Shubhang Sinha

shubhangsinha44@gmail.com | +91-62061-28758 | Leetcode | LinkedIn | GitHub | Kaggle

Education

Dayananda Sagar College Of Engineering Sept 2023 – Aug 2027
BE in Electronics And Communication Engineering CGPA: 8.88

Skills

Languages: Python, C++

Technologies/Methods: Git, Finetuning, SFT, Preference Optimization

Cloud: Docker, GCP, Runpod

Frameworks/Libraries: PyTorch, TensorFlow, scikit-learn, Numpy, MuJoCo, SAM3, COLMAP, GLOMAP

Experience

ML Research Intern, PixelsTwo Oct 2026 – Present

- Built an E2E **Real2Sim** pipeline converting a single monocular/stereo video into a ready Simulation Scene.
- Building Robust **Sim environments** for better Policy training.
- Tools: **MuJoCo, SAM3, COLMAP, GLOMAP, TRELLIS, DA3, VLMs**

LFX Mentorship, LF Decentralized Trust June 2026 – Present

- Building **Talent Angels**, a suite of **AI Graph Agents** that helps users explore skills, tasks, and occupations across global taxonomies (ESCO, O*NET, SFIA, BLS and Lightcast) through a natural language interface.
- Building and traversing **knowledge graphs** to map relationships between skills, tasks, and occupations across multiple global taxonomies.
- Tools: **Python**, **Knowledge Graphs**, **Neo4j**, **LangGraph**, **LLMs**

Google Summer of Code, DeepChem May 2026 – Aug 2026

- Worked on integrating the **DNABERT tokenizer** and **DNABERT-2** model in DeepChem's infrastructure enabling Researchers to perform Research related to Genomic Sequences from within DeepChem.
- Submitted multiple **Pull Requests** adding featurizers, unit tests, and docs while building **benchmarking pipelines** for evaluating the performance of the DNA Foundational Model.
- Tools: **Python**, **HuggingFace Transformers**, **NumPy**, **DeepChem**

ML Research Intern, Indilingo Dec 2025 – Jan 2026

- Worked on improving the efficiency of **translational models** for **rare Indian languages**.
- Employed methods like **ReAct Framework** and **Multi-Shot Prompting** to achieve the goal.
- Tools: **Python**, **HuggingFace**, **Langchain**

Projects

Restrained_ML (On-Device Semantic Photo Search) (Github) *(Python, PyTorch, ExecuTorch, Kotlin)*

- Built a **fully offline Android app** for natural-language photo search: exported Apple's **MobileCLIP-S1** image and text encoders to **ExecuTorch** and ran all inference **on-device**.
- Ported the **CLIP BPE tokenizer** and image preprocessing to **Kotlin**, with **incremental gallery indexing** into a binary embedding store and top-k **cosine-similarity retrieval** at query time.

Open Source Contributions

- **DeepChem** - Added the **DNABERT-2 model wrapper** to DeepChem's infrastructure, along with tests for the integration.
- **DeepChem** - Integrated the **DNABERT-2 tokenizer** into DeepChem, along with related tests.
- **Pandas** - Resolved a **GroupBy API validation issue** by enforcing boolean-only inputs for the `numeric_only` parameter.
- **MLLAM** - Contributed a fix to eliminate **temporary file accumulation** and **file handle leaks** in MLflow logging.

See more: Open Source Contributions Gist

Achievements

- Selected as a Mentee in **Google Summer of Code '26** with **DeepChem**.
- Winner of **Smart India Hackathon '25**.
- Selected as a Mentee in **LFX Mentorship '26** with **LFDT**.