



UNIVERSITY OF LEEDS

HPC in a world of Data Science

Martin Callaghan
Research Computing Manager
University of Leeds



What I'm going to talk about:

20 minutes or so:

- What HPC is (again...)
- What it *used* to be used for
- How we manage it
- Research portfolio and challenges from AI and Data Science



Whoami?

- HPC/ Research Computing is my 5th career:
 - Engineer > Army > Lecturer > Consultant > Lecturer
- Research Computing Manager at the University of Leeds
 - Manage two teams of Research Software Engineers
 - Research Infrastructure Engineers
 - HPC
 - Run the postgrad computational skills training programme
 - Everything from good programming practice to Scientific Programming to Machine Learning and AI
- ‘Jack of all trades’
 - BUT research in Deep Learning for text summarisation



HPC at Leeds

What is HPC- High Performance Computing?

Lots of potential definitions, but for us it is:

- Providing a hardware platform to facilitate large scale parallel jobs
 - Managed by a batch scheduler
 - Running a bunch of Open Source Linux tools, applications and programming languages
- Providing **people** in the form of Research Software Engineers and Research Infrastructure Engineers to:
 - Support researchers in writing better code
 - Using ‘modern’ software development practices
 - Acting as the interface between the research, code and infrastructure



Well, it was...

HPC was happiest in a world of Physics and Engineering simulations:

- Molecular Dynamics
- Climate simulations
- Computational Fluid Dynamics
- All that *Computational Science and Engineering* Julian and Shane mentioned earlier...

Relatively little data, but models that would benefit from parallelism in some form or another.



HPC at Leeds

We have two ~7000 Intel core clusters

- Bunch of P100 and V100 GPUs on each
- ~1PB parallel disk storage on each
- ~1200 research users
- Supporting ~£50M research programmes plus lots of postgrad teaching and research

(Other more specialized research computing platforms, eg. For sensitive data)

All CentOS Linux, applications managed through environment modules and Gridengine scheduler.



