

Adrian Wedd

AI Safety Researcher & Developer

PO Box 331, Cygnet, Tasmania, Australia | +61407081084 | adrian@adrianwedd.com | LinkedIn | GitHub

Summary

Systems builder, AI safety researcher, and adversarial thinker with nearly 45 years across the stack. The failure-first methodology didn't come from papers — it came from coordinating direct actions for Greenpeace's Actions unit against well-resourced opponents where getting the risk assessment wrong had real consequences.

Core Competencies

AI Safety & Evaluation | Frontier AI Models | Risk Assessment | Policy Translation

Professional Experience

Freelance Developer & AI Consultant — Independent

2025 - Present

Client web delivery and independent AI safety research, transitioning from seven years in government.

- Designed and built red-teaming and adversarial testing framework for embodied AI with automated attack adapters (TAP) and cross-modal vulnerability research on VLA models
- Developed evaluation frameworks with hallucination detection, content safety gates, and model monitoring for production LLM pipelines
- Applied prompt engineering, RAG, and retrieval-augmented generation across multiple agentic systems

Systems Analyst / Acting Senior Change Analyst — Homes Tasmania (formerly Department of Communities Tasmania)

2018 - Feb 2026

Led systems integration, cybersecurity, and digital transformation for Tasmania's public housing sector.

- Designed and implemented RESTful API integrations and SFTP data exchange for the Housing Management System
- Led cybersecurity program including vulnerability assessment, penetration testing, and Essential Eight compliance uplift
- Implemented identity and access management (IDAM) controls reducing unauthorised access risk

ITS Client Services Officer — University of Tasmania

2015 - 2018

IT support across the full university user base — researchers, academics, administrative staff.

- Sole support contact for a diverse user base with no tolerance for downtime — learned to triage fast and fix faster
- Built and maintained procedural documentation that actually got used, reducing repeat escalations
- Managed Microsoft 365 and Active Directory environments across a complex multi-faculty structure

Selected Projects

- **This Wasn't in the Brochure** — Practical, neurodiversity-affirming guide for co-parents raising children with ADHD, Autism, PDA, and ODD.
- **NotebookLM Automation** — CLI automation toolkit for Google NotebookLM with research quality controls, deterministic exports, template engine, JSON schema validation, and comprehensive integration

testing.

- **Orbitr** — Multi-track polyphonic AI sequencer with concentric ring interface.

Technical Skills

Programming Languages: Python, JavaScript, TypeScript, PowerShell, FastAPI, Bash

AI & Automation: Large Language Models, AI Safety & Evaluation, Process Automation, ML Libraries, MCP Development, Voice/Telephony AI, Deep Learning, Generative AI, Computer Vision

Infrastructure: Systems Integration, Cybersecurity, Linux/Windows Server, Cloud Platforms

DevOps: Docker, GitHub Actions, Cloudflare Pages, Kubernetes, MLOps

Frontend: React

Data: SQL / PostgreSQL, Data Engineering

Soft Skills: Leadership & Collaboration

Education

Self-Directed AI & Machine Learning Study

Independent Study — 2018 - Present

Computer Science Foundation

Self-Taught & Professional Experience — 2000 - 2018