Academic Curriculum Vitae for

Hamish Haggerty

email: hamish.haggerty@uqconnect.edu.au
Ph: +61409845644

Academic Qualifications

Masters in Statistics 2023
University of New South Wales, Australia

Diploma of Arts 2015
University of Queensland, Australia

Extended Major	Mathematics

Bachelor of Arts 2012
University of Queensland, Australia

Extended Major	Psychology
Minor	Logic and Philosophy of Science
Additional courses	5 Mathematics electives

Academic Highlights

University of Queensland
GPA 6.75 / 7 for Mathematics courses at 3rd and 4th year level
Four Deans Commendation awards for high academic achievement

Australian National University
Summer Research Scholarship 2016

Project	Algebraic Topology and Homotopy Theory
Supervisor	Dr Vigleik Angeltveit

University of NSW
Neural Networks and Deep Learning: 88/100
Advanced Machine Learning: 88/100
Bayesian Inference and Computation: 84/100

Thesis: 2023

Project	*Self-Supervised Framework to Address Limited Data for Cancer Diagnosis*
Supervisor	Dr Rohitash Chandra
Grade	91/100
Github	https://github.com/hamish-haggerty/cancer-proj

WAM: 81.700

Publications:

Haggerty, H., & Chandra, R. (2023). *Self-Supervised Learning for Skin Cancer Diagnosis with Limited Training Data.* Under review at Computer Methods and Programs in Biomedicine.

- In my thesis work at UNSW I studied the problem of cancer image classification in the low labelled data regime. The standard technique in this regime is transfer learning, with the base network usually pretrained on ImageNet through supervised learning. An alternative approach is to pretrain the base network via self-supervised learning (which does not require labels). The basic results of my work are: i) in the low labelled data regime, self-supervised pretraining is superior to supervised pretraining; ii) self-supervised pretraining a *second* time on a large labelled dataset from the target dataset (e.g. unlabelled skin lesion images) before fine tuning can yield additional gains.

Research interests / goals

Until quite recently I was of the view that the most important area of AI to work on was medical applications. This is still an area of interest, and I hope to contribute in the future, particularly to AI applications to human longevity (shockingly still an underfunded area compared to other areas of medicine). My general view on this now though, is that the biggest breakthroughs (from the machine learning point of view) will come from training large foundation models, e.g. AlphaFold. There are also a large number of people working on AI applications to medicine.

On the other hand, it is now clear that AI may exceed human level in the near term. There are various ways this may pose an existential risk, with the most obvious being a mis-alignment between the AIs values and humans values, broadly construed. This can occur for at least two reasons: i) Through failure to specify the right goal, the paradigm example being a paperclip maximiser. ii) Through failure during training to *learn* the intended goal, and instead to learn a different objective not intended by the developers. Such a system is called a mesa-optimiser. A simple example is: a system is trained to reach the door in a maze. On the training distribution, all doors are red. On the test distribution, the doors are all blue, and there are other (non-door objects) that are red. If the system goes to red objects rather than doors. **(This part needs some work….).** i) Is called outer alignment and ii) is called inner alignment and both are unsolved problems at present. A potential solution to ii) is to ensure an AI is always truthful, although this also has subtleties associated with it. (e.g. an AI may be truthful at training but not test time).

AI mis-alignment is an existential risk. This is an area that: a) few people are working on, compared to e.g. capability research or narrow AI applications such as medicine; b) it appears possible for individuals or small groups (with limited compute budgets) to make outsized progress through both conceptual clarifications and empirical work. Some examples include: Hubinger et al conceptual work on clarifying inner vs outer alignment; work by Olah, Neel and others on mechanistic interpretability; work by Burns et al on making large language models "truthful."

Another paragraph or two on stuff here....

Employment

I am self employed via managing investments.

Additional education and skills

Programming Languages: Python (proficient), Pytorch (proficient), MATLAB (basic), R (basic)
Completed > 50 Project Euler problems (https://projecteuler.net/)
Audited Coursera Introduction to Machine Learning by Andrew Ng (grade: 100/100)

Tutoring

Private high school mathematics tutor 2013-2015

Additional interests

Philosophy of mind and the mind-body problem; Meditation; Ethics