

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: Python developer and open-source maintainer with 1.5+ years of remote experience on US-based teams. Two published **PyPI** packages (**10K+ downloads**), **7,275+** tests written across projects, and a strong quality culture: typed code (mypy), linted code (ruff), benchmarked hot paths, and CI-gated releases.

TECHNICAL SKILLS

- CORE PYTHON: Python 3.11+, asyncio, typing/protocols, concurrency (thread pools, subprocess isolation), dataclasses, context managers
- PACKAGING & CLI: setuptools, uv, PyPI publishing, Click, Rich, Sphinx documentation, versioning & release workflows
- QUALITY & TESTING: pytest (7,275+ test functions authored), pytest-asyncio, pytest-benchmark, coverage, ruff, mypy, pre-commit, GitHub Actions CI
- DATA & APIS: pandas, NumPy, requests/httpx, REST, WebSocket, Scrapy, Playwright, SQL
- TOOLING: Git, Docker, docker-compose, Make, Linux, CI/CD

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.

TECHNICAL PROJECTS

NotebookLLM — Open-Source Python Library, CLI & MCP Server 2025 - 2026

- Designed, packaged, and published to **PyPI** (10K+ downloads): unified API across **8+ notebook formats**, 4-mode agent-optimized output with tiktoken token budgeting, **ijson streaming** for 10 MB+ notebooks, atomic crash-safe writes, Click CLI with Rich output, pytest suite with benchmarks, ruff + mypy gates, Sphinx docs, and CI.

Kausal — Production AI Agent Platform (private) 2026

- Extended the ii-agent framework in **pure Python** (no LangGraph): custom **RAG**, layered memory, **Open-Telemetry** observability, **LiveKit** voice/video, **LiteLLM** routing, agent skills, S3 storage — with a **7,275-test** behavioral suite across 16 test categories (unit, integration, load, e2e).

Forge — Code Generation Engine 2024 - 2026

- Built emitters producing **Python SDKs** (Pydantic-validated), CLI tools, MCP servers, and **17 output types** from one OpenAPI 3.x IR; multi-registry publishing pipeline (PyPI, npm, Cargo, Maven, NuGet), byte-stable output tests, and 13 compile gates in CI.

Cerberus — Browser Control Plane 2026

- Engineered an idempotent **Make**-driven setup/verify system (**86 automated checks**), a 60-tool MCP bridge, and diagnostic tooling with actionable failure logs across a multi-process Python/Node stack.

Math2Code — Verified Data & Test Infrastructure 2025 - 2026

- Maintained **209 pytest tests** plus ruff, mypy strict, package-build, Docker, and split-integrity checks in CI; wrote an oracle that verifies generated code on fresh jittered inputs (**200/200 gold solutions**).

Archon 2.0 — Agent Security Platform 2026

- **2,295 tests** at a 93% coverage gate across a py3.11–3.13 CI matrix; schema migrations, durable results store, checkpoint/resume for long battles, and cosign-signed SBOM releases.

Chatterbox-Optimized — TTS Engine 2025

- Production **RunPod serverless** worker (rp_handler + CUDA Dockerfile) with a reproducible benchmarking suite; up to **4.5× faster** (RTF 0.13).

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**