Youssef Cherkaoui

Phone: +1: (613) 617-8911 **Email**: youssef_cherkaoui@hotmail.com **GitHub:** github.com/Yorkchk

Website: youssefcherkaoui.com **LinkedIn:** linkedin.com/in/youssef-cherkaoui-8a85652a8

Education

- **University of Ottawa** - BASc in Computer Engineering (September 2023 – Present)
- 4th year | **CGPA:** 9.85/10.0 | **Expected Graduation:** May 2028

Technical Skills

- **Languages:** Python, Java, JavaScript, C, C++, SQL, Bash, HTML, CSS.
- **Backend & Frameworks:** FastAPI, Spring Boot, React, RESTful APIs, Microservices Architecture, MVC.
- **Data Engineering & Cloud:** Apache Spark (PySpark, Spark SQL), Microsoft Fabric, Azure Databricks, Azure Data Lake, Azure Data Factory, Medallion Architecture, ETL/ELT Pipelines, SAP BTP, ERP Systems (SAP Ariba).
- **Machine Learning & Analytics:** Exploratory Data Analysis (EDA), Pandas, Numpy, XGBoost, Supervised Learning (Regression, Classification), Recommender Systems, Feature Engineering, Matplotlib, Seaborn
- **Databases:** PostgreSQL, MySQL, SQLite, Firebase, Redis, Delta Lake, Data Modeling, Star Schema.

Work Experience

Data Consultant Intern – Kyndryl — May 2026 – December 2026

Digital Evidence Management System (DEMS) | *Client: Royal Canadian Mounted Police (RCMP)*

- **Architecting and building an Azure platform** that lets the RCMP and Ontario prosecutors securely share digital evidence for pre-charge consultations, owning the architecture, cloud infrastructure, development and hosting.
- **Provisioned the full Azure environment:** App Service for hosting, ADLS Gen2 for evidence files, Azure SQL for structured data, Microsoft Entra ID for authentication, Key Vault for secrets, and backup and recovery.

Appeals Analysis Pipeline | *Client: Health Standards Organization (HSO)*

- **Built an automated Python pipeline** using Microsoft Graph APIs to crawl nested SharePoint directories, extracting key metrics from 1,000+ hospital and clinic appeal files into a consolidated Excel summary.
- **Led client meetings end to end:** requirements gathering, solution architecture, development, UAT and delivery.
- **Designed an incremental sync engine** that reprocesses only modified files, cutting execution time by 70%, with validation, extraction-accuracy checks and deduplication ensuring 100% data consistency.

Data Support Analyst – Mindbridge AI — January 2026 – April 2026

- **Engineered automated Python data workflows** using Pandas to parse, clean, and standardize large-scale General Ledger datasets from various enterprise accounting systems.
- **Architected a modular Python script library** to convert varied client file formats into uniform schemas, ensuring smooth data ingestion into the core AI analytics engine.
- **Set up automated data quality and error-handling checks**, preventing corrupted records from crashing downstream modeling pipelines.

SAP Trainee – Atos — May 2025 – August 2025

- **Configured and customized SAP Ariba Buying workflows** and ERP procurement modules, translating enterprise business requirements into functional system specifications.
- **Executed regression testing and system validation across SAP BTP** purchase order modules, diagnosing cross-system integration issues and coordinating end-to-end fixes with backend teams.

Full stack developer - Netopia — May 2024– July 2024

- **Developed backend RESTful APIs** using Java and Spring Boot for an internal web application that manages employee leave requests and approvals.
- **Designed relational database tables in MySQL,** adding indexes and optimizing queries so the application could quickly look up employee records, balances, and request histories.

Personal & Independent Projects

End-to-End Bitcoin Data Pipeline & Analytics Engine — Live Demo | GitHub

- **Architected a cloud-based ETL pipeline** to ingest, validate, and store live cryptocurrency market data from REST APIs.
- **Implemented PySpark jobs** for schema enforcement and deduplication across Bronze, Silver, and Gold Delta Lake layers.
- **Deployed a containerized FastAPI backend** with Docker to serve Gold-layer data for real-time analytics queries.
- **Optimized query speeds**, indexing key analytical fields and caching frequent queries to ensure fast response times for user.

Med Rational – Production Mobile Application — Demo Video | API | Android App

- **Built a centralized mobile application** with a medical student association to streamline access to study guides.
- **Developed backend REST APIs** using Spring Boot for full-text search, resource filtering, and document retrieval.
- **Architected data schemas** and storage layers to support scalable document downloads and fast query responses.

Certificates:

- Databricks Certified Associate Developer for Apache Spark
- AWS Cloud Practitioner Essentials