ANGSHUMAN CHAKRAVERTTY

ML Research Engineer — *End-to-end pipelines · Benchmarking frameworks · Intelligent agents*

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

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, FastAPI, Flask, Streamlit

PUBLICATIONS & SERVICE

A Missing Latent, Not a Missing Simulator: Radius-Augmented Inference for Real JWST Retrieval

Chakravertty, A., Raj, P.V.V. arXiv:2610.02245 [astro-ph.EP] | 2026

- **Diagnosed the sim-to-real collapse of an amortized flow-matching posterior estimator** on real JWST transmission spectra (importance-sampling ESS $= 1$, best-fit $\chi^2/N = 301$) by testing candidate forward-model physics with an independent nested sampler before the network — a temperature gradient, SO_2 opacity, and an $R = 15000$ opacity set each left the fit unchanged at $\chi^2/N \approx 38$, **isolating an omitted inference parameter** rather than simulator fidelity as the cause.
- **Evaluated across three real JWST targets** spanning two instruments and both published and self-reduced data, including one spectrum **reduced end-to-end from raw MAST archive products** with exoTEDRF (280 integrations; white-light depth 0.01421 vs. 0.0142 literature; 1.4 µm water feature recovered at 5.8σ); **recovered planet radii matched published values on all three targets** (1.227 / 1.196 / 0.232 vs. 1.28 / 1.20 / 0.235 R_J).
- **Reported a negative transfer result honestly:** covariance-conditioned noise training that improved coverage 0.587 → 0.687 on synthetic benchmarks did not transfer to real instrument noise, and a learned CycleGAN domain-translation alternative gave no benefit over structured domain randomization.

Argus: A Real-EKS Study of When Predicting Spot Interruptions Beats Simple Checkpointing

Chakravertty, A., Rayyan, M. arXiv:2609.39067 [cs.LG] | 2026

- **Built Argus, a Kubernetes operator** (custom `SpotResilientJob` CRD + `kopf` reconcile loop) on a 5-service AWS stack (EKS, S3, SQS, Lambda, IAM) provisioned as Terraform IaC with IRSA/OIDC zero-static-credential access; validated it surviving a **real EC2 Spot interruption** end-to-end — draining in **10 s**, checkpointing a live PyTorch job to S3 on `SIGTERM`, and resuming on a healthy node in **~75 s** with one epoch lost.
- **Ran an 80-trial benchmark** (4 checkpoint policies × 4 interruption rates) isolating when predicting interruptions beats simple checkpointing: showed a free 2-minute notice degrades toward no protection as interruptions outpace it, that an ML-free periodic baseline stays strong, and **derived a lead-time sensitivity guideline** — instrumented with Prometheus + Grafana, with the learned risk model characterized honestly as advisory.

VESTIGE: A Knowledge-Guided Masking Strategy for Corruption-Aware Fine-Tuning of Genomic Transformers, Validated on Ancient DNA Reconstruction

Chakravertty, A., Maheshwari, R. arXiv:2607.27712 [cs.LG] | 2026

- Introduced a parameter-free masking strategy that aligns training with an empirically measured per-position corruption profile, replacing the standard assumption that reconstruction difficulty is position-agnostic.
- **Cut validation cross-entropy 13 %** (3.274 vs. 3.757) and led standard MLM at every damage width tested (Δ = +4.18 to +10.35 pp, all $p < 10^{-8}$), with ESMFold reconstructions holding **TM-score > 0.95** even under damage amplified up to 30× beyond authentic rates.

Benchmarking LLMs for Verilog Design Flows

Chakravertty, A., Koshti, R., Sharma, B.P., Chamola, V. arXiv:2607.22759 [cs.AR] | 2026

- Built a reproducible benchmarking platform evaluating open-source LLMs on Verilog RTL generation across 50 curated tasks spanning combinational, sequential, FSM, and mixed designs, validated via Verilator compilation and Icarus simulation.
- **Raised syntax validity from 0 % to 70.4 % average** and **simulation pass rate to 51.8 %** across 1,610 runs on three open-source LLMs (Llama-3-8B, StarCoder2-7B, TinyLlama-1.1B); currently under review at IEEE Design & Test.

