Mostafa Malmir

malmir.edumail@gmail.com (210) 350–9263 15651 Chase Hill Blvd, Apt 1110, San Antonio, TX
linkedin.com/in/mostafa-malmir github.com/msmalmir Google Scholar

SUMMARY

PhD candidate specializing in development and application of Deep Learning (DL), Machine Learning (ML), and statistical models for high-impact real-world datasets. Ability to build data pipelines, engineer features, and deploy predictive models in domains such as computational biology. Seeking opportunities to drive data-driven solutions on cross-functional teams.

EXPERIENCE

Research Assistant, University of Texas at San Antonio 2022–Present

- **Automated ML Pipeline for Single-Cell Clustering (sc2Assign R Package):** Packaged an end-to-end R workflow implementing PCA/UMAP for feature reduction, Pearson Correlation Coefficient for feature scoring and Euclidean-based classification, significantly boosting F1 scores on benchmark datasets especially for rare cell populations.
- **Transformer Classifier for Cell-Type Annotation (scTRaCT Python Package):** Built a supervised PyTorch transformer that fuses log-normalized gene expression with MCA-derived distance metrics. Leveraged multi-head self-attention and Focal Loss to capture subtle transcriptional differences and boost rare-cell classification accuracy.
- **Data-driven analysis on large single-cell datasets:** Extract key features that guide biologist partners in clustering and prioritizing critical cell populations (e.g. pain-related neuron subtypes and mammary stem-cell candidates)
- **Transformer-Based DFU Healing Predictor:** Collaborated on a PyTorch transformer that models single-cell expression as gene-sequence inputs, uses multi-head self-attention to capture gene-gene dependencies, highlights top attribution-weighted markers, and achieves over 90 % accuracy on fibroblast and SMC test sets.
- **RAG Chatbot:** Built a domain-specific chatbot using LangChain and Hugging Face models with a Retrieval-Augmented Generation (RAG) pipeline. Designed to extract accurate answers from a 37-page legal document provided by Chen Immigration Law Associates.
- **Dialogue Summarization Fine-Tuning:** Fine-tuned Flan-T5 Base, Flan-T5 Large, and Deepseek 1.5B models on the DialogSum dataset using LoRA for efficient training, and model evaluation based on ROUGE metrics.
- **Personalized Housing Assistant:** Created an apartment-hunting assistant using GPT-3.5-turbo with prompt engineering. The chatbot provided rental suggestions based on user preferences within a specific housing complex.
- **Face Recognition CNN model:** Built and benchmarked GoogleNet vs. SqueezeNet on the 25K-image LFW dataset, incorporating data augmentation, hyperparameter tuning, and GPU-accelerated training.

TECHNICAL SKILLS

Languages and Libraries: Python (NumPy, Pandas, SciPy, statsmodels), R (Tidyverse, ggplot2, lme4, survival, caret), Linux, SQL, CQL, MQL, Git, MATLAB, Jupyter, VS Code
Statistical Methods: PCA, UMAP, MCA distance metrics, Pearson/Spearman correlation, t-tests, ANOVA, chi-square, GLMs (linear/logistic),ARIMA, Bayesian inference (PyMC3, Stan)
ML/DL Frameworks: Scikit-learn, PyTorch, TensorFlow, Transformers/LLMs, GenAI, NLP, RLHF, RAG (LangChain, Hugging Face), LoRA fine-tuning, XGBoost, K-means, CNNs, DNNs
Databases & Big-Data Tools: MongoDB, Cassandra, NoSQL, CRUD operations
Data Viz & EDA: Plotly, seaborn, ggplot2
Cloud: AWS (intro), GitHub

EDUCATION

University of Texas at San Antonio, PhD Electrical and Computer Engineering 2022–Present
Tarbiat Modares University, MS Electrical and Computer Engineering 2013–2016
Enghelab Eslami Technical University, BS Electrical Engineering 2009–2013

RELATED COURSES AND CERTIFICATS

Deep Learning • Machine Learning • NLP Specialization • Generative AI with Large Language Models • Mathematics for ML • Data Analysis with Python • SQL for Data Science • Introduction to NoSQL Databases • Git and GitHub • AWS Cloud Essentials

SELECTED PUBLICATIONS

- **Malmir, M.**, et al. “sc2Assign: Automated marker- and non-marker-gene based cell-type assignment from scRNA-seq datasets at single-cell level.” *iScience* (under review).
- **Malmir, M.**, et al. “scTRaCT: Transformer-based single-cell RNA-seq cell-type identification with a focus on rare populations.” In *ICIBM 2025* (under review).
- Jamil, U., **Malmir, M.**, et al. “Developing an eco-driving strategy in a hybrid traffic network using reinforcement learning.” *Science Progress* (2024).

LEADERSHIP & MENTORSHIP

- **President (Aug 2023-Sep 2024) & Vice President (Oct 2024-Present), Iranian Student Organization, UTSA**: Led a team of six to fundraise and run cultural events for a 200-member community, fostering Iranian heritage and alumni connections.
- **Graduate Mentor, NSF REU Program (Summer 2023 & 2024)**: Guided two undergraduates in computational-biology research-advising on data analysis, pipeline development, and experimental design.