The Alan Turing Institute

Collaborative Computing: JADE's Impact at the Alan Turing Institute

David Llewellyn-Jones, Tomas Lazauskas, Iain Stenson

The Alan Turing Institute

Founded in 2015

Purpose

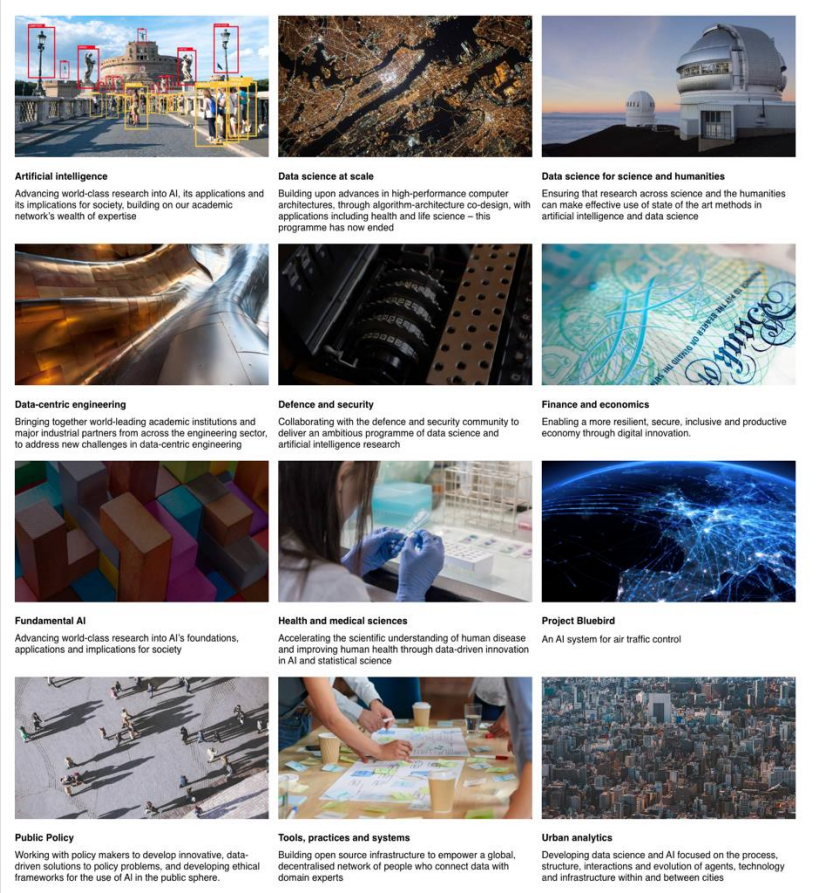
To make great leaps in the development and use of data science and AI in order to change the world for the better.

Goals

- Advance world-class research and apply it to national and global challenges
- Build skills for the future
- Drive an informed public conversation

Research Programmes

- Artificial Intelligence
- Data Science at Scale
- Data Science for Science and Humanities
- Data-Centric Engineering
- Defence and Security
- Finance and Economics
- Fundamental AI
- Health and Medical Sciences
- Public Policy
- Tools, Practices and Systems
- Urban Analytics



The Alan Turing Institute

Defence and national security

To protect the UK, its people and the places they inhabit

Environment and sustainability

To address the climate and biodiversity crisis and the need for greater sustainability.

Transformation of health

To transform health and enable better outcomes for all.

Research software engineering capability

Research Application Managers and Community Managers

Fundamental AI

Open-source infrastructure

Cloud First Approach

– Vision to be responsive to the needs of a new organisation
– The Turing's needs were still evolving
– Azure remains the default

Cloud++

- When the research teams were established it became clear Cloud couldn't fulfil all our compute needs
- Didn't provide GPUs at the required value given the scale
- Cloud was a learning curve for researchers used to HPC
- Interesting that now it's often the reverse
- HPC partnerships were a way to complete the picture

