How to Build Trust in AI: Human-Centered Design of Trusted AI Smart Agents for Collective Intelligence

Prof Phil Morgan: Human Factors & Cognitive Science

School of Psychology, Cardiff University, UK

morganphil@cardiff.ac.uk AI for Collective Intelligence (AI4CI) |

Co-Director – Human Factors Excellence Research Group (HuFEx)

Director of Research – Centre for AI, Robotics, & Human-Machine Systems (IROHMS)

Transportation Lead: CU Digital Transformation Innovation Institute (DTII)

Director – Airbus Centre of Excellence in Human-Centric Cyber Security & Co-Director (H2CS) & Cardiff University Partnership

Guest Professor – Luleå University of Technology; University of Canberra, Australia

School of Psychology

HuFEx

Human Factors Excellence Research Group

HuFEx

Augmented & Assistive Systems
Cyberpsychology
Defence & Security
Emergency Services & Healthcare
Humans in AI & Automation
Transportation Human Factors

14 staff (HF, cog sci, social cog, neuroscience)
12 PhD students (AI, automation, cyber security, defence, emergency services, HRI, transport)

~£15m (30+ grants 2017+); ~£40m grants under review;
Phil Morgan ~£40m (50+ grants).

IROHMS

Human-like AI

- Affective computing
- Augmented cognition
- Computational semantics
- Contextual reasoning

Ethical and Explainable AI

- Ethical AI
- Explainable AI
- Explainable robotics
- Trusted autonomy

Human-centred Technologies and Society

- Human-centred computing
- Human-centred cyber security
- Emerging technology and society

Humans and Robots

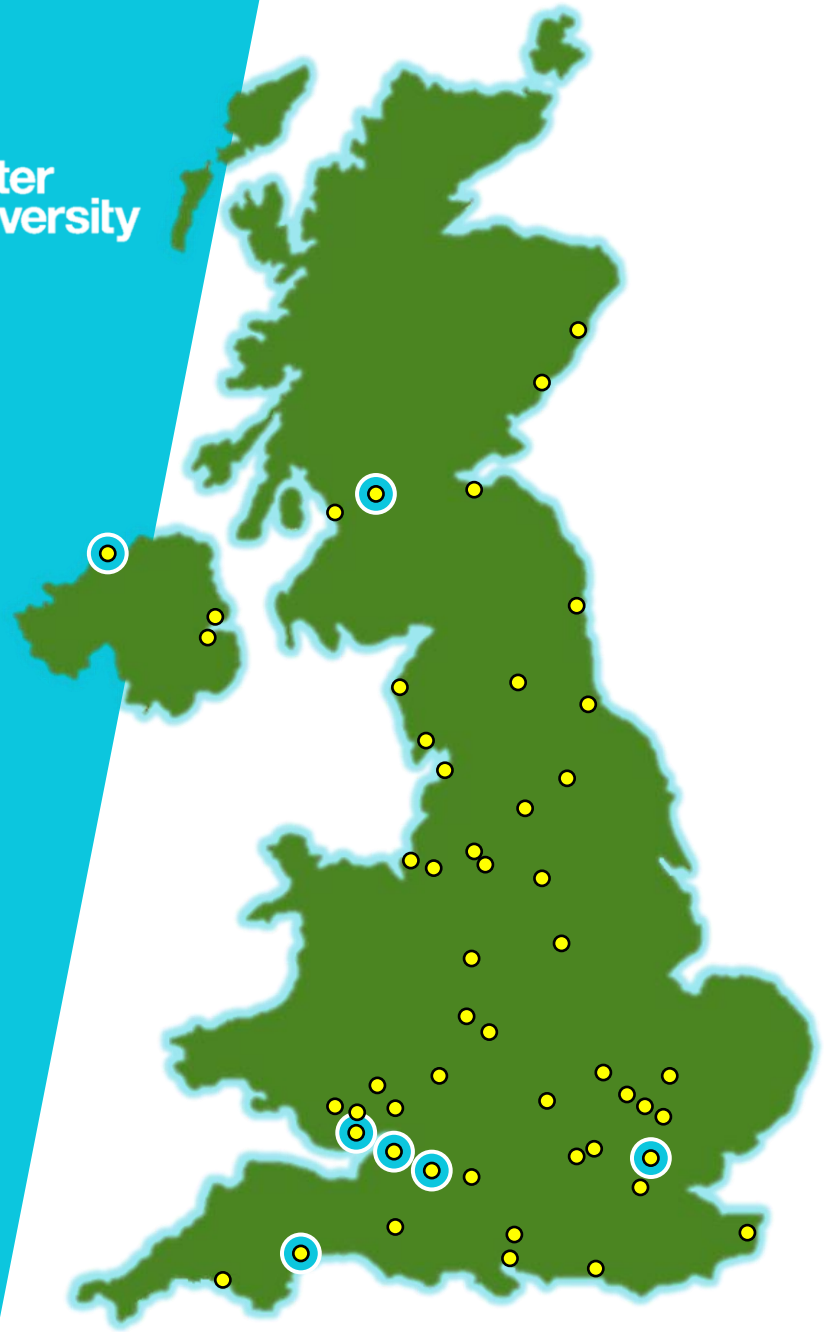
- Human-centred robotics
- Social robotics
- Robot perception/learning

