Mario Marcolongo

AI Evaluation, Red Teaming & Information Integrity

Italy · EU citizen · Native Italian, fluent English · Open to relocation & remote work · Available now
me@mariomarcolongo.com · mariomarcolongo.com · LinkedIn · ORCID · GitHub

Summary

Adversarial LLM testing, including jailbreaks and prompt injection; top 5% of Gray Swan Proving Ground; member, Syntony Research Trusted Red Team Network. Paid scientific fact-checking and AI workflow configuration. Atlas preprint coauthor; merged a Basic Memory fix that stops AI agents from being told to edit or delete notes unasked.

AI red-team competitions

Gray Swan
Independent evaluation · Final competition record, 29 October 2026

- Top 5% of Gray Swan Proving Ground: #81, 127 recorded breaks across chat, image, agent and indirect prompt injection (13 Oct 2026).
- 49th of 390 in Lucky Break: Level Up (Oct 2026), 15 recorded breaks. Refine tests from response patterns and check recurrence across scenarios.

Professional experience

Entropy for Life
Science Writer, Fact-Checker & Website Manager · Paid contractor · Jun 2023–Present · Hybrid, primarily remote · Bassano del Grappa, Italy
Client: Italian science YouTube channel · 272K subscribers (Oct 2026)

- Research primary literature and adapt English-language evidence for Italian content; acknowledged for literature research and error detection in Italiani veri (Mondadori, a major European publishing group, 2025).
- Since Oct 2026, configure Antigravity/Gemini workflows, review outputs and train the client, who used them to rename and tag thousands of images and hundreds of videos.

Projects & public contributions

AI Safety Formalization Atlas
Open-source contributor & preprint coauthor · Sep 2026–Present · Remote

- Five merged PRs: source-metadata audit, 15 citation locators, macOS validation and agent instructions.
- Coauthor, The AI Safety Formalization Atlas (Zenodo preprint, 2026).

Wikipedia, Wikidata & Wikimedia Commons
Public contributor · Since 2018 · Remote

- Review citations, reconcile conflicting sources and check biomedical terminology through attributable public revisions.

Education & credentials

Metropolia University of Applied Sciences — Information Technology — Open UAS Path Studies
Aug 2026–Present
Fully online degree-credit pathway toward a BEng; self-paced coursework and assessments. Python Programming: 3 ECTS, Pass (Sep 2026).

University of Campania Luigi Vanvitelli — Medicine coursework
2020–approx. 2023 (attendance)

Microsoft Applied Skills: Build an agent in Microsoft Copilot Studio · Oct 2026

Skills

Demonstrated work: LLM adversarial testing, jailbreaks, prompt injection, response-pattern analysis, agent instructions, technical documentation and client workflows.
Independent investigations: Source tracing, web archives, domain and website checks, image verification and cross-platform research; distinguish observed connections from attribution hypotheses.
Current study: Python, SQL, systems and quantitative foundations.

Languages

Italian native · English fluent (EF SET C1, Mar 2024) · Certificate

---

Independent software projects use AI-assisted implementation. Full experience and source records.