

# Roki Seydi

roki@rokiseydi.com | rokiseydi.com | github.com/RokiSeydi | London

AI safety researcher and founding product engineer delivering frontier model evaluation systems for high-stakes, real-world deployments. I scope, build, and operationalise evaluation frameworks that test whether AI systems perform safely under real-world conditions — not just benchmarks. My work combines hands-on technical delivery (production systems serving millions of users in regulated environments) with applied AI safety research (frontier model evaluations, cross-model grading protocols, and multi-epoch validation). With an early background in UX/UI design, I bring a strong product lens — ensuring systems are not only technically sound, but usable and effective in real-world decision-making contexts.

---

## AI SAFETY & EVALUATION EXPERIENCE

**Holding Health** — Founder / Technical Lead Aug 2025 – Present

- Built an [open-source frontier model evaluation framework](#) on **UK AISI Inspect Evals**, producing statistically significant evidence of deployment-critical reasoning failures across three frontier model families (Claude Sonnet 4, GPT-4o, Gemini 2.5 Pro; N=72 observations, Cohen's d 0.95–1.82,  $p < 0.01$ ).
- Identified systematic gaps between benchmark performance and real-world reasoning, demonstrating that models fail in deployment-relevant contexts despite passing standard evaluations.
- Designed cross-model grading protocols to mitigate model-as-judge bias, identifying structural limitations in automated evaluation pipelines (“The Auditor’s Blindspot”).
- Built evaluation pipelines designed for integration into real-world AI deployment workflows, enabling systematic testing under contextual uncertainty.
- Translating qualitative cultural and contextual insights into structured, testable evaluation systems for frontier AI models.
- Scoping a **biosecurity and dual-use governance evaluation domain** (Domain 4), operationalising Fink Report categories, WHO Risk Groups, and information hazard taxonomies into scorable dimensions.
- Managing research partnerships with the **Community Health Impact Coalition** and the **Alan Turing Institute** Clinical AI Interest Group.

**Independent Researcher** — AI Safety & Cultural Reasoning Aug 2019 – Present

- Authored working papers defining novel AI failure modes (“Cultural Confabulation”) and structural evaluation gaps (“The Auditor’s Blindspot”).
- Conducted multilingual red-teaming across frontier models, testing reasoning behaviour under contextual uncertainty.
- Translated anthropological and qualitative frameworks into structured, reproducible evaluation protocols.

**Bromley Healthcare** — Lived Experience Advisor & DEI Partner Sep 2024 – Present

- Advising on integration of lived experience into healthcare service design and evaluation.
- Bridging operational requirements with real user needs in a public-sector delivery context.

---

## TECHNICAL DELIVERY & ENGINEERING

**tell.money** — Founding Product Engineer (Full Stack) Apr 2022 – Sep 2024

- Led end-to-end delivery of open-banking infrastructure used by **5M+ users** across major UK retail banks in a highly regulated (PSD2/FCA) environment.
- Acted as the bridge between stakeholders, product, and engineering — translating regulatory and client requirements into clear user stories and production-ready systems.
- Worked directly with stakeholders to scope requirements and align technical delivery with business and compliance needs.
- Owned frontend delivery, shipping iteratively in agile sprints within fast-paced, ambiguous environments.

**Good Boost** — Founding Product Engineer (Backend) Nov 2020 – Apr 2022

- Sole backend engineer delivering a digital health platform supporting **20,000+ patients** across web and mobile (GDPR/HIPAA compliant).
- Partnered with clinicians and product teams to translate complex healthcare needs into backend systems enabling user-facing decision support tools.
- Built secure, scalable systems for handling sensitive patient data in real-world healthcare workflows.

**Samara AI (Antler)** — CTO 2025

- Developed agentic workflow prototypes using LLM orchestration across enterprise systems (Microsoft 365).
- Scoped and built bespoke AI tools for operational use cases, aligning technical capabilities with real user workflows.

## EDUCATION

---

**Imperial College London** — MPH Global Public Health

Jan 2020 – Dec 2022

- International Distinction Award.
- Dissertation: *Integrating Lived Experience in NHS Services to Improve Digital Adoption*.

**King's College London** — BA Global Health and Social Medicine

Sep 2016 – Dec 2019

- King's Sanctuary Award (Research Excellence in Migration & Health).
- Selected to present and chair at Oxford University Anthropology Conference.

## RESEARCH PAPERS

---

**Cultural Confabulation** — 2026

**The Auditor's Blindspot** — 2026

**Contextual Validity Evaluations** — 2026

**Governing the Boundary: Dual-Use Evaluation Frameworks for LLMs** — 2026

## CONFERENCES & PRESENTATIONS

---

**AIES 2026** — AAAI/ACM Conference on AI, Ethics, and Society, Malmö

**TAIGR Workshop, ICML 2026** — Technical AI Governance Research, Seoul

**AIMI Symposium, Stanford Health AI Week 2026** — Poster presentation, Stanford

**Technical AI Safety Conference 2026** — Presentation, Oxford

**Skoll World Forum 2026** — Invited roundtable facilitator, Oxford

**King's College London 2025** — Invited panellist, NHS Reform & Disability Rights, with Tom Shakespeare and MP Liam Conlon

## WRITING

---

**Artificial Intelligence in Plain English** (40K followers) — [AI Doesn't Hallucinate My Mother's Culture. It Rewrites It.](#)

## TRAINING & SCHOLARSHIPS

---

**BlueDot Impact** — AGI Safety Fundamentals (2026)

Bertelsmann Technology Scholarship — Product & AI

International Distinction Award — Imperial College London

Macquarie Group Scholarship — Cloud Computing

King's Sanctuary Award — King's College London