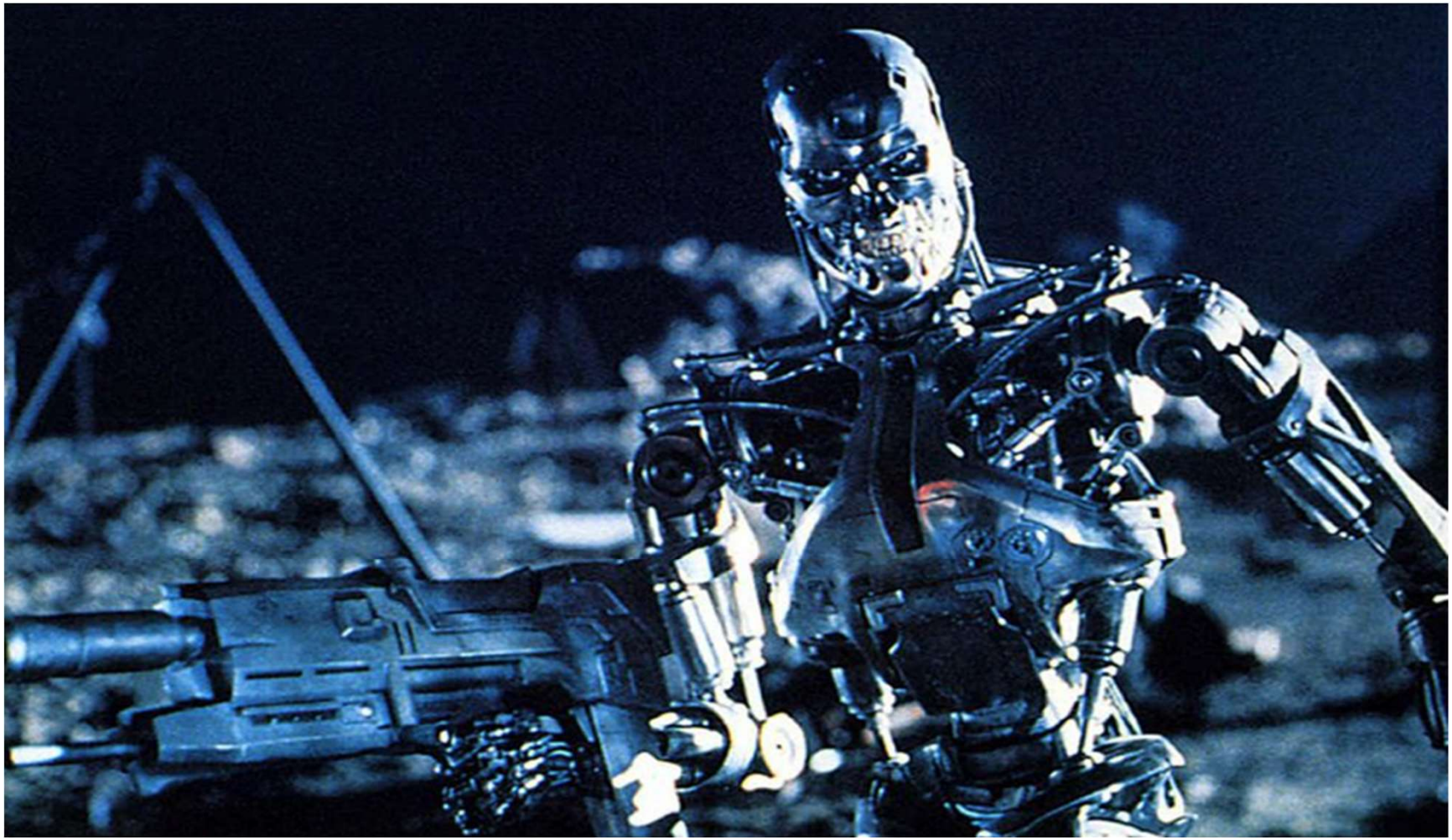
THE UNIVERSITY of EDINBURGH
informatics

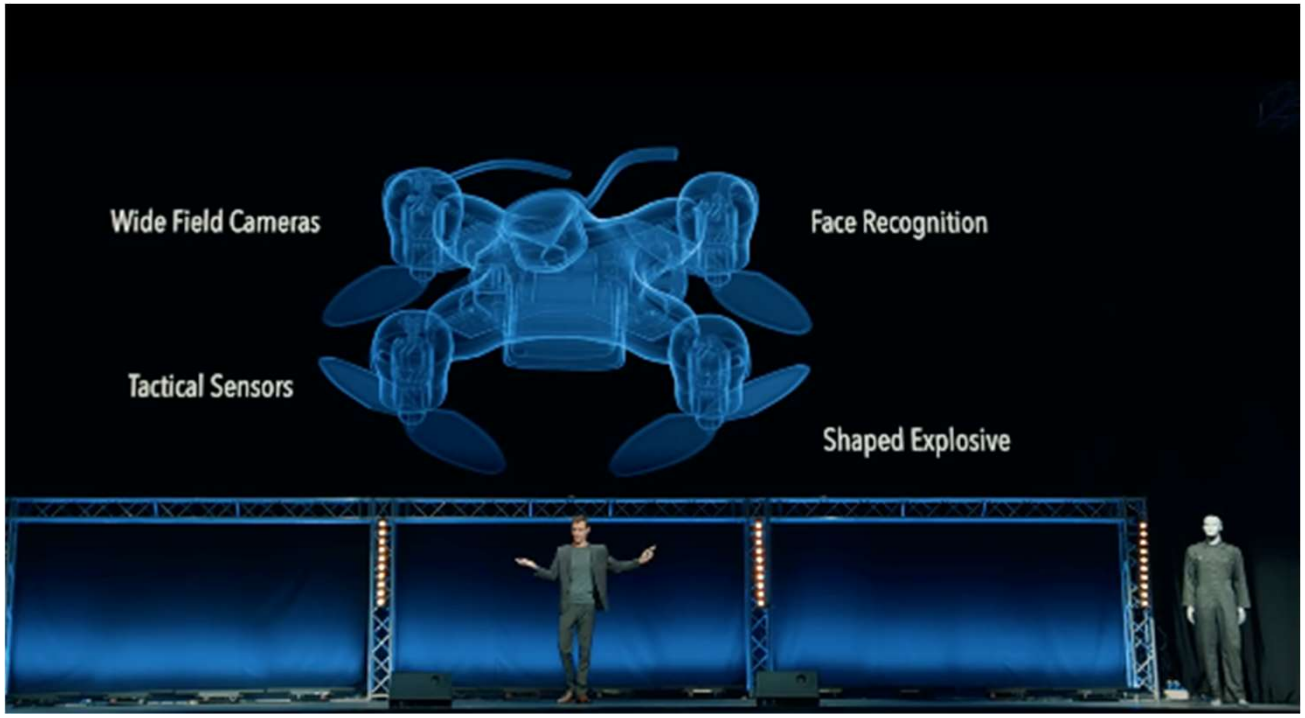
AI for Social Good

Speaker: Jane Hillston, Head of School of Informatics

Visions of AI



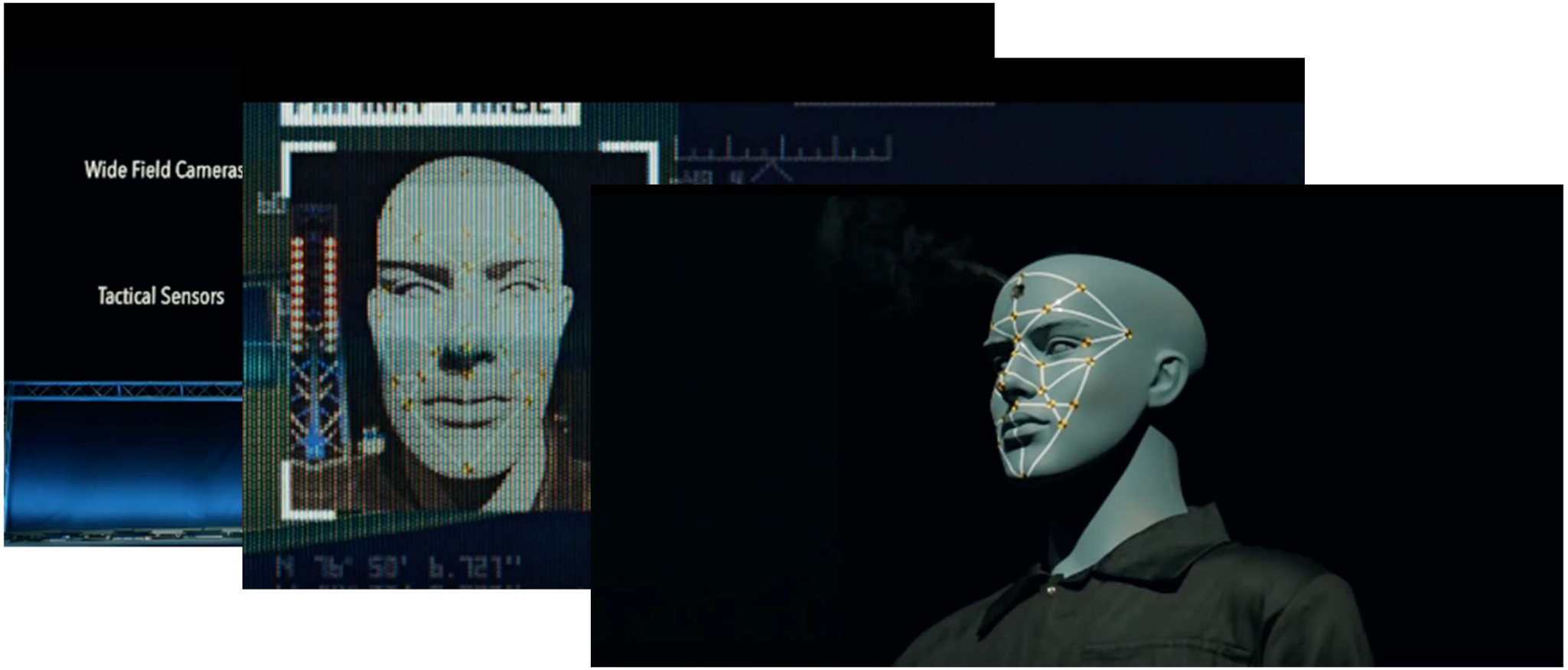
Visions of AI: Slaughterbots



Visions of AI: Slaughterbots

THE UNIVERSITY of EDINBURGH
informatics

Visions of AI: Slaughterbots



Algorithmic (un)Fairness

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 algorithms 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

