Arnav Sonavane

Mumbai, Maharashtra, India | +91 7506129737 | sonavane.arnav2@gmail.com

github.com/w2sg-arnav | x.com/w2sgarnav

EDUCATION

University of Mumbai *Jul. 2023 – May 2027*
*B.Tech. in Electronics & Computer Science; GPA: **8.35**/10.0* *Mumbai, India*
- **Core CSE Topics**: Computer Architecture, Computer Networks, Operating Systems, Deep Learning, Data Structures & Algorithms
- **Core Math Topics**: Differential Equations, Linear Algebra, Probability, Matrix Computation

EXPERIENCE

Harvard University (Remote) *Sep. 2024 – Dec. 2024*
Research Intern, Prof. Devashree Tripathy *USA*
- **Reduced** LLM inference time by 20% via VIDUR control-knob optimizations.
- **Integrated** Meta-Llama and CodeLlama, boosting tool-usage by 15%.
- **Generated** 10,000+ synthetic samples to enhance testing diversity.
- **Accelerated** distributed data processing pipelines by 10%.

MIT Camera Culture Group, MIT Media Lab (Remote) *Jan. 2025 – Mar. 2025*
Machine Learning Intern *USA*
- **Designed** obfuscation methods to safeguard ML model privacy.
- **Developed** attack models for security evaluation.
- **Contributed** new benchmark datasets and extended testing framework.

National Institutes of Health (Hybrid) *Aug. 2024 – Present*
Research Intern, Dr. Chris Grunseich *USA*
- **Analyzed** 100+ gene-expression datasets; achieved 90% serum and 99% plasma classification accuracy.
- **Conducted** Bartlett's, Levene's, and ANOVA tests to confirm biomarker reliability.
- **Optimized** pipelines to reduce processing time by 30%.
- **Identified** biomarkers for early-stage lung cancer detection.

Entrepreneurs First, Bengaluru *Mar. 2025 – Present*
Founding Engineer *India*
- **Building** AI-safety tools and LLM evaluation frameworks under EF VC.

PROJECTS

Click Through Rate Model in DCR [GitHub] *Jan. 2024 – Jun. 2024*
- **Developed** a privacy-preserving data-sharing system, cutting model-build time by 40% across two partners.
- **Reduced** data prep time by 50%, boosting CTR analysis throughput.
- **Processed** 6,000,000+ interaction records, improving prediction accuracy by 12%.

Hierarchical Vision Transformer for Diseases [GitHub] *Jan. 2025 – Present*
- **Implemented** cross-attention between transformer blocks to highlight disease-specific regions.
- **Designed** a curriculum-learning schedule: 128×128 → 384×384 input resolutions.
- **Adapted** Masked Autoencoders for leaf disease pattern encoding.
- **Generated** synthetic rare-class images via GAN augmentation.

TECHNICAL SKILLS

Languages: Python, JavaScript, Java, C/C++, Go
Web: HTML5, CSS3, React.js, Vue.js, Node.js, Flask/Django, FastAPI
Databases: SQLite, PostgreSQL, MongoDB
ML & Data: PyTorch, TensorFlow, scikit-learn, NumPy, Pandas, LangChain, NLTK
DevOps: Git, Docker, Kubernetes, Linux, Shell Scripting, Jenkins
Other: Selenium, Web Scraping, Excel, Postman