Reviewer, NeurIPS 2026 Workshop on Sim2Science (Sim2Sci) 2026

- Reviewed 2 submissions on machine learning with imperfect scientific models, assessing technical validity, experimental design, and reproducibility.

EXPERIENCE

AI Backend Engineering Intern | Asset Telematics GitHub → 2026

- **Extended a Gemini-powered OCR microservice** with multi-language output (English, Arabic, Hindi, French, German) via prompt engineering and a per-request `output_lang` parameter across 10 document types; diagnosed and fixed a Unicode data-loss bug where Latin-only regex silently stripped Arabic/Devanagari — replaced with Unicode-category-aware cleaning preserving combining marks.

- **Built an async batch-processing pipeline** (`/extract_batch` + `/batch_status` polling API) with bounded parallelism via `asyncio` semaphore and per-file partial-failure isolation; engineered concurrency around a blocking vendor SDK using a multi-key round-robin thread-pool, validating graceful degradation under live API rate limits.
- **Integrated OpenRouter (Qwen3-VL)** as an isolated document Q&A chatbot module with upload-once/ask-many session design; built a benchmarking harness measuring throughput, latency, and extraction accuracy — achieving **~86 % name-extraction accuracy** on passports with structured validation rejecting incomplete documents.

PROJECTS

Anvil: Adversarial ML Red-Teaming | *Python · PyTorch · LangGraph · FAISS · UMAP · HDBSCAN · FastAPI* GitHub → 2026

- **Implemented 4 adversarial attacks from scratch** (FGSM, PGD, Adversarial Patch, Semantic) in PyTorch autograd with no external attack libraries, paired with an autonomous patching engine testing up to 3 fix strategies per cluster against a composite safety gate (score ≥ 0.70, accuracy drop $\leq 3\,\%$).
- **Clustered adversarial failure modes** by extracting penultimate-layer activations and projecting through UMAP + HDBSCAN without a fixed cluster count, separating structurally distinct vulnerability classes that gradient-only analysis cannot distinguish.
- **Built a RAG-grounded LangGraph agent** over a FAISS index of 10 adversarial ML papers to generate citation-backed per-cluster vulnerability explanations via Gemini 2.5 Flash with coherence-gated retry logic rejecting low-quality outputs.
- **Deployed the full 8-phase pipeline as an async FastAPI service** on HuggingFace Spaces via Docker with a background job queue, polling endpoint, streamed ReportLab PDF output, and a React frontend on GitHub Pages.

Obscuron: Calibrated Novelty Detection for Microbial Dark Matter | *Python · PyTorch · ESM-2/ProstT5 · GTDB/Pfam · EVT* GitHub → 2026

- **Built a calibrated open-set novelty detector** for microbial "dark matter" (genes matching no reference), scoring each sequence by distance from characterized space in protein-language-model embeddings and calibrating the tail with Extreme Value Theory; achieved **0.962 held-out-family AUROC across 903 families** with a validated calibration fit (KS $p = 0.984$).
- **Solved the missing-ground-truth problem with a leakage-controlled retrospective benchmark** — froze a historical GTDB/Pfam snapshot, scored what was dark then, graded against the current release — surfacing **4,138 validated positives from 1.34M proteins**, with all metrics and a go/no-go data gate pre-registered before modeling to prevent post-hoc fishing.
- **Ran a leakage-controlled protein-vs-genomic model comparison** (ESM-2 0.988 vs Genos-m 0.739), added immune-inspired negative-selection and three-axis convergence as independent corroborating signals, and diagnosed an inverted Precision@K as a retrospective-label selection-bias artifact, not a model defect.

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: 7.3 / 10 | Sem 6 GPA: 3.28 / 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)