Algorithmic (un)Fairness

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."

THE UNIVERSITY of EDINBURGH
informatics

Face Recognition and Surveillance

THE UNIVERSITY of EDINBURGH
informatics

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

THE UNIVERSITY of EDINBURGH
informatics

Data Ethics/AI Ethics

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

THE UNIVERSITY of EDINBURGH
informatics

Data Ethics/AI Ethics

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

Reliability & Safety
AI systems should perform reliably and safely

Privacy & Security
AI systems should be secure and respect privacy

Inclusiveness
AI systems should empower everyone and engage people

Transparency
AI systems should be understandable

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.

THE UNIVERSITY of EDINBURGH
informatics

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).

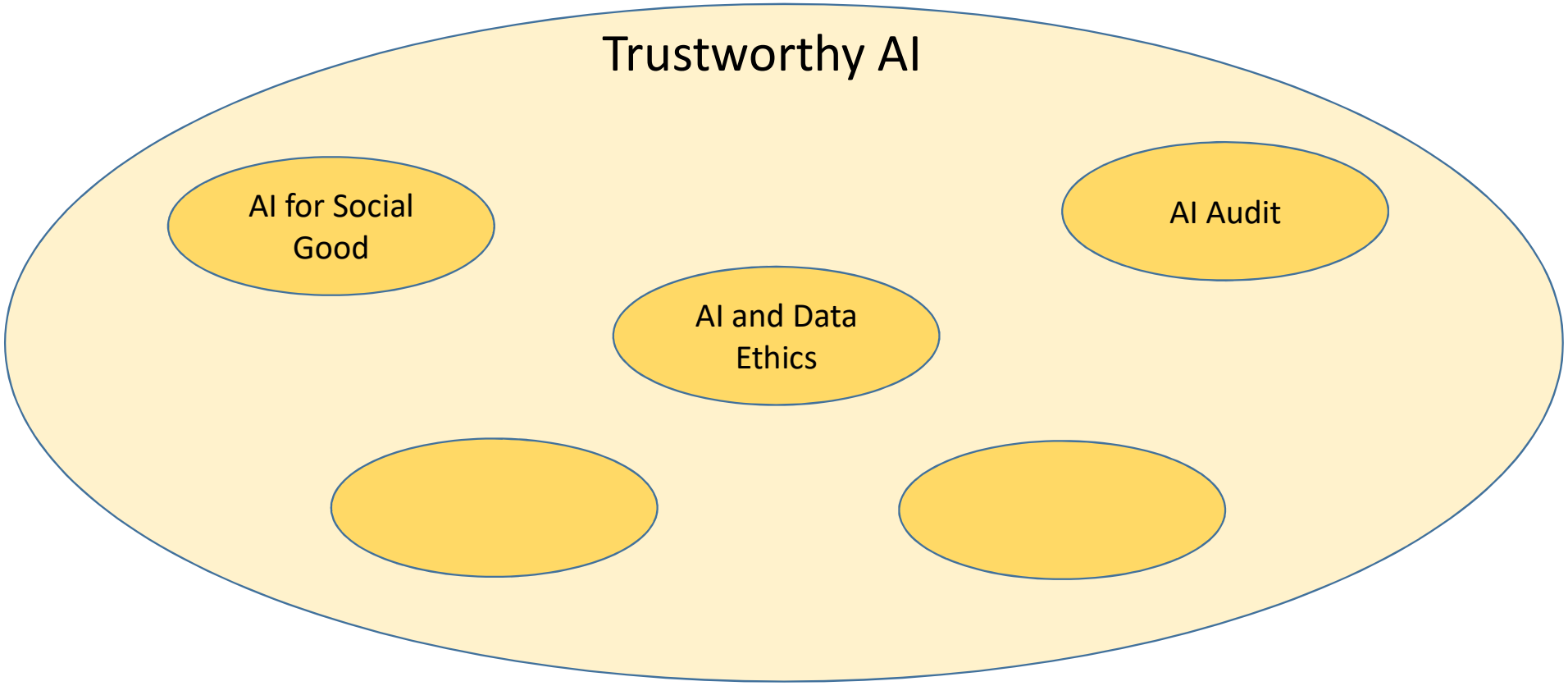
Key appointment

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.

