Training for postgraduate research students

Gert Aarts

JENAS, Madrid, May 2022

Introduction

- over the years the expectation on training for postgraduate research students has changed substantially

- from attending PhD schools and seminars

  → detailed annual reporting on cohort and individual activities

Disclaimer:

- my experience is mostly based on the UK system
- in the UK doctoral researchers are postgraduate research (PGR) **students**
- Research Councils play a major role in shaping the training structure

Outline

- brief outline of Centres for Doctoral Training (CDTs) in the UK
- example: cohort training across disciplines
- support and infrastructure, sustainability
- take-away message and questions

UK Centres for Doctoral Training

- each CDT: 50+ PhD positions over 5 cohorts
- fully funded by Research Council: stipend and fee for 4 year PhD positions
- funding for Centre delivery costs (essential!): manager, administrator, events, …
- may involve multiple institutions and disciplines
- match funding from universities and external partners required

Example: AIMLAC

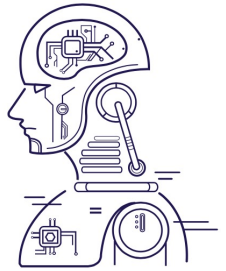
UKRI CDT in Artificial Intelligence, Machine Learning and Advanced Computing

(of which I am the Director)

cdt-aimlac.org

AIMLAC

UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING



AIMLAC
UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING

AIMLAC in short

- 5 universities (4 in Wales, 1 in England) + Supercomputing Wales (infrastructure)
- 3 research themes:
  - ✓ data from large science facilities (particle physics, astronomy, cosmology)
  - ✓ biological, health and clinical sciences
  - ✓ novel mathematical, physical, and computer science approaches

Prifysgol Aberystwyth University · Prifysgol Bangor University · University of Bristol · Cardiff University / Prifysgol Caerdydd · Swansea University / Prifysgol Abertawe · Supercomputing Wales / Uwchgyfrifiadura Cymru

2021 cohort

Name	University	Project Title	Theme	Supervisor(s)
Myles Clayton	Aberystwyth	A deep learning framework for agricultural plant breeding that predicts genotype-phenotype associations	T2	Martin Swain, Chuan Lu
Ong Ding Sheng	Aberystwyth	Few-shot Learning for Environment Adaptive Multi-modal Vision System	T3	Jungong Han
Leena Sarah Farhat	Bangor	Bringing big-data to social science	T3	Simon Willcock, William Teahan
Dan Farmer	Bangor	Edge-based object recognition for immersive analytics in Web-based XR	T3	Panagiotis (Panos) Ritsos
Sam Hennessey	Bangor	Ensembles of Deep Neural Networks for Semi-supervised Learning	T3	Lucy Kuncheva
Fergus Baker	Bristol	Machine learning to study accretion flows around black holes	T1	Andy Young
Laura Ballisat	Bristol	Advanced computational methods for dosimetry, planning and verification in emergent radiotherapy treatments	T1, T2	Jaap Velthuis, Richard Hugtenburg (Swansea)
Matthew Powell	Cardiff	Real-time Situational Understanding using Deep Neural Networks and Knowledge Graphs	T3	Alun Preece
Zara Siddique	Cardiff	Evolving Ethical Deep Neural Networks	T3	Roger Whitaker
Luke Golby	Swansea	AI based approaches multi-dimensional functional genomics	T2	Steve Conlan
Tabitha Lewis	Swansea	ML-guided dynamical systems modelling of sepsis	T2, T3	Noemi Picco
Shobhna Singh*	Cardiff	Dimer models on quasicrystals	T1	Felix Flicker

