ANGSHUMAN CHAKRAVERTTY

ML Systems & Data Science Engineer — *End-to-end pipelines · Benchmarking frameworks · Intelligent agents*

angshumanchakravertty2@gmail.com • (+91) 9885289622 • GitHub • LinkedIn

SKILLS

- **Languages:** Python, C++, SQL
- **ML / DL:** PyTorch, TensorFlow, Scikit-Learn, HuggingFace Transformers
- **Generative AI & LLMs:** LangChain, LangGraph, RAG Pipelines, Ollama
- **Data & Viz:** Pandas, NumPy, Matplotlib
- **MLOps / DevOps:** Docker, Kubernetes, Git & GitHub, CI/CD (AWS CodePipeline), Linux/Bash
- **Cloud & Deployment:** AWS (EC2, S3, SageMaker, Elastic Beanstalk), FastAPI, Flask, Streamlit

PROJECTS

Benchmarking Generative AI in EDA Workflows | *Python · Ollama · Verilog · Verilator · Yosys · pyverilog · Docker* GitHub → 2026

- **Built a full evaluation framework** to test how well LLMs write hardware description code (Verilog) — ran **3 LLMs across 50 tasks** with a 5-metric scoring system (Syntax Validity, Functional Correctness, Synthesis Quality, Testbench Detection Rate, Generation Time), producing automated comparative reports.
- **Designed an auto-repair pipeline** that feeds simulation failures back to the model with structured error context — improved the passing rate from **0 % to 51.8 %** across 1,610 runs using formal verification and syntax-tree patching to validate each fix.
- **Containerised the entire stack** with Docker for one-command reproducibility; all 3 models (Llama 3 8B, StarCoder2 7B, TinyLlama 1.1B) served locally via Ollama with no external API dependency.
- **Authored a full automation suite** (5+ scripts) covering dataset loading, benchmarking runs, statistical analysis, and visualisation — all results auto-exported with per-model comparison figures.

VESTIGE — Ancient DNA Reconstruction | *Python · DNABERT-2 · HuggingFace · ESMFold · mapDamage2 · PyTorch* GitHub → 2026

- **Fine-tuned a genomic language model (DNABERT-2)** on 44,800-year-old woolly mammoth DNA using a custom masking strategy targeting chemically degraded sites — **13 % lower loss** than standard training (3.274 vs 3.757).
- Proved the approach statistically: standard training **performs worse than random guessing** at the most damaged sites (~20.5 %), while the custom method **recovers 30.8 %** — validated across 626 genomic windows with $p < 0.001$.
- **Verified reconstructed proteins fold correctly** via ESMFold — TM-score > 0.95 against living elephant reference proteins, confirming biological plausibility beyond statistical generation.
- **Added a biosecurity classifier (CNN, AUC 0.934)** to flag sequences against known pathogenic profiles before downstream use — 98.2 % of outputs cleared, establishing a responsible-AI checkpoint for ancient genome work.

Student Exam Performance Prediction | *Python · Scikit-Learn · Flask · AWS Elastic Beanstalk · CodePipeline* GitHub → 2025

- **Deployed a Flask ML API to AWS Elastic Beanstalk** with CI/CD via CodePipeline — **300 ms avg response, 99.9 % uptime**; structured logging and 24/7 monitoring for production reliability.
- Compared 7 regression models on 1,000+ records; reduced RMSE by **20 %** and error by **15 %** over baseline via hyperparameter tuning.

OPEN SOURCE — ML4SCI / DeepLense

- PR #194 (#191) — Fixed a breaking **torchmetrics API incompatibility** affecting the full evaluation suite: audited all notebook cells, migrated metric instantiation to the correct API, and refactored shared imports to prevent regression.
- PR #195 (#192) — Replaced hardcoded W&B credentials with a **dynamic –entity CLI argument**; added config validation and standardised run-naming — enabling reproducible multi-user training runs.
- PR #190 — Authored **dataset setup docs** (download links, folder structure) for Models I–IV in root and subfolder READMEs, removing the main contributor onboarding blocker.

EDUCATION

B.Tech CSE (Data Science) — SVKM's NMIMS, Hyderabad Jul 2023 – May 2027

CGPA: 7.1 / 10 | Sem 5 GPA: 3.14 / 4.0

ACHIEVEMENTS & CERTIFICATIONS

- **1st Place, NMIMS Hackathon NMTF** — Team of 6, 24-hr sprint 2025
- CS50's Introduction to Artificial Intelligence with Python (2024) GitHub Repo →
- NPTEL Python for Data Science (2024)
- NPTEL Introduction to Machine Learning (2024)