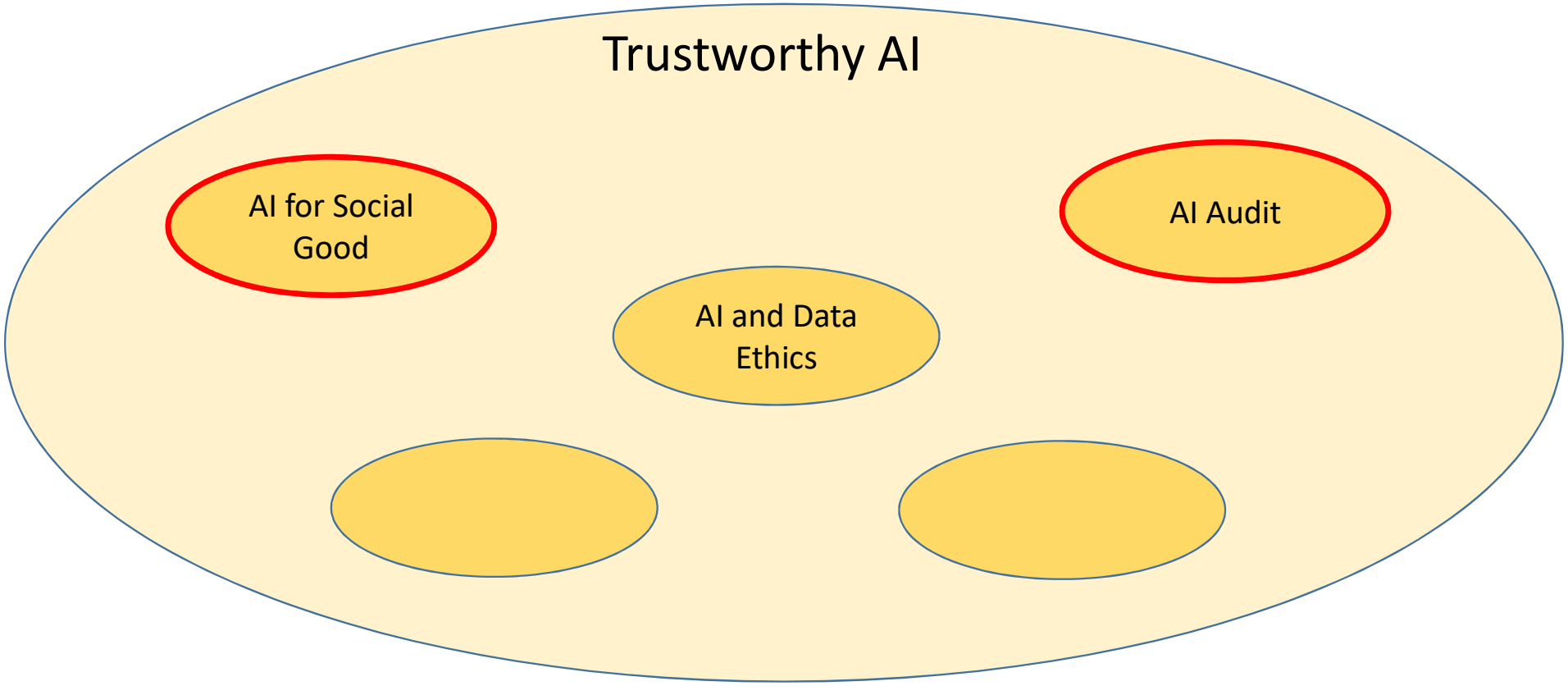
The Professorial Chair is supported by global investment firm Baillie Gifford as part of its £5m pledge to support University research into the challenges and opportunities around emerging technologies – including machine learning, accelerated automation, and financial innovation.

Professor Shannon Vallor

Trustworthy AI



Trustworthy AI



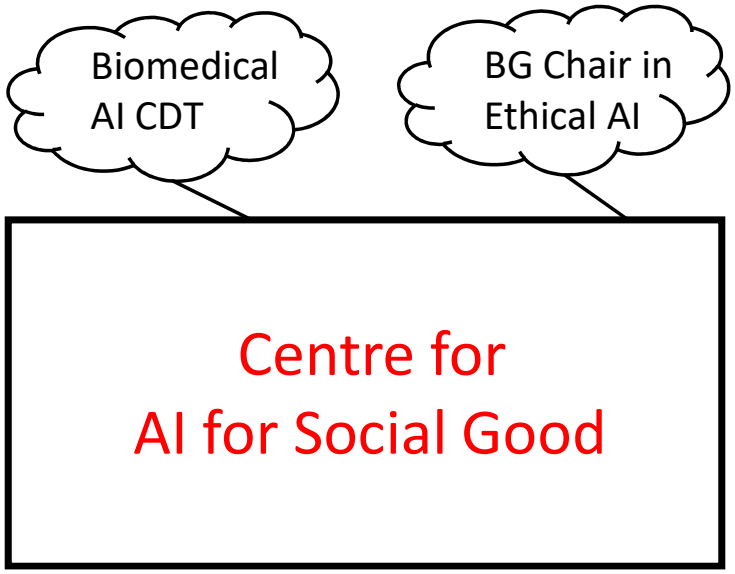
So what is our responsibility as a School of Informatics?

