



THE UNIVERSITY *of* EDINBURGH
informatics

AI for Social Good

Speaker: Jane Hillston, Head of School of Informatics



Some History

AI activity at Edinburgh started with the research group of **Donald Michie** in 1963, and the founding of the **Dept of Machine Intelligence and Perception** in 1966, which became the **Dept of Artificial Intelligence** in 1974.

The University's **Computer Unit** was founded in 1963 by **Sidney Michaelson** and this became the **Dept of Computer Science** in 1969.

The **Centre for Speech Technology Research** was founded in 1984 and the **Dept of Cognitive Science** in 1985.

These came together to form what is now the **School of Informatics** in 1998.

School of Informatics Research Institutes

icsa | Institute for Computing
Systems Architecture

**COMPUTER SCIENCE: Parallel Computing, Micro Architectures,
Wireless Protocols & Apps, Iterative Compilation, Self Timed
Circuits**

aiai | Artificial Intelligence
and its Applications Institute

**ARTIFICIAL INTELLIGENCE: Intelligent Planning, Proof
Planning, Security Engineering, Applied Computational
Logic, Knowledge Engineering, Virtual Worlds**

ipab | Institute of Perception,
Action and Behaviour

ROBOTICS: Robotics, Vision

informatics

anc | Institute for Adaptive
and Neural Computation

BRAIN: NeuroInformatics, Machine Learning

ifcs | Laboratory for Foundations
of Computer Science

**THEORY: Databases Languages, Semantics, Complexity
& Alg Concurrency & Modelling, S/W Engineering
Theory**

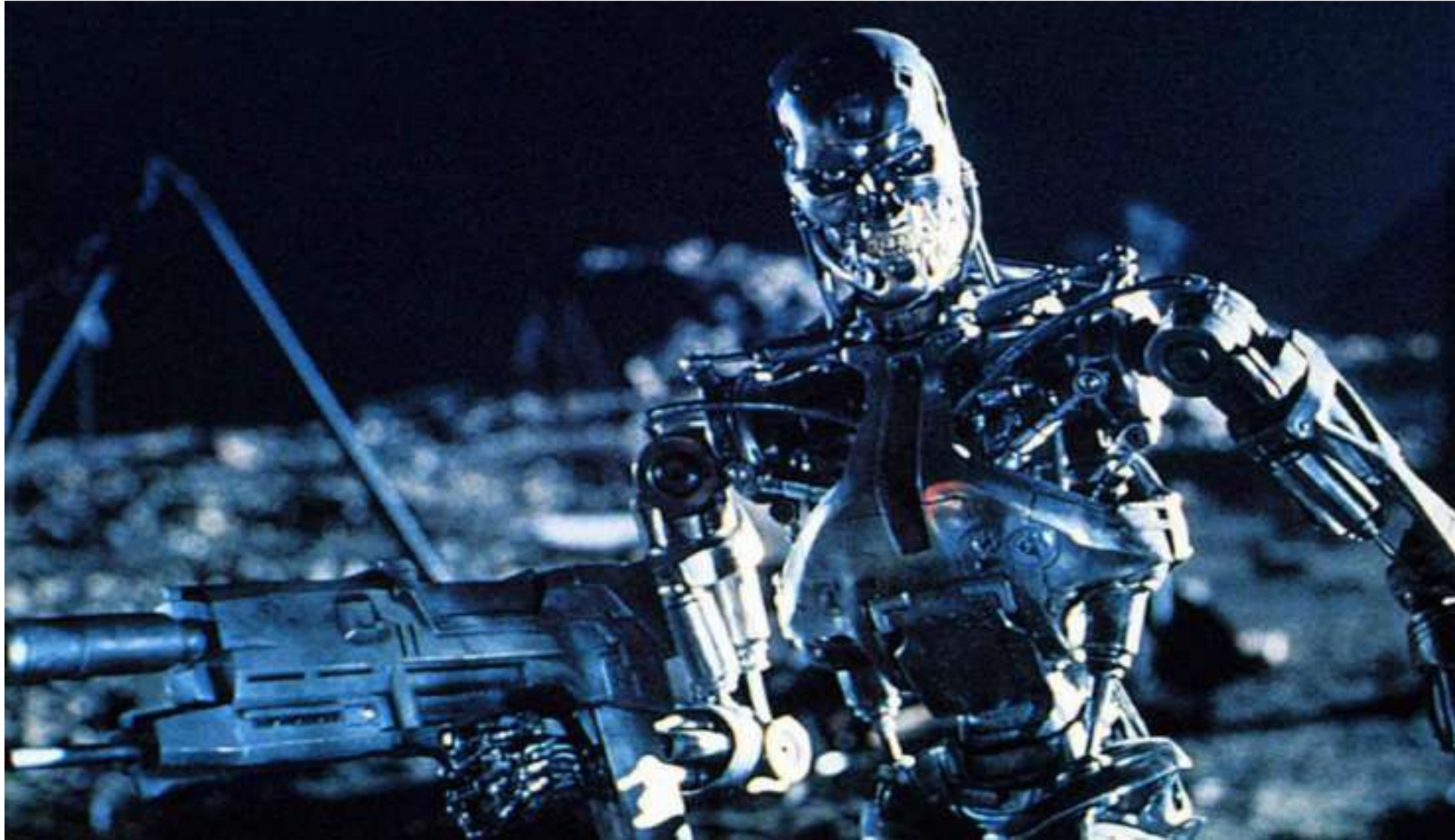
ilcc | Institute for Language,
Cognition and Computation

