

# 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 backend engineer with 1.5+ years of remote experience on US-based teams, building API and agent-serving infrastructure: **FastAPI** services, WebSocket servers, serverless GPU workers, MCP/HTTP control planes, async PostgreSQL, and S3 storage. Author of two published **PyPI** packages (10K+ downloads).

## TECHNICAL SKILLS

APIs & SERVICES: FastAPI, WebSocket, REST design, MCP servers, HTTP control planes, token authentication, async request handling

DATA & STORAGE: PostgreSQL (asyncpg), Alembic migrations, Redis, S3, ETL pipelines, caching

ASYNCH & SYSTEMS: asyncio, thread pools, lazy resource management, subprocess isolation, streaming (ijson), multiprocessing

INFRA & DEPLOYMENT: Docker (multi-stage, CUDA images), docker-compose, RunPod serverless, CI/CD (GitHub Actions), Make, Linux

LANGUAGES: Python, TypeScript, SQL, Bash

## WORK EXPERIENCE

TEAMLIFT — UNITED STATES (REMOTE).....

### Junior Data Scientist

Jan 2024 - Dec 2024

- Scraped **thousands of AI tools** from public tool directories; built an automated **verification pipeline** (HTTP status checks, screenshot capture, freshness validation) feeding Teamlift's **Teamo** recommendation engine.
- Preprocessed and normalized thousands of tool records into model-ready features via reproducible ingestion workflows.

### Data Science Intern

Jul 2023 - Dec 2023

- Engineered an automated ingestion service collecting **thousands of job postings** from dice.com with **deduplication**, delivering clean, queryable datasets.
- Trained **time-series models** to forecast skill demand extracted from job descriptions.

## TECHNICAL PROJECTS

### Kausal — Production AI Agent Platform (private)

2026

- Extended the ii-agent framework in **pure Python** (no LangGraph): custom **RAG** pipeline, layered memory, **Open-Telemetry** observability (tracing, metrics, SigNoz), realtime **voice/video via LiveKit**, **LiteLLM** multi-model routing, agent skills, and **S3** storage — backed by PostgreSQL/Redis, Docker microservices, and a **7,275-test** suite.

### Chatterbox-Optimized — Serverless TTS API

2025

- Shipped a production **RunPod serverless** worker (rp\_handler + CUDA Dockerfile) for text-to-speech: up to **4.5× faster inference** (RTF 0.13), long-text chunking with stitched audio, Whisper-validated candidate selection, and a locally testable API harness.

### Cerberus — Multi-Agent Browser Control Plane

2026

- Built an HTTP API + dashboard (**127.0.0.1:9867**) with token authentication, a multi-agent isolation gateway with worker leases and routing, parallel instance management, persistent per-profile session storage, and a **60-tool MCP bridge** — verified by 86 automated checks.

### NotebookLLM — MCP Server & Async Execution

2025 - 2026

- Published a **PyPI** package + MCP server (**10K+ downloads**): async Jupyter kernel execution with lazy per-session kernel lifecycle and thread-pool responsiveness, streaming ingestion for 10 MB+ notebooks, and atomic crash-safe writes.

## 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