Prannav Shankar

571-447-6969 | prannav.shankar@gmail.com | linkedin.com/in/prannav-s | github.com/prannav-s | Portfolio

Education

Duke University Durham, NC

Biomedical Engineering (B.S.E) and Computer Science (B.S) *August 2022 – May 2026*

- GPA: 3.828 (Dean's List with Distinction)
- Relevant Coursework: Linear Algebra, Multivariable Calculus, Differential Equations, Data Structures, Algorithms, Databases, Computer Architecture, Machine Learning, Generative AI for Protein Design, Data Science

Experience

Full-Stack SWE Intern Jun. 2025 – Present

MITRE *McLean, VA*

- Designed feature-extraction pipeline and optimized LightGBM model with Bayesian search, placing 2nd in intern hackathon and leading the public leaderboard; presented results to the cohort
- Reworked Flask Blueprint endpoints to integrate Keycloak API for user management, replacing direct SQLAlchemy queries with API calls and ensuring UI consistency
- Developed and refactored Flask endpoints for group-assignment to users and modal dialogs, enabling AJAX-driven add/remove workflows, JSON data fetch, and dynamic rendering of colored group tags without full-page reloads
- Optimized modal performance by consolidating initial data fetch, parallelizing API calls (3× speedup), and using targeted jQuery DOM insertions to reduce load time from 10s to instant
- Built dynamic selector component with AJAX and jQuery change listeners to filter displayed lists based on user selection, persisting state in Flask session for seamless updates
- Migrated data tracking from Keycloak API to PostgreSQL via SQLAlchemy and Alembic migrations, removing obsolete tables and adding new attributes to enable efficient cascade deletions

Machine Learning Engineer Sep. 2023 – Feb. 2025

Duke University (MIT CSAIL and Princeton) *Durham, NC*

- Integrated GASTON into GLACIER to generate directed acyclic graphs (DAGs) from spatial omics data
- Reduced KNN model creation runtime from $O(N^2)$ to $O(N)$; implemented accuracy evaluation across k-values
- Built command-line functionality in Velorama for DAG creation from high-dimensional datasets
- Developed tools to extract and summarize key analysis results from completed DAG runs

Projects

SAGEdiff | *PyTorch, Python, AlphaFold, UniProt, MMseqs2* Jan. 2024 – May 2024

- Built a discrete denoising diffusion transformer for species-conditioned protein sequence generation
- Used class-token conditioning and residue-level tokenization to capture species-specific constraints
- Trained models on curated UniProt datasets using masked cross-entropy loss and one-cycle LR scheduling
- Benchmarked against EvoDiff using Shannon Entropy, pLDDT, pTM, Edit Distance, and Jaccard Similarity
- Achieved higher sequence diversity and species clustering while maintaining comparable structural fidelity

GLACIER | *Python, Jupyter Notebooks* Jan. 2024 – Feb. 2025

- Built ML tools to extract features from spatial transcriptomics data and construct DAGs using Velorama
- Accelerated training using gradient clipping and learning rate scheduling
- Created tools for splicing ADATA and comparing results across runs for testing and evaluation
- Built functionality for analyzing ligand-receptor and gene–transcription factor interactions
- Research paper accepted to RECOMB-SEQ 2025

Perspectify | *Python, Django, HTML/CSS* Aug. 2023 – Sep. 2023

- Built full-stack web app using Django with frontend in HTML/CSS
- Used BeautifulSoup4 to scrape article text and passed content to OpenAI API for summarization
- Used OpenAI API to generate summaries and Google API to retrieve related articles
- Provided users with up to 10 diverse article links per topic for broader perspectives

Skills and Interests

Languages: Python, Java, C, SQL, JavaScript, HTML, CSS
Frameworks and Libraries: Flask, Django, SQLAlchemy, PyTorch, TensorFlow, Keras, Scikit-learn, Jinja, jQuery
Tools and Technologies: Pandas, Numpy, Jupyter, SQLite, PostgreSQL