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

Centre for AI for Social Good

Focusing existing research activity within the School and stimulating new activity

Centre for AI for Social Good

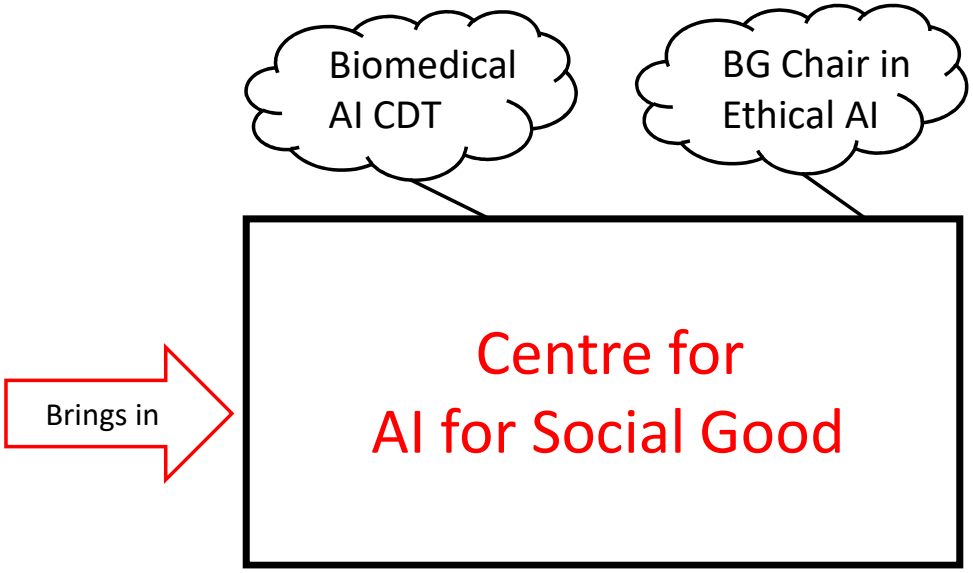


Focusing existing research activity within the School and stimulating new activity

THE UNIVERSITY of EDINBURGH informatics

Centre for AI for Social Good

- Motivated staff and students
- Research Opportunities
- Philantropic donations



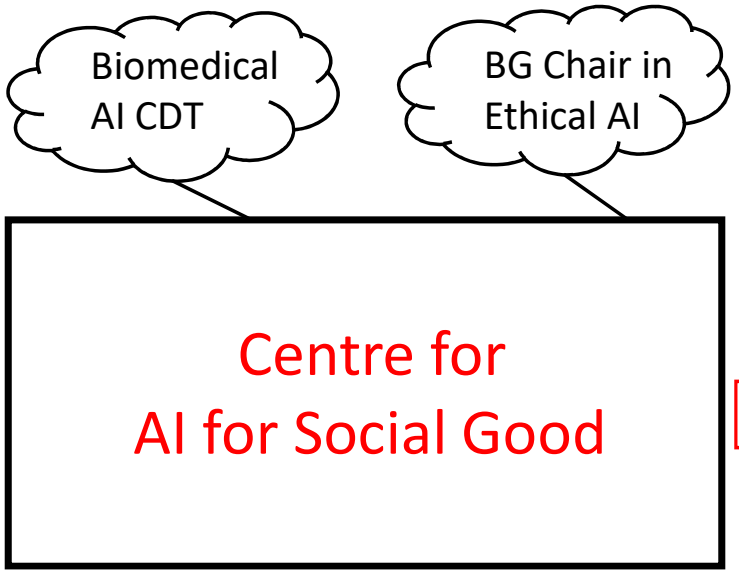
Focusing existing research activity within the School and stimulating new activity

