

Kevin Joseph

647-202-9324 | kevnjoseph42@gmail.com | linkedin.com/in/kevin-joseph/ | github.com/kevin-joseph05 | kevin-joseph.me/

EDUCATION

University of Toronto

Bachelor of Science in Computer Science and Statistics

Toronto, ON

Sept. 2023 – May 2027

EXPERIENCE

Riipen Level UP Intern - Fanverse

Mar 2026 – Present

Data Science Intern

Toronto, ON

- Engineered **SHA-256 deduplicated pipeline** scraping social media and industry reports for female fan data.
- Built **NLP signal extraction** with VADER and DistilRoBERTa to classify fan sentiment, emotions, & behavior.
- Deployed **K-Means segmentation** with PCA and Streamlit dashboard utilizing Gemini API for fan insights.

IMI Big Data & Artificial Intelligence Competition UTM

Nov. 2025 – Feb.2026

Big Data & AI Case Competitor

Mississauga, ON

- Architected a multi-stage AML risk-scoring pipeline combining **38 expert rules**, three unsupervised detectors (Isolation Forest, LOF), and a supervised Random Forest model to classify accounts into **explainable risk tiers**.
- Built a rule-based system with 38 red-flag indicators, requiring multi-category evidence for HIGH classification.
- Engineered behavioral features from transaction data, balancing anomaly detection with supervised predictions.

CIBC

May 2025 – August 2025

Application Developer

Toronto, ON

- Engineered a robust asynchronous data processing pipeline (Python, Azure SDK) using a producer-consumer model, **improving throughput by 40%** and enabling **efficient handling of 1,000+ data blobs** per run.
- Designed and implemented a targeted ETL pipeline in Python to extract, filter by date/criteria, and load relevant log events from Azure Blob Storage into structured text files, **reducing manual data retrieval time by 60%**.
- Automated the **aggregation of data from 50+ API endpoints** into standardized Excel reports using Pandas and fuzzy matching, directly contributing to a **risk deficiency rating improvement from High to Medium**.
- Enhanced **data reliability and completeness 30%** by implementing retry logic and handling for API failures.

PROJECTS

Prove It! (EmberHacks Winner 2025) | *TypeScript, React, MongoDB, DigitalOcean, Auth0*

- Architected a multi-panel, resizable UI with shadcn/ui, boosting user productivity and **engagement by 30%**.
- Developed an AI-powered tutor with Gemini 2.5 Flash, reducing **user navigation time by 50% per concept**.
- Streamlined infrastructure management for seamless scaling, **winning "Best Use of DigitalOcean"** award.

Hangman Game | *Java, Java-Websocket, Java Swing*

- Engineered real-time multiplayer via WebSockets **with sub-500ms latency** and Firebase live leaderboards.
- Developed streamlined multiplayer loop, **cutting player wait by 40%** with instant state sync across all clients.
- Designed dynamic Java Swing UI with letter grid and real-time visualization, **boosting engagement by 30%**.

Parka (SpurHacks) | *Python, Next.js, TypeScript, Tailwind*

- Engineered a CV pipeline with OpenCV & Gemini, **achieving 95% accuracy** in parking space identification.
- Architected a full-stack proof-of-concept to process images and suggest optimal spots in **less than 5 seconds**.
- Designed a Next.js map UI to render real-time parking availability and optimal spots for **10+ simulated lots**.

TalkBox (DeerHacks) | *Python, Whisper, TKinter*

- Built a real-time speech-to-subtitle pipeline using Whisper and DeepL APIs for live cross-lingual captioning.
- Engineered a **multi-threaded backend** to process dozens of audio chunks per minute without performance lag.
- Supported **20+ languages with 90%+ transcription accuracy** across diverse accents, maximizing accessibility.

TECHNICAL SKILLS

Languages: Java, Python, JavaScript, HTML/CSS, R, SQL (Postgres, MySQL)

Frameworks: React, Node.js, Material-UI, FastAPI

Developer Tools: Git, Docker, Azure SDK, MongoDB, Supabase, Shell, VS Code, IntelliJ, Vim

Libraries: Pandas, NumPy, Matplotlib, Scikit-learn