Sameer Zaheer

smrzhr@gmail.com | 905-782-9601 | Toronto |linkedin.com/in/sameerzaheer/

SUMMARY: Experienced software developer seeking opportunities in AI, machine learning, algorithms and modelling. I am a fast learner, excellent communicator and value good software engineering practices.

LANGUAGES: python, TypeScript, C++, C#, bash/shell, MySQL

LIBRARIES: pyTorch, pandas, Eigen, numPy, sklearn

OTHER: git, Docker, kubernetes, Jenkins, linux

ENGINEERING EXPERIENCE

Forward Deployed Engineer, Palantir Technologies, Toronto 2023/02 – 2025/05

- Designed and implemented an end-to-end pipeline scheduling workflow for a Canadian midstream client, simulating the entire system, optimizing for commodity value and power consumption, and respecting constraints (tank sizes etc). The workflow reduced human labour ~3x and generated ~2x in energy savings.
- Designed, implemented and patented a machine-learning and LLM based workflow to find and visualize the most optimal parcels of land in North America for greenfield wind power development. Reduced land finding time from a month to an hour.
- Created a cost forecasting workflow for a construction client that used both machine-learning and manual data entry to accurately forecast costs and profits on projects.
- Wrote a computer vision script that processed ~20,000 parts blueprints of an automotive manufacturer to find interchangeable parts, in order to reduce bottlenecks and leverage economies of scale.

Software developer, SoundHound Inc, Toronto 2018/02 – 2022/11

- Developed an end-to-end machine learning (ML) pipeline for keyword spotting, consisting of data processing, training neural networks (NNs), model deployment, scaling pipeline to produce 100+ models semi-automatically.
- Built scalable data pipelines (Python, Bash) and automated ML model training, reducing development time by 2 months through synthetic data generation and augmentation.
- Innovated and rapidly prototyped ML solutions using PyTorch, including automating hyperparameter search for DNNs, CNNs, and RNNs, leading to a 30% reduction in error rates.
- Optimized inference systems in C++ (Eigen, perf tools) to cut CPU and memory usage by 40%, enabling deployment on resource-constrained devices (e.g., Raspberry Pi).
- Led cloud deployment of training/testing infrastructure using Kubernetes and Docker, freeing up GPU resources
- Demonstrated strong leadership and cross-functional impact by mentoring developers and linguists, onboarding team members, resolving critical client-facing ML issues, and running agile standups and sprints
- Streamlined workflow by building an intuitive YAML interface for ML training, empowering non-developers and resolving post-release bugs with quick turnaround to client satisfaction.
- Improved the accuracy of French speech recognition (ASR) by 5% by modifying the language model to infer liaison pronunciations

Algorithmics engineer, Claron Technology, Toronto 2013/01 – 2017/12

- Developed a fully automatic and accurate algorithm in C# to segment the spine and vertebrae in CT images. Evaluated and optimized the algorithm on ~200 datasets. Improved this software's time performance by optimizing low-level algorithms, eliminating bottlenecks, and parallelization. Productized this software in accordance with ISO 13485.
- Maintained the Withinsight Framework (WIF), a versatile SDK for image processing and visualization. This included debugging, maintaining a usable API and implementing automated integration and unit tests.
- Rapidly prototyped the following applications: automatic finding of landmarks on the pelvis in CBCT images; automatic detection of femur head and acetabulum cup in CT images; and automatic detection of vertebra in 2D sagittal MR images.

Research Assistant *2009-2010*

Defence Research and Development Canada – Toronto

- Designed and implemented a synchronized experiment measuring facial expression (using 3D cameras), electroencephalography and electrocardiography.
- Developed data processing algorithms (filtering, artifact detection and correction or rejection) in MATLAB, analyzed the data in both time and frequency domains and found EEG frequency and ERP indicators of task complexity, in tasks with increasing memory load

---

EDUCATION

Master of Health Science, *2011 – 2013*

Institute of Biomaterials and Biomedical Engineering, University of Toronto, Canada
Thesis: Development of a knowledge-guided, automatic, near real-time algorithm for segmenting vessels from MR volumes of the liver
Supervisors: Dr. James Drake and Dr. Edward Huang

Bachelor of Applied Science, University of Toronto, Canada *2006 – 2011*

Division: Engineering science, Biomedical Engineering option
GPA: 3.52

PATENTS

N. Nainani, **S. Zaheer**, G. Seite, E. Haddad, D. Wai, "Machine Learning and Language Model-Assisted Geospatial Data Analysis and Visualization." U.S. Patent 20240394296, published Nov 28, 2024.

N. Jain, **S. S. Zaheer**, "Server supported recognition of wake phrases." U.S. Patent 20220148572, published May 12, 2022.

I. Bitter, R. Chandrashekara, S. Stapleton, **S. Zaheer**, "System and Methods of Segmenting Vessels from Medical Imaging Data. " U.S. Patent 20170154435, published Jun 1, 2017.