THE UNIVERSITY of EDINBURGH informatics

Centre for AI for Social Good

- Motivated staff and students
- Research Opportunities
- Philantropic donations

Brings in



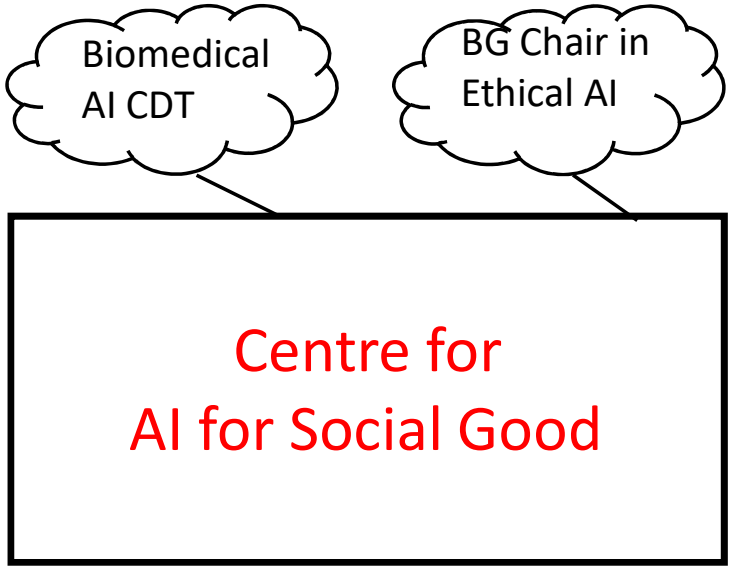
Sends out

- Well-informed graduates
- Understandable public profile
- Research output
- Influence in society and government

Focusing existing research activity within the School and stimulating new activity

THE UNIVERSITY of EDINBURGH
informatics

Centre for AI for Social Good



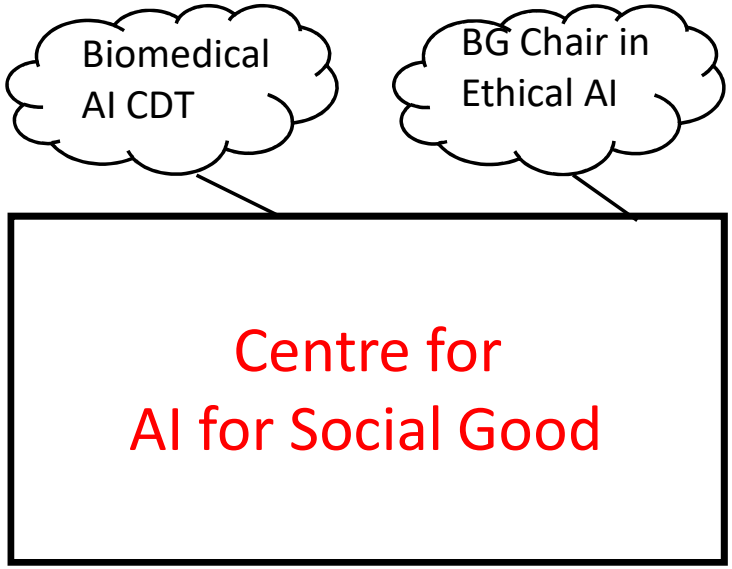
Focusing existing research activity within the School and stimulating new activity

Longer Term Opportunities

- MSc programme in AI for Social Good
- CDT programme
- Physical centre within EFI
- Visitor programme (cf Newton Institute) centred on social challenges

THE UNIVERSITY of EDINBURGH informatics

Centre for AI for Social Good



Focusing existing research activity within the School and stimulating new activity

Longer Term Opportunities

- MSc programme in AI for Social Good
- CDT programme
- Physical centre within EFI
- Visitor programme (cf Newton Institute) centred on social challenges

In discussions with the Turing Institute about alignment with their programme on Data Science for Social Good (DSSG)

Focusing existing research activity within the School and stimulating new activity

- Using data science to support charities working with rough sleepers in London (DSSG)
- 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

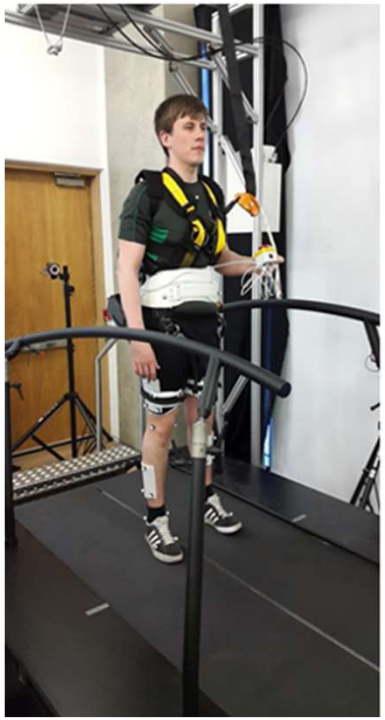
Focusing existing research activity within the School and stimulating new activity

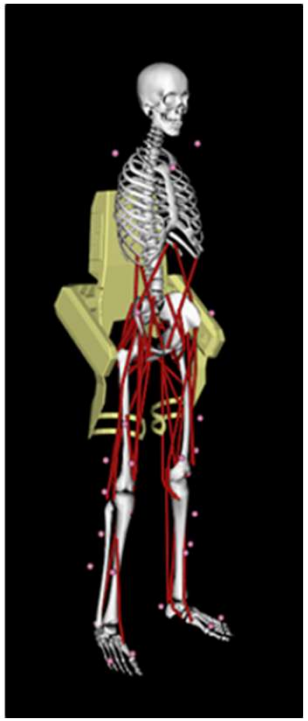
- Safe autonomous driving in urban environments
- Estimating bacterial and cellular load in lung images
- Trusted answers from incomplete data
- Encouraging active lifestyles by continuous tracking of activity and exercise
- Promoting engagement in citizen science using intelligent interventions

THE UNIVERSITY of EDINBURGH
informatics

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.





