

JAMES CHOI

647-978-8220 | jinhchoi128@gmail.com | linkedin.com/in/jameschoi01/ | github.com/jinhchoii

EDUCATION:

Queen's University - Bachelor of Computing - Software Design Specialization April 2027

- Coursework: Data Structures, Algorithms, Machine Learning, Software Architecture, Operating Systems
- GPA: 3.7

EXPERIENCE:

Data Engineer Intern

May 2025 - present

Scotiabank

Toronto, ON

- Developed **10+ reusable Python scripts** for **data validation** and duplicate removal on **GCS**, querying **BigQuery** tables to **improve pipeline reliability** and cut **reducing manual work by 10 hours/week**.
- Built and deployed a **Cloud Composer DAG** on **GCP** to **orchestrate Python validation scripts**, run **dbt** test and published lineage/docs to **Confluence**, providing an **end-to-end workflow** accelerating **stakeholder onboarding by 70%**.

Software Engineer Intern

May 2023 - April 2024

FGF Brands

Toronto, ON

- **Architected** and implemented **scalable Python- and SQL-based backend microservices** using **Azure Logic Apps**, **Azure Functions** and **Cosmos DB** to ingest, parse, and normalize master-data feeds (**JSON, XML, EDI**) from **20+ EDI partners**, **10+ direct API customers**, and **5+ 3PL partners**, processing over **1500+ records daily** and **reducing downstream data errors by 30%**.
- Engineered **data-transformation pipelines** in **C# (.NET)** and **XSLT**, producing **20+ EDI/XML mappings**, complete with **unit tests** and **CI/CD** via **Azure DevOps**.

Machine Learning Engineer Intern

June 2021 - October 2021

ZeroOne AI

Toronto, ON

- Developed and deployed **LSTM-, CNN-, and GAN-based anomaly-detection models** for **KT Corporation** using **Python/TensorFlow**, and **ML algorithms: XGBoost, LightGBM, CatBoost**, achieving **95% test accuracy**.
- Developed **Python scripts** for data preprocessing—**cleaning and normalizing ~50K records for testing** and creating feature routines to **improve data quality by 25%** and speed up model training.
- Delivered **5+ Matplotlib/Seaborn dashboards**, **visualizing** anomaly trends to help stakeholders triage issues **50% faster**.

PROJECTS:

DataDream | *React JS, Python, AWS (Bedrock, Lambda, S3, DynamoDB), Streamlit*

May 2025 - Jul 2025

- Built an **AI-powered synthetic data platform** for the **GenAI Hackathon by Impetus & AWS** using **React**, **FastAPI**, and **AWS** services, achieving **85% data quality similarity** and processing **500+ records per minute**.
- Architected a **serverless data generation** system using **AWS Lambda, S3, and DynamoDB**, implementing real-time quality monitoring with **88% correlation maintenance** and integrating **AWS Bedrock** for **AI-powered synthetic data generation**.

automate-gcs-bq | *Python, JavaScript, GCP*

Jan 2025 - March 2025

- Developed **Python Cloud Functions** triggered by **GCS uploads** to dynamically create **BigQuery** tables (via **schema JSON and UDF files**) and kick off **Dataflow jobs—automating end-to-end CSV ingestion** with zero manual intervention.
- Orchestrated and versioned the workflow using **Cloud Storage, Cloud Functions, Dataflow, BigQuery, and Cloud Run**, **managing code artifacts** (Python scripts, UDFs, schema definitions) in **GitHub** for **reproducible, scalable** deployments.

ACHIEVEMENTS/CERTIFICATES

Microsoft Certified: Azure Fundamentals (AZ-900)

Jan 2025

AWS Certified: Cloud Practitioner (CLF-C02)

July 2025

AWS Certified: Solutions Architect (SAA-C03)

In Progress

SKILLS

Languages: Python, Java, C#, SQL, YAML, JavaScript, HTML/CSS, TypeScript

Frameworks & Platforms: Azure: Logic Apps, Data Factory, Functions, BizTalk, Apache Airflow, Docker, dbt, GCP: Google Cloud Storage, BigQuery, Cloud Composer, AWS: Lambda, S3, DynamoDB, API Gateway, Bedrock, CloudFormation), FastAPI, RESTful API, React JS

Developer Tools: Git, VS Code, Visual Studio, Azure DevOps, JIRA, Confluence, Docker, Kubernetes, AWS Cli

Libraries & SDKs: TensorFlow, pandas, NumPy, scikit-learn, OpenAI SDK, matplotlib