Chinmay Kapoor

Vancouver, BC | 250-986-5497 | chinmaykapoor2301@gmail.com | LinkedIn | GitHub

EDUCATION

MEng in Electrical & Computer Engineering, *University of Victoria* — Sep 2021 – Aug 2024

BEng in Computer Science, *Amity University* — Aug 2017 – Apr 2021

SKILLS

Data Analysis/Engineering Tools : **MS Azure & Fabric**(ADF2, DataFlow2, ADLS2, Synapse, PowerBI, Azure Storage), **GCP**(BigQuery), **AWS**(RDS, S3), **Databricks**, Tableau, T-SQL, NoSQL, PySpark, Grafana

AI/ML Tools : **Azure**(Foundry, DevOps), **AWS**(SageMaker, Bedrock, EC2, Boto3, Lambda), Qdrant, **GCP**(Vertex, A2A), PydanticAI, **Kubernetes**(GKE, AKS), **Docker**(GCS, ACI), llama.cpp, LangChain, LangGraph, ChromaDB

Frameworks & Languages : Python, C/C++, JavaScript, KQL, DAX, Pandas, MCP, MQTT, EDGE-AI, OpenCV, Jenkins, RKNN-Toolkit, Hailo, CUDA, ROCm, TensorFlow, GitLab, Pytorch, GStreamer, FastAPI, Streamlit, ONNX

EXPERIENCE

Data Scientist - AI Consultant, *Swissvault Global (Vancouver, BC) - Hybrid* — Jan 2024 - Present

– **COMPIO**: Client-facing tool to benchmark hybrid systems for AI/ML workloads at edge, guiding optimal solution selection.
  - Designed COMPIOs end-to-end benchmarking pipeline to replicate client-specific workflows from real-time data ingestion setup (APIs, web scraping, object storage, on-prem, and cloud) through ML compute and inference enabling short-scale feasibility runs that identify the optimal hybrid architecture for AI/ML workloads.
  - Built integrations for hybrid storage solutions (VaultFS, AWS S3, GCS, Azure Blob) and connected them to LLMs hosted locally (Llama 3.2, Mistral, Qwen-Phi) or via APIs (OpenAI, Gemini2.0), with embedding workloads powered by OpenAI or on prem with Ollama, Hugging Face, also utilizing vector stores like Qdrant, ChromaDB, or FAISS.
  - Evaluated inference latency & retrieval efficiency across diverse compute environments ranging from AWS serverless (Lambda) & EC2 Trainium instances to on-prem/edge deployments using GPUs, NPUs, accelerators, and CPUs.

– **SVSI**: Swiss Vault Storage Intelligence RAG-grounded local LLM chat interface for secure client data access.
  - Developed a revenue-generating feature enabling business clients to query sensitive data in VaultFS with GPT-like capabilities, supporting secure on-premise Edge-AI adoption with optional cloud integration (OpenAI, Gemini).
  - Optimized performance through model quantization, load balancing, and offline embedding retrieval for RAG, reducing inference latency on resource-constrained edge devices, with optional offloading of non-sensitive tasks to the cloud based on client requirements. *Skills/Tools: llama.cpp, FastAPI, vLLM, Streamlit, LangChain, S3, boto3*
  - Integrated seamless API connections to cloud analytics platforms such as MS Fabric for hybrid data solutions, ensuring proprietary data remains undisclosed to uncensored platforms. *Skills/Tools: MS Fabric, PowerBI*

– Led client engagement, taking opportunities from coffee room conversations to data room negotiations, and from POC to MVP delivery where needed. Represented the organization at high-profile networking and industry events such as VSW 2025 and Web Summit 2025, building strategic partnerships to support business growth and scaling.

– Conducted ML benchmarking with detailed performance reports across hardware platforms, evaluating inference trade-offs, system-level efficiency, and optimization strategies. *Skills/Tools: Critical Analysis, Regex, Plotly, PowerBI, SQL*

– Built **Imagilect** features like object detection & LPR integrated with VaultFS storage, using YOLOv5 for bounding box detection and DeepSORT(tracking), OCR (text recognition) & ResNet50(classification).*Skills/Tools: OpenCV, Gstreamer*

– Deployed and optimized batch & streaming ETL pipelines for VaultFS to process semi-structured telemetry and system performance data, improving tracking of system health & reduced load and transformation times by over 30%. *Skills/Tools: MS Fabric, Databricks, DevOps, Lakehouse, PySpark, SSMS, ADLS2, T-SQL, Grafana, Prometheus*

Research Assistant (ML Team Lead), *SOLIDS Lab, UVIC (Victoria, BC)* — Sep 2021 - Aug 2024

– Developed FUSTT, improving time-series forecasting accuracy by 16% and compute time by 19%, currently used in trend prediction for Blue Carbon Marine Ecosystems, Sponsored by National Research Council of Canada (NRC) *Skills/Tools: Research, Publication, Transformers, Python, Spatio-temporal time-series forecasting, Model Training, AI/ML Inference*

– Published post research & review of building predictive pipelines and estimating physicochemical properties of concrete, presented findings at CSCE 2022. *Skills/Tools: XGBoost, AdaBoost, ANFIS, Random Forest, SVR, MLP*

– Recognized as a Mentor(2023) and Co-Judge(2024) for CleanTech in OCEAN; Data-driven projects at UVic Tech Challenge Events. *Skills/Tools: Mentoring, Collaboration, Team Building, Tutorial Instruction, Knowledge Sharing*

PUBLICATIONS & BLOGS

- **FUSTT** - Spatio-Temporal Transformer for supporting sustainable Blue Carbon Ecosystems. *Springer, Accepted*.
- **Concrete Characteristics Evaluation**↑- ML prediction for concrete properties for carbon emission reduction |
- **Human Activity Recognition**↑- Evaluating human activity detection via wearable sensor data (LSTMs). |

INDEPENDENT PROJECTS

- **COMPIO**↑ - LLM web-app built to scrape/ingest, structure & summarize company information for financial profiling. |
- **Fraud Detection**↑- Random Forest & Logistic Regression Based Fraudulent transaction detection and alert system |
- **Stock Price Prediction**↑- Sequence forecasting of public stock movement using stacked LSTM & XGBoost models. |