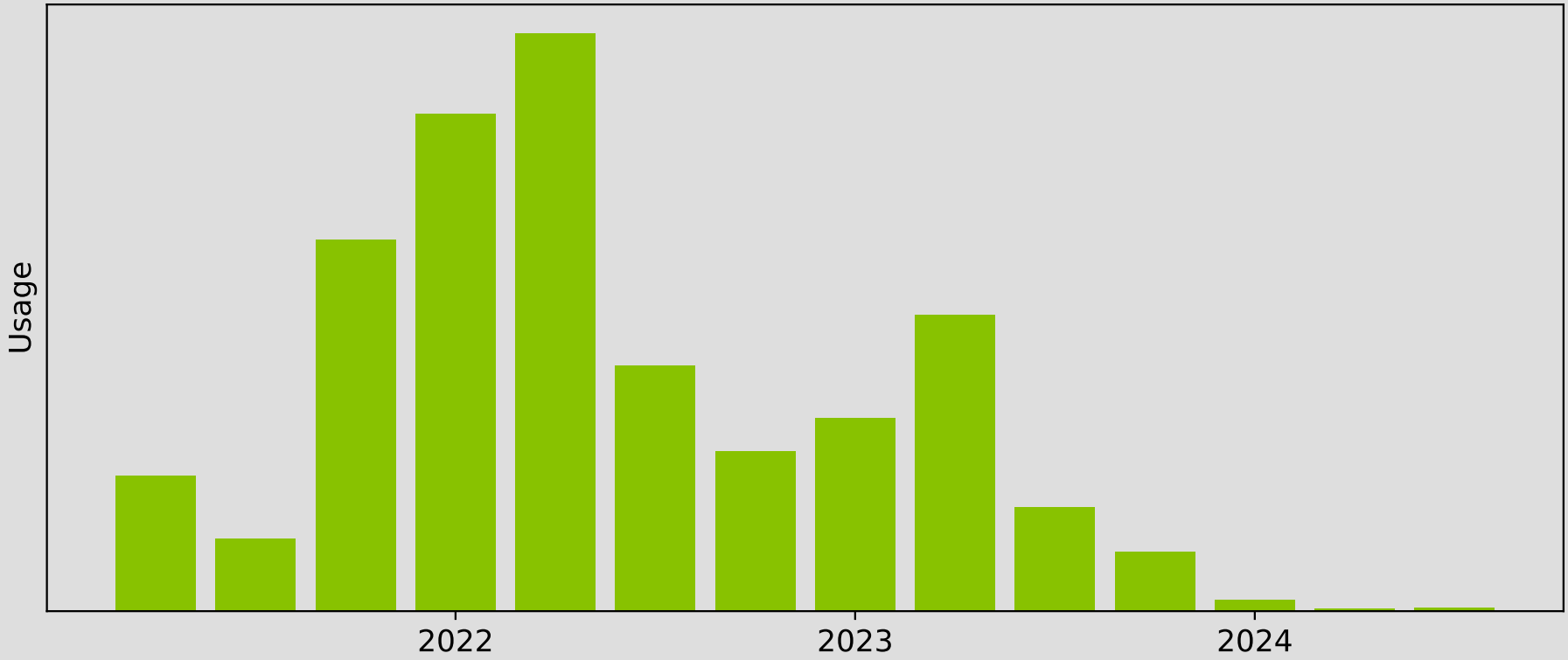
Establishing Collaborations

– First iteration of JADE
  - Turing helped to upgrade the P100s to V100s – around 2019
– Gradually our researchers started using JADE

Learning from JADE

- At this point we were a new organisation, not familiar with HPC
- A great learning experience for us
- Understanding DGX boxes with NVLink connections was a great learning experience
- We're now affiliated with several Tier-2 HPC systems in the UK

