

# Adrian Wedd

AI Safety Engineer & Independent Researcher

Cygnet, Tasmania, Australia

[GitHub](#) [LinkedIn](#) [Email](#) [PDF](#) [Print](#) [Watch Me Work](#)

---

## ABOUT

---

AI Safety Engineer and Independent Researcher. Three years empirical research on frontier AI models, focused on red-teaming, evaluation frameworks, and failure-first methodology. Seven years translating complex technical findings into actionable insights for government decision-makers.

---

## CORE COMPETENCIES

### AI Safety & Evaluation

Red-teaming, adversarial testing, failure-first methodology

### Frontier AI Models

Claude API, multi-agent systems, evaluation frameworks

### Risk Assessment

Pre-mortem analysis, FMEA, failure mode identification

### Policy Translation

Technical findings to actionable governance for decision-makers

## EXPERIENCE

---

### Applications Specialist / Acting Senior Change Analyst

Homes Tasmania 2018 - Present

Complex socio-technical systems analysis for Tasmania's housing and community services portfolio. Developed Homes Tasmania's first Generative AI usage policy in 2023.

- Developed Homes Tasmania's GenAI usage policy, procedure, and training materials
- Led cybersecurity initiatives improving system security
- Systems integration and API development using RESTful APIs and Python

Python, PowerShell, JavaScript, RESTful APIs, Azure

## PROJECTS

---

### ADHDo

GitHub

AI safety system with Claude integration, confidence gating, and multi-tool orchestration.

Python Claude API Safety Systems

---

### Agentic Research Engine

GitHub

Multi-agent evaluation framework with LangGraph orchestration and self-correction loops.

Python LangGraph Multi-Agent Systems

SKILLS

---

Programming Languages

---

Python	Primary
--------	---------

JavaScript / TypeScript	Primary
-------------------------	---------

AI & Safety

---

Frontier AI Models	Primary
--------------------	---------

Red-Teaming	Primary
-------------	---------

Evaluation Frameworks	Primary
-----------------------	---------

Infrastructure

---

Systems Integration	Primary
---------------------	---------

Cybersecurity	Primary
---------------	---------

Research Methods

---

Risk Assessment	Primary
-----------------	---------

Technical Writing	Primary
-------------------	---------

ACHIEVEMENTS

---

**Failure-First Research Program**

Developed comprehensive failure-first evaluation methodology for agentic AI systems.

2022-Present

---

**Published Research: Organisational AI Governance**

Published research on structural barriers to acting on AI safety evidence in organisations.

2025

---

**AI Governance Policy Development**

Developed Homes Tasmania's first Generative AI usage policy and training materials.

2023

---

**Adrian Wedd**

AI Safety Engineer & Independent Researcher

Last updated: Feb 16, 2026, 10:07 PM