

Shorya Consul

✉ shoryaconsul@utexas.edu | in shorya-consul | 🌐 shoryaconsul

EXPERIENCE

Quantum-Si

Dec 2024 – Present

Data Science & Algorithms Team

San Diego, CA

- Redesigned QSI's **Bayesian network-based** protein-inference algorithm for **nanowell** sequencing data via **statistical filtering**, delivering a **5× speedup**, **> 10% accuracy gain**, and **90% recall** with **calibrated confidence tiers** extensible to higher-multiplexity samples.
- Major contributor to migration of core workflows to **Rust** for effective parallelization, achieving a further **10× runtime reduction** and scaling processing to **40× more data per analysis** (raw file sizes **> 500GB**).
- Designed a new, more **interpretable** protein-inference workflow matching prior performance at **40× scale**.
- Built an interactive **motif dashboard** for visualizing sequencing results by recurring motifs, adopted by R&D scientists and customers.
- Leading early-stage development of a **neural network** model with **structural priors** to predict amino acid/binder kinetics amid QSI's **data-scarce** instrumentation regime.
- Regularly brief **C-suite leadership** on algorithm performance and cross-functional roadmap decisions.

CognitiveScale

Jun – Aug 2019

Machine Learning Team

Austin, TX

- Designed the **regression-model audit framework** within Cortex Certifai, a model-trust toolkit scoring black-box models on **4 metrics** - robustness, fairness, explainability, and performance.
- Won **first place** in a company-wide Shark Tank competition for responsible-AI product pitches.

RESEARCH

XHap: Transformer-Based Haplotype Assembly from Long-Read Correlations

- Built XHap, a **transformer-based** model that learns long-range correlations between sequencing reads, producing **> 40% denser** correlation graphs than existing graph-based methods.
- Used this richer representation to reconstruct haplotypes **300x longer** than prior methods at **state-of-the-art accuracy**, unblocking assembly of previously intractable regions.

NextVir: Oncoviral Classification with Genomic Foundation Models

- Developed NextVir, a **genomic foundation model** framework - fine-tuning **DNABERT-S, Nucleotide Transformer, and HyenaDNA** via **LoRA** - the **first** method to jointly identify and classify oncoviral sequencing reads in a single model.
- Achieved **≈ 95% accuracy** across 7 oncoviral families and human background, outperforming **3 SOTA baselines** on realistic, non-uniform-coverage sequencing data.

XVir: Transformer-Based Identification of Oncoviral Reads

- Devised XVir, a transformer-based sequence encoder using **representation learning** to embed biological sequencing reads, achieving **99% accuracy** on oncoviral read identification.
- Robust to up to **15% sequence diversity** within oncoviral families and trains **8x faster** than SOTA methods, enabling practical deployment on large-scale sequencing data.

EDUCATION

The University of Texas at Austin

Austin, TX

PhD in Electrical and Computer Engineering

- GPA: 4.0 / 4.0

The University of Texas at Austin

Austin, TX

M.S in Electrical and Computer Engineering

- GPA: 4.0 / 4.0

Indian Institute of Technology Bombay

Mumbai, India

B.Tech in Electrical Engineering (Minor in Computer Science)

- GPA: 9.86 / 10

SKILLS & COURSES

Skills: Python, Rust, C++, MATLAB, Bash, PyTorch, TensorFlow, Transformers/LLMs, Samtools, Docker, Cwltool

Domains: Biological Sequence Modeling, Representation Learning, Self-supervised Learning, Bayesian Inference, Computational Biology, Genomics

AWARDS

- 2017 Awarded **four-year fellowship** from the Graduate School at UT Austin
- 2017 Awarded the **Institute Silver Medal** for ranking **1st** in department
- 2017 Received **Prof. K.C. Mukherjee Award** for the best senior project
- 2016 Awarded **Urvish Medh Memorial Prize** for top academic performance

SELECTED PUBLICATIONS

- Consul, Shorya et al. (2025). “XVir: A Transformer-Based Architecture for Identifying Viral Reads from Cancer Samples”. In: *Journal of Computational Biology*.
- Robertson, John et al. (2025). “NextVir: Enabling classification of tumor-causing viruses with genomic foundation models”. In: *PLOS Computational Biology* 21.8, e1013360.
- Consul, Shorya et al. (2023). “XHap: haplotype assembly using long-distance read correlations learned by transformers”. In: *Bioinformatics Advances* 3.1, vbad169.
- Consul, Shorya et al. (2020). “Balance is key: Private median splits yield high-utility random trees”. In: *arXiv preprint arXiv:2006.08795*.
- Henderson, Jette et al. (2020). “Certifai: A toolkit for building trust in AI systems”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20), Demonstrations Track*, pp. 5249–5251.
- Consul, Shorya et al. (2019). “Reconstructing intra-tumor heterogeneity via convex optimization and branch-and-bound search”. In: *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 524–529.