

Mohamed Hafez

Machine Learning Engineer | AI Engineer | Applied ML Engineer

RAG & Vector Search · LLM Evaluation · NLP & Document Intelligence · Generative AI

Toronto, ON, Canada | +1 647 929 2480 | mohamed.m.nader555@gmail.com | linkedin.com/in/mohamed-nader555 | github.com/Mohamed-Nader555

Legally authorized to work in Canada for any employer — no sponsorship required. Open PGWP valid to June 2029.

Summary

Machine Learning Engineer with an **M.A. in Information Systems & Technology** and **3+ years** of professional engineering experience building end-to-end **ML and NLP** systems. Thesis research produced **ASC-PIE**, a PII-aware named-entity-recognition evaluation framework benchmarking encoder, encoder-decoder, and decoder-only models, and **SPRINT-PP**, a privacy-safe method for mitigating catastrophic forgetting. Recently built a containerized **RAG** service delivering grounded, citation-backed answers with an automated evaluation harness. Delivered **10+** applied ML projects spanning forecasting, classification, recommendation, and computer vision.

Technical Skills

Languages: Python, SQL, Java, Kotlin, Bash

Machine Learning: PyTorch, TensorFlow, scikit-learn, pandas, NumPy, feature engineering, cross-validation, deep learning, classification, regression, clustering, recommendation, imbalanced learning, model selection and evaluation

NLP & LLMs: Named-Entity Recognition, token classification, Transformers, Hugging Face Transformers/Datasets, instruction prompting, in-context learning, LLM evaluation, model benchmarking, span-level evaluation, class-incremental learning

RAG & Vector Search: Retrieval-Augmented Generation, LangChain, Chroma, vector databases, sentence-transformers, embedding models, semantic retrieval, document chunking, top-k retrieval, grounded prompting, source citation, hallucination mitigation, RAGAS

MLOps & Deployment: Docker, CI/CD, model packaging and versioning, reproducible experimentation, FastAPI, Flask, REST APIs, batch and near-real-time inference, Git, Linux

Cloud & Data: AWS, Google Cloud Platform, Microsoft Azure, MS SQL Server, dataset construction, data validation, label-quality checks, matplotlib, seaborn

Ways of Working: Agile, Scrum, weekly sprints, daily standups, Trello, Azure DevOps, code review, technical documentation, stakeholder communication

Professional Experience

Graduate Researcher — Machine Learning & NLP

Sep 2024 – Apr 2026

York University, CERAS Lab (Supervisor: Prof. Marin Litoiu)

Toronto, ON

- Designed end-to-end **ML/NLP pipelines** for privacy-aware information extraction — dataset construction, preprocessing, feature engineering, fine-tuning, inference, and evaluation — using **Python**, **PyTorch**, and **Hugging Face Transformers**.
- Built **ASC-PIE**, a unified English PII corpus and evaluation framework, standardizing heterogeneous datasets under a shared **19-entity schema** with data-validation and label-quality controls.
- Benchmarked **encoder**, **encoder-decoder**, and **decoder-only** architectures under a shared protocol, comparing supervised fine-tuning against in-context learning on span-level **precision, recall, and F1** alongside latency, output validity, and robustness.
- Developed **SPRINT-PP**, a privacy-safe method improving retention in class-incremental PII extraction without storing raw historical PII; evaluated against replay and distillation baselines.
- Built reproducible experimentation workflows with versioned schemas, configurations, prompts, and evaluation artifacts; presented findings to faculty and research collaborators.

Software Engineer — ECM Consultant

Aug 2023 – Sep 2024

BASS, an enterprise content management consultancy

Cairo, Egypt

- Developed **microservices** in **Java** and **Spring Boot** exposing **REST** and **SOAP** APIs, standardizing **JSON** schemas, authentication flows, and interface contracts consumed by client systems.
- Built validation scripts and metadata-quality controls for document workflows in **PII-aware, regulated environments**; performed root-cause analysis, performance tuning, and log-based investigation.
- Delivered in **Agile** weekly sprints with daily standups, tracking work in **Trello** and **Azure DevOps**.

AI & ML Specialist

Jun 2021 – Aug 2023

DotPy

Cairo, Egypt

- Delivered applied **machine learning** solutions across forecasting, customer segmentation, recommendation, sentiment and emotion analysis, fraud detection, and image classification using **Python**, **SQL**, **scikit-learn**, **TensorFlow**, and **PyTorch**.
- Owned projects end to end — scoping, data preparation, feature engineering, cross-validation, model training and selection, evaluation, packaging, and deployment — delivering notebooks, API-based services, and **Dockerized** prototypes with handover documentation.
- Applied **MLOps** practice including CI/CD workflows, model packaging, versioning, code review, and reproducible experiment tracking.

