

Yasir RAZA

✉ yasirabdali6@gmail.com | 🌐 github.com/Yasirrazaa | 🔗 linkedin.com/in/yasir-raza | 📍 Islamabad



☎ +92 314 5363379 | 📦 pypi.org/user/yasirabdali6 | 🏆 kaggle.com/yasirabdaali | 🌐 yasirrazaa.github.io

SUMMARY: Research Engineer focused on LLM training and evaluation methodology: rule-based RL (**GRPO/RLVR**), execution-based reward design, contamination-checked benchmarks, and adversarial robustness. Built an oracle-verified RL training harness; my Archon security platform ranks **13th/100+ overall** (49.7% avg; 85.1% defense / 14.4% attack) in Berkeley's AgentBeats Arena.

TECHNICAL SKILLS

LLM TRAINING: GRPO / RLVR (TRL), supervised fine-tuning, reward design & shaping, curriculum construction, LoRA/QLoRA (PEFT, Unsloth), PyTorch

EVALUATION METHODOLOGY: Execution-based harnesses, SHA-256 frozen splits, bootstrap 95% CIs, pass@k, benchmark design (AgentDojo, GAIA)

ADVERSARIAL ROBUSTNESS: Prompt-injection attack/defense, red-teaming, GOAT/PAIR/Crescendo taxonomies, defense gateways, OWASP ASI01-10

SYSTEMS & INFRA: Sandboxed code execution (AST allowlists, RLIMIT), CUDA graph optimization, Docker, CI/CD, FastAPI

LANGUAGES: Python, SQL, Bash, TypeScript

WORK EXPERIENCE

TEAMLIFT — UNITED STATES (REMOTE)

Junior Data Scientist

Jan 2024 - Dec 2024

- Scraped **thousands of AI tools** from public tool directories, building the structured dataset that powers **Teamo**, Teamlift's recommendation engine.
- Built an automated **verification pipeline** (HTTP status checks, screenshot capture, freshness validation) that kept listings accurate and trustworthy — eliminating stale and duplicate entries at scale.
- Preprocessed and normalized thousands of tool records into model-ready features for the recommendation engine.
- Worked directly with US-based stakeholders remotely, shipping dataset improvements on tight iteration cycles.

Data Science Intern

Jul 2023 - Dec 2023

- Built an automated pipeline collecting **thousands of job postings** from dice.com with **deduplication** — hardened with retries, rate limits, and robust parsers.
- Extracted skill-demand signals from job descriptions at scale; trained **time-series models** to forecast demand, informing product and market decisions.

RESEARCH PROJECTS

Math2Code — Rule-Based RLVR for LaTeX → Python

2025 - 2026

- Built the complete verified training + evaluation stack for a math LLM: **22,796 deduplicated samples** with a SHA-256 frozen split, oracle-verified gold solutions (**397/397**), bootstrap-CI execution metrics, **no LLM-judged rewards** (GRPO rewards on freshly resampled inputs). Sandbox: **10,000 execs in 26s**.

Archon 1.0 — Adversarial Agent Security (Competition)

2026

- Scenario-agnostic attacker/defender agents for the Lambda × Berkeley AgentBeats Security Arena (A2A + Google ADK): **13th/100+ overall (49.7%)**, **6th Defense (Blue, 85.1%)**, **18th Attack (14.4%)** on the private leaderboard.

Archon 2.0 — Measurable Attack-and-Defense Platform

2026

- The only open platform where an adaptive attacker and a measurable defense fight in the same loop: **222 probes** across 10 packs, full **OWASP ASI01–10** coverage, AgentDojo v1 benchmark (0% ASR on direct overrides at the zero-cost tier), CVSS-style severity — enforced by **2,295 tests** (93% gate).

PRISM-Bench — Cultural Intelligence (CQ) Benchmark 2026

- Authored **PRISM: Pluralistic Reasoning & Identity-Specific Modeling**, an open benchmark measuring cultural intelligence in AI systems across pluralistic scenarios.

Chatterbox-Optimized — Systems Research for TTS 2025

- Researched and productionized **bucketed CUDA Graphs** for autoregressive TTS and bfloat16 + KV-cache tuning — up to **4.5× speedup (RTF 0.13)** — plus a Whisper-validated candidate-selection scheme eliminating hallucinated audio.

NotebookLLM — Token-Efficient Notebook Tooling 2025 - 2026

- Open-source package + MCP server (**10K+ PyPI downloads**): token-budget modes drop up to 80% of context waste, with tiktoken measurement per cell.

EDUCATION

B.Sc. Software Engineering, National University of Modern Languages (NUML), Islamabad Aug 2026

CERTIFICATIONS

Datasets Expert · Notebooks Expert (13 medals; Datasets rank 731/11,277, best #109), Kaggle **2024 - 2026**

Associate Data Scientist in Python · Data Analyst with Python (Career Tracks), DataCamp **2025 - 2026**

Generative AI Fundamentals · Intro to LLMs · Intro to Generative AI · Intro to Responsible AI, Google Cloud Skill Badges **2023**

Python · Intro to Programming (Professional Certificates), Kaggle Learn **2025**

Advanced LLM Bootcamp, NUST & Sky Electric **2024**