AI/ML training BoF

Link back to the meeting agenda and notes

Panel members

Jingbo Wang (NCI)

Slava Kitaeff (Monash)

Patrick Tung (UNSW)

Mike Laverick (UoA)

Maxime Rio (NeSI / NIWA)

Mitchell Hargreaves (Monash / MLeRP)

Mark Crowe (QCIF)

Abdullah Shaikh (Intersect)

Oliver Cairncross (UQ)

Intersect Australia - Machine Learning Courses

Introduction to Machine Learning: Introduction & Linear Regression (1 day course)

→ Introduction
Supervised and Unsupervised
Linear Regression and Model Training
Machine Learning modelling workflows
Scikit-learn

Introduction to Machine Learning: Classification (1 day course)

→ Logistic Regression, Decision Trees and Ensemble Learning
Understand the Machine Learning modelling workflows.

Introduction to Machine Learning: Support Vector Machine & Unsupervised Learning (0.5 day course)

→ Support Vector Machine, K-Nearest Neighbor and Dimensionality Reduction

INTERSECT
Advanced Analytics & AI Platform

NeSI – New Zealand eScience Infrastructure

Our ML/DL workshops

- Audience: Aotearoa – NZ researchers
- ML 101
  - intro to Machine Learning – 6 hours
  - Jupyter notebooks, Scikit-learn based
  - started in 2021
  - 7 workshops (in person, online)
  - 127 attendees in total (from 10 to 32)
- ML 102
  - intro to Deep Learning (CNNs) – 3 hours
  - Jupyter notebooks, Tensorflow based
  - started in 2022
  - 2 workshops (online)
  - 44 attendees in total (20 and 24)

Challenges and Solutions

- cf. our talk (which was just before 😁)
- Find the right balance theory/practice
  - Reuse and adapt content
  - Hands-on and notebook based
- Minimise time for technical issues
  - Use of our JupyterHub platform
  - Have support helpers, use breakout rooms
- Ensure people show up to the workshop
  - EOIs and selection
  - Pre-workshop communication

UoA - University of Auckland

Gained funding to deliver first "Introduction to ML/DL" workshops this year

Carpentries-style workshop using modified-carpentries material
Delivered to UoA researchers

1st workshop (March - 4h x 5 days)

Mon - Intro to Python
Tues - Intro to Python
Wed - Intro to ML (Alpha)
Thur - Intro to DL (Beta)
Fri - Intro to DL (Beta)

2nd workshop (Sept - 4h x 4 days)

Mon - Intro to ML (Alpha)
Tues - Intro to ML (Alpha)
Wed - Intro to DL (Beta)
Thur - Intro to DL (Beta)
Fri - X

Feeding changes back into carpentry material (ML especially)

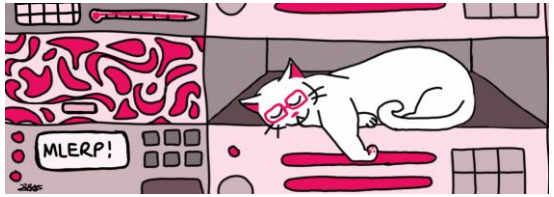
MONASH University

Identifying gaps

- Data Science degrees and courses (mostly FIT) provide a foundation for ML and AI, but not necessarily the development techniques or technologies.
- Science and Engineering HDR students needing ML/AI for their projects do not necessarily have the foundations, and not the technologies, which may limit the choice of methodologies in research projects.
- Efficient development of AI and development of advanced algorithms that use AI require in depth understanding of technologies that come in different shapes and colours.
- Non-technical researchers (e.g. HASS) are actively exploring AI and AI tools. The lack of foundational training might be a barrier to adoption or present a risk of incorrect assumptions or expectations from the tools using one of another methods.

Monash

- Monash Data Fluency Program
    - Python, Command Line, Intro HPC
- NCI, DUG, other external training (not AI/ML)
- Monash eResearch - started thinking about how best we can help researchers (probably HDR students mostly)
- Hope - we train some users, and they become our future trainers.



File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ··· / environments / dsks /

Name	Last Modified
dsks_2023....	1 hour ago
dsks_2023....	1 hour ago

2_dask_pytorch.ipynb

Raw

Python 3 (ipykernel)

```
---
title: PyTorch and Dask
format:
  html:
    code-fold: false
  ipynb: default
execute:
    freeze: true
format-links: [ipynb]
---
```

Defining models, datasets and functions

If you're doing something relatively simple, Dask has integrations with Scikit-Learn and XGBoost. You can also pass PyTorch models into Scikit-Learn with Skorch and TensorFlow models with SciKeras.

But if you need to do something more complex, Dask clusters can have python functions submitted to them to remotely execute code. This gives us the low level control to implement whatever bespoke algorithm we want and have it accelerated by remote GPUs.

In this example we're going to write our own PyTorch functions to train a custom model on the CIFAR dataset. While we could do this with Skorch, we hope that this example gives you some idea of how Dask can be flexible enough for any applications that you need.

Content adapted from the PyTorch CIFAR10 Tutorial

MLeRP

Select a site

mhar0048@mlerp-qcif

mhar0048@mlerp-monash

- Account Info
- Jupyter Lab
- DSKS Jupyter Lab
- Terminal

INSTRUCTIONS/DESCRIPTION

This app will allow you to run an instance of Jupyter Lab using one of our provided conda environments, or one that you provide. If you're looking for an envionment that's not listed here, check that its path is included in the **.conda/environments.txt** file in your home directory.

