Sam Clarke

Research Experience

March-May 2020	*Independent AI strategy research* at AI Safety Camp
	I will be compiling the strongest arguments about why advanced AI might pose an existential threat. The focus will be on making the various claims and their relations to one another as clear as possible, and finding the weakest inferences and assumptions.
May-Sep 2019	*Master's Thesis in Computer Science* at University of Oxford
	Applied Deep Bayesian Active Learning to the Reward Modelling approach to agent alignment in order to improve its sample efficiency. Thesis available here.
Aug-Sep 2018	*Summer Research Fellow* at Centre for Effective Altruism, Oxford
	Developed a method for quantifying uncertainties in the ITN Framework. Produced a report which weakly recommends not to use ITN results when choosing between top cause areas.
Nov 2016 - Jan 2017	*Summer Research Scholarship* at University of Auckland, New Zealand
	Completed a 10-week Natural Language Processing research project. Reviewed the applications of word embedding algorithms and investigated using Word2vec to model a "train of thought".

Education

Oct 2018 - Sep 2019	Master of Science in Computer Science **University of Oxford**, United Kingdom Master's Thesis: *Active Reward Modeling for Agent Alignment* Supervisors: Yarin Gal & Zachary Kenton | List of Exams
Jan 2015 - June 2018	Bachelor of Science in Philosophy and Computer Science **University of Canterbury**, New Zealand Gpa: 9.0/9.0 | List of Exams
Oct 2017 - June 2018	Visiting Student in Philosophy and Computer Science **University of Oxford**, United Kingdom Average Grade: A | List of Exams
2010-2014	High School, **King's College**, Auckland, New Zealand Dux and Deputy Head Boy | List of Exams

Other Academic Achievements and Experiences

March 2020	Attending *MIRI AIRCS Workshop* with full funding
December 2019	Visits to MIRI, CHAI, PAI
September 2019	Attended *CFAR Workshop* with 90% scholarship
October 2018	Participant in *MIRIx* Oxford
August 2018	Attended *Human-aligned AI Summer School*, Prague
2017-2018	Organiser and member of *Oxford AI Safety Reading and Discussion Group*
2016-2017	*Tutor in Computer Science*, University of Canterbury, New Zealand
March 2014	National Champion at *New Zealand Young Physicists' Tournament*

Work Experience & Leadership

Aug-Sep 2018	*Summer Research Fellow* at Centre for Effective Altruism, Oxford
Nov 2016 - Jan 2017	*Summer Research Scholarship* at University of Auckland, New Zealand
2014 - 2018	*University and Private Tutor* in Computer Science, Maths and English
	Held regular tutorials at my university college and mentored individual students
Sep 2017	*Teaching Assistant*, Lycée de La Sauque, Bordeaux
	Assisted with English lessons at French lycée; taught classes on New Zealand life and culture
July 2017	*Volunteer*, Wahoe NGO, New Delhi
	Taught English at slum school; wrote material for teacher and future volunteers to use
June 2015	*Class Tutor*, Campus Link Foundation, Auckland
	Prepared and presented two-day intensive courses to secondary students as part of a non-profit striving for educational equality in New Zealand

Awards

April 2018	Book Prize for "Logic & Proof" Coursework, St. Catherine's College, Oxford
2016	Top student in Engineering 1st year, University of Canterbury, New Zealand
2015 - 2017	Undergraduate and College awards: Madam Tiong Guok Hua Memorial Prize; Emerging Leaders' Scholarship; Lyall Holmes Memorial Scholarship for academic excellence; George T Weston Scholarship for community contribution
2014	Top in New Zealand in CIE A Level Chemistry

Contact

Nationality:	New Zealand and Ireland
Address:	16 Seaview Road, Auckland 1050, New Zealand
Phone:	+64 21 1898177
email:	samckclarke@gmail.com

Master of Science in COMPUTER SCIENCE

Exam	Grade
Foundations of Computer Science	93
Computer-Aided Formal Verification	72
Computational Learning Theory	82
Computational Complexity	59
Advanced Machine Learning	75
Probability and Computing	83
MSc Thesis	65
AVERAGE	76

Maximum grade: 100. Distinction: 70-100

Visiting Student in PHILOSOPHY AND COMPUTER SCIENCE

Course	Grade
Kant and the Critique of Pure Reason	A
Computational Game Theory	A
Machine Learning	A
Logic & Proof	A
Computers in Society	not graded
Philosophy of Cognitive Science	A
Frege, Russell and Wittgenstein	A

Maximum grade: A

Bachelor of Science in PHILOSOPHY AND COMPUTER SCIENCE

Exam	Grade
Introduction to Computer Programming	A+
Mathematical Modelling and Computation	A+
Engineering Mathematics 2	A+
Engineering Mechanics	A+
French Culture and French Language	A+
God, Mind, and Freedom	A+
Cyberspace, Cyborgs, and the Meaning of Life	A+
Algorithms	A+
Computer Systems	A+
Principles of Electronics and Devices	A+
Stress, Strain and Deformation in Machine Elements	A+
Cognition	A+
Artificial Intelligence	A+
Alan Turing	A+
Paradoxes	A+
US Politics	A+
Computer Graphics	A+
Formal Languages and Compilers	A+
Philosophy of Religion	A+
Discrete Mathematics and Cryptography	A+
FINAL GPA	9.0/9.0

Maximum grade: A+

High School - CIE A Level

Exam	Grade
Mathematics	A*
French	A*
Physics	A*
Chemistry	A*
Further Mathematics	A

Maximum grade: A*