Tanay Kumar

Markham, ON | tanaykumar29@gmail.com | linkedin.com/in/tanay1kumar | github.com/tanay1kumar

Education

Queen's University, *Bachelor of Computing (Honours)* *Sep 2024 – May 2028*

- **Specialization:** Computing and Mathematics
- **Relevant Coursework:** Data Structures, Linear Algebra, Computer Architecture, System-Level Programming

Work Experience

AI/ML Developer *Toronto, ON*
The Home Depot Canada *Sep 2026 – Present*

- Evaluate and improve an approximate nearest-neighbor retrieval model (ScaNN) over a 700K+ product catalog, running experiments against a golden dataset to measure recommendation quality
- Tune anomaly detection thresholds across 50+ microservices, benchmarking IQR and other dispersion methods against a Prophet baseline to cut false positives
- Diagnosed root causes of false-positive alerts in Kong API gateway metrics, tracing noise to low-traffic periods and stale alerts that failed to auto-resolve
- Curate a golden evaluation dataset in BigQuery with anchor products per leaf category, defining ground truth to benchmark a multimodal (text + image embedding) substitute-product model
- Refine Grafana/Prometheus alert rules with SRE and ML engineers, validating changes against historical traffic

Software Engineer Intern *Toronto, ON*
Impressico Business Solutions *Jun 2026 – Sep 2026*

- Built a production-grade LangChain RAG pipeline over SEC compliance documents with ChromaDB retrieval, semantic chunking, and citations, achieving 94% faithfulness and 100% out-of-scope refusal in LLM-as-judge evals
- Engineered a LangGraph browser agent that reads config files to scrape sites without hardcoded selectors
- Developed LLM agents with tool calling for a Fortune 500 beverage client's supply chain forecasting platform
- Trained Prophet and XGBoost forecasting models on historical sales data, benchmarking variants on MAPE

Software Engineer Intern *Toronto, ON*
SharpStakes *Sep 2023 – Feb 2024*

- Shipped production features within an Agile scrum team for a web application serving 2,500+ active paying users
- Implemented automated end-to-end testing workflows with Playwright and Appium to validate releases
- Developed ingestion pipelines integrating the Sportradar API to process sports data for personalized analytics
- Optimized PostgreSQL schemas and queries, reducing server response times

Projects

Pigeonhole — Swift, SwiftUI, AppKit, Vision, NaturalLanguage, Google Drive API *Jul 2025 – Present*

- Built a macOS notch app (11.5K lines of Swift) that suggests a Google Drive folder for each dropped file
- Designed a nearest-neighbor classifier over Vision labels, OCR, and sentence embeddings that ranked the correct folder first for 36/36 real test files with zero confident mistakes
- Cut suggestion latency ∼5× to 44–64 ms, idle CPU to 0%, and upload peak memory 8× (200 MB to 25 MB)
- Tested with 273 unit tests and 37 automated scenarios that drive the real app, including Drive failures

Starter — Python, TypeScript, Next.js, Gemini API, PostgreSQL, pgvector *Oct 2026*

- Built a recruiting tool that ranks student developers by their attributed code across 1,000+ hackathon repos
- Cut repo analysis time ∼4.5× (171 s to 38 s per repo) by switching from partial to full no-checkout git clones
- Wrote a Python pipeline that classifies code via rules and Gemini, rejecting LLM text with unverified numbers
- Built topic search over local embeddings in Postgres + pgvector, scoring chunks by z-score; 220+ automated tests

Technical Skills

Programming Languages: Python, Java, JavaScript, TypeScript, Go, C++, SQL, Swift, Bash
Tools & Methodologies: Git, React, Docker, REST APIs, GraphQL, Playwright, Cursor, Claude Code, Copilot
Data & Cloud: PostgreSQL, BigQuery, MongoDB, AWS, GCP (Vertex AI), Data Pipelines
AI/ML: Recommender Systems, Vector Search (ScaNN), Embeddings, RAG, AI Agents, LangChain