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

AI and SWE Intern Jun. 2025 – Present

MITRE *McLean, VA*

- Developing secure internal tools using Flask, SQLAlchemy, PostgreSQL, and Alembic for a classified ML project
- Building interactive web interfaces with jQuery and Foundation 6 to support data workflows and model interaction
- Writing modular backend code and integrating ML components into production-grade Flask APIs

Machine Learning Engineer Sep. 2023 – Feb. 2025

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

- Integrated Princeton's GASTON into GLACIER to generate directed acyclic graphs (DAGs) from high-dimensional cell expression and location 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
- Maintained experiment logs and tracked results in structured spreadsheets

ML and AI SWE Intern Jul. 2023 – Aug. 2023

SureStart & MIT *Remote*

- Implemented CNNs, MLPs, and other deep learning models in TensorFlow for various ML tasks
- Developed full-stack features for Perspectify, a platform that summarizes current events and finds articles with alternative viewpoints
- Integrated OpenAI and Google APIs through Django backend to automate information retrieval and summarization

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