

Abdul Haseeb Shaikh

AI Engineer, RL Environments and LLM Evaluation

shaikh.haseeb.work@gmail.com · +92 336 2844576 · Islamabad, Pakistan (UTC+5), remote · linkedin.com/in/haseebshaikh03 · github.com/haseebshaikh03 · haseebshaikh03.github.io

SUMMARY

I am part of Turing's RL team in the Frontier AI Labs department, where for the last two years I have built the RL environments, verifiers and reward signals that frontier labs use to train and evaluate their models. I lead quality and tooling across several projects. I write the first environments on each project myself, then set the QC bar and review every task before it ships, more than 2,000 tasks and trajectories so far across evaluations of 10 models. Before that I spent four years in backend engineering, including running my own software company, working mostly in Python, TypeScript, Docker and AWS. I also compete in CTFs at a national level, which is why I design rewards and verifiers that are hard to game and quick to expose real model failures. I am looking for a remote role in RL environments, LLM evaluation or applied AI engineering.

SKILLS

RL environments: Agent environment design with tasks, data, tools and automated verifiers, reward design, reward hacking and answer leakage checks, solution and verifier parity, reproducible Docker environments

LLM evaluation: Benchmark and task design, failure mode analysis, rubric calibration, multi-turn trajectory review, SFT and RLHF data quality

Agentic AI: Tool-calling agents on the OpenAI and Anthropic APIs, structured outputs, prompt pipelines, sandboxed agent environments

Eval tooling: Harbor tasks (task.toml, instruction.md, tests/test.sh, solution/solve.sh), BrowseComp-style web research tasks, SkillsBench, pytest, Docker, CI

Engineering: Python (Flask, Django), TypeScript and Node.js (Express), REST APIs, microservices, Docker, CI/CD, AWS EC2, Git

Security: Web pentesting, reverse engineering and CTF competition, which I use for adversarial and anti-leakage review of tasks

EXPERIENCE

Senior AI Engineer, RL Team (Frontier AI Labs) | Turing (remote, US based)

Nov 2024 to present

- Member of Turing's RL team in the Frontier AI Labs department. Build RL environments and evaluation tasks for frontier labs to train and evaluate their models, covering multi-turn software engineering with coding agents, BrowseComp-style deep web research, SkillsBench skill-use tasks, tool-use environments and an RL Gym benchmark for security tools.
- Start each project by writing the first Dockerized Harbor environments myself, each with an agent-facing instruction, tools and data, a reference solution and an automated verifier, then turn what I learn into the task anatomy, reward and verifier rules and QC checklists the team works from.
- QC every environment before it ships. Reviewed and recalibrated more than 2,000 tasks and multi-turn trajectories for verifier correctness, solution and verifier parity, answer leakage and reward problems where an agent could pass without solving the task.
- Built and maintain the review tooling, Python pipelines and Docker environments (custom Dockerfiles, pytest-based verifiers, reward files) so that environments run the same way locally, in CI and on lab infrastructure.
- Evaluated 10 frontier models on reasoning, code understanding and tool use. Found recurring failure modes such as spurious tool calls, hallucinated file state and under-specified verification, and fed those findings back into environment and reward design.
- Worked as Engineering Manager and Delivery Manager on several of these projects, running technical reviews and calibration sessions across a distributed team of 60 engineers and grading SFT and preference (RLHF) data against lab-specific rubrics.

CEO and Head of Engineering | Cloud Mesh Solutions (Pvt) Ltd, Islamabad

Nov 2022 to Dec 2024

- Ran a company of around 30 people and led its engineering team across several SaaS and marketplace products for international clients. Owned architecture, delivery, hiring and client relationships, all while completing a BS in Cyber Security in parallel.
- Products I led and coded on include [Kaufes](#) (a Swiss classifieds and auction marketplace with a Node.js backend, Next.js web app, Flutter mobile app, Wallee payments and Swiss QR bills), [Park4Fly](#) (airport parking management with a driver app), [Büro 365](#) (task, appointment and project management for small businesses) and [Offerten Helden](#) (a Swiss quote and lead management platform for service businesses).
- Designed the backend systems in Node.js with TypeScript and Python with Django. Introduced clear service boundaries, containerization and CI/CD pipelines so that releases became routine instead of risky.
- Set the standards for code review, automated testing and observability, and mentored engineers into senior roles.

Lead Software Engineer (started as Software Engineer) | Ripple Intra Technologies, Karachi

Aug 2021 to May 2022

- Worked on apps for Concord Healthcare Group, including the [Concord Pharma](#) sales team app (backend and mobile), [Toticell](#), [EW Villa Medica](#) and DMFR, owning backend and DevOps with Python (Flask) and TypeScript (Express), Docker-based microservices and AWS EC2. Towards the end I built the Brand Emirate backend (be-backend) in TypeScript with MongoDB and Redis, along with a shared microservices base layer.
- Promoted to lead in March 2022, where I set up the delivery pipelines and engineering workflow for the team.