**LANGUAGE: Natural Language Processing, Multi Modal Interaction,
Information Extraction, Speech Synthesis**



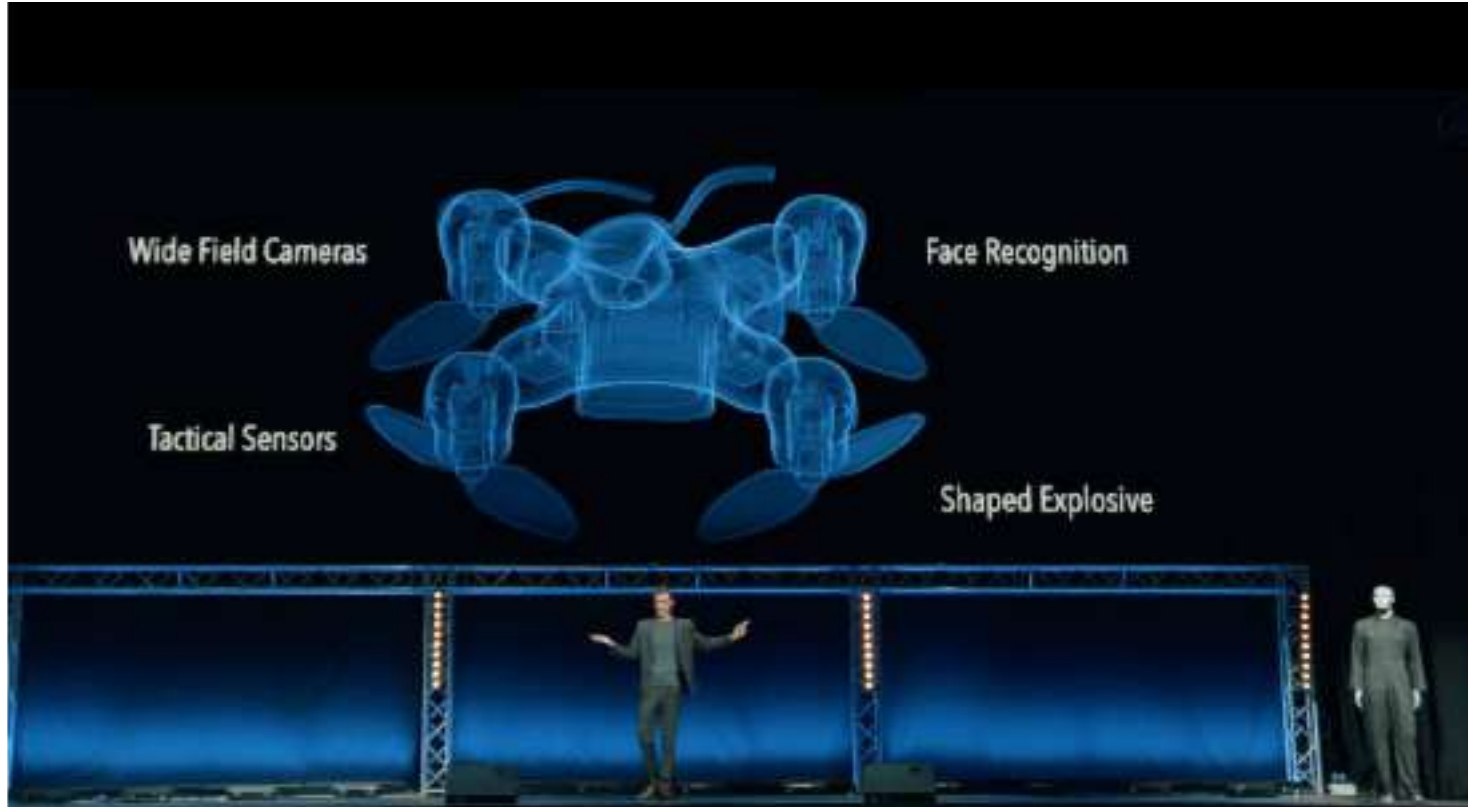
THE UNIVERSITY of EDINBURGH
informatics

Visions of AI





Visions of AI: Slaughterbots





Visions of AI: Slaughterbots





Visions of AI: Slaughterbots





Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia

Songül Tolan*
Marius Miron*
Joint Research Centre,
European Commission
firstname.lastname@ec.europa.eu

Emilia Gómez
Joint Research Centre,
European Commission
Universitat Pompeu Fabra, Barcelona
emilia.gomez@upf.edu

Carlos Castillo
Universitat Pompeu Fabra, Barcelona
chato@acm.org

ABSTRACT

In this paper we study the limitations of Machine Learning (ML) algorithms for predicting juvenile recidivism. Particularly, we are interested in analyzing the trade-off between predictive performance and fairness. To that extent, we evaluate fairness of ML models in conjunction with SAVRY, a structured professional risk assessment framework, on a novel dataset originated in Catalonia. In terms of accuracy on the prediction of recidivism, the ML models slightly outperform SAVRY; the results improve with more data or more features available for training (AUCROC of 0.64 with SAVRY vs. AUCROC of 0.71 with ML models). However, across three fairness metrics used in other studies, we find that SAVRY is in general

QC, Canada. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3322640.3326705>

1 INTRODUCTION