https://www.cardiff.ac.uk/artificial-intelligence-robotics-and-human-machine-systems

University of Bath
University of Bristol
Cardiff University
Ulster University

AI4CI

AI for Collective Intelligence



University of Exeter
University of Glasgow
UCL

AI4CI Management Team

Vicky Walter

Hub Manager

Lauren Clarke

Partnership Manager

Amy Ardon

Hub Administrator

Nirav Ajmeri

Deputy Director

(Intelligent Agents)

Cangxiong Chen

ECR Lead

(Machine Learning)

Seth Bullock

Hub Director

(Simulation Modelling)

AI4CI Investigators

Maggie Chen
(Cardiff)
Theme Lead
Finance

Phil Morgan
(Cardiff)
Theme Lead
Design

Michaela Black
(Ulster)
Theme Lead
Governance

Joan Condell
(Ulster)
Theme Lead
Governance

Hywel Williams
(Exeter)
Theme Lead
Environment

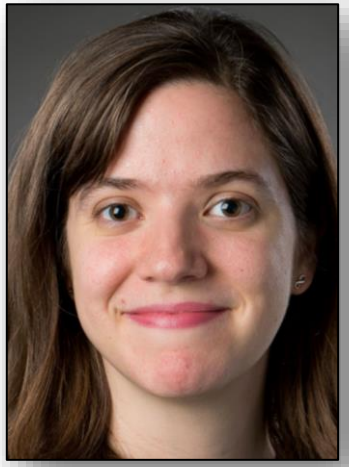
Alison
Heppenstall
(Glasgow)
Theme Lead
Smart Cities

Adam Dennett
(UCL)
Theme Lead
Smart Cities

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

School of Psychology

AI4CI Investigators



Theresa Smith
(Bath)
Theme Lead
Healthcare

Cangxiong Chen
(Bath)
ECR
Lead

Laura Smith
(Bath)
Theme Lead
Design

Nirav Ajmeri
(Bristol)
Deputy
Director

Aisling O'Kane
(Bristol)
Theme Lead
Healthcare

John Cartlidge
(Bristol)
Theme Lead
Finance

Lenon Danon
(Bristol)
Theme Lead
Pandemics

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
School of Psychology

The Wider UK AI Ecosystem

Northumbria University NEWCASTLE

MANCHESTER 1824

Imperial College London

AI

University of BRISTOL

UNIVERSITY OF OXFORD

CARDIFF UNIVERSITY

PRIFYSGOL CAERDYDD

RAiUK

The Alan Turing Institute

open data institute

Cabinet Office

Office for Artificial Intelligence

FOUNDATION MODEL TASKFORCE

NetZero

GALLANT

Ada Lovelace Institute

Our EPSRC AI4CI Hub (~£12m UK, 5-Years)

