

Ricardo Nahuel Prieto

AI SECURITY | RED TEAMING | AGENTIC SYSTEMS

Professional summary

01

Offensive security and AI security engineer with 7+ years of hands-on experience across penetration testing, application security, security research and AI systems. Currently focused on threat modeling LLM applications, agents, multi-agent systems and MCP integrations.

Recent work connects conventional security with emerging AI risks: mapping trust boundaries and attack paths, identifying failure modes, designing adversarial assessments and translating findings into technical controls. Increasingly focused on AI control, monitoring and loss-of-control problems introduced by autonomous agents.

Selected AI security & research work

02

Agentic Threat Modeler

IN DEVELOPMENT

Agentic tool using multiple AI agents to analyze system architectures, trust boundaries and attack surfaces, then structure the resulting threat-model findings.

[Project overview](#)

AI Red Teaming Field Guide

PUBLIC DRAFTS UNDER REVIEW

Long-form practical methodology for scoping, testing and reporting adversarial assessments of LLM, agentic, RAG and MCP systems.

[Read the guide](#)

Deceiving the Reviewer

TECHNICAL ANALYSIS

Analysis of the AISI agent-deception incident, including oversight failure, multi-agent coordination, context compaction and human review boundaries.

[Read the analysis](#)

MCP Security for Enterprise Organizations

RESEARCH + CONFERENCE TALK

Technical work on MCP trust boundaries, tool poisoning, prompt injection, privileged actions and isolation controls, presented at Ekoparty and UTN.

[Open the talk](#)

Professional experience

03

Senior AI Security Engineer

Jul 2026 - Present | Remote, USA

CookUnity

- Build and lead the AI Security practice for production GenAI systems, defining security requirements for LLM and agentic architectures.
- Threat model LLM pipelines, agent workflows and MCP/tool integrations, translating trust boundaries, privileged actions and prompt-injection paths into controls and secure defaults.
- Use Cursor, Codex and Claude Code daily; adversarially assess in-house agentic tooling for tool misuse, authorization gaps and scope overreach.

Senior AI Security Engineer

Oct 2025 - Jul 2026 | Remote, Argentina

itti

- Built the organization's AI Security function across architecture review, threat modeling, secure adoption and assessment of GenAI systems.
- Threat modeled multi-agent and tool-using architectures, covering trust boundaries, prompt injection, excessive agency, sensitive-data exposure and potential tool misuse.
- Defined guardrails and human-oversight boundaries for higher-risk workflows; reviewed Kore.ai and AWS Bedrock configurations.
- Developed reusable AI threat-modeling approaches and incorporated prompt-injection testing into application assessments.

Senior Penetration Tester & Security Consultant

Jun 2023 - Dec 2024 | Wellington, New Zealand

Bastion Security Group

- Delivered 40+ penetration tests and security assessments across web applications, APIs, cloud infrastructure and networks.
- Identified and documented 250+ vulnerabilities with reproducible evidence, attack-path analysis and actionable remediation.

Earlier security experience

04

Senior Application Security Engineer

Jun 2025 - Oct 2025

itti

- Led the Security Champions program, initiated GenAI threat-modeling work and integrated SAST, DAST and SCA into CI/CD pipelines.

Cybersecurity Engineer

Sep 2022 - Apr 2023

Mercado Libre

- Coordinated vulnerability remediation across 15 development teams and 200+ applications; automated dependency-security analysis with Python and SQL.

Cybersecurity Researcher

Dec 2020 - Sep 2022

Onapsis

- Performed 30+ vulnerability assessments and SAST/DAST reviews for enterprise systems, producing bilingual technical and executive reports.

Offensive Security Analyst

Dec 2019 - Dec 2020

Deloitte

- Built a Python-based security validation framework and reusable modules for enterprise SaaS authorization testing.

Additional experience

05

Cybersecurity Tutor, TripleTen (2025 - Present): offensive-security mentoring through live and one-to-one sessions.

Hack The Box Ambassador & Organizer (2024 - Present): co-founded the New Zealand meetup and later joined the Argentina chapter.

Co-founder, ThreatX Security (2025 - Present): Spanish-speaking cybersecurity community grown from HTB Meetup Argentina, with nearly 2,000 members and 34 events.

Selected technical writing & talks

06

Agentic Threat Modeler

Presentation of a multi-agent tool for structured system threat modeling.

OWASP Agentic Skills Top 10

Technical analysis of risks introduced by reusable instructions and tool-enabled agent skills.

Prompt Injection Attacks

Attack taxonomy, real incidents and layered controls for direct and indirect instruction attacks.

MCP Security

Conference talks at Ekoparty and UTN on tool trust boundaries and enterprise defenses.

Technical skills

07

AI & AGENT SECURITY

LLM application threat modeling, multi-agent and tool-using systems, MCP/tool security, prompt injection, agent authorization, AI red teaming

OFFENSIVE SECURITY

Web/API penetration testing, attack-path analysis, adversarial testing, vulnerability research, code review

ENGINEERING

Python, TypeScript, security automation, CI/CD security, AWS; daily use of Cursor, Codex and Claude Code, with experience using OpenCode and Antigravity

AI SAFETY / CONTROL

Agent failure modes, monitoring and control concepts, evaluation awareness, human oversight

Education & credentials

08

EDUCATION

M.S. Information Security - University of Buenos Aires (expected Dec 2026)
Diploma in Computer Security - National Technological University (2020)

CERTIFICATIONS

Selected: HTB CBBH, eWPT, Certified AI Security Professional (CAISP), Trusted AI Safety Expert (TAISE)

LANGUAGES / LOCATION

Spanish (native) | English (professional)
Argentina / remote worldwide / open to relocation