Mansour Saffar

Machine Learning, Generative AI, MLOps, NLP, AdTech

+1 (587) 937 0770 | linkedin.com/in/msaffarm | msaffarmehrjardy@gmail.com

Work Experience

07/24 - Present **Senior Machine Learning Engineer** Pinterest, Toronto

• Leading the GenAI content detection, disclosure, and distribution initiative across Pinterest. This is a strategic project closely tied to user retention KPIs.[Pinterest blog][TechCrunch blog]

• **Technologies:** *Python, Java, PyTorch, Transformers(HF), PEFT, Airflow, PySpark, QWEN VLM, Querybook, MCP, RAG, LangChain*

01/24 - 07/24 **Lead Machine Learning Engineer (Unity Ads)** Unity, Toronto

• Led Unity Ads Audience Pinpointers (APP) ML team to develop end-to-end data, ML, and MLOps solutions/services for APP Ad products.

• Led the data pipeline optimization initiative with data team at Unity Ads which led to **$XX million increase in yearly revenue** due to more accurate data for model training.

05/22 - 01/24 **Senior Machine Learning Engineer (Unity Ads)**

• Worked as a senior MLE/engineering tech lead in the app-event team. My duties included sprint planning for engineering, developing and optimizing backend/data/ML services, and writing/reviewing design documents where we **developed multiple app-event campaign types for the Unity Ads network**.

• Developed and optimized end2end pipelines including data ingestion pipeline, streaming/batch data processing pipeline, data validation pipeline, model training pipeline, model serving pipeline/microservice, and data/model monitoring pipelines to support app-event campaigns.

• **Technologies:** *Python, Go, Tensorflow, Kubeflow, Airflow, PySpark, BigQuery, BigTable, Docker, Kubernetes, GKE, Dataproc, Composer, Kafka, Prometheus, Grafana, Terraform, Jira, Jenkins, GitHub Actions*

01/22 - 05/22 **Lead Machine Learning Engineer** AltaML, Toronto

• Served as ML Solution Architect and team lead in multiple machine learning projects.

• Designed the architecture and led a team of 5 ML devs to build a **customizable object detection solution** built on top of Azure.

03/21 - 12/21 **Senior Machine Learning Engineer**

• Led multiple teams of machine learning engineers (2-3 team members) in multiple projects, including recommender systems and time-series forecasting.

• Designed the architecture and implemented a **machine learning model deployment pipeline** built on top of AWS and Kubernetes (EKS). This pipeline is used internally to deploy ML models.

03/18 - 03/21 **Machine Learning Engineer**

• Developed **NLU** modules (intent recognition and entity detection), **extractive text summarization**, **AdWords generation**, **keyword extraction**, and **OCR** in various NLP projects.

• Designed and developed multiple task-oriented chatbot projects including **synthetic data generation tool for NLU and dialogue management**, **chatbot to answer complex natural queries from a financial database**, **customer service** and **FAQ chatbots**.

• **Technologies:** *Python, PyTorch, Tensorflow (tensor2tensor), Tf-Serving, Rasa, spaCy, NLTK, Gensim, Scikit-learn, MongoDB, Pandas, Dask, PySpark, AWS, Azure, Git, DVC, Docker, Kubernetes, MLflow*

Education

2016 - 2019 **M.Sc in Computer Science (GPA: 4/4)** University of Alberta, Edmonton

• With a focus on Machine Learning, Natural Language Processing and Chatbots

2011 - 2016 **B.Sc in Electrical Engineering (GPA: 3.67/4)** University of Tehran, Tehran

• With a focus on Machine Learning and Medical Image Processing

Select Projects

09/17 - 11/18 **Deep Learning Models for Task-oriented Chatbots** Graduate Research Assistant (Master's Thesis)
• Researched usage of **self-attentional models** for training end-to-end task-oriented chatbots. The results showed faster training with comparable accuracy. [Source Code] [Publication Link]
• Developed **ChatSim**, an **architecture agnostic evaluation framework** for task-oriented chatbots that can model **user characteristics** and behaviour in chatbot evaluation. [Source Code] [Publication Link]
• **Technologies:** *Python, Tensorflow (tensor2tensor), Rasa, spaCy, Git*

09/16 - 11/16 **Retinal Image Segmentation** Machine Learning Course
• Developed a segmentation model by applying **ensemble and SVM models** on retinal images. Faced with the problem of small dataset size, we achieved good results by using bagging methods. [Report Link]
• **Technologies:** *Python, MATLAB*

Technical Skills

Languages **Programming Languages:**
Python (5+ years), Go (2 years), C++ and Java

ML/DL/XAI **Machine Learning, Deep Learning, and Explainable AI:**
Scikit-learn, H2O, PyTorch, Tensorflow, SHAP

GenAI/NLP **Natural Language Processing and Generative AI:**
ms-swift, Transformers, spaCy, NLTK, Gensim, Model distillation, PEFT, LoRA, MCP, LangChain

Data **Big Data Analysis Frameworks and Databases:**
MySQL, BigTable, BigQuery, Pandas, Dask, MongoDB, Redis, PySpark, Kafka

Cloud **Cloud Computing Platforms:**
AWS (Certified ML-Speciality, EKS, Lambda, SageMaker), Azure (Azure ML, App Service, AKS), GCP (GKE, BigQuery, BigTable, Composer, Dataproc)

MLOps/DevOps **Machine Learning Model Deployment and MLOps:**
Docker, Kubeflow, Ariflow, MLflow, Streamlit, Flask, FastAPI, Terraform, Kubernetes, TF-Serving, Grafana, Prometheus, Jenkins, GitHub Actions

Publications

August 2019 **Mansour Saffar, Amine Trabelsi, Osmar R. Zaiane**
• *Self-Attentional Models Application in Task-Oriented Dialogue Generation Systems*
Recent Advances in Natural Language Processing (RANLP 2019) [Publication Link]

February 2019 **Ghazal Sahebzamani, Mansour Saffar, Hamid Soltanian-Zadeh**
• *Machine Learning Based Analysis of Structural MRI for Epilepsy Diagnosis*
International Conference on Pattern Recognition and Image Analysis (IPRIA 2019) [Publication Link]

Volunteering

October 2020 **Public Speaker** Data Science Meetup, Edmonton
• Presented a talk, **From RNNs to GPT-3**, about the progression of deep learning for NLP [Slides]

03/19 - 10/19 **Machine Learning Engineer in Ana Project** Joint Project with Amii
• Collaborated with Prof. Zaiane's team in architecture design and developing NLU modules for Ana, an intelligent chatbot with emotional capabilities designed to help older adults. [Mentioned on CBC News]

October 2019 **Public Speaker** Data Science Meetup, Edmonton
• Presented a talk about **explainable AI (XAI)** and its application in industrial ML. [YouTube Video] [Slides]