School of Psychology

Director: **Seth Bullock**
Deputy Director: **Nirav Ajmeri**
ECR Lead: **Cangxiong Chen**

A. Heppenstall (Glasgow)
Adam Dennett (UCL)
AI for Smart City Design
With: Arup, UK Govt. …

Leon Danon (Bristol)
AI for Pandemic Resilience
With: Pfizer, DHSC, NHS, Sanofi, Cabinet Office

Academic Pool: 50+ AI/ML co-supervisors from across six AI4CI partner institutions.

Laura Smith (Bath)
Phil Morgan (Cardiff)
Human-Centred Design

Michaela Black (Ulster)
Infrastructure & Governance

AI4CI CORE
Community & Capacity
Cross-Cutting AI/ML
EDI Team
Engagement Team

Hywel Williams (Exeter)
Environmental Intelligence
With: Wikipedia, PML, Met Office, IBM

External Advisory Board
Science Council

John Cartlidge (Bristol)
Maggie Chen (Cardiff)
AI for Financial Stability
With: HSBC, Coinbase…

Aisling O'Kane (Bristol)
Theresa Smith (Bath)
Joan Condell (Ulster)
AI for Health Ecosystems
With: NHS, Mayden

AI4CI Hub

What Will Our Hub Do?

Swarm Intelligence

Collective Behaviour

Complex Systems

Organisational Behaviour

AI for Collective Intelligence

Collective Dynamics

Economics

(Behavior Change at Scale)

Robotics

Cognitive Psychology

Multi-Agent Systems

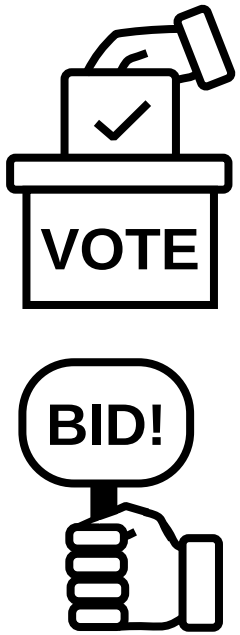
Social Psychology

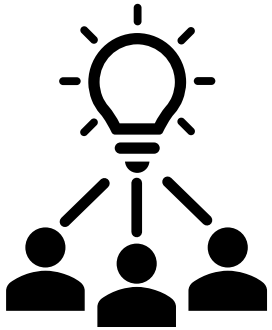
AI

Animal Behaviour

AI4CI Vision

Collective Intelligence: Pool diverse experiences for centralised decision support.