If you're new to the service, we reccommend the DSKS environment which contains many commonly used datascience and machine learning libraries so that you can get started without needing to make your own. However, if you need specialised packages which aren't available in our provided environments, see our Custom Conda Environments documentation for more details.

LAUNCH JUPYTER LAB

Time (hours)
1

Launch

Service*
Small Lion

Conda Environment*
dsks_2023.10 | /apps/mambaforge/envs/dsks_2023.10

QCIF ARDC Nectar Research Cloud

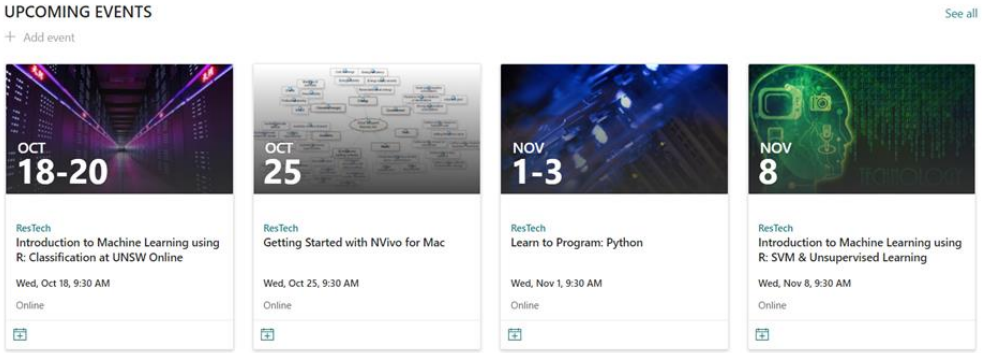
The University Of Queensland AUSTRALIA · MONASH University · ARDC Australian Research Data Commons

NCI's ML course offering (2022-2023)

AI/ML Applications on Gadi - Computational Biology (1 day)	66
AI/ML Applications on Gadi - Astronomy (1 day)	66
AI/ML Applications on Gadi - Natural Language Processing (1 day)	80
AI/ML Applications on Gadi - Materials Science (1 day)	24
NCI RMIT HPC/AI Training Day (2-3 days)	107
NCI UNSW HPC/AI Training Day (2-3 days)	40
NCI Monash HPC/AI Training Day (2-3 days)	70
NCI AAPP/UTAS HPC/AI Training Day (2-3 days)	54
AI for Science Bootcamp (with NVIDIA) (2-3 days)	165
AI hackathon - two weeks intensive coding (with NVIDIA)	13

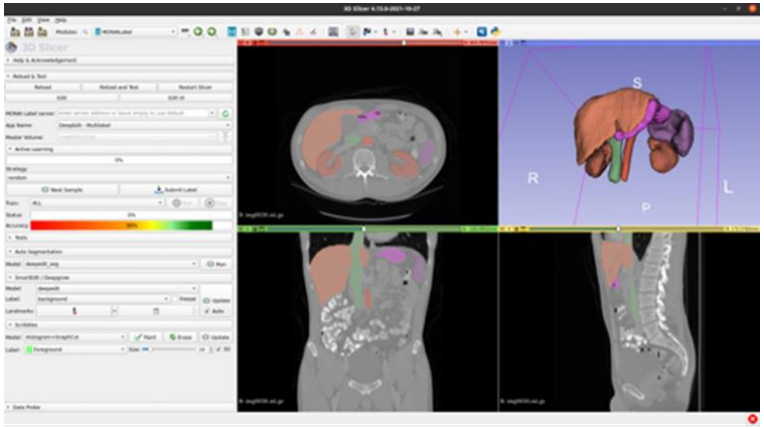
UNSW Research Technology Services (ResTech)

Training courses



UNSW AI Institute

AI imaging collaborations



Affiliated groups

Affiliated Group	Academic Contact	Associated Faculty
AI for characterisation of solar cells Research Group	Dr Priya Dwivedi	Engineering
Algorithmic Decision Theory Research Group	Prof Toby Walsh	Engineering
ANFF-NSW (Australian National Fabrication Facility)	Prof Andrew Dzurak	MWAC
ARC Centre of Excellence for Automated Decision-Making and Society	Prof Jackie Leach Scully	ADA
ARC Centre of Excellence in Nanoscale Biophotonics	Prof Ewa Goldys	Engineering
ARC Research Hub for Connected Sensors for Health	Prof Chun Hui Wang	Engineering
ARC Research Hub for Resilient and Intelligent Infrastructure Systems (RIIS)	Prof Nasser Khalili	Engineering
ARC Training Centre for Next-Gen Architectural Manufacturing	A/Prof Matthias Hank Haeusler	ADA
Australasian Legal Information Institute (AustLII)	Dr Philip Chung	Law & Justice
Business School AI Hub	Prof Mary-Ann Williams	Business School
Centre for Big Data Research in Health	Prof Louisa Jorm	Medicine and Health

1

Training Resource Gap:

- Lack of skilled instructors
- Training materials - develop and update frequently
- Fit for diverse and individual needs
- Lack of dedicated funding to pay the course development and delivery

2

Community efforts

- Mentorship and support
- Collaboration and partnership
- Industry support and opportunities
- Measuring Success and Impact

3

Educational initiative for underrepresented disciplines

- Share examples of programs and organizations that promote AI education among underrepresented groups
- Discuss their impact and scalability

4

Measure success, impact which leads to opportunities

- Evaluate the effectiveness of AI skill training programs, especially in terms of inclusivity.
- Highlight success stories and impact metrics