Ma'alona Mafaufau

✉ maalona.mafaufau@approxiomresearch.com   in www.linkedin.com/in/maalonamafaufau/

○ www.github.com/Lona44   🌐 www.approxiomresearch.com

PROFESSIONAL SUMMARY

Full Stack Developer and AI Safety Researcher with 6+ years of experience across cybersecurity, data analytics, and software development. 3rd place winner of the Palisade Research AI Misalignment Bounty, demonstrating expertise in systematic vulnerability discovery in frontier AI systems. Combines strong full stack development skills (React, TypeScript, Node.js, Azure, MongoDB) with deep understanding of AI agent architectures and safety evaluation methodologies. Passionate about developing robust AI-integrated applications and advancing AI safety research.

TECHNICAL SKILLS

- **Frontend:** React, TypeScript, JavaScript, HTML5, CSS3, Tailwind
- **Backend:** Node.js, Express.js, FastAPI, REST APIs, Python
- **Databases:** MongoDB, Mongoose, SQL Server, PostgreSQL, Firestore
- **Cloud:** Azure, GCP (Cloud Run, Build, Firestore), AWS, Docker, Git, CI/CD
- **AI/ML:** OpenAI API, Anthropic API, Gemini API
- **Testing:** Jest, Supertest, MongoDB Memory Server, TDD
- **Security:** SIEM (Splunk), Threat Detection, APT Hunting, Blue Team

PROJECTS

REDD - AI Safety Testing Platform

- redd.approxiomresearch.com
- github.com/Lona44/unified-ai-misalignment-framework

Full-stack platform and containerised framework for systematic AI safety evaluation, built on methodologies from the Palisade Research AI Misalignment Bounty

- Designed production platform testing 12+ frontier AI models (GPT-5, Claude Opus/Sonnet, Gemini, o3) for misalignment behaviors through scenario-based experiments

- Built React frontend with Material-UI; architected FastAPI backend on GCP (Cloud Run, Firestore, Cloud Build, Secret Manager)

- Developed reproducible containerised testing framework supporting multiple AI providers; captured reasoning traces from OpenAI and Anthropic APIs

- **Technologies:** React, FastAPI, GCP, Docker, Python, LiteLLM, OpenAI SDK, Anthropic SDK, OpenRouter, GitHub Actions

Trade Me Marketplace — Mission Ready HQ (Team Project)

- https://github.com/Lona44/Mission-Ready-L5-Mission-5-Phase-2

Full-stack auction marketplace with AI-powered features

- Team of 3 developers and 2 UX designers developing an auction marketplace with bidding, watchlists, product comparison, and AI-powered Q&A (Gemini/Ollama)

- Developed RESTful API with MongoDB; achieved 56+ passing tests using Jest with TDD methodology

- **Technologies:** React 19, Node.js, Express.js, MongoDB, Google Gemini API, Jest

Turners Insurance AI Chatbot — Mission Ready HQ
- https://github.com/Lona44/Mission-Ready-L5-Mission-3-Gemini-Chat-Bot
Conversational AI insurance consultant with multi-turn dialogue
- Built AI chatbot using Google Gemini 2.5 Flash with session-based conversation management and business rule enforcement
- **Technologies:** React 19, Node.js, Express.js, Google Gemini API, Vite

WORK EXPERIENCE	

Independent Researcher   **Approxiom Research**
April 2025 - Present   Auckland, New Zealand
Independent AI Safety researcher specialising in systematic vulnerability discovery in frontier AI systems.
- **3rd place winner** of the Palisade Research AI Misalignment Bounty demonstrating distinct misalignment patterns in o3 and GPT-5 models. Work published in peer-reviewed research paper (arXiv:2510.19738) and released publicly on Hugging Face
- Built full-stack web platform (React, FastAPI, GCP) for experiment management and results visualisation - redd.approxiomresearch.com
- Created reproducible experiment pipelines identifying misalignment behaviors in GPT-5, Claude Opus, and o3 models
- Implemented secure API key management with pre-flight validation and encrypted storage

Assistant Trainer   **Mission Ready HQ**
October 2025 - Present   Auckland, New Zealand
Mentored full-stack developer students in React, Node.js, and MongoDB. Provided code reviews and facilitated learning sessions on AI integration and API development.

Cybersecurity Engineer / Operations   **Workday**
Analyst
January 2022 - December 2024   Auckland, New Zealand
Promoted from analyst to engineer. Focus on threat hunting and incident response with global teams (Ireland, U.S.A.).
- Developed custom detection rules using Splunk SPL; built automated threat intelligence pipelines in Python
- Led APT threat hunting initiatives; created SIEM efficacy analysis tool adopted globally by security team

Cybersecurity Analyst/Intern   **Datacom**
April 2021 - January 2022   Auckland, New Zealand
Monitored customer environments as triage analyst for SIEM detections. Escalated tuning suggestions to engineering team.

Data Analyst   **Halter Limited**
September 2019 - February 2021   Auckland, New Zealand
Analysed agricultural IoT data from smart collar systems. Built statistical models and ML-based products for customer needs.

Actuarial Analyst   **Ministry of Social Development**
April 2018 - August 2019   Wellington, New Zealand
Developed forecasting models and analytical tools for ministry performance monitoring.

ADDITIONAL EXPERIENCE

Pasifika Tech Tutor/Speaker   **Pasifika Tech Education Charity**
September 2022 - November 2024   Auckland, New Zealand
Facilitated technology workshops for Pasifika communities. Organised workplace visits for Pacific high school students at major tech companies.

EDUCATION

Mission Ready HQ, Auckland, NZ
Diploma in Digital Technology Development and Design (L5/L6), Expected April 2026

University of Auckland, Auckland, NZ
Bachelor of Science (Honours), Statistics, May 2016
Bachelor of Science, Double Major Mathematics and Statistics, May 2015

AWARDS & CERTS

3rd Place - Palisade Research AI Misalignment Bounty — 2025
Demonstrated reproducible misalignment in frontier AI models. Published: arXiv:2510.19738

Full Stack Developer Certificate — Mission Ready HQ, August 2025
`https://www.credential.net/6f647d31-cbd0-4b57-ac98-9a658d993c31#acc.k27D3BG6`

Palo Alto Networks Cybersecurity Professional Certificate — Coursera, February 2025
`https://coursera.org/share/a2f350d91f2567156406af86e1076ba1`

Blue Team Level 1 (BTL1) — Security Blue Team, September 2022
`https://credly.com/badges/74f5a12a-e8e4-4ae8-8823-f7bc0f81b78a`

AWS Cloud Quest: Cloud Practitioner — Amazon Web Services, August 2023
`https://www.credly.com/badges/673233f8-3b1e-4c96-846b-ed15de3867b1/public_url`

Splunk Fundamentals 1 — Splunk, September 2021

Unitec CTF 2021 Winner — Kawaiicon 2021 prize for Boot2Root challenges

References available on request.