

ANURAG NAVIN BHANDARY

+91 8169370520 | bhandaryanurag@gmail.com | anuragbhandary.github.io/ai | linkedin.com/in/bhandary-anurag | github.com/AnuragBhandary | Mumbai, India

SUMMARY

AI/ML engineer with a year on the founding team of an early-stage AI sports startup, building **LLM** and **text-to-speech** pipelines for AI sports commentary. Works in **Python** with **PyTorch**, **Hugging Face**, **scikit-learn**, and **LightGBM**, from **RAG** and embedding fine-tuning to leakage-safe, calibrated models tracked in **MLflow** and served behind **FastAPI**.

TECHNICAL SKILLS

Languages: **Python** (primary), **SQL**, **Java** (core language, data structures and algorithms)

Machine Learning: **PyTorch**, **scikit-learn**, **LightGBM**, **Optuna**, **SHAP**, **pandas**, imbalanced classification, calibration, A/B testing, causal inference

LLMs and GenAI: **RAG**, hybrid search (BM25 + embeddings), **pgvector**, cross-encoder reranking, embedding fine-tuning, **LangGraph** agents, **MCP**, **LLM evaluation**, **Gemini**, **Groq**, **Ollama**

MLOps and Serving: **MLflow** (tracking, model registry), **FastAPI**, **Docker**, **GitHub Actions**, drift monitoring (PSI)

Data and Infra: **PostgreSQL**, **BigQuery**, **dbt**, **Apache Kafka**, **Redis**, **AWS S3**, **OCR**, **ffmpeg**

EXPERIENCE

Software Developer (Founding Team) | **Fayble Inc.** | **Oakland, CA, USA** | **Jan 2026 - Jun 2026**

Software Development Intern | **Fayble Inc.** | **Oakland, CA, USA** | **Jul 2025 - Dec 2025**

AI sports-simulation platform

- Engineered the **audio pipeline** for an AI sports-commentary system, turning **LLM**-generated play-by-play into **dual-commentator** speech with **ElevenLabs** and stitching per-play clips into continuous broadcast output.
- Moved commentary text generation from **Gemini** to **Groq** and chose a one-emotion-per-play voice design over per-segment emotion stitching, keeping full-game audio generation fast and within free-tier limits.
- Automated archery match production with **ffmpeg**, synchronizing AI commentary, crossfades, **7.1 (8-channel) audio** with 1 kHz broadcast trigger tones, and **structured per-play metadata** into finished match videos.
- Integrated **Apache Kafka** into the platform's animation flow, replaying **19 games** (about **1,274 messages**) through an ordered, de-duplicated ingestion path served through **FastAPI**.

PROJECTS

Retrieval-Augmented QA with Measured Retrieval | *Python, pgvector, PyTorch, sentence-transformers, LangGraph, FastAPI*

- Built a **RAG** service over the **FiQA** financial QA corpus (**57,638 documents**): **hybrid retrieval** (BM25 + **pgvector** HNSW embeddings) with rank fusion, **cross-encoder reranking**, and answers that cite a source for every sentence.
- Raised **nDCG@10** from **0.25** (BM25) to **0.43** (+72%, $p < 0.001$) on **648 held-out queries** by **fine-tuning the embedding model**; measured that an off-the-shelf reranker cut nDCG by 14% and fused its ranking with the retriever's instead.
- Lifted LLM-judged answer **faithfulness** from **69% to 80%**, and to **93%** with a **LangGraph** agent that self-checks, served by **FastAPI** streaming and an **MCP** server.

Card Fraud Detection with MLOps | *Python, LightGBM, scikit-learn, Optuna, MLflow, SHAP, FastAPI, Polars*

- Trained **LightGBM** on **590,540** IEEE-CIS card transactions (**3.5% fraud**) with an **out-of-time split**: test **PR-AUC 0.64** vs 0.53 for logistic regression; a random split would have claimed 0.87.
- Engineered **leakage-safe** customer features, including known fraud counted only after a **30-day chargeback delay** (+0.09 PR-AUC); a **calibrated expected-cost rule** cut test-month fraud cost by **68%**, vs 34% for a fixed cutoff.
- Promoted models through an **MLflow** registry (champion/challenger) and served them with top-5 **SHAP** reasons via **FastAPI** at **p95 8 ms**; online features matched offline on all **14M values**, with a **PSI drift** report.

Product Analytics and A/B Testing | *BigQuery, dbt, SQL, Python, SciPy, scikit-learn, Evidence*

- Modeled Google's **GA4** ecommerce export (**4.3M events**) in **BigQuery** with **dbt**; 27 data tests reconcile revenue and caught **314 duplicated purchases** (6.3% of revenue).
- Validated an **A/B testing** toolkit (SRM, power, **CUPED**) by simulation: daily peeking gave **29.7%** false positives, **mSPRT** 1.5%.
- Read out **Criteo's 14M-user** randomized experiment: covariates were imbalanced, and **doubly robust (AIPW)** estimation corrected the visit lift from **+27.1% to +19.5%**.

EDUCATION

M.S. Computer Science | **University of Texas at Arlington** | **Aug 2023 - May 2025**

GPA 3.42/4.0 | Focus: Artificial Intelligence and Software Engineering | Arlington, TX

B.E. Electronics Engineering | **University of Mumbai** | **Aug 2019 - May 2023**

GPA 8.36/10 | Mumbai, India

Competitive Coding: LeetCode 450+ problems in Python and Java (data structures and algorithms) | leetcode.com/anuragb2901