FAREES SIDDIQUI

+1 (647) 554-7513 || fareesuniversity@gmail.com || github.com/FareesSiddiqui

TECHNICAL SKILLS

Programming Languages: Python (7+ years), JavaScript/TypeScript (5+ years), SQL (5+ years), Java (5+ years), C++ (3+ years), C (3+ years)

ML & NLP: PyTorch, TensorFlow, Hugging Face Transformers, spaCy, NLTK, scikit-learn, NumPy, Pandas, CUDA, OpenCV

Web Development: React.js, Node.js, Vue.js, Flask, MongoDB, PostgreSQL

DevOps & MLOps: Git, Docker, Linux/UNIX, MLflow, TensorBoard, Weights & Biases, CI/CD, AWS

EXPERIENCE

Teaching Assistant – Machine Learning II
Jan 2026 – Apr 2026

Ontario Tech University — Oshawa, Ontario

- **Contributed to the curriculum material** covering transformer architectures, LLM fine-tuning, prompt engineering, and reinforcement learning, enabling students to build production-grade NLP applications
- **Developed hands-on lab assignments** implementing text generation pipelines, attention mechanisms, and sequence-to-sequence models using PyTorch and Hugging Face Transformers
- **Mentored students** on advanced NLP concepts including tokenization strategies, embedding models, and retrieval-augmented generation (RAG) systems

Research Assistant – Document Intelligence
Jan 2025 – Apr 2025

Ontario Tech University — Oshawa, Ontario

- **Developed automated document processing pipelines** for parsing PDF academic papers, implementing OCR and layout detection for accurate equation extraction and section segmentation
- **Applied NLP techniques** to structurally segment documents into semantic components (abstract, methodology, equations, references) for downstream text extraction and analysis
- **Benchmarked state-of-the-art LLMs** including GPT-4, Google Gemini, and Meta's Nougat for document understanding tasks, evaluating text extraction accuracy across 500+ technical documents

Software Engineer (Contract)
Sep 2024 – Dec 2024

Bombardier Aerospace — Mississauga, Ontario

- **Designed and built web application** using Python, Flask, and JavaScript to automate configuration document generation using template-based text synthesis, reducing manual effort by 80%
- **Implemented automated data processing workflows** for network infrastructure decommissioning, ensuring secure data retention and compliance with organizational standards

Research Assistant – Database Lab
Sep 2023 – Apr 2024

Ontario Tech University — Oshawa, Ontario

- **Conducted research on neural network similarity metrics** through embedding vector analysis, developing novel methods to quantify semantic similarity across different model architectures
- **Built large-scale data pipelines** using Hugging Face, PyTorch, and scikit-learn to extract and compare model embeddings, processing millions of vectors for similarity analysis
- **Optimized training workflows** with CUDA acceleration, reducing experiment iteration time from 8+ hours to under 2 hours

Software Engineer (Full-time Co-op)
Sep 2023 – Aug 2024

Bombardier Aerospace — Mississauga, Ontario

- **Engineered automation solutions** to configure 1,500+ network switches programmatically, delivering $30,000+ in cost savings
- **Built data analysis pipelines** using Python, Selenium, and Pandas to query network devices and generate actionable insights for management decision-making
- **Created data visualization tools** using Matplotlib and Pandas to generate calendar heatmaps for task scheduling optimization

Machine Learning Engineer   Jun 2020 – Aug 2020
Recruit For Me   Toronto, Ontario

- **Developed NLP-based document matching system** to automatically match job descriptions with candidate resumes, implementing semantic similarity scoring using cosine similarity on text embeddings
- **Built named entity recognition (NER) pipeline** using spaCy to extract and tag key entities (skills, qualifications, experience) from unstructured resume and job posting text
- **Designed ranking algorithm** to score and rank candidate-job matches based on extracted entities and semantic similarity, improving recruiter efficiency in candidate screening
- **Utilized PyTorch, spaCy, NumPy, and Matplotlib** to build end-to-end ML pipeline from data preprocessing to model evaluation and visualization

EDUCATION

Ontario Tech University – Database Lab   Sep 2025 – Present
Master of Science: Computer Science   Oshawa, Ontario

- **Research Focus:** Agentic AI systems for workflow automation, tool-use learning in LLM agents, retrieval-augmented generation (RAG) optimization through reinforcement learning, document layout analysis, and OCR pipelines for structured information extraction
- **Thesis:** Investigating reinforcement learning approaches to improve RAG system performance, including reward modeling for retrieval relevance, policy optimization for query reformulation, and agent-environment interaction for multi-step document reasoning tasks

Ontario Tech University   Sep 2020 – Apr 2025
Bachelor of Science: Honours Computer Science (Co-op), Comprehensive   Oshawa, Ontario

- **Relevant Coursework:** Machine Learning I (A+), Machine Learning II (A+), Algorithms (A-), Databases (A), Massively Parallel Programming (A)
- **Awards:** Tribute Communities Scholarship ($10,000), President's List (4.06/4.3 GPA)

HONOURS THESIS

Model Similarity through Embedding Vector Analysis   Sep 2023 – Apr 2024
Ontario Tech University – Database Lab   Oshawa, Ontario

- **Developed novel quantification framework** for measuring semantic similarity between neural network architectures by analyzing their learned embedding representations
- **Designed and implemented embedding extraction pipelines** to compare 50+ pre-trained models from Hugging Face, computing pairwise similarity metrics across model families
- **Applied dimensionality reduction and clustering techniques** to visualize model relationships and identify architectural patterns in embedding space
- **Research conducted under the Database Lab**, contributing to ongoing work in model interpretability and transfer learning analysis

PROJECTS

Chess Position Classification & RL-based Chess Engine | GitHub

- Built vision-based chess analysis pipeline by fine-tuning YOLOv5 to detect chess pieces from board images and convert positions into FEN notation
- Integrated Proximal Policy Optimization (PPO) reinforcement learning agent using OpenAI Gym framework to suggest optimal moves based on detected positions