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 contractors, Jul 2025 – May 2026). Named author on the ICML 2026-accepted LinuxArena paper (here). Built the Media Processing environment end-to-end (here): ∼4000 LOC; synthetic-meeting generation pipeline; five side tasks (`access_other_teams_media`, `hijack_hls_playback`, etc.); monitor-evading attack trajectories. Owned technical-requirements communication between Redwood and the engineering team; reviewed engineering outputs.
- TPM for EU AI Office Loss of Control project (36-month European Commission contract, Jan 2026 – Dec 2028, here); Equistamp is consortium lead with METR and Epoch AI. Primary technical contact for the AI Office. Selected the team and built the budget. Chair weekly syncs, align technical tickets with enforcement requirements, own delivery of all evaluations and infrastructure.
- TPM for UK AISI's Control Arena project (Feb – Jul 2025); assisted with the development of the K8s training-cluster infrastructure setting (see here, here). Coordinated 15 engineers on main-task development for the setting (here). Led the configurable-presets workstream in the same setting, providing AISI with togglable K8s security measures of varying complexity across training-cluster deployments, via Helm and Cookiecutter (here).
- Further parallel workstreams: LinuxArena ↔ Hawk integration with METR (leading Hawk bug-fixing and LinuxArena-to-Hawk port); Sabotage Blue-Teaming projects with Redwood Research (paper here; primary technical contact; recruited and oversaw a 12-person team); 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 AISI; mapped container setup, task logic and scoring across both organisations' implementations to streamline engineer workflows. Covered multi-stage Docker builds, sandbox debugging and concurrent evaluation logic.
- Onboarded MMMU to `inspect_evals` (here); oversaw nine further evaluation implementations with high-fidelity replication of reported performance. Worked on `tau_bench` onboarding to METR's Vivaria.
- Secured AISI funding to evaluate software-engineering uplifts from multi-agent systems. Enhanced `swe_bench` 2.X/3.X compatibility in `inspect_evals` (here); designed baselined single- vs. multi-agent experiments quantifying test-time-compute uplifts; built a multi-agent tool over AutoGen's event-driven messaging API that lets single-agent systems invoke SWE-Agent-equipped sub-agents (report here).

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 UK-brand pitches. Reduced computation time by up to 80% by migrating to JAX with Ray on GPU nodes. Won a WPP-wide hackathon by modelling event-attendee paths from geo-spatial 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, LLaMA2 fine-tuning, YOLOv1 and Stable Diffusion in PyTorch over eight weeks; deployed via FastAPI. Picked up Kubernetes, Docker and multi-GPU training; applied at Choreograph for the JAX migration.

Utrecht University – MSc Thesis *Mar 2020 – Apr 2021*

- Designed a novel agent-based model calibration scheme using high-dimensional Bayesian optimisation on a large-scale macroeconomic ABM; outperformed random search.

Education

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

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