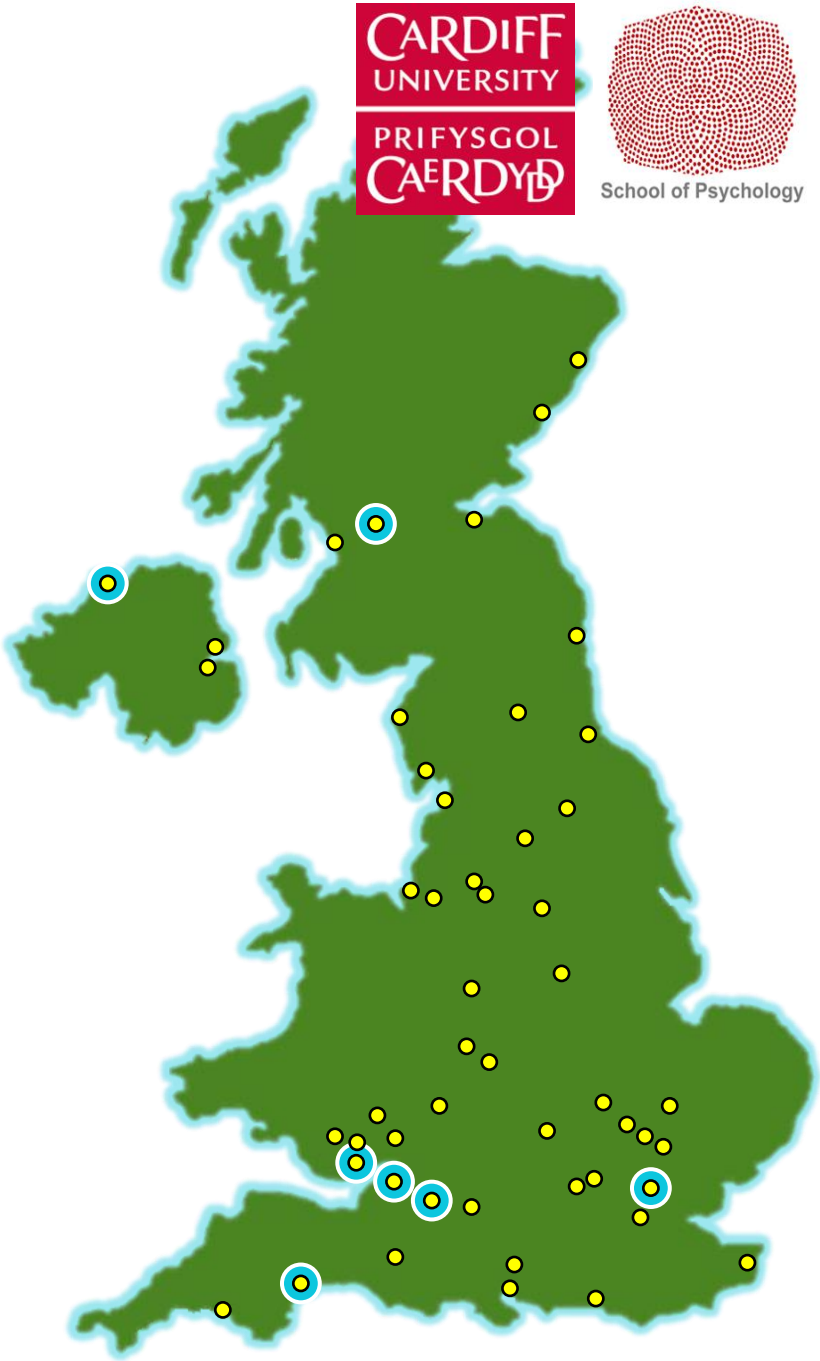




Dataminr

citizenlab

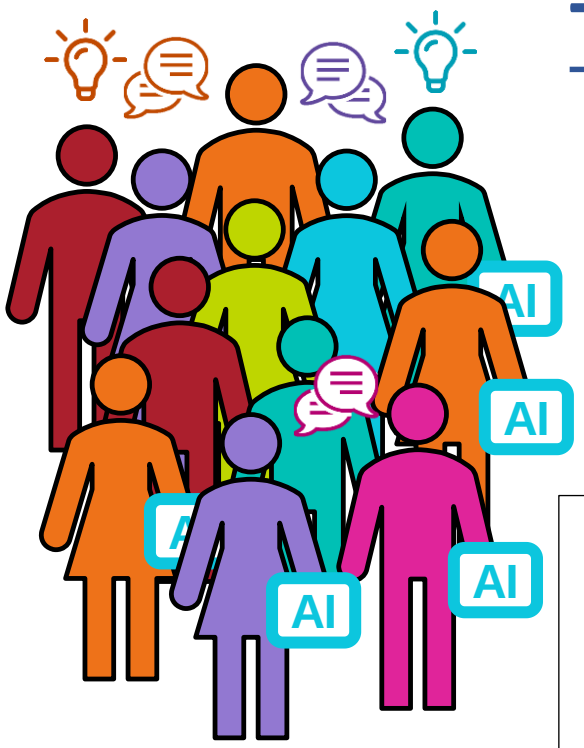
FACTMATA



AI4CI Vision

AI For Collective Intelligence – the loop:

1. Design Smart Agents with end-users in mind – key focus on trust



The AI4CI Loop

Gather Intelligence

2. 'Speaking' to the Smart Agents with end-users in mind

ML Analytics

Collective Intelligence Policies

Smart Agents

3. Implement Smart Agents with end-users in mind

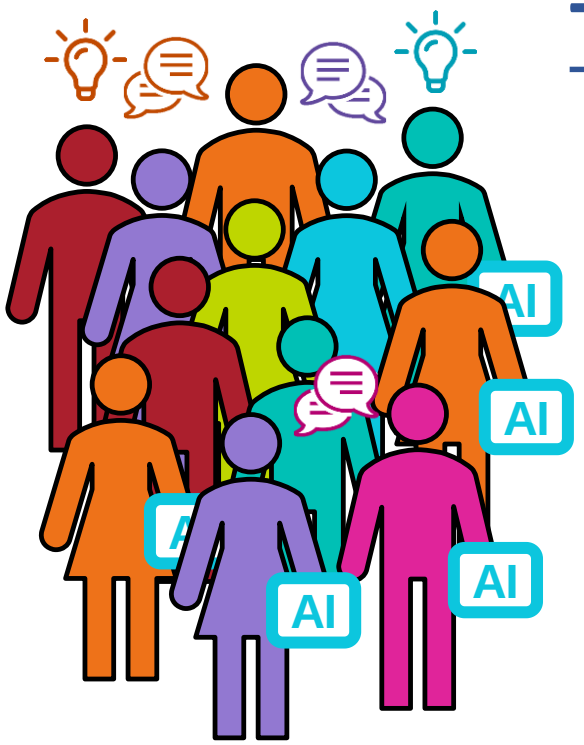
Inform Behaviour

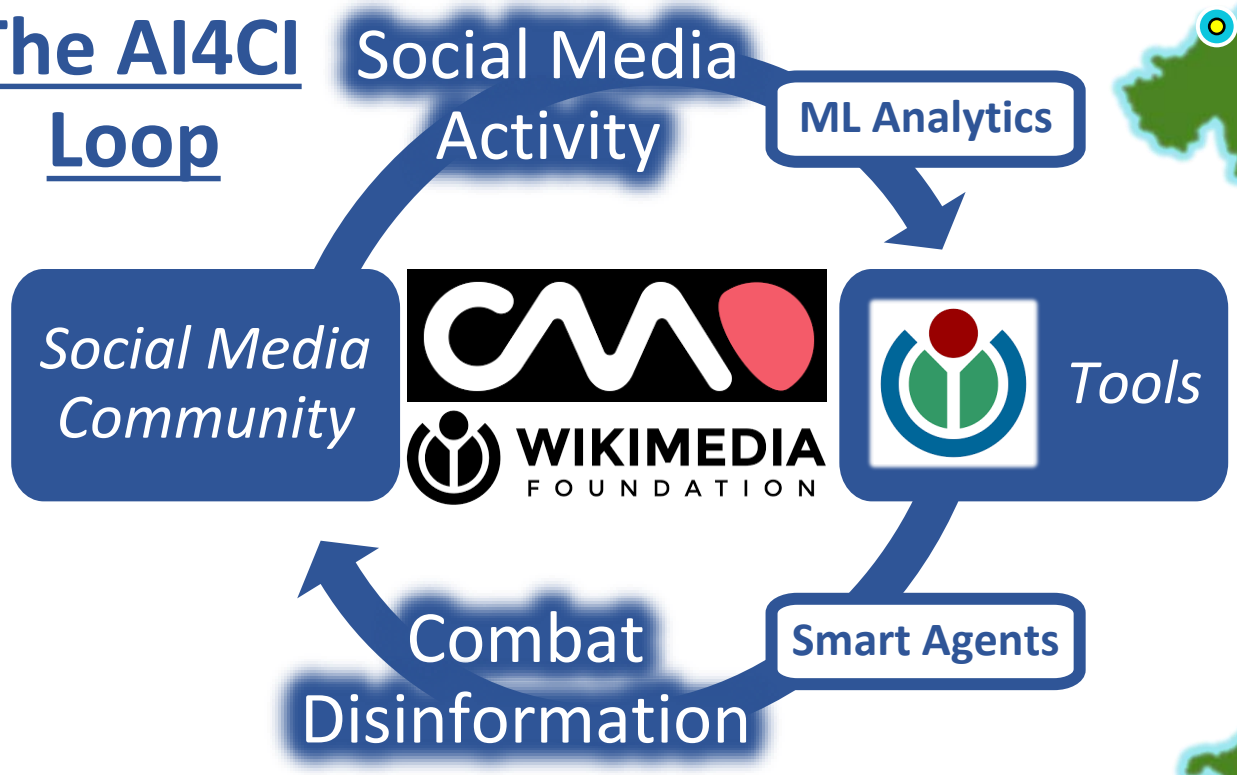
4. Iterate (based on findings) Smart Agents with end-users in mind

