

Judith Begeer

London, UK • judihbegeer1@gmail.com • +447568201607 • linkedin.com/in/judith-begeer

PROFESSIONAL PROFILE

Trust & Safety leader and AI red-teaming specialist with 9+ years of experience across emerging threats, LLM safety, social media platform integrity, and global team leadership. Proven record of partnering with Tier-1 AI labs and platforms, including OpenAI, Google, and Amazon AGI, to identify vulnerabilities, architect safety policies, and build resilient content operations. Skilled at translating complex threat intelligence into concrete, automated testing protocols and managing multi-disciplinary, globally distributed teams.

CORE COMPETENCIES

LLM Safety & Red-Teaming • Trust & Safety Operations • AI Policy Architecture • Multimodal Threat Assessment • Social Media Platform Integrity • Intelligence Analysis • People & Team Leadership • Budget & Resource Management • Cross-functional Stakeholder Engagement • Operational Efficiency & Automation • Cyber Security

PROFESSIONAL EXPERIENCE

Head of Emerging Trends (interim) | Alice (formerly ActiveFence) | January 2022 – Present

Leading a cross-functional team of senior analysts delivering threat intelligence-led red-teaming to leading AI labs and social media platforms, with direct client responsibility for some of the world's most prominent technology organisations.

AI Safety & LLM Red-Teaming

- Provided threat intelligence and collaborated on multimodal red-teaming (text, image, video) to surface vulnerabilities in next-generation LLMs and AI systems prior to and following launch.
- Partnered with Amazon AGI to build their foundational safety policy framework, identifying critical implementation gaps, stress-testing models at key milestones, and directly shaping their core safety policies.
- Maintained collaborative relationships across a Tier-1 client portfolio including OpenAI, Google, Apple, Newsbreak, TSMC, Vimeo, and LinkTree.

Intelligence Operations

- Established an intelligence collation hub aggregating cross-functional data from specialist teams (CSAM, Counter-Terrorism, Hate Speech) to stress-test AI systems across a comprehensive range of abuse domains.
- Led a strategic intelligence partnership with Google, surfacing critical violations across Search, Maps, and AdSense.

People Management & Operational Leadership

- Managed three globally distributed teams across time zones and work cultures, achieving a 30% increase in intelligence delivery output.
- Served as Interim Vertical Lead (Emerging Threats): restructured workflows via AI-assisted automation and a freelancer network, reducing core headcount by 31% (13 → 9) while maintaining quality and volume targets.
- Directly managed team budgets including freelancer allocations, welfare provisions, and professional development investment.

Cyber Security Desk Lead & Subject Matter Expert | Dataminr (London) | July 2017 – January 2022

- Led and scaled the first Cyber Security desk in Dataminr's London office, growing the team to 8 analysts and generating the company's first cyber-focused revenue streams.
- Project-managed a rotating global team of 15 across three offices, providing real-time threat intelligence to public sector and corporate security executives.

Cyber Security Intern | Atos Consulting (Amstelveen, NL) | 2017

- Co-designed an executive crisis simulation for C-suite training on hacking incidents and internal fraud response.
- Prepared high-level policy advisories for healthcare boards and supported GDPR rollout for enterprise clients.

EDUCATION & QUALIFICATIONS

- MSc Crisis and Security Management, Leiden University (Top grade; thesis on Media-Induced Terrorism)
- BA International Studies, Leiden University (Distinction)

Certifications

- CompTIA Security+
- Beginners Python for Managers
- Agile Management

LANGUAGES

Dutch (Native) • English (Fluent) • German (A2) • Spanish (A2)