[

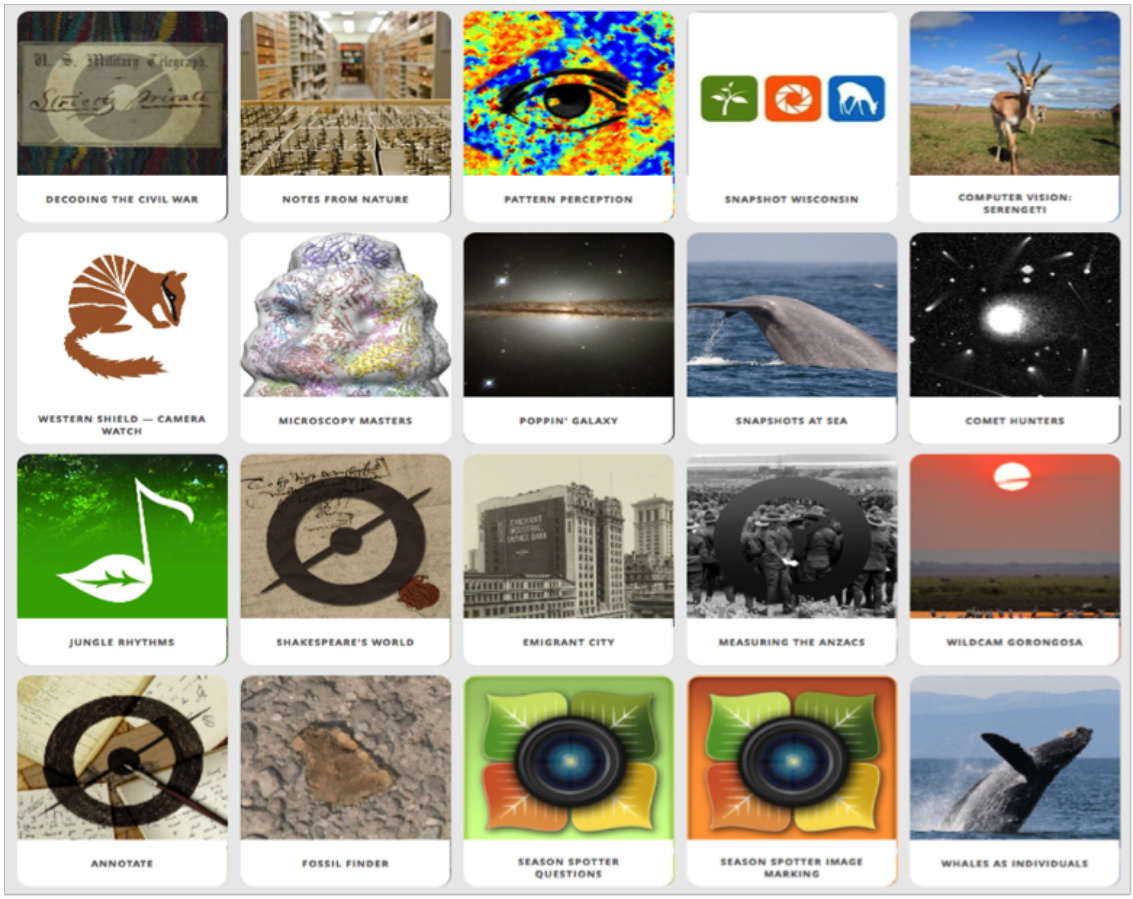
THE UNIVERSITY of EDINBURGH
informatics

Promoting engagement in citizen science

Citizen science engages people in scientific research.

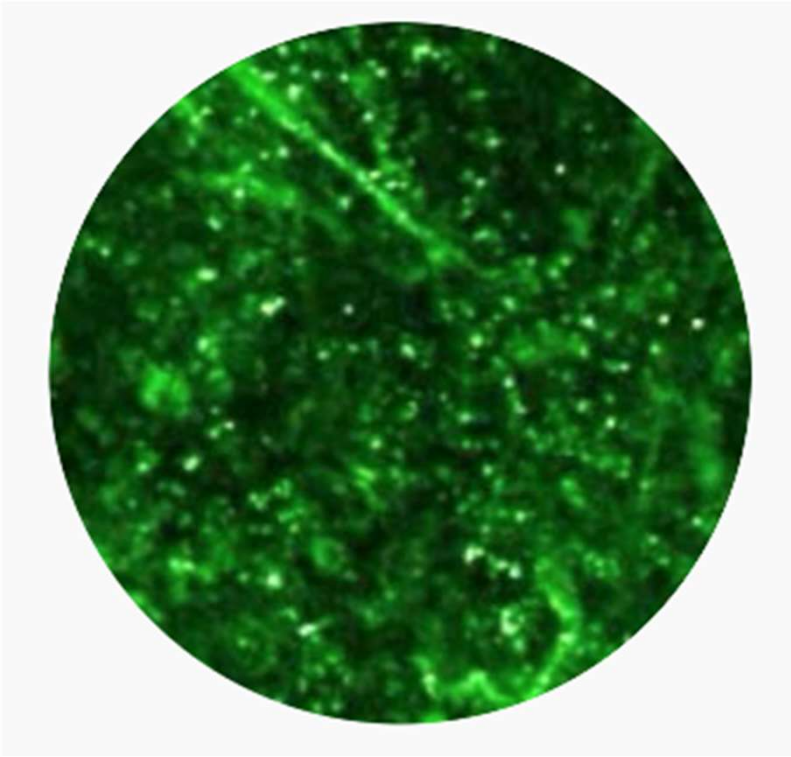
Machine learning is used to predict when users are becoming disengaged and intelligent intervention design guides personalised feedback in the form of motivational messages that pop up on the volunteer's screen.