Real World Collective System

Example: Environment / Climate

- *Federated and Privacy-Preserving ML*
- *Anomaly and Change-Point Detection*
- *Ensuring Trust and Algorithmic Fairness*

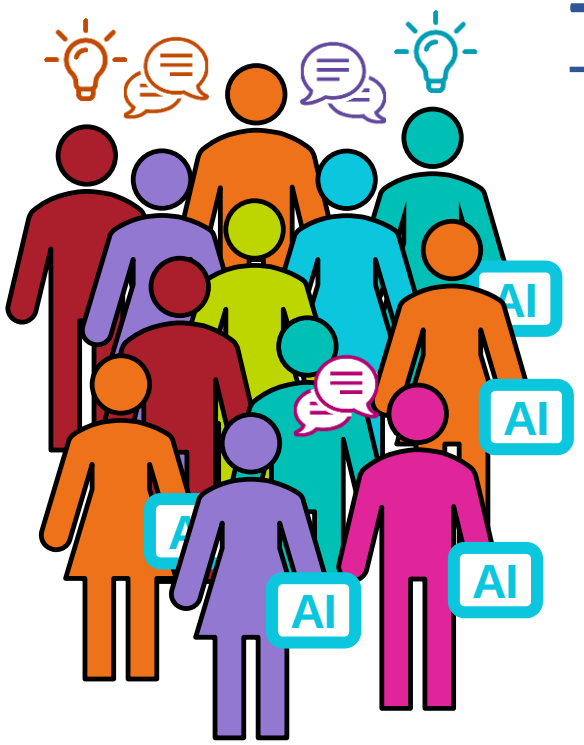


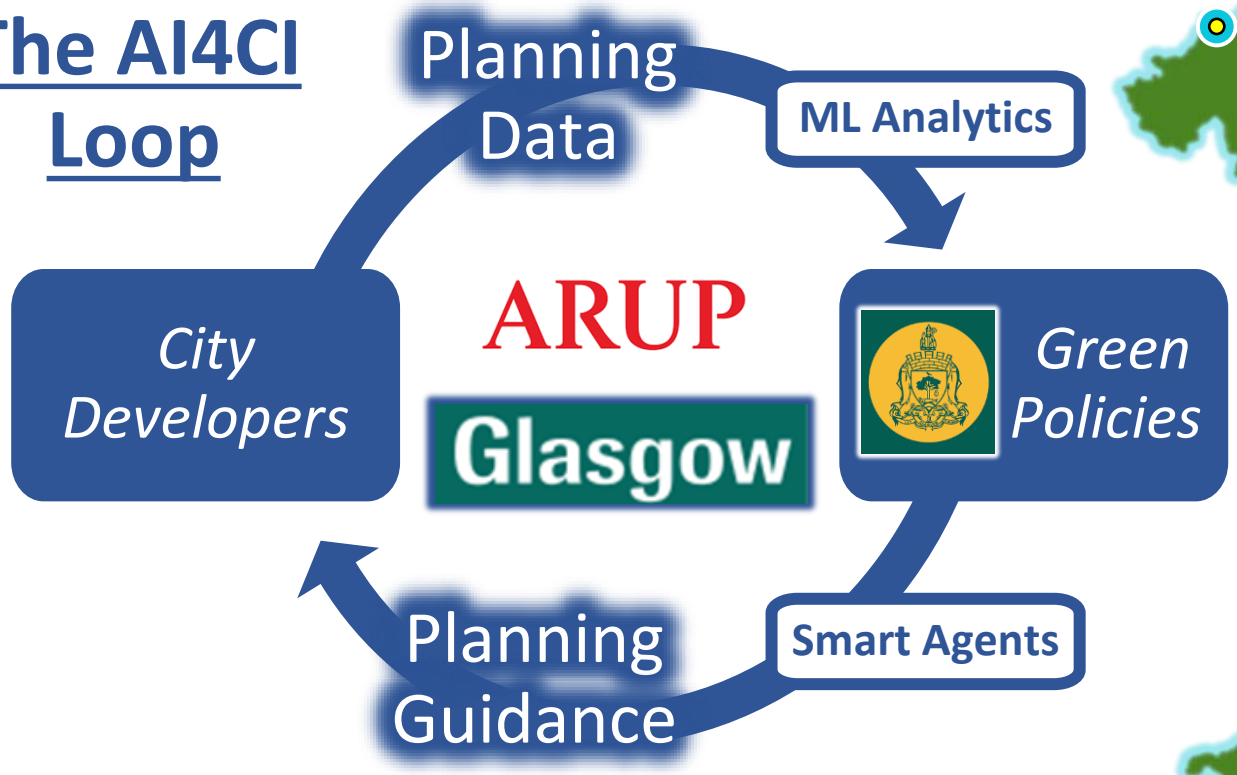


CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
School of Psychology

Example: Cities

- *Federated and Privacy-Preserving ML*
- *Anomaly and Change-Point Detection*
- *Ensuring Trust and Algorithmic Fairness*





CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
School of Psychology

Human-Centered Design

Prof Phil Morgan (Cardiff), Prof Laura Smith (Bath), Dr Sachini Weerawardhana (Bath), Dr Thom Griffith (Cardiff), Jessica Matthews (Cardiff – Placement Student)

