

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(SSMS, SSIS, Dataflow Gen2, Data Factory, ADLS2, Synapse, PowerBI, Azure Storage, DAX Studio), **GCP**(BigQuery), **AWS**(RDS, S3), **Databricks**, Tableau, Terraform, Grafana

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

Frameworks & Languages : Python, C/C++, JavaScript, T-SQL, NoSQL, KQL, DAX, Pandas, Statsmodels, scikit-learn, OpenCV, MLlib, Jenkins, Plotly, SparkML, PySpark, TensorFlow, GitLab, Pytorch, GStreamer, FastAPI, Cuda, ONNX

EXPERIENCE

Data Scientist - AI Consultant, Swissvault Global (Vancouver, BC) - Hybrid

Jan 2024 - Present

- Led development of COMPIO for automated AI workload-system benchmarking. Integrated RAG pipelines, real-time data ingestion (API, Web Scraping) & embedding workloads tested across Cloud (Gemini-2.0, AWS S3), On-Prem (Llama3.2, VaultFS) & Hybrid architectures. Evaluated inference latency, retrieval efficiency & performance across hardware setups including NPUs & CPUs. *Skills/Tools: LangChain, FastAPI, llama.cpp, Gemini-API, HuggingFace, QDrant, boto3, bs4*
- Built SVSI, local AI chat interface enabling file interaction on VaultFS, optimizing load balancing, offline embeddings retrieval for RAG and inference latency at resource constrained Edge devices. *Skills/Tools: Llama3.2, llama.cpp, FastAPI*
- Supported AI democratization by developing quantized inference pipelines using Llama.cpp and HuggingFace models to improve latency & resource efficiency for LLM workloads deployed on Edge devices (Hailo, NPU, ARM), & conducted ML benchmarking with detailed performance reporting across hardware platforms, evaluating inference trade-offs, system-level efficiency, and optimization strategies. *Skills/Tools: Critical Analysis, Docker, 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*

PROJECTS, PUBLICATIONS & BLOG[†]

- **COMPIO[†]** - LLM web-app built to scrape/ingest, structure & summarize company information for financial profiling. | 🐙
- **SVSI** - Local Embeddings + RAG + Llama3.2-11B based interface to interact with files in VaultFS (proprietary file system).
- **Imagilect** Real-time LPR on streaming video feed during vaultFS storage for tracking highway vehicle frequency.
- **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). | 🐙