

SHEK MAN (VINCENT) CHU

vincentc0823@gmail.com | 613-890-6742 | [LinkedIn](#) | [Personal Website](#)

Professional Experience

Accelex Technology Limited, Toronto, ON

May 2022 – July 2023

Data Scientist

- Led the training and deployment of a multimodal transformer for document classification; managed data annotation, transformation, Docker containerization, and Amazon EKS configuration.
- Enhanced model accuracy by implementing post-processing techniques, evaluated outputs from transformer-based models for document classification and table understanding, and contextual embedding models for question answering.
- Designed and built a dashboard on Databricks for continuous monitoring and visualization of our machine learning-based document parsing product's performance over time.
- Implemented an automated data pipeline that integrates AWS S3, ArangoDB, and our proprietary parser to regularly assess the performance of the latest document parser.
- Introduced the 'Talk to Your Investment Report' feature using Retrieval-Augmented Generation technique, leveraging LangChain, vector databases, and OpenAI API for an interactive LLM chatbot offering investment insights.
- Implemented A/B testing and analyzed user journeys in collaboration with the UX team, leading to the introduction of shortcuts and features that streamlined user interactions and lowered user effort.
- Collaborated with the product team to test and maintain ground truth data for performance monitoring.
- Developed and maintained a set of regular expressions to standardized naming conventions from documents.

Data Analyst, Product

- Analyzed production data and formulated an enhancement plan to optimize parser stack performance, achieved an increase of over 30% in data point capture accuracy within two sprints through the identification of high-impact actions.
- Designed and implemented a Key Performance Indicator (KPI) along with the underlying querying logic to effectively capture and showcase the product's performance.
- Collaborated with 30+ customers to develop a templating engine using Python to output data in customized formats for direct integration with clients' data backend.
- Collaborated with the sales team to provide custom data analyses and tailored proof-of-concept presentations that showcased the product's functionality and value proposition, leading to the successful acquisition of several clients.
- Actively involved in the execution and maintenance of a Python CRON job that is responsible for migrating data nightly from unstructured data in Arango to tabulate data in Postgre to power our internal analytics pipeline.

Department of Economics, University of Toronto, Toronto

Sep 2021 – Dec 2022

Research Assistant

- Conducted data processing and exploratory data analysis using R to uncover relationships within experimental data.
- Performed Kolmogorov-Smirnov tests to assess the impact of varying levels of information provided on participant behavior.

Education

Faculty of Arts and Science, University of Toronto

Toronto, ON

Honours Bachelor of Science, Double Major in Economics and Statistics

2017-2022

- **Academics:** cGPA:3.6/4.0, High Distinction
- **Relevant Courses:** Econometrics, Machine Learning, ML Application in Macro Finance, Time Series Analysis

Projects/Contracts

LumberLens

September 2023 – Present

- Evaluated and fine-tuned image super resolution models, including GANs and diffusion models, to facilitate precise lumber stock analysis at trading ports; achieved cost savings by reducing dependency on high-quality, high-cost satellite images.

Additional Information

- **Skills:** Python, R, Excel, Git, PyTorch, TensorFlow, Hugging Face, SQL, AWS, Databricks, Linux, LangChain, Vector Stores, TDD