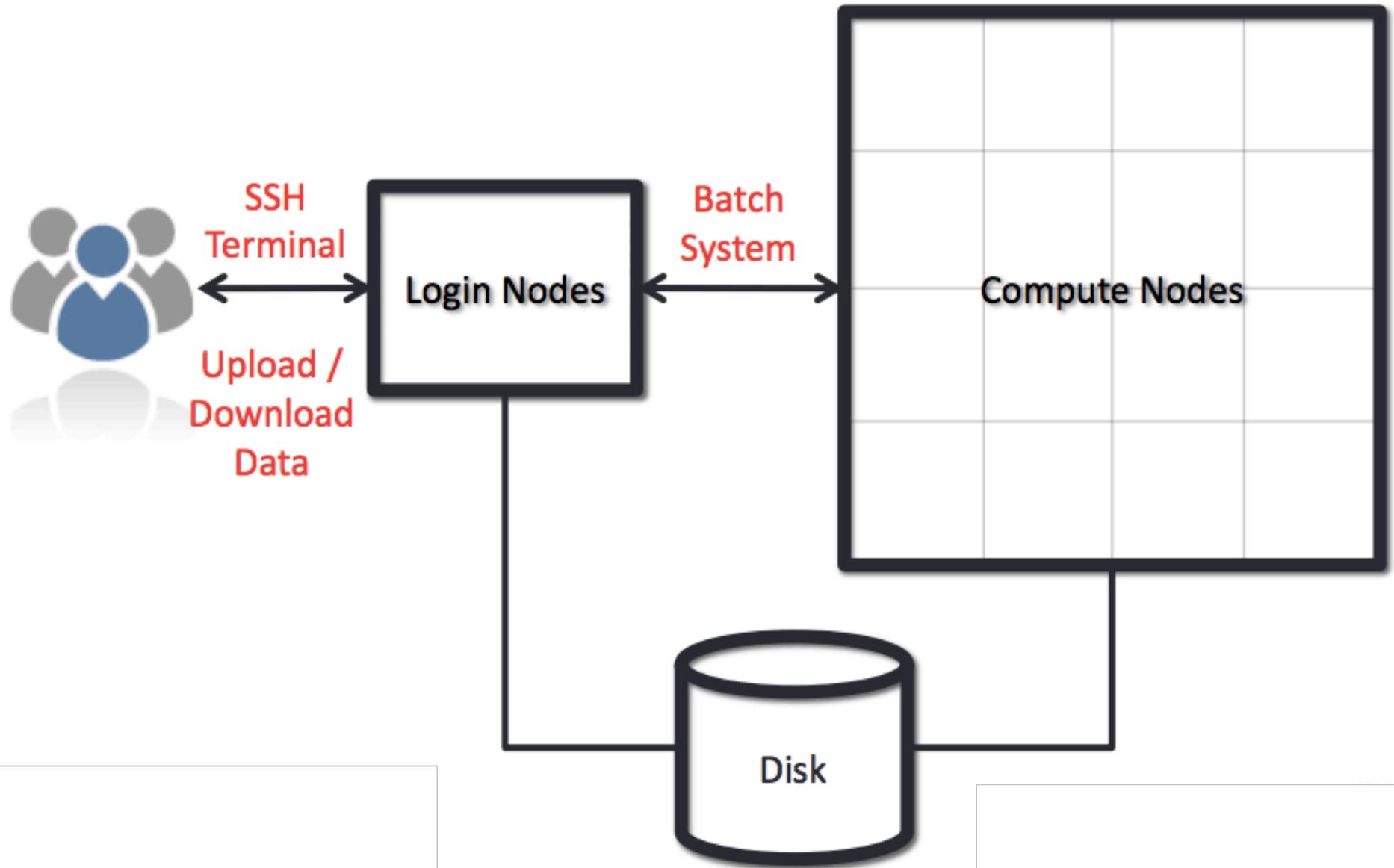
UK HPC landscape

A large number of research intensive Universities run their own HPC machines like ours at Leeds (we call these **Tier 3** machines)

In addition, research funders support a number of *regional* or *domain specific* machines (at **Tier 2**)

And a national service called *Archer* (at **Tier 1**)

With access to other machines across Europe (at **Tier 0**).





Then along came Data Science

By Data Science I mean anything from Bioinformatics to image analysis.

What do they have in common?

- Not quite as parallel any more:
 - Lots of Python and R (as opposed to C/C++ and Fortran)
 - Lots (*lots*) more data
 - Lots more in memory processing (Spark etc.)
- Different hardware requirements
 - Deep Learning > needs GPUs



