

Max McWhae

Perth, Western Australia · maxmcwhae.com · github.com/mcawezome · linkedin.com/in/max-mcwhae-1bb983102

Technical AI safety researcher in training, focused on evaluation awareness and sandbagging in open-weight language models. Founder of the Perth AI Safety Meetup. Mathematics undergraduate at the University of Western Australia; writing code since 2002.

AI SAFETY & RESEARCH

SANDGLASS — Independent research project 2026 – present

A pilot study of evaluation awareness and trained sandbagging in open-weight models: whether models internally represent “I am being evaluated,” and whether deliberate underperformance triggered by that recognition can be detected behaviourally or via linear activation probes. Developed through BlueDot Impact's Technical AI Safety project track; runs on consumer hardware; all outputs open, including negative results.

Founder & Organiser — Perth AI Safety Meetup 2026 – present

Founded Perth's first dedicated AI safety group. Piloted three sessions at UWA and Curtin in collaboration with EA UWA; relaunching as a monthly series at UWA.

Technical AI Safety course — BlueDot Impact 2026

Completed. Currently working through the ARENA curriculum (transformer internals and mechanistic interpretability).

Conferences 2025 – 2026

AJCAI 2025, Canberra (attended); AI Safety Forum 2026, Sydney (attended).

EDUCATION

Bachelor of Mathematics — University of Western Australia 2026 – present

Diploma of IT (Advanced Programming — Python) — North Metropolitan TAFE 2025

Including the AI Skill Set and Data Analysis Skill Set.

Certificate IV in IT (Advanced Programming) — North Metropolitan TAFE 2022

BEng (Mining Engineering) — Curtin University 2012 – 2016

EXPERIENCE

Professional Poker Player — Self-employed, Perth Jun 2025 – Mar 2026

Supported myself through live cash-game play; cashed both Crown Perth monthly tournament entries (2nd and 4th).

Retail Assistant — ALDI, Busselton 2023 – 2024

Pick & Packer / Delivery Driver — Galvin's Plumbing Supplies 2019 – 2020

Sales & Warehouse All-rounder — YHI Power, Canning Vale 2018

TECHNICAL

Python (primary language) · PyTorch (in training via ARENA) · data analysis · Git/GitHub · local LLM tooling · Linux