Ankur Batta

Vancouver, BC | 778-927-4164 | ankurbatta.30@gmail.com | Eligible to work in Canada (Open Work Permit)
linkedin.com/in/ankurbatta | github.com/ankurxbatta | iankurbatta.web.app

SUMMARY

Machine learning engineer who builds ML and LLM systems that run in production, and holds them to the numbers. Built QuizMark, a RAG auto-marking platform on AWS that ran a real graded quiz for **~150 students** on **~$22** of model spend, agreeing with the instructor's independent grading at **r = 0.80**.

EDUCATION

Langara College — Vancouver, BC
Post-Degree Diploma, Data Analytics — May 2025 – Aug 2026

- **4.11 CGPA**; graduated with distinction; Dean's Honour Roll for three consecutive semesters
- Coursework: Machine Learning, Cluster Analysis, Statistical Inference, Data Mining

IK Gujral Punjab Technical University — Punjab, India
Bachelor of Technology, Computer Science and Engineering — Distinction (85.2%) — Aug 2020 – Jul 2024

TECHNICAL SKILLS

ML & Modelling: Deep Learning, Natural Language Processing (NLP), scikit-learn, CatBoost, XGBoost, TensorFlow/Keras, Hugging Face Transformers (BERT fine-tuning)

LLMs & Retrieval: RAG, embeddings and vector search, Reciprocal Rank Fusion, hybrid reranking, LangChain, Model Context Protocol (MCP), tool calling, LLM cost optimization; OpenAI, Anthropic, Gemini APIs

Evaluation: Recall@k, MRR, nDCG@k, ROC/PR AUC, calibration and Brier scores, stratified k-fold CV, class imbalance

MLOps & Cloud: AWS (EC2, S3, SageMaker), Docker, Kubernetes, MLflow, FastAPI, Celery/Redis, GitHub Actions CI, pytest, Linux

Languages & Data: Python, SQL, R, TypeScript, C++; pandas, NumPy, MongoDB Atlas Vector Search, PostgreSQL

PROJECTS

QuizMark — RAG Auto-Marking Platform — *Python, FastAPI, LangChain, MongoDB Atlas, Celery, Docker, AWS*
Langara Capstone · quizmark.app — May 2026 – Jul 2026

- Built a RAG platform that ingests 600+ page PDF textbooks, writes exam questions and auto-marks answers; ran a real graded quiz for **~150 students** on **~$22** of model spend
- Validated marking against the instructor's independent grading of 358 answers from 91 students: **r = 0.80**, within one mark on **92.5%**; traced the remaining gap to rubric literalness
- Designed multi-index retrieval over 768-dim embeddings in MongoDB Atlas with intent routing, Reciprocal Rank Fusion and hybrid reranking; scored **MRR 0.84** on a hand-built golden query set
- Benchmarked four LLMs and chose gpt-4o-mini over gpt-4o, which cost 16× more for near-identical quality, holding marking to **~$0.002 per answer**
- Designed a four-tier marking triage (answer key, keyword and similarity check, LLM, human); in one exported sitting, 52 of 78 answers were marked with no model call
- Deployed 14 Docker services and 8 Celery workers on AWS EC2 with failover from OpenAI to Anthropic to Gemini, which fired mid-pilot when one provider's credits ran out

Clinical Multiclass Classification — 3-Class Parkinsonism — *Python, scikit-learn, XGBoost, CatBoost, MLflow*
Project Lead, Langara College (DANA 4830)

- Separated Parkinson's, atypical parkinsonism and controls across **3,433 patients** and 163 features; CatBoost reached **0.899 macro AUC** and 0.707 balanced accuracy on a stratified holdout
- Benchmarked against XGBoost, Random Forest, ExtraTrees, logistic regression and SVM-RBF under stratified 5-fold CV, with every run tracked in MLflow on AWS SageMaker
- Handled a 16:1 class imbalance by weighting the rare class, and found and removed a feature that leaked the diagnosis; reported calibration curves, Brier scores and per-class sensitivity

AI-Driven Penetration Testing — MCP Server *Python, FastMCP, Docker, Kali Linux*

github.com/ankurxbatta/AI-driven-penetration-testing

- Built a Model Context Protocol server exposing **7 typed tools** (nmap, nikto, sqlmap, wpscan, dirb, searchsploit, a system check), so the model decides what to run and the server controls how it runs
- Placed the guardrails in the server, not the prompt: input validation on every scan tool, a controlled command wrapper with no raw shell, and scans run as a non-root user in Docker

iScore — Automated IELTS Essay Scoring *Python, TensorFlow/Keras, Hugging Face Transformers*

github.com/ankurxbatta/iScore

- Fine-tuned BERT (bert-base-uncased) for NLP regression, pairing essay text with hand-built numerical features to predict IELTS Writing Task 2 band scores; optimized for MAE and RMSE
- Fronted the model with a rule-based validator that rejects essays too short or too repetitive to score

HACKATHONS

StormHacks 2026 — Simon Fraser University *Python, TypeScript, Three.js, Snowflake, SQL*

Search and Rescue Drone Swarm (team of 2) · rescue-drone-swarm.vercel.app Oct 2026

- Added census data and map parameters (population, land cover, terrain) to the search layer so the swarm's A* pathfinding could score search sectors
- Contributed to the rescue simulation's engine, drone-cam view, challenge game mode and mission-history backend

Web Summit Vancouver Hackathon *Next.js 14, TypeScript, OpenAI GPT-4o, Vercel*

EventOps AI (team of 2) · event-bridge-ivory.vercel.app May 2026

- Set up the Next.js app and built the main flow (event wizard, API routes to GPT-4o, deployment) that turns an event idea into a vendor-ready brief

EXPERIENCE

Langara College Vancouver, BC

Full-Stack Developer, Student Work Assistance Jan 2026 – Apr 2026

- Built and deployed internal full-stack web apps over relational databases, turning vague requests from non-technical staff into maintainable systems
- Wrote REST APIs and database-backed services, then profiled and tuned the slow queries and endpoints

Nebero Systems Pvt. Ltd. Mohali, India

IT Infrastructure & Cybersecurity Intern Jan 2024 – Jul 2024

- Built anomaly-detection models on live network and server data to catch traffic and machine behaviour that existing rule-based alerts missed, trained on the company's own data
- Ran the work end to end: pulled and cleaned data, built features, then compared and evaluated models
- Tuned the alert threshold to the cost of a false alarm for the security team rather than raw accuracy, on rare-event data, messy logs and shifting infrastructure behaviour

Freelance Remote

Web Developer & Designer Jun 2020 – Jul 2023

- Delivered full-stack applications with REST APIs and ETL ingestion (transformation, validation, error handling), owning build, deployment and client hand-off

LEADERSHIP

Google Developer Student Club – BBSBEC Punjab, India

Web Development Team Lead Aug 2022 – Jul 2023

- Led a team of **12 developers** building user-centric web applications focused on data handling and integration

CERTIFICATIONS

Model Context Protocol: Introduction & Advanced Topics (Anthropic) · Generative Models for Developers (Infosys) · Behavior of Large Language Models (Infosys Springboard) · Deep Learning Onramp (MathWorks)