

MO SADEK

Vulnerability Disclosure & Security Community Leadership Coordinated disclosure & bug bounty programs · Offensive AI security research · Standards & community leadership

San Francisco Bay Area · mo@makesecurity.fun [linkedin.com/in/msadek1](https://www.linkedin.com/in/msadek1) · blog.makesecurity.fun

SUMMARY

A technical leader defined by range rather than a single lane. Deep enough in security and AI systems to do the work personally, from threat modeling to evaluation to prototyping, and organizationally fluent enough to build the structures around it: governance, cross-functional programs, executive alignment. That same investment extends beyond the day job, into open-source tooling, industry standards, and the policy conversations shaping how AI gets deployed. What ties it together is translation, getting researchers, engineers, and executives moving on the same problem, with a standing preference for showing an idea rather than describing it.

EXPERIENCE

Alice (formerly ActiveFence) — Technical Director

Jul 2025 – Present · San Francisco Bay Area *AI security company. Stood up and runs the coordinated-disclosure and offensive-research function.*

- Established a **security-research function** running offensive testing against AI systems and **coordinated vulnerability disclosure end to end** — from intake through remediation and public writeup.
- Designed and shipped **Caterpillar**, an open-source scanner auditing **AI agent skills and MCP tools** for malicious behavior — configurable **LLM-as-a-judge or rule-based** analysis, offline-capable, CI/CD gating.
- Co-led the **Rogue Agents** risk area of the **OWASP Top 10 for Agentic Applications (2026)** — a globally peer-reviewed framework for securing autonomous AI agents, referenced by major vendors and NIST.
- Built and led the company's public presence across **OWASP, CoSAI, and ISACA**; represents technical direction at industry conferences and hosts an AI-security podcast, *Curiouser & Curiouser*.
- Personally researches disclosure-worthy findings in frontier systems — a named finding against **Perplexity's Comet browser**, prompt-injection testing against **Gemini 2.5 Pro and GPT-5** — so the program is run by someone who has stood on the researcher's side of it.

Roblox — Sr. Application Security Engineer (Security Partner)

Sep 2019 – Jun 2024 *Early security hire through the company's hypergrowth years and DPO; owned bug bounty and external disclosure for a platform at consumer scale.*

- Took over Roblox's **bug-bounty and vulnerability-disclosure program** in 2020 and turned it around against hundreds of incoming vulnerabilities: response rate from ~60% to **90%**, average closing time from a month-plus to **~2 weeks** — while owning HackerOne and coordinating external disclosures.
- Led and supported **incident response** across account takeover, insider threat (including a 5,000+ person outsourced workforce), fraud, and platform-abuse domains — the judgment calls a disclosure queue depends on: telling a bad-faith actor from a good-faith researcher.
- Built the **application-security practice from scratch** and a partnership model that made security the front door for engineering; grew the function **3 → 12** and stewarded the org **7 → 90**, leading the first ~40 hires and owning onboarding and mentorship.
- Established org-wide **open-source security governance** ("Trust by Design," 120+ reviews) and built (in Go) a **security-scanning platform** orchestrating SAST, DAST, secret, dependency, and container scanning across internal and external tooling.
- Ran **threat modeling and architecture reviews** for high-impact systems — a voice-chat safety system, region-specific apps for China and Vietnam, and an **RBAC re-architecture** rationalizing 400+ roles across that same 5,000+ person workforce.

Roblox — Chief of Staff to the Chief Scientist

Jun 2024 – May 2025 *Executive strategy seat: governance for how research and new AI capability get evaluated before they ship.*

- Established and chaired the companywide **AI Review Board** — cross-functional reviews of **80+ AI initiatives** against a risk framework spanning cost, ethics, performance, infrastructure, and user safety, with automated intake and triage across security, privacy, patent, and IP.
- Led evaluation of third-party AI integrations and open-source models against internal security standards, and partnered with legal and compliance as regulation evolved — the same translation work a disclosure program needs between what a researcher finds and what a release decision requires.

NBCUniversal — Sr. Application Security Analyst

Sep 2016 – Mar 2018 No security program existed; built one from the ground up as a junior engineer, including a **bug-bounty program**, security testing across broadcast and on-air systems with a PCI compliance track, and a mobile-application security program.

Bosch — Security Engineer, Center for AI Research

Nov 2018 – Aug 2019 Security engineer for Bosch's AI research and enablement division: threat-modeled autonomous-driving systems and their software supply chain, built Kubernetes RBAC for a GPU research cluster, and stood up an early ML-for-SOC threat-detection capability.

Warner Brothers — Lead Application Security Engineer

Apr 2018 – Oct 2018 Restructured the security-assessment program around risk-based architecture reviews and strengthened third-party risk processes.

Cigital (now Synopsys) — Security Consultant

2015 – 2016 Delivered OWASP Top 10 vulnerability assessments for client applications — the pentesting work that shaped everything since: breaking things to prove you could versus building the structure that actually gets things fixed.

DISCLOSURE, RESEARCH & COMMUNITY

Program operations bug bounty & VDP program ownership (HackerOne) · disclosure queue triage · incident response & abuse investigation · coordinated public disclosure **Research** offensive AI/LLM testing · prompt-injection research · agentic-system & MCP security · threat modeling **Standards & community** OWASP (Rogue Agents/ASI10 co-lead) · CoSAI · ISACA · conference speaking · podcast host

SELECTED PROJECTS & RECOGNITION

- **Caterpillar** — open-source AI-skill / MCP-tool security scanner (TypeScript, MIT).
- **OWASP Top 10 for Agentic Applications (2026)** — Co-Lead, *Rogue Agents* risk area.
- **Independent security research** — named finding against Perplexity's Comet browser; prompt-injection testing against Gemini 2.5 Pro and GPT-5.
- **Speaking & writing** — RSAC 2025; HumanX 2026 ("What AI Red Teaming Looks Like Outside the Lab"); GOSIM Paris 2026; The Official Cybersecurity Summit 2025; host, *Curiouser & Curiouser* AI-security podcast; *makeSecurityFun* blog.

EDUCATION

B.S., Computer & Information Sciences — SUNY College at Fredonia (2011–2015)