# Workshop on Korea-UK AI/ML Research in Fundamental Sciences

Gert Aarts (ECT* & Swansea University)

# Meetings

*Artificial Intelligence & Machine Learning for Fundamental Science*

- Nov 15-17, 2021 (online)
- pyweb.swan.ac.uk/~aarts/ai-uk-korea.html

*2022* AIMLAC *Artificial Intelligence Conference*

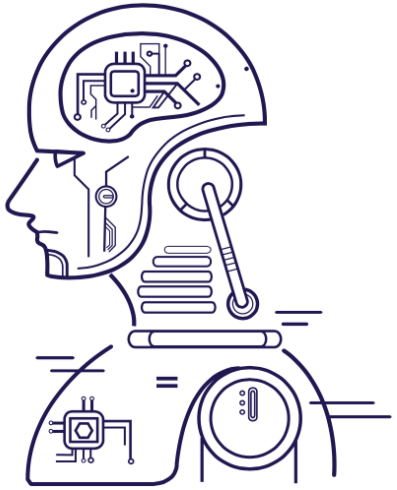
- June 23-24, 2022 (Swansea University)
- cdt-aimlac.org/cdt-events.html

*Workshop on Korea-UK AI/ML Research in Fundamental Sciences*

- Oct 31- Nov 4, 2022 (Sejong University)
- indico.cern.ch/event/1203114/

# AIMLAC

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



PRIFYSGOL ABERYSTWYTH UNIVERSITY

PRIFYSGOL BANGOR UNIVERSITY

University of BRISTOL

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Swansea University
Prifysgol Abertawe

SUPERCOMPUTING WALES
UWCHGYFRIFIADURA CYMRU

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

# UKRI Centres for Doctoral Training

- UKRI launched a call for proposals for **Centres for Doctoral Training** in **AI** in 2018
- each CDT will train about 50-55 students across 5 cohorts
- fully funded, 4-year PhD positions, with external partner engagement
- 16 UKRI CDTs in AI were funded, across a wide range of topics

**Table 4**
Focus areas of the 16 funded Centres for Doctoral Training in the 2018 UKRI AI CDT funding call.

| Focus area | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Healthcare, biomedical | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | | | | | | |
| Responsible AI, human | | | | | | ✓ | ✓ | ✓ | ✓ | | | | | | | |
| Core AI research | | | | | ✓ | | | | | ✓ | | | | | | |
| Fundamental science | | | | | ✓ | | | | | | | | | | | |
| Language processing | | | | | | | | | | | ✓ | ✓ | | | | |
| Environment, sustainability | | | | | | | | | | | | | ✓ | ✓ | | |
| Engineering, nano-devices | | | | | | | | | | | | | | | ✓ | |
| Creative industries, music | | | | | | | | | | | | | | | | ✓ |

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

# Our proposal

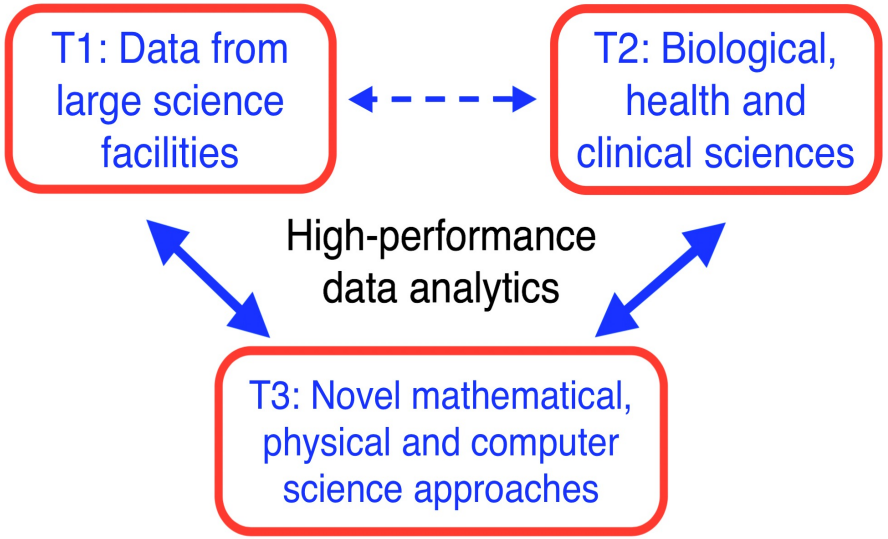
## UKRI CDT in Artificial Intelligence, Machine Learning and Advanced Computing

- 5 universities: 4 in Wales + Bristol
- Supercomputing Wales provides essential contributions, especially in training
- built on experience of STFC CDT in Data-Intensive Science (Cardiff, Bristol, Swansea)
- first cohort started in October 2019, recruitment of cohort 4 is complete (Oct 2022 start)

cdt-aimlac.org

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

# AIMLAC research themes




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

# 2022 cohort

| Name | University | Project Title | Theme | Supervisor(s) |
|---|---|---|---|---|
| Michael Casaletto | Aberystwyth | Prediction of facial growth for children with cleft lip and palate using 3D data mining and machine learning | T2, T3 | Richard Jensen |
| Luke Williams | Aberystwyth | Collaborative mapping of large scale outdoor environments | T3 | Myra Wilson |
| Jamie Pointon | Bangor | X-ray simulation and deep learning: Application to Automatic segmentation of defects in CT images corrupted by artefacts | T3 | Franck Vidal |
| Preben Vangberg | Bangor | Automatically Analysing Big Language Data | T3 | William Teahan |
| Rhys Shaw | Bristol | Machine learning and radio source multiplicity | T1 | Mark Birkinshaw |
| Tanya Kushwahaa | Cardiff | Exploiting GAIA data and understanding the galaxies' past histories with machine learning | T1 | Mikako Matsuura |
| Sama Al-Shammari | Cardiff | Simulation-based Inference of gravitational waves signals from black holes and neutron stars | T1 | Vivien Raymond |
| Chanju Park | Swansea | Learning (from) lattice field theory | T1 | Gert Aarts, Biagio Lucini |
| Vasiles Balabanis | Swansea | Multimodal analysis of Anatomical and Functional features to enhance the understanding of Brain Processing Phenomena: A Machine Learning Approach. | T2, T3 | Scott Yang |

First-Year Students

# 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

# 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 |
| 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 aspects

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

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
- delivery by **Research Software Engineers** (RSEs), supported by Supercomputing Wales and Swansea Academy for Advanced Computing
- ongoing dialogue with students/Student Board/Management Board

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/

# Cohort aspects

# DataAid

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

- student-led activity: Data Aid
- unites the analytical skills and expertise of our students with charities in need of data support
- generate valuable new insights and perspectives on the charity's data
- annual 2-day event to analyse the data and provide feedback to charities

✓ 2021: The Fairtrade Foundation, The Diana Award and Chance to Shine

✓ 2022: Go Beyond, Carers Trust, SERICC

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

# External partner engagement

- participate in cohort events
- contribute to training activities
- **six-months placements for 2nd/3rd year students**

  - ✓ emphasis on skills, employability, not on PhD research project
  - ✓ partner: access to highly skilled PhD researcher
  - ✓ student: gain experience in real-world data science employment
  - ✓ for details: cdt-aimlac@swansea.ac.uk or cdt-aimlac.org

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

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

# Scientific Research

all students carry out research and will obtain (hopefully!) a PhD

AI/ML applications in:

- theoretical/experimental particle physics
- astronomy/astrophysics
- health and medicine
- computer science/maths

# Questions?