UNIVERSITY OF LEEDS

AI at Leeds

Over 50 AI researchers in Computer Science

UKRI Centre for Doctoral Training in AI for Medical Diagnosis and Care

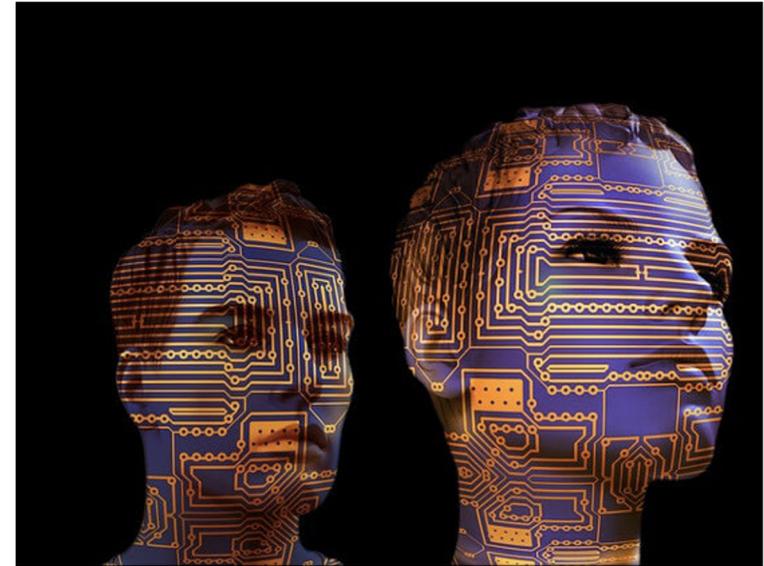
- 50 PhD students, £5.8m

Northern Pathology Imaging Co-operative

- £17m from UKRI and industry

Partner in The Alan Turing Institute

EPSRC National Facility for Innovative Robotics Systems



**The
Alan Turing
Institute**

NHS
The Leeds
Teaching Hospitals
NHS Trust



AI in science and engineering

Environmental sciences

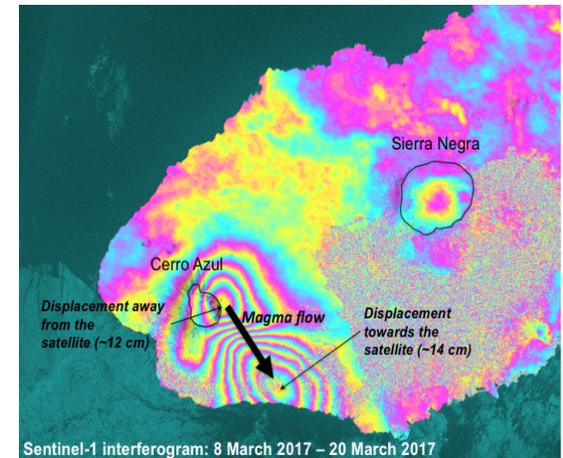
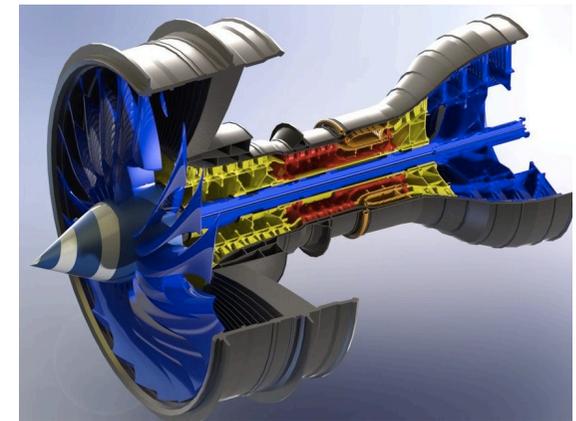
Biological sciences

Engineering design

Healthcare

Process engineering

Molecular modelling





Working with multi-modal data

| Recipe | |
|------------------|---|
| @type | Recipe |
| name | Strawberry-Mango Mesclun Recipe |
| image | http://images.media-allrecipes.com/userphotos/600x600/1116471.jpg |
| datePublished | 2008-03-03 |
| description | Mango, strawberries, and sweetened dried cranberries are a vibrant addition to mixed greens tossed with an oil and balsamic vinegar dressing. |
| prepTime | PT15M |
| totalTime | PT14M |
| recipeYield | 12 servings |
| recipeIngredient | 1/2 cup sugar |

Structured data



Video

WUN is committed to investing in research that makes a difference to the world. We bring the researchers and resources of our world leading member universities together to develop solutions that no one university or nation can solve alone. Our international, interdisciplinary research teams are collaborating with universities, research centres, civil society, government and industry to solve some of societies most vexing concerns.