Machine learning (ML) systems detect patterns in data and are able to predict complex outputs under high uncertainty [37]. Medicine, finance and law, are a few domains where humans rely on an algorithm to solve expert tasks [28]. In these cases ML systems can surpass human capabilities, particularly when dealing with large datasets or a high number of input features. One example where ML algorithms and expert systems can better inform human decisions



[Google](#) / [Press Releases](#) / [Tech](#) / [Top Stories](#)

Project Veritas Re-Uploads Google Exposé Taken Down By YouTube Ahead of White House Social Media Summit

5 months ago by [Staff Report](#)

The report shows a Google executive discussing Google's plans for the 2020 elections, making reference to "prevent[ing]" the next "Trump situation."



Face Recognition and Surveillance





Face Recognition and Surveillance



Report: Use of AI surveillance is growing around the world

20 SEP 2019

1

Government security, Law & order, Machine Learning, Privacy



THE UNIVERSITY of EDINBURGH
informatics

Data Ethics/AI Ethics

 **GOV.UK**

Search



Home > Data Ethics Framework



Department for
Digital, Culture,
Media & Sport

Guidance

Data Ethics Framework

Updated 30 August 2018



 GOV.UK

Home > Data.Ethics.Framework



Department for
Digital, Culture,
Media & Sport

Guidance

Data Ethics

Updated 30 August 2018

The Alan Turing Institute

Understanding artificial intelligence ethics and safety

A guide for the responsible
design and implementation of AI
systems in the public sector





GOV.UK

Home > Data Ethics Framework



Department for
Digital, Culture,
Media & Sport

Guidance

Data Ethics

Updated 30 August 2018

The Alan Turing Institute



Microsoft AI principles

Designing AI to be trustworthy requires creating solutions that reflect ethical principles that are deeply rooted in important and timeless values.