- Built and maintained Python and Django backend systems, and improved the performance and maintainability of production services.

LEADERSHIP & COMMUNITY

President | [CUI Tech Society](#), COMSATS University Islamabad

2025 to 2026

- Led the umbrella society for the Computer Science department, coordinating all of its student programs and technical societies.
- Organized [CUI Tech Fest 2025](#), a six-day national event hosted by the CS department with 1,200 participants across 12 modules, from a CTF, an AI agents arena and a web hackathon to gaming tournaments, workshops and project exhibitions.

Co-founder and President | [Cyber Hacktivators Club \(CHC\)](#), COMSATS University Islamabad

2023 to 2025

- Founded the university's cybersecurity society and grew it into the most active student body in the department, delivering more than a dozen events in two years. Highlights include CodeCon-CUI'24, a three-day national event with 350 participants from 14+ universities, an inter-university CTF hackathon with 30 teams, and ComsiCon'25 with 300 participants.
- Built and ran the technical training program, including a structured CTF course for 50 students, hands-on workshops for around 100 students each, and a software engineering career session. Authored CTF challenges across web, pwn, reversing and forensics and operated the CTFd infrastructure with Docker-based challenge containers on Azure and DigitalOcean.
- Secured sponsorship and coordinated 30 CHC teams for Black Hat MEA 2024. Two teams reached the finals in Riyadh.

PROJECTS

AVDA.ai, AI voice agent platform (independent contractor for a US client, Sep 2025 to Jan 2026)

- Designed and built the backend for a multi-tenant SaaS that runs AI phone agents for inbound and outbound calls. Real-time audio streams over WebSockets through Twilio Media Streams, speech goes through Google Cloud, AssemblyAI or ElevenLabs, and the conversation is driven by GPT-4o or Gemini with function calling for appointment booking, business hours and hand-off to a human.
- Node.js and TypeScript, MongoDB, Redis, AWS S3, Stripe subscriptions and minute credits, GoHighLevel CRM integration over OAuth, call recording, transcripts and an analytics dashboard. Interrupt detection keeps the agent responsive in under 300 ms when the caller talks over it. Solo build over roughly 170 commits.

Freelance backend work through FuzionDev LLC, US (Dec 2025 to Mar 2026)

- [DrayMatch](#), a US container logistics platform. Built the Node.js and PostgreSQL backend and the scrapers that pull import and export manifests and draymen registry data, and specified the DrayTrends analytics layer (trade lanes, equipment mix, container-to-draymen index).
- Also shipped backends for Condo Whisper, FinancelQ, ABCard, Soul Story and TimeCloud, mostly Node.js and TypeScript with Prisma or Sequelize, PostgreSQL, AWS S3 and SES, and Docker.

ExploitOps, a multi-cloud attack simulator (final year project, 2025 to 2026)

- An offensive cloud security framework that safely simulates real attack paths across AWS, Azure and GCP inside sandboxed environments provisioned with Terraform. It covers open bucket exploitation, serverless function abuse, SSRF and metadata endpoint attacks, IAM privilege escalation and pivoting from one cloud to another.
- Python and Flask backend with both a CLI and a web dashboard, packaged with Docker, with centralized logging and PDF and CSV reporting. Two-person team. Delivered and approved end to end, from SRS and SDS through implementation and evaluation.

EDUCATION

BS Cyber Security | COMSATS University Islamabad

2022 to 2026

Completed while working full time in the engineering roles above. Final year project was ExploitOps (see Projects).

Diploma in Computer Science | Aptech Pakistan

2019 to 2022

Three-year software development diploma, completed alongside school and my first developer roles.

Intermediate, Pre-Engineering | Government National College, Karachi

2020 to 2022

ACHIEVEMENTS

- Organized [CUI Tech Fest 2025](#), the largest event the CS department has run, with 1,200 participants across 12 competition, workshop and exhibition modules.
- Organized [CodeCon-CUI'24](#), a three-day national event with 350 participants from 14+ universities, combining a 30-team CTF with gaming and coding competitions, run with Ignite and Trillium (COMSATS Annual Report 2024 to 2025, page 102).
- Black Hat MEA 2024 Finals, Riyadh. Placed 40th of 250 finalist teams with Team PsychoTherapist, after qualifying from a field of 9,000 teams.
- Ignite National Cyber Hackathon (Ministry of IT, 700 teams). 2nd place nationally in Red Teaming and 1st in the regional round, both rounds.
- DevConCTF'25 (NUST). 1st place nationally. Team PsychoTherapist is ranked 3rd in Pakistan on ctftime.org.