

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: ML Engineer with 1.5+ years of remote experience on US-based teams, owning the full ML lifecycle: data pipelines, training, rigorous evaluation, and GPU deployment. Built the verified training + evaluation stack for an RL (**GRPO**) math LLM; shipped a **4.5×** TTS inference speedup; processed thousands of tools and job postings into production datasets.

TECHNICAL SKILLS

MODELING & TRAINING: PyTorch, Transformers, scikit-learn, LoRA/QLoRA, Unsloth, PEFT, GRPO (TRL), supervised fine-tuning, reward design

DATA ENGINEERING: pandas, NumPy, SQL, Scrappy, Playwright, deduplication at scale, feature pipelines, S3

EVALUATION & STATS: Execution-based harnesses, frozen/contamination-checked splits, bootstrap 95% CIs, WER, RTF benchmarks, time-series forecasting

MLOPS & DEPLOYMENT: Docker, FastAPI, CI/CD (GitHub Actions), RunPod/AWS serverless, model serving, GPU cost optimization

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 and built an automated verification pipeline (HTTP status checks, screenshot capture, freshness validation) powering **Teamo**, Teamlift's recommendation engine.
- Preprocessed and normalized thousands of tool records into model-ready features, eliminating stale and duplicate listings from the dataset.

Data Science Intern

Jul 2023 - Dec 2023

- Built an automated pipeline collecting **thousands of job postings** from dice.com with deduplication, guaranteeing a clean non-redundant dataset.
- Trained **time-series models** to forecast skill demand extracted from job descriptions, informing product and market decisions.

TECHNICAL PROJECTS

Math2Code — LaTeX → Verified Python via Rule-Based RL

2025 - 2026

- Built the complete verified training + evaluation stack for a math LLM (**GRPO/RLVR** on TRL): **22,796 deduplicated samples**, SHA-256 frozen split, oracle-verified gold solutions (**397/397**), and a sandbox pool sustaining **389 exec/s**. Total budget under **\$100**.

Kausal — Production AI Agent Platform (private)

2026

- Extended the ii-agent framework in **pure Python** (no LangGraph): custom **RAG**, layered memory, **OpenTelemetry** observability (tracing, metrics, SigNoz dashboards), realtime **voice/video via LiveKit**, **LiteLLM** routing, agent skills, S3 storage — guarded by a **7,275-test** suite.

Chatterbox-Optimized — Production TTS Engine

2025

- Up to **4.5×** **faster inference** (RTF 0.59 → 0.13) via bucketed CUDA Graphs, bfloat16, and KV-cache tuning; Whisper-validated candidate selection and a reproducible benchmarking suite on difficult text corpora.

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, tiktoken measurement per cell, and output summarization for DataFrames, plots, and tracebacks.

EDUCATION

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

Aug 2026

CERTIFICATIONS

CERTIFICATES: Kaggle **2×** **Expert** (Datasets · Notebooks, 13 medals) · DataCamp (Associate Data Scientist in Python & Data Analyst career tracks) · Google Cloud skill badges (4 Gen AI badges) · NUST Advanced LLM Bootcamp