Fairness

AI systems should treat all people fairly

Inclusiveness

AI systems should empower everyone and engage people

Reliability & Safety

AI systems should perform reliably and safely

Transparency

AI systems should be understandable

Privacy & Security

AI systems should be secure and respect privacy

Accountability

AI systems should have algorithmic accountability



Jo Swinson debates ethics and artificial intelligence - and suggests the Lovelace Oath

By [Caron Lindsay](#) | Fri 19th January 2018 - 1:30 pm

[Follow @caronmlindsay](#)

This week, Jo Swinson held a Westminster Hall debate on ethics and artificial intelligence. While recognising the huge advantages of AI, there are some ethical challenges we need to do something about. Jo looked at this from a very liberal perspective, as you would imagine. Here are some of the highlights of her speech. You can read the whole debate [here](#).



Ethics of data and AI are focus of appointment

One of the US's leading experts on the impact of artificial intelligence and other innovative technologies on people's lives is set to join the University.

Professor Shannon Vallor – who joins the University in February 2020 from Santa Clara University in California's Silicon Valley – has been appointed as the first Baillie Gifford Chair in the Ethics of Data and Artificial Intelligence at the Edinburgh Futures Institute (EFI).

Key appointment

Professor Vallor's appointment will help to establish Edinburgh as a leader in harnessing the fast-moving developments in data and artificial intelligence to benefit society.

The Professorial Chair is supported by global investment firm Baillie Gifford as part of its £5m pledge to support University research into the challenges and opportunities around emerging technologies – including machine learning, accelerated automation, and financial innovation.