Text



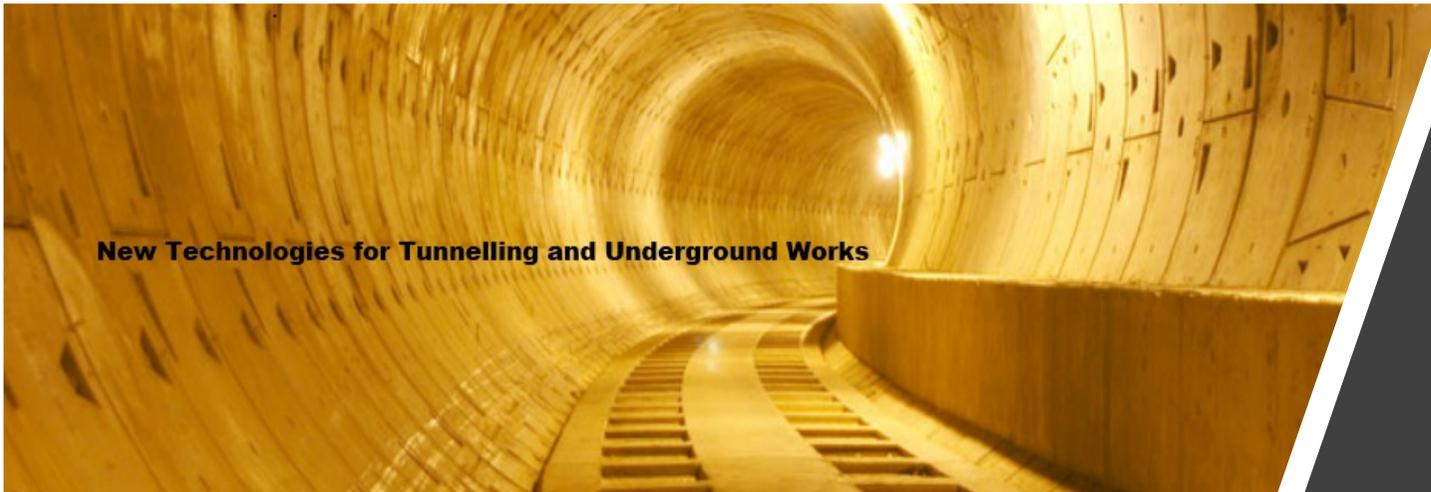
Genomic data



Images



Qualitative representations and reasoning about underground cables, pipes and tunnels

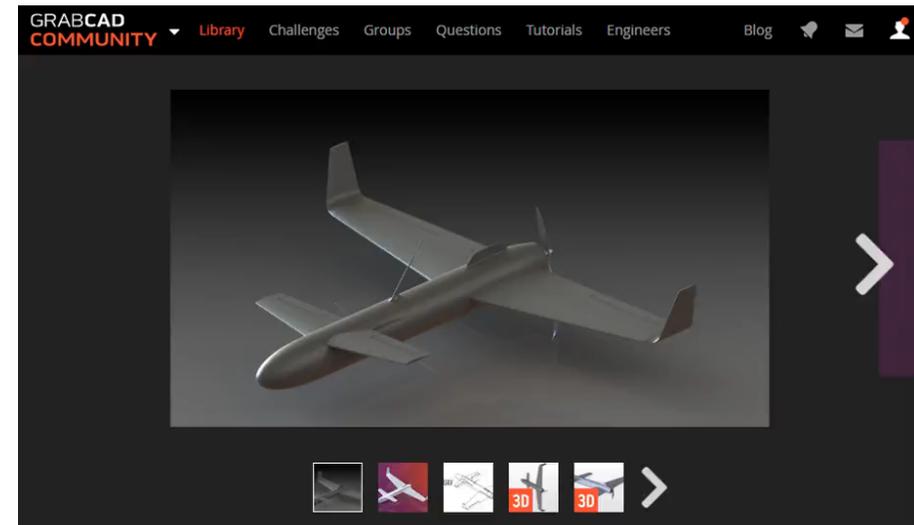
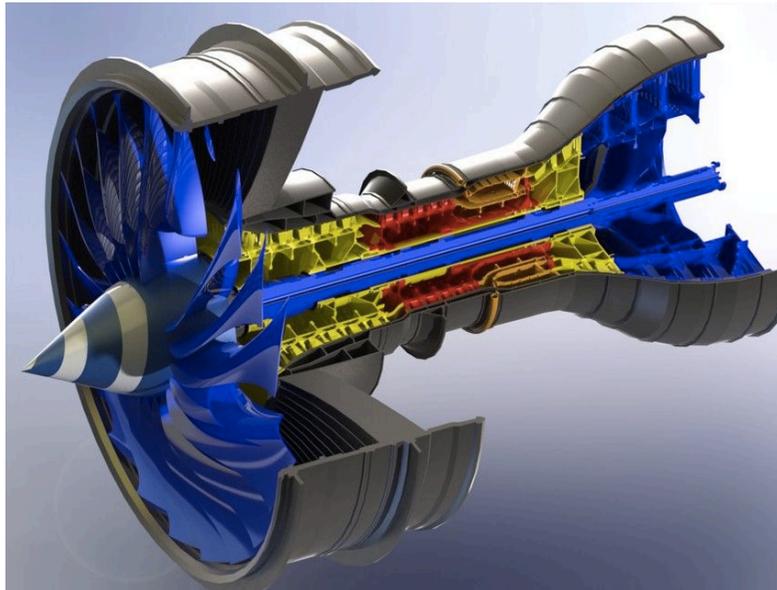


