Shaheen Ahmed-Chowdhury

shaheen.ahmed@live.co.uk | linkedin.com/in/shaheen-ahmed-chowdhury | github.com/shaheenahmedc | shaheenahmedc.com

Professional Experience

Equistamp – Technical Project Manager / Research Engineer *Apr 2025 – Present*

- TPM for Redwood Research's LinuxArena project (25–30 contributors, Jul 2025 – May 2026). Now used by Anthropic to test frontier models. Planned weekly sprints and distributed work across the team (environment creation, side-task auditing, sabotage trajectory creation). Reviewed technical deliverables across the project. Managed the LinuxArena dev-tooling workstream. Co-author of the LinuxArena paper accepted at ICML 2026 (arXiv). Built the Media Processing environment end-to-end (GitHub): ∼4000 LOC; synthetic-meeting generation pipeline; five sabotage side tasks (`access_other_teams_media`, `hijack_hls_playback`, etc.); monitor-evading attack trajectories. Owned communication of technical requirements between Redwood and Equistamp.
- Equistamp's TPM for the EU AI Office Loss of Control project (36-month European Commission contract, Jan 2026 – Dec 2028, tender specs); Equistamp is the consortium lead, working with METR and Epoch AI. Entrusted with full ownership of this contract nine months after joining Equistamp. Primary technical contact for the AI Office. Selected the team and built the budget. Chair weekly syncs, write technical tickets to align with enforcement requirements, and own delivery of all evaluations and infrastructure.
- TPM for Control Arena project with UK AISI (Feb – Jul 2025) – helped develop a Kubernetes-native AI-control evaluation setting for testing whether frontier agents can subvert control measures on training- and inference-cluster infrastructure (docs, GitHub). Coordinated 15 engineers on main-task development, reviewed outputs, and delivered team-authored code upstream to AISI's repository (GitHub). Led the configurable-presets workstream, providing AISI with toggleable K8s security measures of varying complexity across training-cluster deployments via Helm and Cookiecutter (GitHub).
- Run several other workstreams in parallel, coordinating with researchers and engineers across clients (Redwood, METR, UK AISI) and Equistamp. Lead Hawk bug-fixing (METR's K8s-native Inspect orchestrator on AWS) and LinuxArena–Hawk integration, including debugging deployment issues in Hawk's Datadog monitoring stack and verifying Datadog's functionality during complex evals like LinuxArena on Hawk. TPM on Sabotage Blue-Teaming project with Redwood Research (arXiv; primary technical contact; recruited and ran a 12-person auditor pool). Manage new technical leads on Control Tower (extracting LinuxArena code into a reusable AI-control evaluation library). Contributed to multiple grant applications for government evaluation tenders.

Arcadia Impact – Technical Project Manager / Software Engineer *Oct 2024 – Apr 2025*

- Led engineers porting five RE-Bench tasks from METR to UK AISI; wrote documentation mapping container setup, task logic, and scoring to guide engineers. Covered multi-stage Docker builds, sandbox debugging, and concurrent evaluation logic. Participated in the first round of Arcadia's ASET programme (LinkedIn), onboarding MMMU to `inspect_evals` (GitHub). Became TPM for the second round. Oversaw nine evaluation implementations with high-fidelity replication of reported performance.
- Secured UK AISI funding to evaluate engineering uplifts from multi-agent systems. Enhanced `swe_bench` 2.X/3.X compatibility in `inspect_evals` (GitHub); designed single- vs. multi-agent experiments quantifying test-time-compute uplifts against sensible baselines; built a multi-agent tool with AutoGen that lets single-agent systems invoke SWE-Agent-equipped sub-agents (report).

WPP (Choreograph) – Data Scientist *Jan 2022 – Jan 2025*

- Built and maintained marketing-focused agent-based models in NumPy; led data scientists in delivering new model features for time-sensitive brand pitches. Migrated simulations to JAX with Ray on GPU nodes, cutting computation time by up to 80%. Won a WPP-wide hackathon by modelling event-attendee paths from geospatial POI data, improving billboard pricing accuracy.

ERIKS Digital – Data Scientist *Jan 2019 – Dec 2021*

- Led intermittent demand forecasting across 800,000+ products; implemented Python forecasting models including Amazon's DeepAR: +15% accuracy, −10% inventory cost.

Projects

Open Source – Software Engineer *Jan 2024 – Apr 2024*

- Built an open-source mechanistic-interpretability library re-implementing Google PAIR's PatchScopes via `nnsight`; published on PyPI. Designed extensible base classes for clean iteration on PatchScope configurations.

Founders and Coders – ML Engineering Apprentice *Oct 2023 – Dec 2023*

- Re-implemented Word2Vec, GPT-2, Llama 2 fine-tuning, YOLOv1, and Stable Diffusion in PyTorch over eight weeks; deployed via FastAPI. Built experience with Kubernetes, Docker, and multi-GPU training; applied this at Choreograph for the JAX migration.

Education

Utrecht University – MSc in Mathematical Sciences *Sep 2018 – Apr 2021*

Durham University – MPhys in Theoretical Physics *Sep 2013 – Jun 2017*