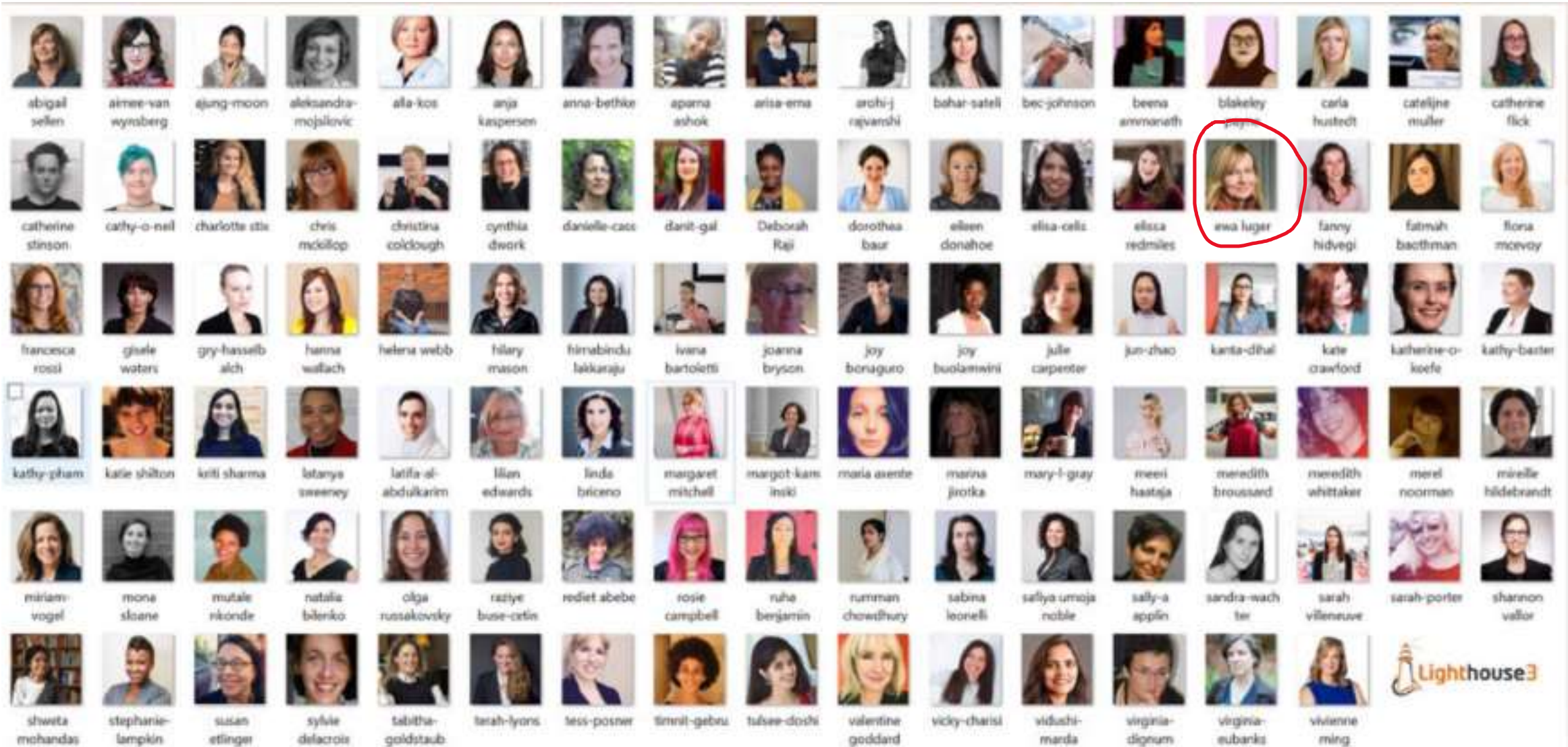


Professor Shannon Vallor





Diversity and Ethics in AI -- 2020





THE UNIVERSITY *of* EDINBURGH
informatics

So what is our responsibility as a School of Informatics?



- Lead by example
- Educate the next generation
- Enable others
- Inform the public

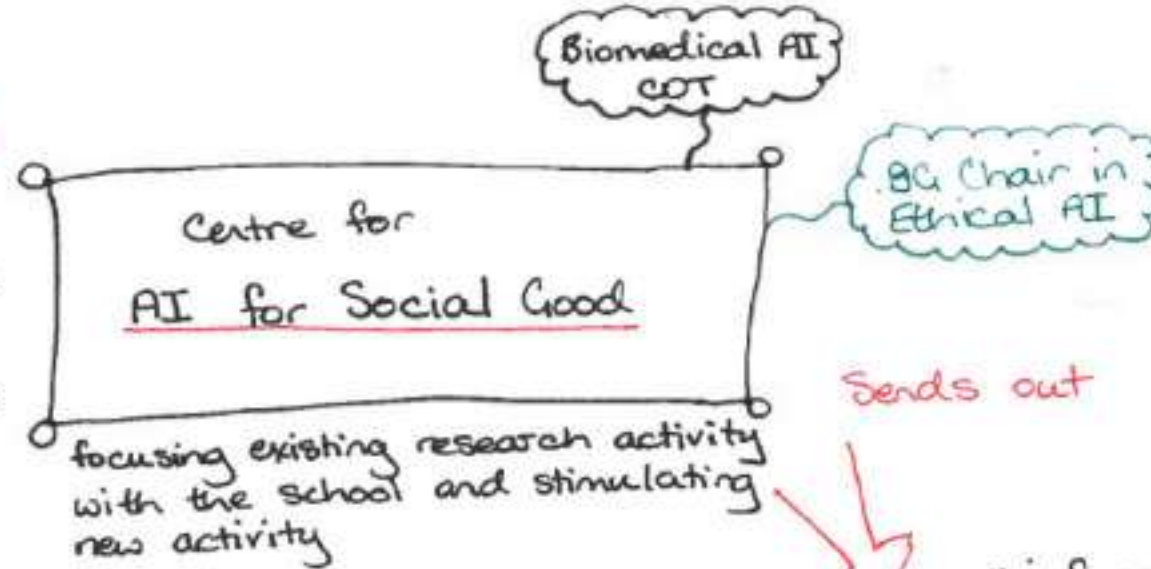


Brings in

Motivated staff
and students

Research opportunities

Philanthropic donations



Sends out

Well-informed graduates

Understandable public
profile

Research output

Longer term opportunities:

- MSc programme in AI for Social Good
- COT programme
- Physical centre within EFI
- Visitor programme (cf Newton Institute)
centred on societal challenges (Turing?)



Focusing existing research activity within the School and stimulating new activity

- Using data science to support charities working with rough sleepers in London
- Applying speech, vision and machine learning to speech and language therapy
- Using AI planning to support emergency response by teams of people and robots
- Knowledge-based system for assessing workplace exposure to potentially hazardous substances
- Identifying Malaria parasites in images



Focusing existing research activity within the School and stimulating new activity

- Safe autonomous driving in urban environments
- Estimating bacterial and cellular load in lung images
- Trusted answers from incomplete data
- Encouraging active lifestyles by continuous tracking of activity and exercise
- Promoting engagement in citizen science using intelligent interventions



Shared Control of Exoskeletons and Prostheses

Millions of people worldwide suffer from disabilities or injuries which affect mobility and decades of research has gone into the development of assistive robotic devices, such as exoskeletons and prosthetic limbs, in an effort to restore movement potential to these individuals.



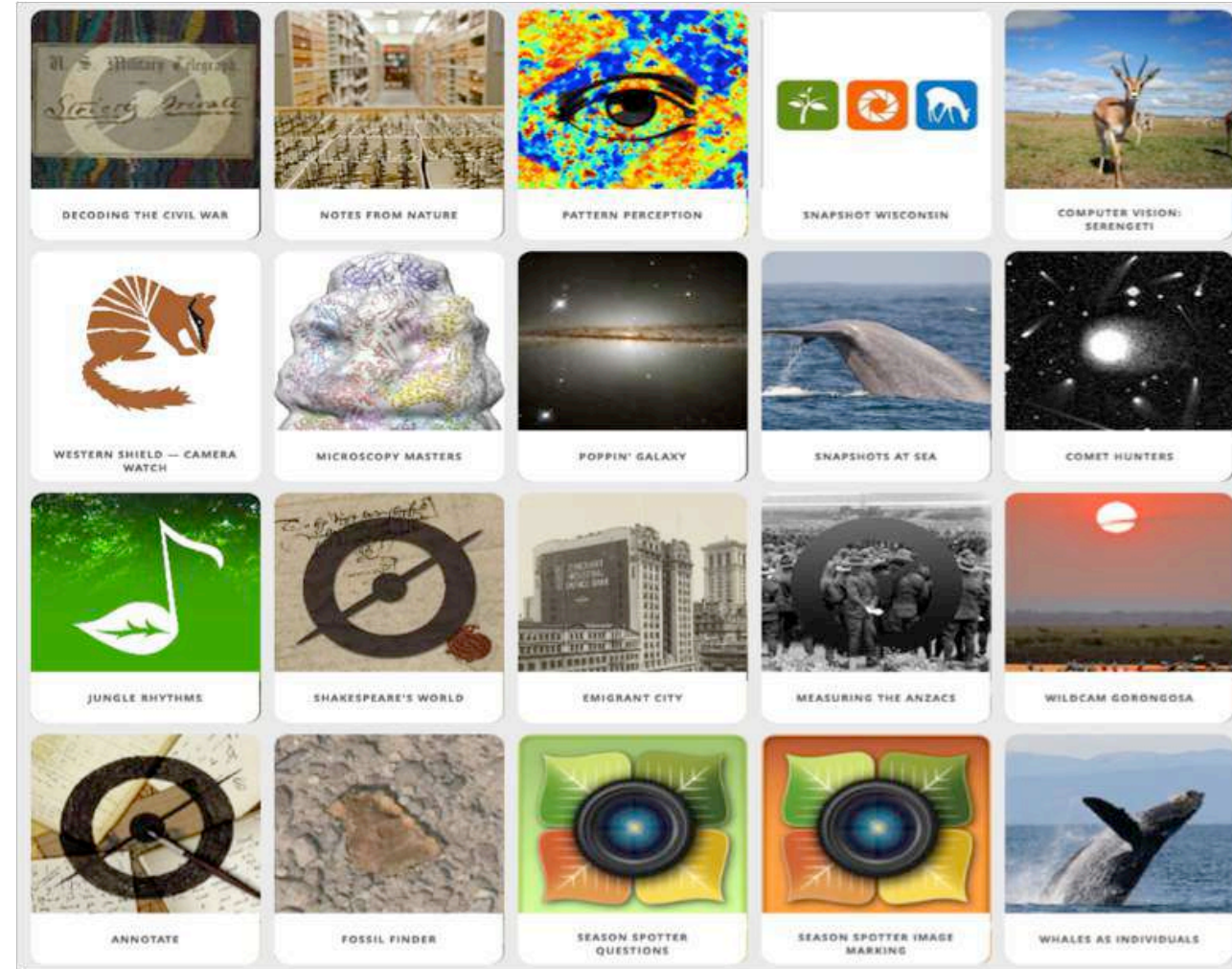


Promoting engagement in citizen science

Citizen science engages people in scientific research.

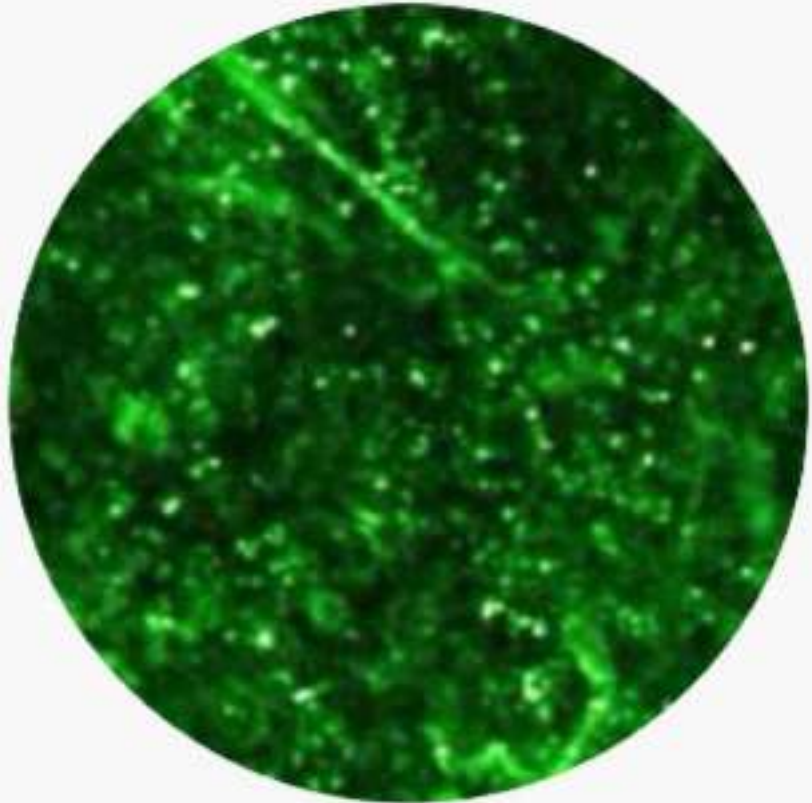
Machine learning is used to predict when users are becoming disengaged and intelligent intervention design guides personalised feedback in the form of motivational messages that pop up on the volunteer's screen.

