MATEENAH JAHAN

jahanmateenah55@gmail.com | mateenahjahan.github.io | GitHub: MateenahJAHAN | Research map

Research interests: computational genomics · machine learning for healthcare · model evaluation and mechanistic interpretability

RESEARCH EXPERIENCE

Fatima Fellowship, *Predoctoral Research Fellow* **Remote**
Mentor: Dr. Marouen Ben Guebila, Viswanathan Lab, Dana-Farber Cancer Institute *2026 – Present*

- **Research focus:** How sex-chromosome expression in cultured cells relates to recorded donor sex and to drug response.
- **CMap/GTEx screen:** Used the 33 GB CMap matrix (720,216 signatures) to compare 36,523 drug-response signatures for 50 compounds from 135 cell lines (20 lineages) with sex-biased genes in 44 GTEx tissues: 39,996 drug–lineage–tissue tests using GSEA, Mann–Whitney U and Fisher's exact tests with FDR correction. Notebook | Code
- **Draft manuscript with Dr. Ben Guebila:** In 2.2 million vehicle-treated cells from 50 Tahoe-100M lines, recorded donor sex did not match sex-chromosome expression in 27 of 49 lines; the measured patterns agreed with external RNA-seq for 47 of 49. A LINCS2020 drug-response comparison (114 compounds, 41 lines) was inconclusive. Manuscript | Supplement

Technical Alignment Research Accelerator (TARA), *Technical AI Safety Fellow, Round 2*
Model evaluation and mechanistic interpretability *Sep – Dec 2026*

- **Tool-calling study:** In Qwen2.5-1.5B-Instruct on BFCL tasks, native tools made more false calls on irrelevant requests than a text tool list (16/32 vs. 9/32), and the gap held on 20 request pairs (Qwen2.5: 19/20 vs. 14/20; Qwen3.5: 10/20 vs. 0/20). Activation steering did not reduce these false calls. Results | Code

CSIR–Institute of Genomics and Integrative Biology (CSIR-IGIB), *Research Trainee* **New Delhi, India**
CAR-T / TCR repertoire analysis; supervised by Dr. Sherry Bhalla *Aug 2025 – Jan 2026*

- **TCR repertoires and CAR-T response:** Applied DeepTCR models to 239,637 TCRβ sequences from 34 patients with B-cell lymphoma (18 CD19 CAR-T responders, 16 non-responders). The original split mixed each patient's sequences across training and test; holding out whole patients across five folds gave AUROC 0.32 (95% CI 0.14–0.51), no predictive signal. Results | Code
- **Research software:** Helped update DeepTCR for Python 3.10 and TensorFlow 2.12 (credited in release 2.1.28); maintained a pinned environment and 11 updated tutorial notebooks. Upstream credit | Notebooks

EDUCATION

M.Sc. in Computer Science: Artificial Intelligence and Machine Learning, Woolf Awarded Aug 2026
Delivered through Scaler Neovarsity · GPA 4.0/4.0 · 90 ECTS · Capstone (Applied Computer Science Project): 98/100

B.Sc., University of Kashmir (chemistry, botany, zoology) 2014 – 2017

SELECTED RESEARCH AND SOFTWARE OUTPUTS

- **Biomni-AD drug repurposing (Mar 2026):** Built two Biomni agent workflows on the ADDI AD Workbench that pulled 2,331 Alzheimer's risk genes from the GWAS Catalog and queried Open Targets and ChEMBL for drug targets, producing 10 ranked repurposing hypotheses for experimental testing. Code | Results
- **MedGemma dataset (Apr 2026):** Released 13 hematologic-oncology test prompts for MedGemma-4B (11 run on the pretrained and instruction-tuned models, plus 2 controls) with all 24 outputs, reference answers and annotations (CC BY 4.0). Dataset
- **Refusal directions by harm category (BlueDot Impact, 2026):** Tested whether four harm categories have separate refusal directions in Qwen2.5-1.5B-Instruct. The category directions were less similar than chance ($p = 0.003$), but no candidate layer passed the standard KL < 0.1 check (values 0.13–1.68), so the question remains open. Code | Results

AWARD

Round 1 winner (one of three), Biomni-AD Discovery Prize, ADA Consortium challenge. Announcement

SKILLS

Programming: Python (pandas, NumPy, SciPy, statsmodels, scikit-learn), SQL (MySQL, BigQuery), C++, Bash, Git, Docker
Deep learning: PyTorch, TensorFlow/Keras, Hugging Face Transformers, DeepTCR; CNNs, RNNs, transformers, autoencoders
Machine learning: PCA, t-SNE, UMAP, clustering; ensemble methods; NLP (BERT, NER, topic models); time-series forecasting
Genomics: CMap/LINCS L1000, GTEx, Tahoe-100M single-cell data, Scanpy/AnnData, GSEApy, TCRβ repertoires, GWAS Catalog
Statistics: Mann–Whitney U, Fisher's exact, sign and permutation tests; FDR correction; bootstrap intervals; cross-validation
Interpretability and evaluation: Difference-in-means directions, directional ablation, activation steering with PyTorch hooks, linear probes, KL-divergence checks; tool-calling (BFCL) and clinical LLM evaluation
Tools: Jupyter, Google Colab, GitHub, FastMCP (MCP servers), Biomni agents, Hugging Face Hub, Tableau, Looker Studio