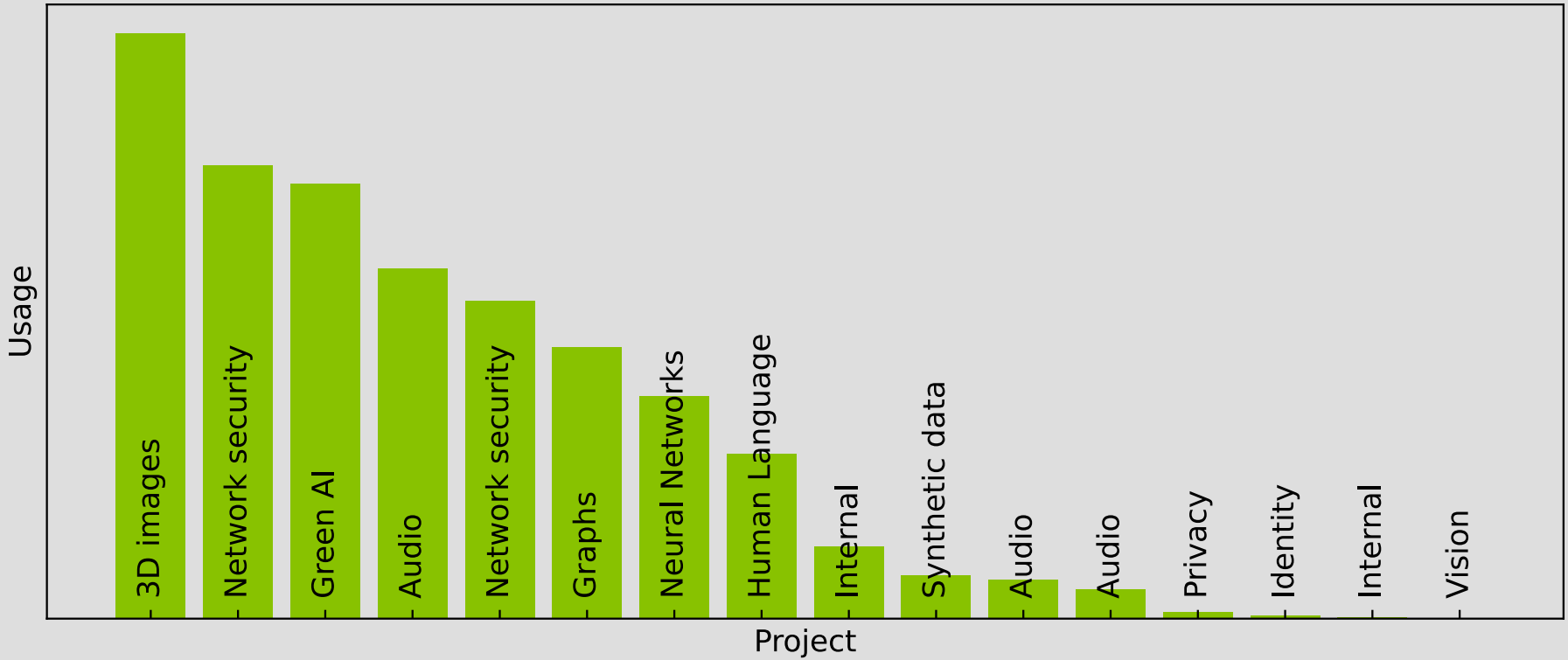
Usage over time



Turing’s research on JADE2

– 24 projects
– Average usage of 4 639 GPUh per month across 42 months
– 8 projects with over 10 000 GPUh use

Usage by project



Music Ensemble Separation

Saurjya Sarkar, Louise Thorpe,
Emmanouil Benetos, Mark Sandler

Time domain Music Source Separation

- Use synthesised dataset **EnsembleSet** to train instrument-agnostic music source separation models
- Models are trained to separate mixtures of any fixed number monophonic sources using permutation invariant training
- Monophonic sources present in the mixture may even be identical, and can be instrument or singing voice

Best Student Paper Award - IEEE WASPAA 2023
S. Sarkar, L. Thorpe, E. Benetos and M. Sandler, "Leveraging Synthetic Data for Improving Chamber Ensemble Separation," 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2023

Emmanouil Benetos

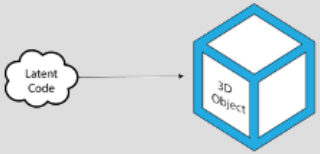
> this work used time-domain convolutional neural networks for training, and it would not have been possible to use such large models without support from JADE