New Technologies for Tunnelling and Underground Works

Engineering design



UNIVERSITY OF LEEDS



The CAD files and renderings posted to this website are created, uploaded and managed by third-party community members. This content and associated text is in no way sponsored by or affiliated with any company, organization, or real-world good that it may purport to portray.

Griffin-60 UAV

 **Ik Ik**
August 10th, 2017

Design by Kavian Niazi

[Download files](#)

[Like](#) [Share](#)

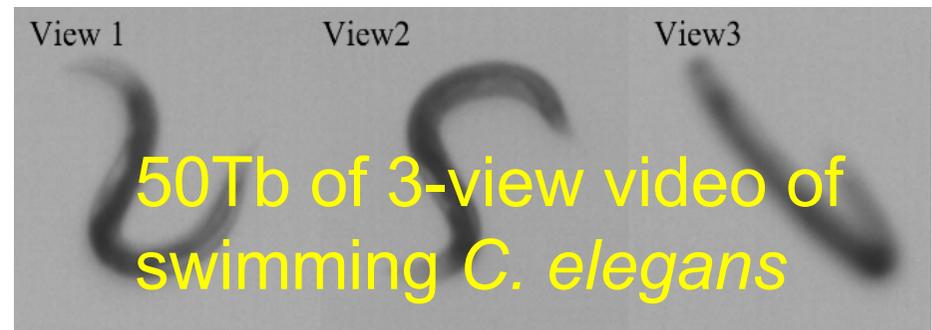
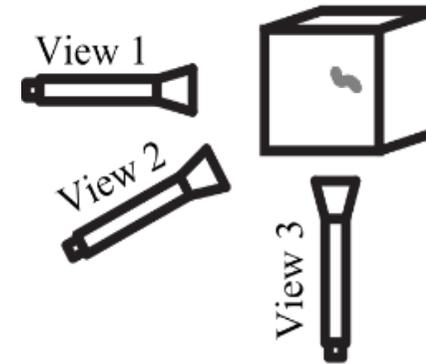
[480 Downloads](#) [59 Likes](#) [Comments](#)

Files (6)



Science discovery

How do soft bodies generate movement through deformation?



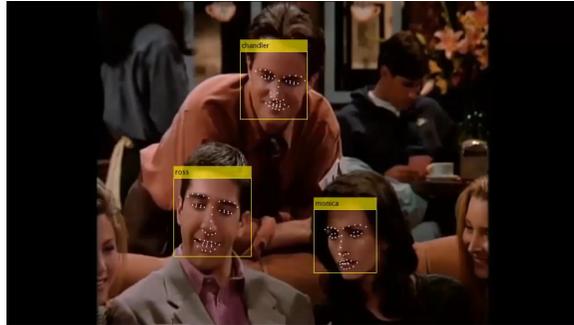
Digital humans



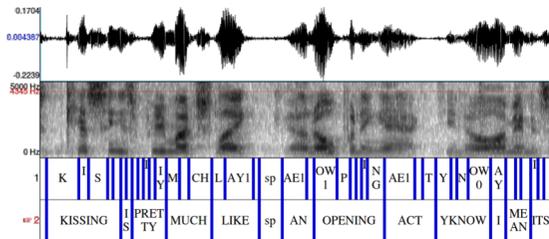
UNIVERSITY OF LEEDS



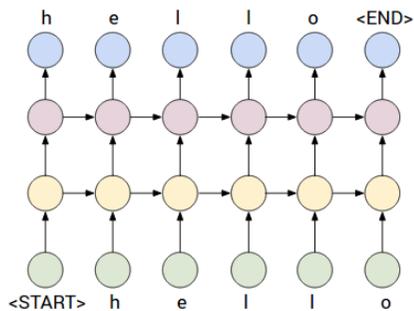
Warner Bros. Television and
Bright/Kauffman/Crane Productions:
Friends, seasons 1-10 (1994-2004)



