# Aarushi Gajri

aarushigajri@gmail.com | linkedin.com/aarushigajri | Graduate Student

## Education

**Georgia Institute of Technology**                            Aug 2024 – Present
M.S. in Bioinformatics                                           *GPA: 4.0/4.0*
**Award:** Graduate Research Award for AI-driven MoRF research (Torres Lab)

**Indian Institute of Technology (IIT), Mandi**                 Jul 2019 – Jul 2023
B.Tech (Honours) Bioengineering, Minor in Computer Science     *GPA: 8.78/10 (Institute Silver Medalist)*
Thesis: *Neural Network-based GWAS Analysis for ALS (generalizable method)*

## Research and Professional Experience

**Regeneron Pharmaceuticals**                                  May 2025 – Aug 2025
*Graduate Intern, Molecular Profiling and Data Science*

- Designed scalable graph-based algorithms to cluster B-cell receptor families based on VDJ gene usage.
- Explored hierarchical structures within immune repertoires to uncover lineage and clonal selection patterns.

**Georgia Institute of Technology - Torres Lab**               Aug 2024 – Present
*Graduate Research Assistant*

- Developed advanced NLP pipelines using LLMs to mine biomedical abstracts, extract MoRF triggers, and classify them into molecular categories.
- Constructed a curated knowledge base of MoRF-trigger interactions and induced secondary structure transitions using prompt engineering and regex parsing.
- Engineered deep-learning models (ESM2 embeddings + BiLSTM, dual-head PLM fine-tuning) to predict residue-level binding sites and disorder propensities.

**MedGenome Labs**                                             Jul 2023 – Jul 2024
*Bioinformatics Analyst, Genomics and Personalized Medicine*

- Built an ensemble machine learning pipeline to predict polygenic risk scores (PRS) for cardiac health using clinical genomic datasets.
- Developed deep-learning frameworks for low-pass whole-genome imputation; replicated and extended results from GLIMPSE 2.0.

**IIT Mandi - Undergraduate Thesis**                           Jan 2023 – Jul 2023
*Researcher, ALS GWAS Project*

- Implemented neural network-based genome-wide association models to identify ALS-associated variants from WGS data.
- Proposed a statistical benchmarking framework to validate GWAS pipelines without reliance on clinical metadata.

**University of British Columbia - Mitacs Globalink**          May 2022 – Aug 2022
*Research Intern, Heart and Lung Institute (Dr. Denise Daley)*

- Automated annotation workflows for aging- and sex-associated CpG methylation loci across population-scale datasets.
- Designed ML models to detect combinatorial CpG patterns underlying phenotypic aging and sex-specific biomarkers.

**Elucidata and MedGenome**                                    Jun 2021 – Dec 2021
*Data Science Intern*

- Developed an end-to-end somatic variant calling pipeline integrating BWA, GATK, and bedtools on exome-seq data.
- Benchmarked RNA-seq deconvolution methods and trained deep neural networks using the GenNet framework to predict disease phenotypes from variant profiles.

## Technical Skills

**Programming Languages:** Python, R, Bash, C++, CUDA
**ML Frameworks:** PyTorch, TensorFlow, Keras, scikit-learn, XGBoost
**Bioinformatics Tools:** GATK, BWA, bedtools, IGV, PLINK, GLIMPSE 2.0, Bioconductor
**AI and NLP:** Hugging Face Transformers, GPT models, Prompt Engineering, spaCy, regex, BiLSTM
**Data Analysis  Visualization:** NumPy, pandas, matplotlib, seaborn, ggplot2
**Cloud Computing  HPC:** Linux, HPC clusters (SLURM), Docker, Git

## Publications and Presentations

- Poster Presentation: *MoRF Trigger Analysis Using NLP and LLM*, Suddath Symposium, Georgia Tech (2025).
- Manuscript in preparation: *Large Language Model Assisted Extraction and Categorization of MoRF Triggers and Structural Transitions* (2025).