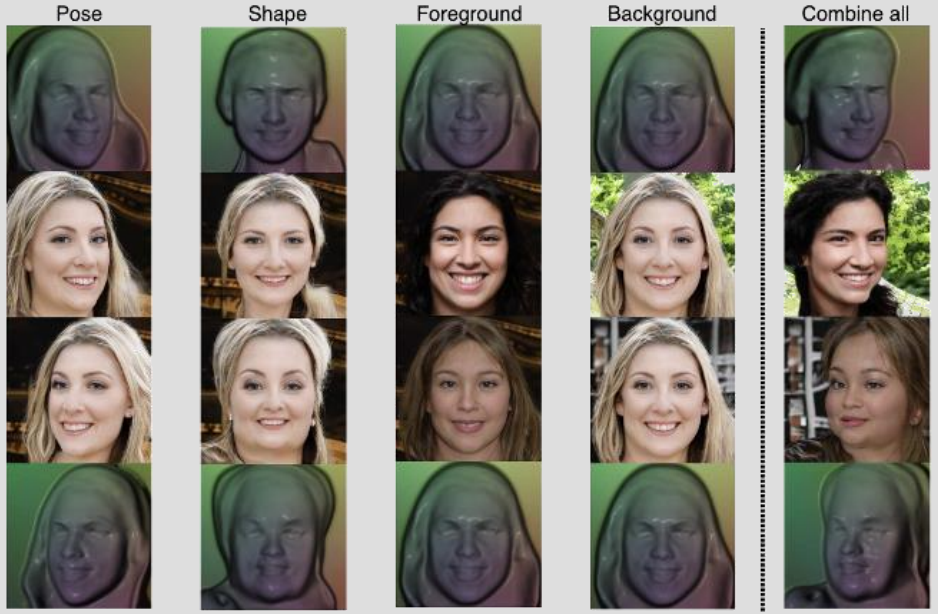
StyleMorph

Eric-Tuan Le, Edward Bartrum, Iasonas Kokkinos

Disentangling Shape, Pose and Appearance through 3D Morphable Image and Geometry Generation

- Project summary:
- We introduce a disentangled, deformable 3D controllable image generator
- Outcome:
- This work was published at ICLR 2023 as a poster presentation





Separately control shape, foreground, background and pose by manipulating different latent codes.

Interpolating latent codes for each separate factor of control

Background Control

Foreground Control

Pose Control

Shape Control

Edward Bartrum

> ...JADE allowed us to run a lot of experiments in parallel so they were really ideal from a research perspective. A project like this one wouldn't really have been feasible using a small number of local gpus... Anyway, we're very grateful that we had access to these resources!

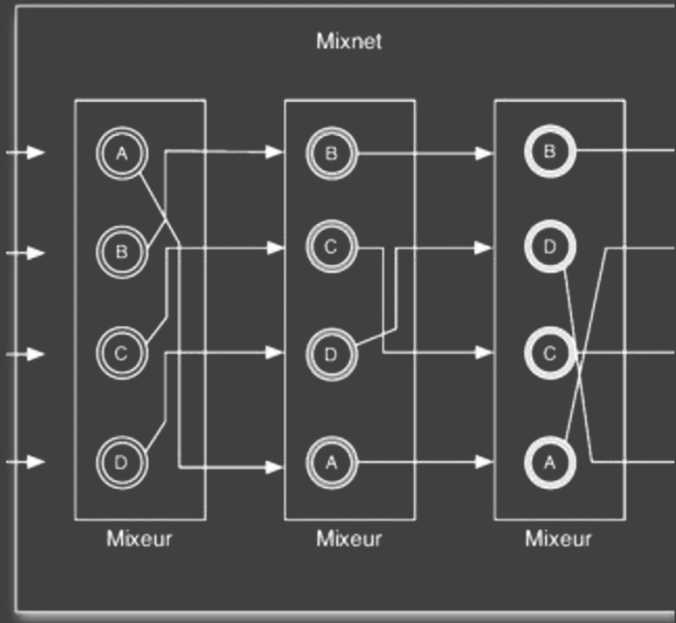
LLMs for Privacy Measurement

Vasilios Mavroudis, AICD

https://www.turing.ac.uk/aicd
aicd@turing.ac.uk

Anonymity Network Privacy Estimation

- Modern mix networks improve over Tor and provide stronger privacy guarantees by robustly obfuscating metadata
- As long as a message is routed through at least one honest mixnode, the privacy of the users involved is safeguarded
- However, the complexity of the mixing mechanisms makes it difficult to estimate the cumulative privacy erosion occurring over time.
- This work uses a generative model trained on mixnet traffic to estimate the loss of privacy when users communicate persistently over a period of time.