facial
appearance



voice



language

Character model



Anomaly detection



UNIVERSITY OF LEEDS



Carried bag detection



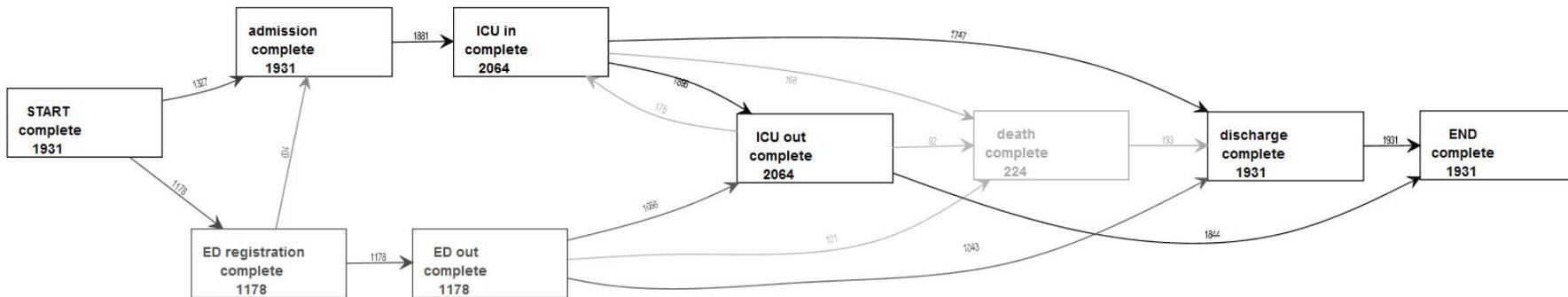
UNIVERSITY OF LEEDS





Understanding patient pathways

UNIVERSITY OF LEEDS



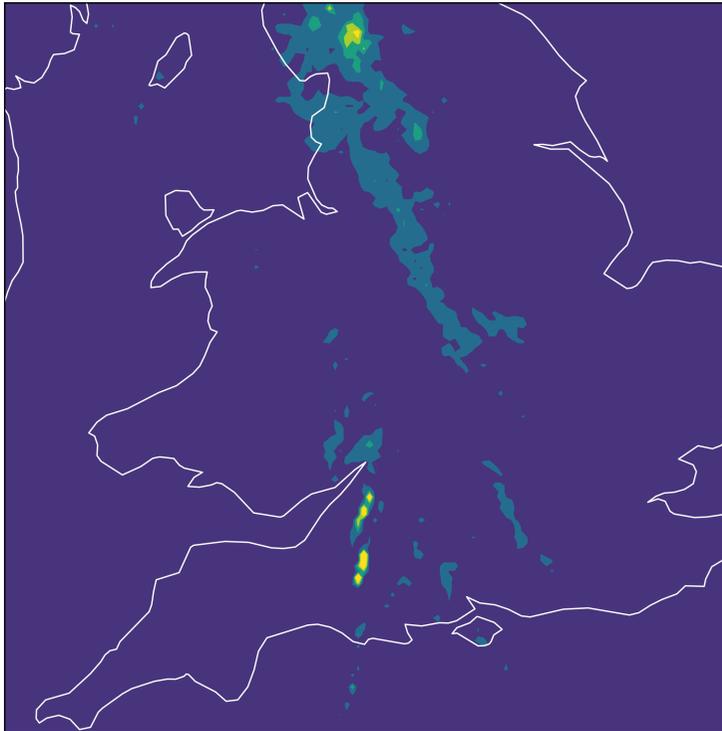
Rainfall forecasting



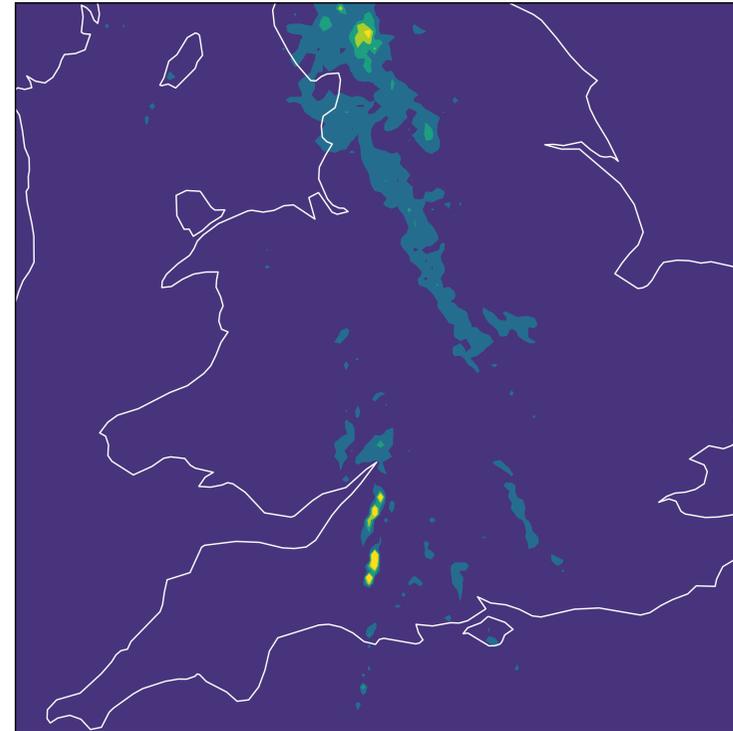
UNIVERSITY OF LEEDS

T+00

Radar rain rate



Neural network prediction

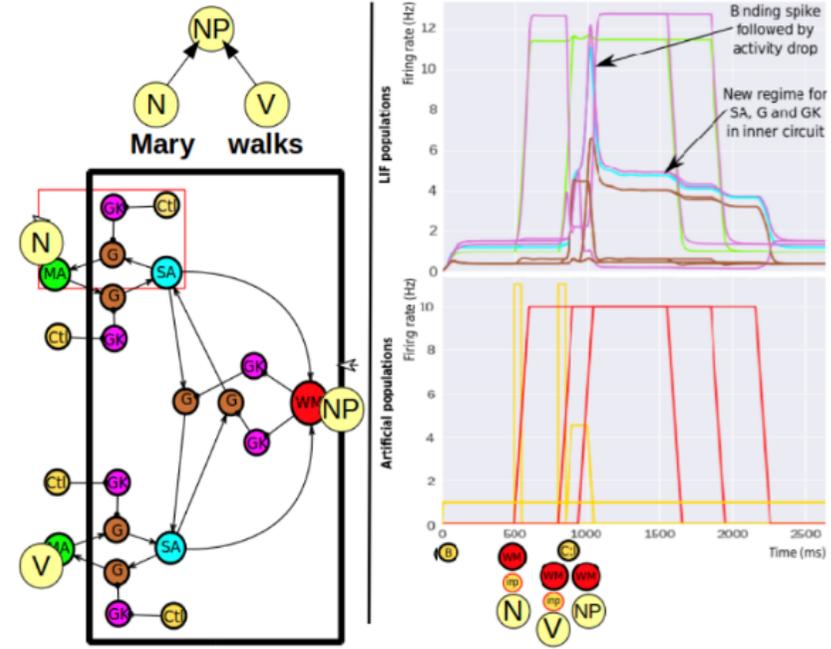


3000 frames, 32 hours to train



Human Brain Project

Binding related neural activity



Neural mechanisms for vision and language



Augmenting Human Intelligence



Helping people take decisions, expand their knowledge, learn from experience, and develop self-regulation skills.



Challenges for Research Computing

- Up to now, infrastructure support has been via HPC platforms with GPU nodes and ad-hoc desktops with a massive variety of consumer-grade GPUs
- A new 'AI for Personalised Medicine' CDT started this September- loads more demand for Deep Learning projects.
- We need more GPUs – and access to other tools – quickly
- **BIG** challenge: Deep Learning with sensitive data