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



THE UNIVERSITY of EDINBURGH
informatics

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 courses 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) and initial discussions have begun in particular for second year course, Working with Data and an AI Ethics course under development

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

THE UNIVERSITY of EDINBURGH
informatics

DTC in AI for Social Good: Vision

- Train a cohort of PhD students to develop new methodologies in AI and study their outcomes in context.
- Counteract media presentation of AI as utopian or dystopian
- Much interest in AI ethics, primarily directed to ethical development processes.
- To bring about genuine social good, there must also be attention on ethical outcomes.
- In this collaboration with Social and Political Science we could give students grounding in sound techniques to not only plan for responsible research and innovation, but also interrogate the outcomes to assess the extent to which objectives have been met.

THE UNIVERSITY of EDINBURGH
informatics

DTC in AI for Social Good: Vision

- Train a cohort of PhD students to develop new methodologies in AI and study their outcomes in context.
- Counteract media presentation of AI as utopian or dystopian
- Much interest in AI ethics, primarily directed to ethical development processes.
- To bring about genuine social good, there must also be attention on ethical outcomes.
- In this collaboration with Social and Political Science we could give students grounding in sound techniques to not only plan for responsible research and innovation, but also interrogate the outcomes to assess the extent to which objectives have been met.

Three focus areas are envisioned (distinct from Biomedical AI to attract funding):

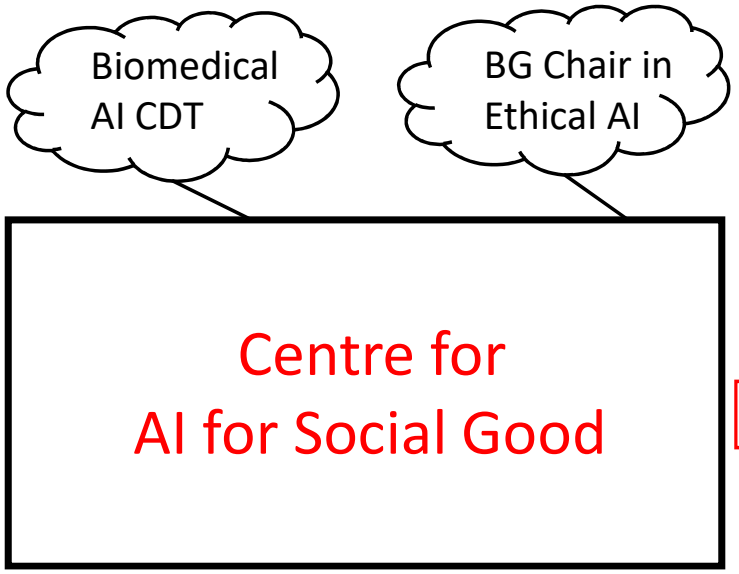
- international development,
- social inclusion and
- sustainability.

THE UNIVERSITY of EDINBURGH informatics

Centre for AI for Social Good

- Motivated staff and students
- Research Opportunities
- Philantropic donations

Brings in



Sends out

- Well-informed graduates
- Understandable public profile
- Research output
- Influence in society and government

Focusing existing research activity within the School and stimulating new activity

AI Audit

Many organisations are espousing good “values” for their AI, eg Microsoft state:

- Fairness
- Inclusiveness
- Reliability and safety
- Transparency
- Privacy and Security
- Accountability

Other organisations variously also mention:

- Robustness
- Interpretability/Explainability

Statements are easy….the need for audit

Whilst intentions are good, there is not yet in place a robust mechanism for checking whether these stated goals are achieved.

This is why we are exploring the idea of AI Audit.

Not yet fully developed but the idea is that Ethics, may inform decisions about what an organisation is setting out to achieve, but for confidence there needs to external assessment that stated values are being adhered to.

This is likely to require precise definitions of the planned properties or values as well as techniques for ensuring that those properties are achieved

Focus on impact on external beneficiaries rather than development methods.

Thank you!