This approach increased the contributions of thousands of participants in Galaxy Zoo by over 70%





The Proteus Project



This work provides a new and fast method for detecting bacteria or cells in the human lungs using fibered confocal fluorescence microscopy (FCFM). Dr Sohan Seth (School of Informatics) trained neural networks to predict whether a each pixel of the FCFM image a bacterium or cell is present. The work forms part of PROTEUS, an EPSRC-funded Interdisciplinary Research Collaboration which aims to develop technology that will provide quick, in situ, in vivo diagnoses and management of lung diseases in the clinical environment.



Educating the next generation

We aim to integrate consideration of ethical issues and social responsibility throughout our curriculum --- in course on Professional Issues and within assistive robot projects, but not just there...

Projects are clearly badged as “AI for Social Good”

All our CDTs include training in responsible research and innovation, but this is particular emphasized in the CDT on Biomedical AI

We look forward to working closely with Shannon Vallor (Baillie Gifford) when she arrives



Enable others

As part of our long term vision we will seek to establish a visitor programme that will allow researchers from around the world to come to spend extended periods with us developing AI techniques to tackle societal problems (cf The Newton Institute)

This will be guided by and contribute to the University's 2030 Vision, addressing UN sustainability goals





Inform the public

One of the major goals of the Centre is to surface work on beneficial AI that is already going on in the School and to provide a beacon for others, both in the discipline and more widely

Our students are already leading the way with a series of public discussions with the title “**We Need to Talk About AI**”

Going forward we will be careful to communicate our work in an accessible way and to engage with government and policy makers to foster greater trust in AI

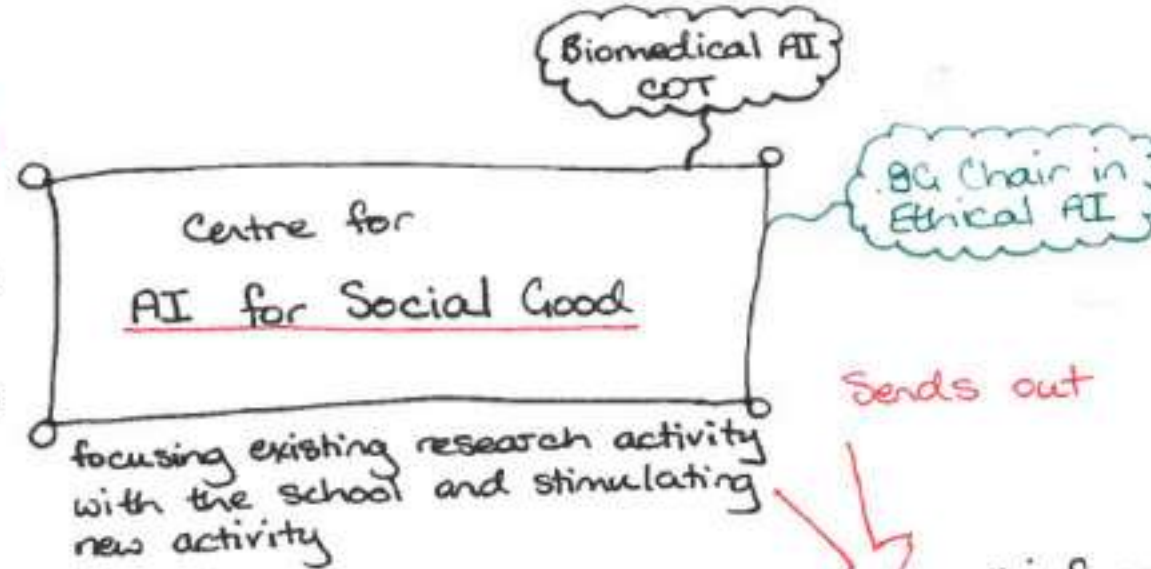


Brings in

Motivated staff
and students

Research opportunities

Philanthropic donations



Sends out

Well-informed graduates

Understandable public
profile

Research output

Longer term opportunities:

- MSc programme in AI for Social Good
- COT programme
- Physical centre within EFI
- Visitor programme (cf Newton Institute)
centred on societal challenges (Turing?)



THE UNIVERSITY *of* EDINBURGH
informatics

Thank you!



THE UNIVERSITY *of* EDINBURGH
informatics