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/momalmir   Google Scholar

SUMMARY

PhD candidate specializing in the development and application of advanced Deep Learning (DL), Machine Learning (ML), and statistical models for high-impact problems in computational biology. Experienced in building transformer-based models for rare cell type identification, gene marker discovery, and clinical data interpretation. Seeking opportunities to contribute to cutting-edge projects at companies focused on impactful, real-world applications.

EXPERIENCE

Personal Projects   2022–Present

- **Smart Job Alerts** (AI-Powered Resume-Based Job Matching): Designed and deployed an automated job matching system using LLMs (via OpenRouter) and the JSearch API. Implemented semantic filtering to reduce LLM cost, added support for preferred application sources, and automated alerts via GitHub Actions and Docker.
- **Smart Resume Optimizer**: Developed a full-stack web application that uses LLMs (via OpenRouter) to generate tailored resumes from user-uploaded PDFs and job descriptions. Built with Streamlit, FastAPI, and Docker, the app supports multiple language models (e.g. Claude, GPT-4, DeepSeek) and generates downloadable PDFs via Pandoc. Deployed on Hugging Face Spaces with GitHub Actions CI/CD integration. Focused on user privacy with no data stored or logged.
- **Conversational 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 LoRA FineTuning:** 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.
- **RentaBot:** Created an interactive apartment-hunting assistant using GPT-3.5-turbo with prompt engineering. The chatbot provided customized rental suggestions based on user preferences within a specific housing complex.
- **Human Face Recognition with Convolutional Neural Networks (Advanced Deep Learning course project):** Built a computer-vision pipeline for the 25 k-image LFW dataset; compared GoogleNet and SqueezeNet CNNs with augmentation, hyper-parameter tuning, and 25 k GPU training iterations.

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

- **scTRaCT (Python Package):** Developed a transformer-based model for cell-type annotation in scRNA-seq, integrating log-normalized expression with MCA-derived distance features. Uses self-attention and Focal Loss to boost accuracy for rare cell types. Outperformed human and mouse benchmarks, offering interpretability via attribution scores.
- **DFU Project:** Collaborated in developing a transformer model that predicts diabetic-foot-ulcer healing from single-cell data with over 90 % accuracy and pinpoints high-impact gene markers for clinical follow-up.
- **sc2Assign (R Package):** Developed an ML tool for automated cell-type annotation from scRNA-seq data. Utilized PCA and UMAP for dimensionality reduction and visualization, Pearson correlation to score marker genes, and Euclidean distance to incorporate non-marker gene data. This workflow boosted F1 scores on benchmark datasets, particularly for rare cell populations.
- **Mouse Trigeminal Neuron Analysis:** Led quality-controlled integration and clustering of single-cell neuron datasets in Seurat, revealing pain-associated cell subtypes and transferring labels to new samples.
- **Mammary Stem Cell Identification (Ongoing):** Integrated single-cell data to flag 216 candidate stem cells, ranked key genes with ML-based analyses, and shortlisted surface-protein markers for follow-up validation.

TECHNICAL SKILLS

Languages and Libraries: Python (NumPy, Pandas, SciPy, statsmodels, PySpark), R (Tidyverse, ggplot2, lme4, survival, caret), Bash, Linux (CLI), SQL, CQL, MQL, Git, MATLAB, Jupyter, VS Code

Statistical Methods: PCA, UMAP, MCA distance metrics, Pearson/Spearman correlation, t-tests, chi-square, GLMs (linear/logistic), Bayesian inference (PyMC3, Stan)

Bioinformatics Tools: scRNA-seq, NGS, Cell Ranger, Seurat, Scanpy, DESeq2, Harmony, DGE analysis, QC (FastQC), Alignment (HISAT2, STAR), Functional Enrichment (GO/KEGG)

ML/DL Frameworks: Scikit-learn, PyTorch, TensorFlow, CNNs, Transformers/LLMs, GenAI, NLP, RAG (LangChain, Hugging Face), LoRA fine-tuning, RLHF, XGBoost, K-means, DNNs

Databases & Big-Data Tools: Apache Hadoop, Apache Spark, MongoDB, Cassandra, NoSQL, CRUD operations, ETL

MLOps & DevOps: CI/CD (GitHub Actions), Docker, Kubernetes, MLflow, Model Deployment, Monitoring, Model Registry, Pipeline Automation

Cloud: AWS , Azure, GitHub

APIs & Web Frameworks: FastAPI, Flask, RESTful API design

Data Viz & EDA: Matplotlib, Seaborn, Plotly, ggplot2

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

DATA SCIENCE RELATED COURSES

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 • MLOps | Machine Learning Operations Specialization • 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).
- Mecklenburg, J., Shein, S. A., **Malmir, M.**, et al. "Transcriptional profiles of non-neuronal and immune cells in mouse trigeminal ganglia." *Frontiers in Pain* (2023).
- Jamil, U., **Malmir, M.**, et al. Developing an eco-driving strategy in a hybrid traffic network using reinforcement learning. *Science Progress* (2024).

For complete list of publications please visit my google scholar.

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.