Daniel Filan – Curriculum Vitæ

daniel_filan@berkeley.edu

AREAS OF INTEREST

AI Alignment (Machine Learning Transparency, Value Learning), Theory of Artificial Intelligence (Reinforcement Learning, Algorithmic Information Theory, Statistical Machine Learning), Economics (Agency Theory)

DEGREES

Doctor of Philosophy in Computer Science, 2016 – present
University of California, Berkeley

- Studying AI alignment, supervised by Stuart Russell.
- Researcher at the Center for Human-Compatible AI.
- GPA: 3.45/4.00

Bachelor of Philosophy (Hons), 2012 – 2015
Australian National University

- Honours in Computer Science, undergraduate studies in Mathematics and Physics.
- Thesis: "Resource-bounded Complexity-based Priors for Agents", supervised by Marcus Hutter.
- GPA: 7.00/7.00, 1st Class Honours.

PUBLICATIONS

- Loss Bounds and Time Complexity for Speed Priors. With Jan Leike and Marcus Hutter. AISTATS 2016.
- Self-modification of Policy and Utility Function in Rational Agents. With Tom Everitt (lead author), Mayank Daswani, and Marcus Hutter. AGI 2016, recipient of Kurzweil Prize for Best Paper.
- Exploring Hierarchy-Aware Inverse Reinforcement Learning. With Chris Cundy (lead author). GoalsRL Workshop at ICML/IJCAI/AAMAS 2018.

SELECTED AWARDS

University Medal, Australian National University 2015

- Prize; awarded to students who have obtained First Class Honours (or Masters Advanced Equivalent) and demonstrated exceptional academic excellence across their studies, the highest academic prize for undergraduates.

Erin Brent Computer Science Prize, Australian National University 2015

- Monetary prize; awarded to the student who achieved the best Honours result in any of the degree programs relating to Computer Science, Software Engineering or Information Technology.

PROJECTS

Writing my PhD thesis Aug 16 2023 – present

- Compiling my research related to cluster structure inside neural networks, writing it up, and explaining how it forms a unified whole.

INTERNSHIPS

Machine Intelligence Research Internship 2019

- Spent 3 months on research engineering team 4 days per week, while supervising a UC Berkeley intern 1 day a week.