Goal: Optimise people in 'Smart' collectives

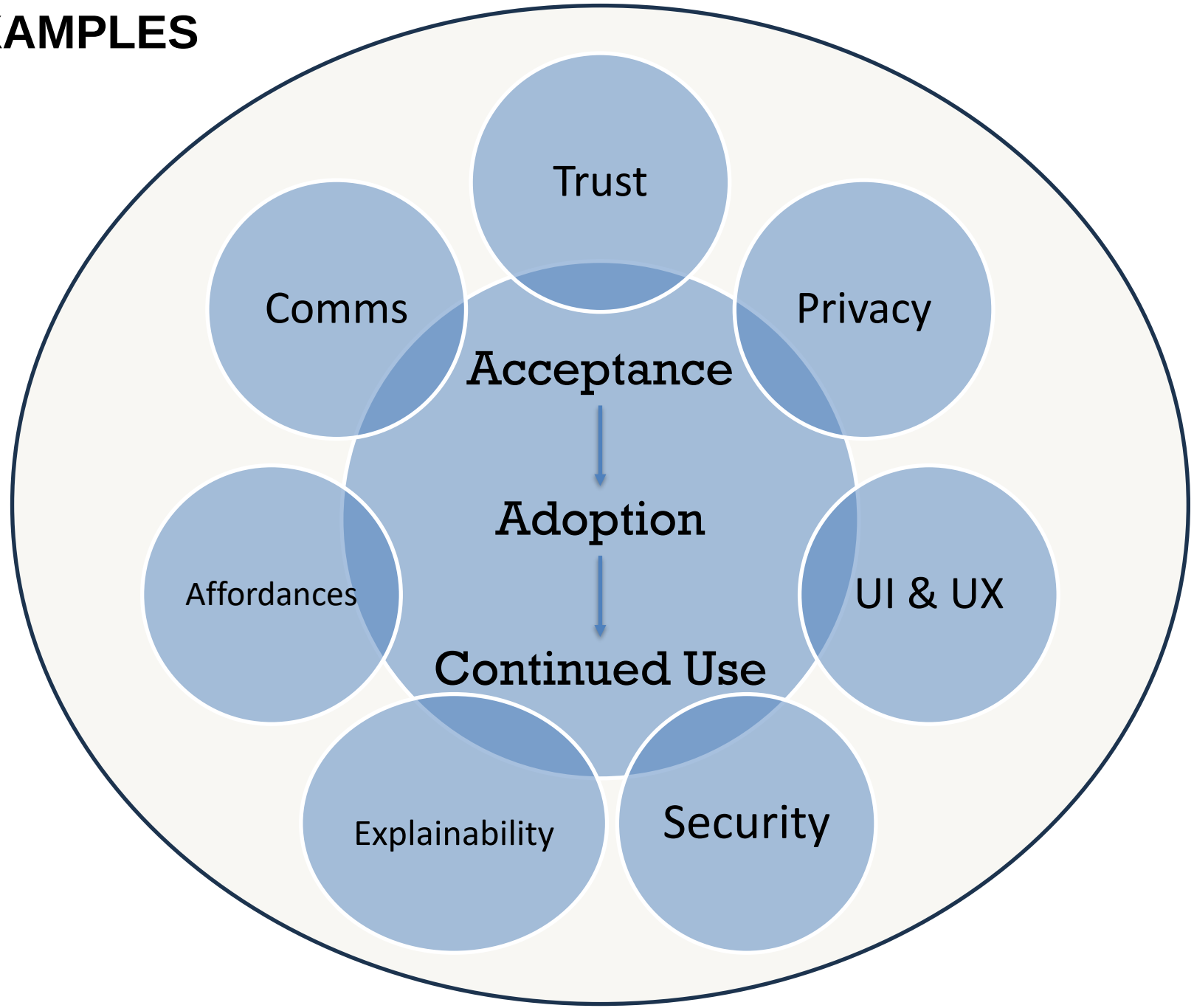
1. ***Review & Participatory Input:*** Embed methods & insights (from psych, HF) into smart agent design: reviews; workshops; recommendations; prototypes. Key focus on e.g. trust, accessibility, usability, acceptance, adoption.

2. ***Embed & Test:*** Design Smart Agents for optimal outcomes - *people* become integrated & dynamic component of usable & trustworthy Smart collective AI systems.
   i. Measurement toolkit
   ii. Calibrate Trust

3. ***Understand:*** How Smart agents can affect behaviour changes, at scale. Focus on trust – e.g. thresholds, tipping points, optimisation

School of Psychology

EXAMPLES



School of Psychology

Experiments: Objective & Subjective Measures

UNIVERSITY OF BATH

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

School of Psychology

Bath

Bath Institute for Digital Security and Behaviour

Cardiff: AI, Robotics & Human-Machine Systems - Human Factors Simulation Laboratory

Igloo 6m VR & Ambisonics Cave

Command & Control Centre

Transport Sim + Motion Platform

Social & Cognitive Robotics

VR, Eye Tracking, EMG, EEG

CyberRith & Virtualizer Platforms

Thank you!
And…AI4CI Global…

15th Nov 2025 - ENFIELD Webinar 2024

morganphil@cardiff.ac.uk

How to Build Trust in AI: Human-Centered Design of Trusted AI Smart Agents for Collective Intelligence

School of Psychology

Prof Phil Morgan: Human Factors & Cognitive Science

School of Psychology, Cardiff University, UK

morganphil@cardiff.ac.uk *https://ai4ci.ac.uk/*

HuFEx

Human Factors Excellence Research Group