*Associate member

First-Year Students

2020 cohort

Name	University	Project Title	Theme	Supervisor(s)
Luke Ian Lunn	Aberystwyth	Approximating the colour of Mars	T1, T3	Helen Miles
Bishnu Paduel	Aberystwyth	Automatic stroke recovery prediction using artificial intelligence	T2	Otar Akanyeti, Reyer Zwiggelaar
Will Robinson	Aberystwyth	Detecting when deep learning goes wrong in medical image analysis	T2	Bernie Tiddeman, Reyer Zwiggelaar
Franciszek Krzyzowski	Bangor	Learning from badly behaving data	T3	Lucy Kuncheva, Franck Vidal
Iwan Mitchell	Bangor	Automated optimisation of industrial X-ray computed tomography	T3	Franck Vidal, Simon Middleburgh
Jake Amey	Bristol	New Physics searches in B and D meson decays with machine learning	T1	Jonas Rademacker, Konstantinos Petridis
Matthew Selwood	Bristol	Using machine learning to explore the evolution of active galaxies with Euclid	T1	Sotiria Fotopoulou, Malcolm Bremer
Drew Barratt	Cardiff	Examination of SARS-CoV-2 severity, transmissibility and spread within Wales through the analysis of linked patient health records and genomic sequence data	T3	Tom Connor
Matthew Walker	Cardiff	Inferring brain tissue microstructure from standard structural imaging	T2	Leandro Beltrachini, Kevin Murphy
Samuel Wincott	Cardiff	AI and neuro-evolution: Exploiting network motifs to enhance prediction of contagion in complex networks	T3	Roger Whitaker, Alun Preece
Natalia Sikora	Swansea	Enhancing the diagnostic performance of a bowel cancer blood test using advanced machine learning algorithms and the incorporation of information from the patient's medical record	T2	Peter Dunstan, Dean Harris
Lukas Golino	Swansea	Machine learning with anti-hydrogen	T1	Niels Madsen, Gert Aarts
Maciej Glowacki*	Bristol	Searches for Beyond-Standard-Model signatures with jets + missing energy	T1	Henning Flaecher
Jacob Elford*	Cardiff	Monsters in the dark: gas, dust and star formation around supermassive black holes	T1	Timothy A. Davis, Mattia Negrello
David Mason*	Swansea	Non-perturbative dynamics and compositeness	T1	Biagio Lucini, Maurizio Piai
Jack Furby**	Cardiff	Human-machine collaboration with deep learning agents	T3	Alun Preece
Paul Murphy**	Cardiff	Adaptive neural networks through epigenetic processes	T3	Roger Whitaker
Ben Page**	Swansea	Studies of thermal QCD using lattice gauge theory	T1	Chris Allton

*STFC CDT on Data-Intensive Science
**Associate member

2019 cohort

Name	University	Project Title	Theme	Supervisor(s)
Lily Major	Aberystwyth	Big Data algorithmics for efficient search and analysis of large collections of genomes	T2	Amanda Clare, Jacqueline Daykin, Benjamin Mora, Christine Zarges
Cory Thomas	Aberystwyth	Modelling the development of breast cancer abnormalities	T2, T3	Reyer Zwiggelaar, Tom Tornsey-Weir, Jason Xie
Vanessa Cassidy	Bangor	Machine learning for narrative data visualisation	T3	Jonathan Roberts, Panos Ritsos
Benjamin Winter	Bangor	The research of neuroevolution algorithms	T3	William Teahan, Franck Vidal
Hattie Stewart	Bristol	AI techniques for extracting source information from Square Kilometre Array (SKA) datasets	T1	Mark Birkingshaw
Robbie Webbe	Bristol	X-Ray Astronomy, concerning the identification and classification of highly variable AGN	T1	Andy Young
Christopher Wright	Bristol	Multi-channel waveform reconstruction for dark matter searches with LUX-ZEPLIN	T1	Henning Flaecher, Stephen Fairhurst
Michael Norman	Cardiff	Deep learning for real-time gravitational wave detection	T1	Patrick Sutton
Bradley Ward	Cardiff	Investigating the epoch of galaxy formation using artificial intelligence	T1	Steve Eales
Tonicha Crook	Swansea	Game theory	T3	Arno Pauly, Edwin Beggs
Jamie Duell	Swansea	Machine learning in medical science	T2	Xiuyi Fan, Shangming Zhou, Gert Aarts
Sophie Sadler	Swansea	Visual analytics for explainable graph-based machine learning	T3	Daniel Archambault, Mike Edwards
Raul Stein*	Bristol	FPGA implementation of machine learning for low latency data processing in particle detectors	T1	Jim Brooke
Eleonora Parrag*	Cardiff	Rewinding supernovae with machine learning	T1	Cosimo Inserra
Thomas Spriggs*	Swansea	Spectral features of hadronic states in thermal QCD	T1	Chris Allton, Tim Burns

