com | [LinkedIn](#)

EDUCATION

Manipal Academy of Higher Education

Bachelor of Technology in Computer Science (Artificial Intelligence)

Bangalore, KA

Sep. 2022 – Present

Central Board of Secondary Education

Intermediate (XII) : 93.6%, Matriculation (X) : 95.1%

Golaghat, AS

Jul. 2021

EXPERIENCE

AI ML Engineer Intern

Barsys

Jun. 2025 - Present

Coimbatore, TN

- Built and deployed scalable vector embedding infrastructure with automated data pipelines and RESTful APIs, enabling sub-millisecond semantic search on 200+ recipes using OpenAI embeddings, PostgreSQL with pgvector HNSW indexing, and integrating 95%-accurate AI-powered cocktail recommendations.

Deep Learning Research Intern

Manipal Institute of Technology Bangalore — Dept. of CSE

Sep. 2023 - Mar. 2024

Bangalore, KA

- First-authored two research papers on semantic intelligence and machine learning, including VRSIL (video recommendation with dynamic ontology & LLMs) and SSCKC (crime classification using LLaMA & XGBoost), advancing Web 3.0 technologies through hybrid AI, semantic reasoning, and optimization algorithms.

PROJECTS

Developing a Transformer from "Attention Is All You Need"

Feb. 2025 - Apr. 2025

- Implemented the Transformer architecture from scratch, referencing the seminal "Attention Is All You Need" paper, demonstrating expertise in advanced neural network design.
- Trained the model on a custom dataset using PyTorch, optimizing hyperparameters to achieve a BLEU score improvement of 15% over baseline performance.
- Analyzed model performance through detailed evaluation metrics, such as attention heatmaps and loss curves, gaining insights into Transformer interpretability and optimization strategies.

Fine-Tuning an Open-Source LLM for SQL Generation

Jan. 2024 - Mar. 2025

- Fine-tuned Llama on a dataset of English-to-SQL query pairs using techniques like LoRA and parameter-efficient fine-tuning, reducing inference errors by 25%.
- Developed a preprocessing pipeline to clean, tokenize, and balance the dataset, improving model training efficiency and SQL query accuracy.
- Developed a preprocessing pipeline to clean, tokenize, and balance the dataset, improving model training efficiency and SQL query accuracy.

RAG Workflow with LangChain, ChromaDB & Open-Source LLM

Dec. 2024 - Jan. 2025

- Built a Retrieval-Augmented Generation (RAG) pipeline by integrating LangChain, ChromaDB, and an open-source LLM, enabling context-aware, domain-specific text generation.
- Utilized ChromaDB indexing for efficient retrieval, achieving a 10% reduction in query latency and improving response relevance in real-time applications.
- Utilized comprehensive evaluation frameworks to measure retrieval accuracy, model performance, and end-to-end system reliability, ensuring production-readiness.

TECHNICAL SKILLS

Languages: Python, C, Java, SQL, JavaScript, HTML & CSS, LaTeX

Frameworks: LangChain, MongoDB, React.js, Node.js, Express.js, PostgreSQL(pgvector), Flask, FastAPI

Developer Tools: Linux, Git, Docker, Google Cloud Platform, Firebase, VS Code

Libraries: Pydantic, Streamlit, Pandas, NumPy, SciPy, Matplotlib, TensorFlow 2.0, PyTorch, OpenCV