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 GCP, querying BigQuery tables to improve pipeline reliability and cut manual work by 40%.
- Built and deployed a Cloud Composer DAG on GCP to orchestrate Python validation scripts, run dbt test and published lineage/docs to GCS—providing a end-to-end workflow that accelerated stakeholder onboarding by 50%.

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—proocessing over 1500 records daily and reducing downstream data errors by 30%.
- Engineered data-transformation pipelines in C# (.NET) and XSLT—producing 20+ EDI/XML maps,

Data Scientist 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 and TensorFlow—achieving 95% test accuracy and enabling proactive identification of data irregularities.
- Developed Python scripts for data preprocessing—cleaning and normalizing ~500K records/day and creating 15 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:

Paper2Video | *Python, GPT-4*   *Dec 2024 - Feb 2025*
- Built a Python pipeline to transform academic PDFs into 60-second Korean YouTube Shorts: extracted text with Grobid & PyMuPDF; summarized, translated, and scripted via GPT-4; created avatars with Midjourney & D-ID Studio; and assembled videos using MoviePy.
- Automated the end-to-end workflow with Docker and cron jobs, integrating PyPDF2 for PDF handling, GPT-4 for NLP, and ffmpeg/MoviePy for media generation.

automate-gcs-bq | *Python, JavaScript, GCS*   *May 2025 - June 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 all code artifacts (Python scripts, UDFs, schema definitions) in GitHub for reproducible, scalable deployments.

ACHIEVEMENTS/CERTIFICATES

Dean's List, School of Computing, Queen's University   Fall 2024 - Winter 2025
Microsoft Certified: Azure Fundamentals (AZ-900)   Jan 2025
AWS Certified: Cloud Practitioner (CLF-C02)   July 2025

SKILLS

Languages: Python, Java, C#, SQL, JavaScript, HTML/CSS
Frameworks & Platforms: Azure: Logic Apps, Data Factory, Functions, BizTalk, Apache Airflow, Docker, dbt, GCP: BigQuery, Dataproc, Cloud Composer
Developer Tools: Git, VS Code, Visual Studio, Azure DevOps, JIRA, Confluence, Docker, Kubernetes
Libraries & SDKs: TensorFlow, pandas, NumPy, scikit-learn, OpenAI SDK, matplotlib