Arnav Sonavane

Mumbai, MH, IN |+91 7506129737 | sonavane.arnav2@gmail.com | Linkedin | GitHub | Google Scholar

Profile Summary

A driven and successful student of computer science at the University of Mumbai, with a focus on artificial intelligence, machine learning, and medical imaging, to graduate in 2027. Research internships at esteemed organizations such as Harvard University (LLM inference optimization), JuliaHealth (GPU-accelerated medical imaging containers), and the National Institutes of Health (biomarker identification for lung cancer, achieving 99% accuracy on plasma samples) proving to create and execute novel solutions. Competent in Python and knowledgeable about a variety of machine learning models and tools. Won several honors, including the Live-AI Global Harvard-Duke Hackathon. Looking to contribute innovatively to AI research and development.

Education

Bachelors of Technology in Electronics and Computer Science — **University of Mumbai**
(Mumbai, MH, India - 07/2023 to 05/2027) — **GPA: 7.9/10.0**

- Core CSE topics like: Computer Architecture, Computer Networks, COA, Adv. Distributed Systems, Computer Graphics, DBMS, OS, Deep Learning, DSA, etc. Core Math topics like: Differential Equations, Discrete Calculus, Matrix Computation, Regression, Probability, Linear and Abstract Algebra, etc.

High School Degree in **Computer Science** — Pioneer Jr. College of Science
(Mumbai, MH, India - 05/2021 to 03/2023) — **CET: 91.718, JEE Mains: 89.23 %ile**

- Core Physics Topics like: Accelerator Physics, Nuclear and Particle Physics, Quantum Mechanics, Quantum Electrodynamics, etc.
- Core Biotechnology Topics like: Proteomics, Biostatistics, Cell Biology, Genomics, etc.

Associate Degree in **Computer Science** — **CISCE Board**
(Mumbai, MH, India - 05/2011 to 03/2021) — **Percentage: 95.5 %**

- **Head boy** in School, Collaborated with the **School Council** and the Faculty & Staff
- Active participation in **IIMUN, Debates, Elocutions, and International Quizzes** (eg. **KIQC 2021**).

Work Experience

Harvard University (USA) (09/2024- Present) [Remote] — **Research Intern**

- LLM inference optimization for VIDUR under Professor Devashree Tripathy, quantifying speedups achieved.
- Successfully testing “control knobs” for the LLM inference
- Contributed to particular features like the Chrome Tracing integration, synthetic data generation, & distributed computing aspects.
- Worked on integrating specific LLMs (like Meta-Llama or CodeLlama) into the framework.
- Key contributions: **Performance Optimization**, **Specific Feature Development**, **Profiling**, **Tooling & Infrastructure**

JuliaHealth (under the umbrella of Julia Lang) (09/2024 - Present) [Remote] — **Machine Learning Intern**

- Developing **GPU accelerated containers for medical imaging**
- Improved GPU-accelerated **simulations**, having generated a **new simulation method BlochKernel** with **multi-vendor GPU support**
- **Developed documentation** explaining the new simulation method
- Implemented **automatic pipelines** on Buildkite for testing the simulations across multiple GPU **architectures**.
- Reported **performance improvements** between BlochKernel and Bloch.

National Institute of Health (USA Govt) (08/2024 - Present) [Hybrid] : — **Research Intern**

- Worked under the domain of "Analyze **gene expression data** to identify potential biomarkers for a **disease**" for Dr. Chris Grunseich
- Successfully reviewed and analyzed/identified potential biomarkers for **Lung Cancer**, I have taken datasets of serum and plasma and analyzed using different tests like: **Bartlett's test, Levene's test**, etc to uncover **potential biomarkers**.
- Working on Biomarkers, correlating RNA sequencing and proteomics and working on the datasets, **achieved 90% accuracy on serum samples** and around **99% accuracy on plasma samples**.

University of California (Los Angeles) (03/2024- 06/2024): **Research Intern**

- Worked on 2 datasets of 3 to 4 million rows and 65 columns (apiece) for a **Click Through Rate Model** for the advertisements and feeds of the marketing industry.
- Achieved **70+% accuracy** in a short span of time, and created **Synthetic Datasets** using CTGAN model.
- Did **Remote Attestation**, **used Docker Images** and **Encrypted the files for privacy**, using **Intel TDX and SGX** ,SDK kits

---

Projects

Computational Identification and Validation of Novel Biomarkers for Lung Cancer (01/2024 - 05/2024):

- Successfully reviewed and analyzed/identified **potential biomarkers** for Lung Cancer
- Took **datasets of serum and plasma** and analyzed using different tests like: **Bartlett's test, Levene's test**, etc to uncover potential biomarkers.
- Although I faced a reduced number of metabolites, the prediction accuracy was around 99-100 % for plasma samples and 90 % for serum samples. GitHub Link

Synthia: Synthetic Data Generation with Phi 3 & AI Workbench (06/2024 - 10/2024):

- **Fine-Tuned Phi-3 mini-4k-instruct** to **tokenize query** for **CTGAN model** to read the query .
- Used **Nvidia AI Workbench** for creating **custom containers** from **Nvidia NGC**.
- This model lets you create your own **custom synthetic data** according to your needs. GitHub Link (private repository)

Awards & Scholarships

Winner : Live-AI Global Harvard-Duke Hackathon (01/2024 - 03/2024) **Harvard & Duke University**
2nd Place: GES x UCLA Hackathon(01/2024 - 06/2024) **UCLA University**
Winner : Stanford Biohacks Hackathon (09/2024) **Stanford University**
International Maths/Physics/Science Olympiad Winner (District & States) **SOF**

Skills

Programming Languages: Python, Javascript, Java, C++/C and Go.
Web Development: HTML5, CSS3, React.js, Node.js, Vue.js, Flask/Django, Fast API.
Database: SQLite, PostgreSQL, and MongoDB.
ML - Libraries: Pytorch, Tensorflow, SciPy, Numpy, Sk-Learn, Pandas, LangChain, and NLTK.
DevOps: Git, GitHub, GitHub Actions, Linux, Shell Scripting, Awk, Docker, Docker-Compose, Kubernetes
Interpersonal Skills: Teamwork, Communication skills, Leadership Qualities, Emotional Intelligence.
Others: Web Scraping, Selenium, Excel, Jenkins, Postman.

Publications

Integration of IoT and Quantum Computing: Revolutionizing Manufacturing (Chp. 16) **IGI Global**
HepaScope: Densely Connected UNet for CT Volume-Based Liver and Tumor Segmentation **IEEE**
(Accepted in 13th International Conference on Bioinformatics and Computational Biology)

MOOCS & Certifications

Deep Neural Network With PyTorch
Big Data, Artificial Intelligence and Ethics
Statistics
Time Series Analysis
Quantum Physics 1 to 3 (MIT OCW)

Position of Responsibility & Extracurriculars

Class Representative Of Electronics and Computer Science Branch for First Year
Nationals Badminton Doubles Player & **Represented** College in University Games for **Table Tennis**
Nationals Spelling Bee in **Scholastic Tournament**