- Collaborated across teams on model-serving endpoints, REST/JSON contracts, rate limits, and error handling; worked in **Agile** weekly sprints with daily standups coordinated through **Trello**.
- Translated business requirements into delivery plans, aligned developers with stakeholders, and mentored junior developers.

Teaching Assistant & Technical Instructor

Jun 2019 – Apr 2026

York University; CIC, DotPy Academy, and private instruction

Toronto, ON | Cairo, Egypt

- Taught **200+ hours** across six institutions in Data Structures, Java/OOP, and AI/ML, and served as Teaching Assistant in Data Visualization and Systems Architecture at York University — demonstrated ability to explain complex technical material to non-expert audiences.

Projects

Northstar RAG System | Python, FastAPI, LangChain, Chroma, sentence-transformers, RAGAS, Docker

2026 | GitHub

- Built an end-to-end **retrieval-augmented generation** service over a mixed document corpus (PDF, TXT, Markdown) with **PyMuPDF** extraction, recursive chunking, local **sentence-transformer** embeddings, and a persistent **Chroma** vector store using cosine distance.
- Implemented **grounded generation with hallucination control**: retrieved chunks are filtered against a maximum distance threshold, and the service returns a fixed refusal rather than an unsupported answer when no chunk qualifies.
- Returned full provenance on every response — source document, page, chunk identifier, and retrieval distance — making answers auditable against the corpus.
- Built a **RAGAS** evaluation harness to compare retrieval configurations across faithfulness, answer relevancy, context precision, and context recall; packaged with **Docker Compose** and covered by **pytest**.

ASC-PIE — PII-Aware NER Corpus & Evaluation Framework | Python, PyTorch, Hugging Face, scikit-learn

2024 – 2026

- Constructed a unified English PII corpus standardizing heterogeneous public datasets under a shared **19-entity schema**, with data-preparation, validation, and label-quality controls.
- Built a reproducible benchmarking harness covering fine-tuning, inference, output parsing, and metric reporting across three model families under a single protocol.

Dive Simulation & Safety Profile Planner | Python, scikit-learn, Android (Java), Firebase, GCP

2024

- Trained and compared four classifiers (**SVC, Logistic Regression, Decision Tree, ANN**) to classify dive profiles as safe or unsafe from depth, bottom time, oxygen percentage, and calculated PPO2; the Decision Tree reached **99.5% test accuracy**.
- Served the model to a mobile client over a **POST-based API** hosted on **PythonAnywhere** and **Google Cloud Platform**, with an iterative recommendation routine that reduces depth and bottom time until a safe profile is found.

Mind’s Eye — Assistive Smart-Glasses System | Android (Java), Arduino, OpenCV, Tesseract OCR

Capstone — Grade A+

- Delivered **80%+** of a wearable assistive system integrating mobile, embedded, and cloud components for face, object, and currency detection with OCR-driven text-to-speech; built Egyptian currency recognition entirely from scratch.

Education

York University — M.A. Information Systems & Technology

Sep 2024 – Apr 2026

Thesis: ASC-PIE — An Evaluation Framework for PII-Aware Named-Entity Recognition

Toronto, ON

- Focus areas: Machine Learning, NLP and Named-Entity Recognition, Large Language Models, and privacy-aware AI systems. Supervised by Prof. Marin Litoiu, CERAS Lab.

Helwan University — B.Sc. Computer Science & Artificial Intelligence

Sep 2018 – Jun 2022

Excellent with Honors | Ranked 2nd in class | GPA 3.72/4.0

Cairo, Egypt

- Major: Computer Science. Minor: Information Systems. Capstone graded A+.

Information Technology Institute (ITI) — Mobile Application Development (Android)

Mar 2023 – Jul 2023

Professional certificate: Android engineering, REST integration, MVP/MVVM, Git workflows

Cairo, Egypt

Certifications & Awards

AWS Machine Learning Foundations — Udacity, 2022

Excellent with Honors — 2nd in class, GPA 3.72/4.0

Mobile Application Development (Android) — ITI, 2023

Capstone Grade A+ — Mind’s Eye

Certified Ethical Hacker (CEH) — Training Programme — Inspire Academy, 2022

ICPC Contestant — 2019, 2020, 2021

Languages: English (Professional proficiency), Arabic (Native)