Synthetic Facial Biometrics

Praveen Selvaraj, Roberto Leyva,
Santhosh Narayanan

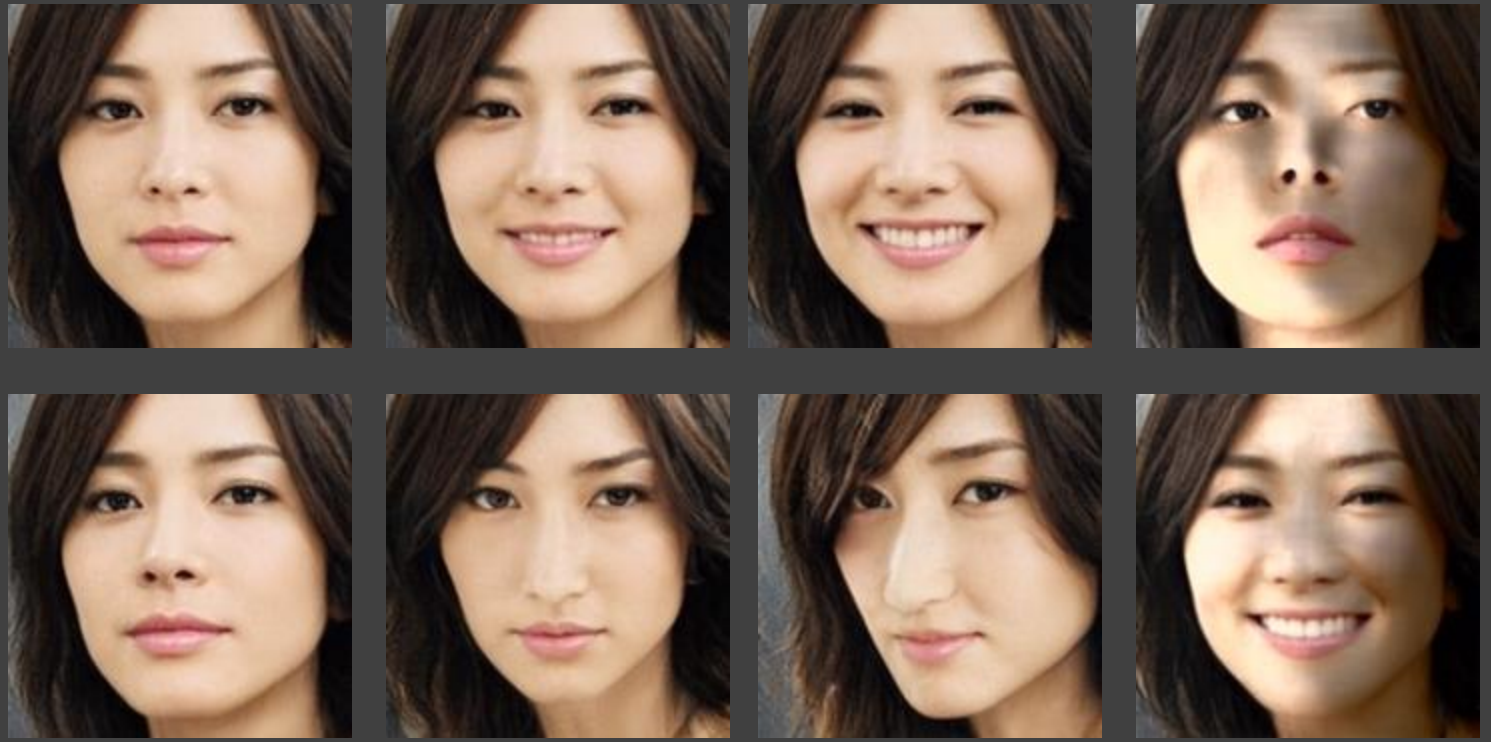
Faces across demographics








Variations for each face



Praveen Selvaraj

“My role is to specifically use synthetic data to investigate different components in a digital ID system.... So, a lot of it involves training, testing different AI models for different tasks and a system like JADE is vital for that and I loved using it. I still use it and plan to do so until it's decommissioned.”

Final Thoughts

– As an institution we're really grateful for
  - What JADE has offered in terms of compute
  - The support provided

– JADE has helped support AI research in the UK

Final Thoughts

- HPC will only become more important for AI research
  - As it becomes classified as Critical Infrastructure
- Our thanks go to Wes Armour and the JADE team
  - For all the hard work that’s enabled great science to happen
- We hope this isn't the end of the JADE legacy

The Alan Turing Institute

Thank you!

David Llewellyn-Jones

Research Data Scientist

JADE Day 2024, Oxford, 3 October 2024