*STFC CDT on Data-Intensive Science

Cohort training

- expectation on training goes (far) beyond usual activities
- develop common base, irrespective of research area
- develop cohort “feel”

implementation:

- 5 cohort events and 2 intra-cohort events each academic year
- typically 3 days/2 nights at one of the partner institutions
- emphasis on transferable skills/computing/best practice
- ongoing dialogue with students/Student Board/Management Board

AIMLAC
UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING

Cohort training delivery

- investment required to deliver training
  - → time (# events/year) and repetition (# cohorts) and know-how
- orthogonal to one-off PhD school

our solution:

- Research Software Engineers (RSEs)
- employed by Supercomputing Wales and Swansea Academy for Advanced Computing
- fully integrated in CDT management and delivery

some examples of RSE training activities

available online under open-source licenses

cdt-aimlac.org/cdt-resources.html

Software Carpentry

- Introduction to the Unix Shell: swcarpentry.github.io/shell-novice
- Introduction to programming with Python: swcarpentry.github.io/python-novice-inflammation
- Introduction to Version Control with Git: swcarpentry.github.io/git-novice

Advanced introductory topics

- Performant Numpy: edbennett.github.io/performant-numpy
- Git: Beyond the Basics: sa2c.github.io/git-demystified/

High Performance Computing

- Introduction to High Performance Computing with Supercomputing Wales: supercomputingwales.github.io/SCW-tutorial

Reproducible environments and containers

- Introduction to reproducible environments with Binder: zenodo.org/record/2598530
- Reproducible computational environments using containers: carpentries-incubator.github.io/docker-introduction

Automated testing and continuous integration in Python

- Automated testing and continuous integration: https://edbennett.github.io/python-testing-ci/

Object-Oriented Programming

- Introduction to Object-Oriented Programming in Python: edbennett.github.io/python-oop-novice

Data Management with SQL

- Data Management with SQL: https://datacarpentry.org/sql-ecology-lesson

How the Web works

- Introduction to the Web and Online APIs: https://edbennett.github.io/web-novice/

Online training delivery

- pandemic forced all of us to redesign delivery
- “hands-on” delivery feasible using zoom breakout rooms, screen sharing, etc

example:

- TALENT school “Machine Learning applied to Nuclear Physics, Experiment and Theory”
- 2-week online intensive course: 20h of lectures, 10h of exercises
- run at ECT* by Morten Hjorth-Jensen in 2020 and 2021
- www.ectstar.eu/talents/talent-school-on-machine-learning/

Online training portfolio

- collate online courses with annotations, e.g. beginner/intermediate/advanced
- make available to wider community via central point of access

example proposal (Maria Paola Lombardo):

- LaVA: Lattice Virtual Academy
- hub for lattice field theory online resources
- in development as part of STRONG-2020
- hosted at ECT*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824093

Experiences

- training across disciplines is stimulating for students and others involved
- enhances PhD experience, silos are removed
- not one-off: ongoing training throughout PhD programme shapes cohort
- exposed to many clever peers in other research domains: broadens horizon
- communication and cooperation outside one's own domain essential for e.g.
  - fellowship applications
  - changing disciplines
  - knowledge transfer
  - employability
  - …

- but ...

Investment

dedicated resources are required to:

- deliver training across disciplines (not owned by a specific department/unit)
- deliver multiple training events/year
- repeat training for cohorts
- make online resources accessible in coordinated manner

Open questions

substantial investment in undergraduate teaching and student experience

- should PGR student training be considered in a similar way?

which aspects of PGR training should be organised by

- universities?
- across (national) institutions?
- international facilities?

who is responsible for sustainability?