Hooman Ramezani

hooman125@gmail.com | www.hoomanramezani.com | github.com/HoomanRamezani

Education

University of Toronto — Toronto, ON
MASc in Machine Learning & Operations Research — *2023 – 2024*
- Thesis: Building healthcare LLM and ViT models with multimodal clinical data for lung cancer treatment
- GPA: 4.0 / 4.0 — Coursework: Large Language Models, Cloud Data (Spark, AWS), Deep Learning Theory

University of Waterloo — Waterloo, ON
BASc in Systems Design Engineering — *2018 – 2023*
- GPA: 3.7 / 4.0 — Coursework: Intro. Deep Learning, Intro. Machine Learning, Pattern Recognition, Neuroscience

Experience

Machine Learning Researcher — May 2023 – Present
University Health Network — *FT – Toronto, ON*
- First-author 'LN-Transformer' paper for lung nodule segmentation, achieving DiceC of 94.2% (SOTA)
- Tuned **multi-modal LLM** for treatment planning with 84% concordance, utilizing CT imagery and clinical notes

Machine Learning Engineer — Sept. 2022 – April 2023
Advanced Micro Devices (AMD) — *Intern – Toronto, ON*
- Accelerated **LLM inference** on AMD CPUs, cutting latency by 80% via pruning, quantization, and distillation
- Designed scalable **data pipelines** to preprocess 5TB dataset with Spark and SQL, enhancing distributed training
- Developed content recommendation system with candidate retrieval and reranking with latency of **200ms**

Machine Learning Engineer — Jan. 2022 – April 2022
DarwinAI (Acquired by Apple) — *Intern – Remote*
- Developed **XGBoost + CNN** model for Pfizer, achieving 91% sensitivity in Liver Fibrosis diagnosis
- Automated MLOps CI/CD pipelines for retraining & monitoring, accelerating deployment time by 50%
- Partnered with 3 **enterprise clients** to translate requirements into scalable deep learning solutions on edge

Computer Vision Engineer — May 2021 – Aug. 2021
Applied Brain Research — *Intern – Waterloo, ON*
- Achieved 94% **drone object detection** accuracy of defects for edge-deployed CNN model for real-time scanning
- Designed Unreal Engine 4 simulation to synthetically generate 50,000 annotated images to overcome lack of data

Research // Projects

Livy Education Chatbot Agent | *GPT-4, RAG, LangChain, Chroma* — 2024
- Developed a RAG-powered AI chatbot for Canvas students, integrating GPT-4 for assignment research assistance
- Utilized Chroma vector database and fine-tuned embeddings for course material reranking, improving relevancy
- Built context-aware **AI agent** to adaptively monitor student engagement and provide real-time support

Advanced Detection of Parkinson's Freezing of Gait | *CNN, Time-Series Data* — 2023
- Achieved 94% accuracy for fall forecasting for early detection of Freezing of Gait (FOG) in Parkinson's patients
- Built novel preprocessing techniques for **EMG, ECG** sensor data, improving time-series feature extraction

Enhanced Robotic Grasping with PointNet | *LiDAR, Robotics, 3D Computer Vision* — 2022
- Trained grasp proposition neural-net achieving 87% grasp success rate, adapting custom PointNet deep learning models to identify optimal grasp points from LiDAR camera, on dataset with **50k+ annotated grasp** scenarios

Technical Skills

Languages: Python (NumPy, pandas, Matplotlib), SQL, C/C++, Java, JavaScript
AI & ML: PyTorch, TensorFlow, Scikit-learn, XGBoost, Random Forest, OpenCV, Deep Learning, RAG, AI Agents
Data & Cloud: Spark, Hadoop, MapReduce, Airflow, SQL, Kafka, AWS (S3, EC2, Lambda, SageMaker), GCP
Deployment: Kubernetes, Docker, RESTful APIs, LangChain, Hugging Face, Flask, CUDA, Git