

# LEO HYAMS

## Curriculum Vitae

Cape Town, South Africa — leo@aisafetysa.com — LinkedIn — Github — (+27) 72 928 3288

### EXPERIENCE

---

**Founder, Executive Director**, AI Safety South Africa January 2024 - present  
Responsible for fundraising, operations, and team management, as well as research and program direction. Recently completed directing the Cooperative AI Research Fellowship and a Science of Evaluations research grant with the UK AI Security Institute (UK AISI). You can see more details on this organisation here. I am also supporting the University of Cape Town's African Hub on AI Safety Peace, Safety, and Security to develop a postgraduate course in technical AI safety.

**Course Facilitator**, Cooperative AI Foundation March 2025 - July 2025  
Ran the first ever in-person cohort for the Intro to Cooperative AI Course. Handled recruitment, selection, course facilitation, and contributed to session guide design.

**Software Engineer**, AI Safety Engineering Taskforce January 2025 - April 2025  
During this contract I was engaged in both a managerial and an engineering role. Here, I was subcontracted by ASET to led a team of four engineers from the AI Safety South Africa network. We developed capability evaluations for the UK AISI and of the four projects that we contributed to, all four were accepted by the UK AISI.

**Course Facilitator and Ambassador**, BlueDot Impact June 2024 - February 2025  
Ran the first ever official in-person cohort of BlueDot Impact's Intro to Transformative AI course. I recruited talent and facilitated the course. As a result of this course, two of the members of my small cohort made significant career transitions into reducing large-scale risk from AI.

**Talent Identification Coordinator**, Impact Academy October 2024 - December 2024  
Recruited talent for the Global AI Safety Fellowship conducted by Impact Academy in 2025. According to Impact Academy, all the referrals I provided were among the top 10% candidates from Africa.

**Operations Contractor**, Condor Initiative August 2024 - December 2024  
Organised an upskilling retreat for 24 of the top AI Safety talent in South Africa, including both technical and governance talent. These were young professionals in ML engineering or students sourced from the top universities in SA (mainly masters students but some PhDs and undergraduates).

**Growth and Partnerships Representative**, Fruitpunch AI September 2023 - January 2024

### EDUCATION

---

**University of Cape Town** January 2020 - June 2023  
Bachelors of Social Science in Computer Science and Psychology

### VOLUNTEERING

---

**Research Fellow**, Apart Research Lab May 2024 - December 2024

**Lead Organizer**, Deep Learning IndabaX Conference South Africa April 2023 - September 2024  
Organised the 2024 Deep Learning IndabaX acting as Partnerships Lead. Additionally, I was involved in organising 2023's Deep Learning IndabaX, where I served as both the Hackathon lead and on the Local Organising board. Both conferences had 300-400 attendees

**Community Lead & AI for Good Engineer**, Fruitpunch AI March 2022 - February 2024

### SELECTED RESEARCH & OUTPUTS

---

**Predicting Long-Horizon Task Performance using Proxy Evals.**

J. Du Toit, L. Hyams, D. Anisimov, B. Chikavanga, J. Stein, and S. Brown, submitted for publication at the 2026 NeurIPS Track for AI Evaluation and Datasets, *OpenReview*: [OpenReview Link](#)

**Precursors, Proxies, and Predictive Models for Long-Horizon Tasks.**

J. Du Toit, L. Hyams, D. Anisimov, and S. Brown, published at the 2025 NeurIPS LLM Evaluation Workshop, *OpenReview*: <https://openreview.net/forum?id=bFX0oa2XDr>

**HumanAgencyBench: Do Language Models Support Human Agency?**

B. Sturgeon, L. Hyams, D. Samuelson, E. Vorster, J. Haimes, and J. R. Anthis, published at the 2025 AAI Workshop on Datasets and Evaluators of AI Safety, *OpenReview*: <https://openreview.net/forum?id=nHp5FquS2R>

## SELECT GRANTS

---

- Organisational Grant**, Coefficient Giving June 2026 - June 2028  
Awarded 2 years of funding for developing the Cape Institute for Safe AI.
- Project Grant**, Cooperative AI Foundation June 2025 - September 2026  
Awarded a grant for work relating to the Cooperative AI Research Fellowship.
- Research Grant**, UK AI Security Institute September 2025 - April 2026  
Grant titled: Proxies for Efficient Evaluation of Autonomous Replication and Adaptation in AI Agents. This work was developed in collaboration with the Science of Evaluations Team.
- Project Grant**, Cooperative AI Foundation June 2025 - April 2026  
Awarded a grant for work relating to the Cooperative AI Research Fellowship.
- Project Grant**, Survival & Flourishing Fund June 2025 - April 2026  
Awarded a grant for work relating to the Cooperative AI Research Fellowship.
- Project Grant**, AI Safety Tactical Opportunities Fund June 2025 - April 2026  
Awarded a grant for work relating to the Cooperative AI Research Fellowship.
- Organisational Grant**, Open Philanthropy (now Coefficient Giving) May 2025 - June 2026  
Awarded a grant for one year of general support of AI Safety South Africa (previously named AI Safety Cape Town).

## SKILLS

---

**Technical:** Python, Git, Inspect AI Evaluation Framework, Pytorch, NumPy, Matplotlib, OverLeaf, basic web.

**Interpersonal:** Communication, Conflict Management, Stakeholder Management, Strategy, Culture Building