AI ethics from the ground up:

Cultivating interdisciplinary capabilities (for care)

Cian O'Donovan UCL

Reflections from ongoing collaborations with **Carlos Cuevas-Garcia** Technical University Munich, **Ola Michalec** University of Bristol, **Joshua Moon** University of Sussex, **Mehdi Sobhani** Bristol Robotics Laboratory

c.o'donovan@ucl.ac.uk | @cian | slides at cianodonovan.com

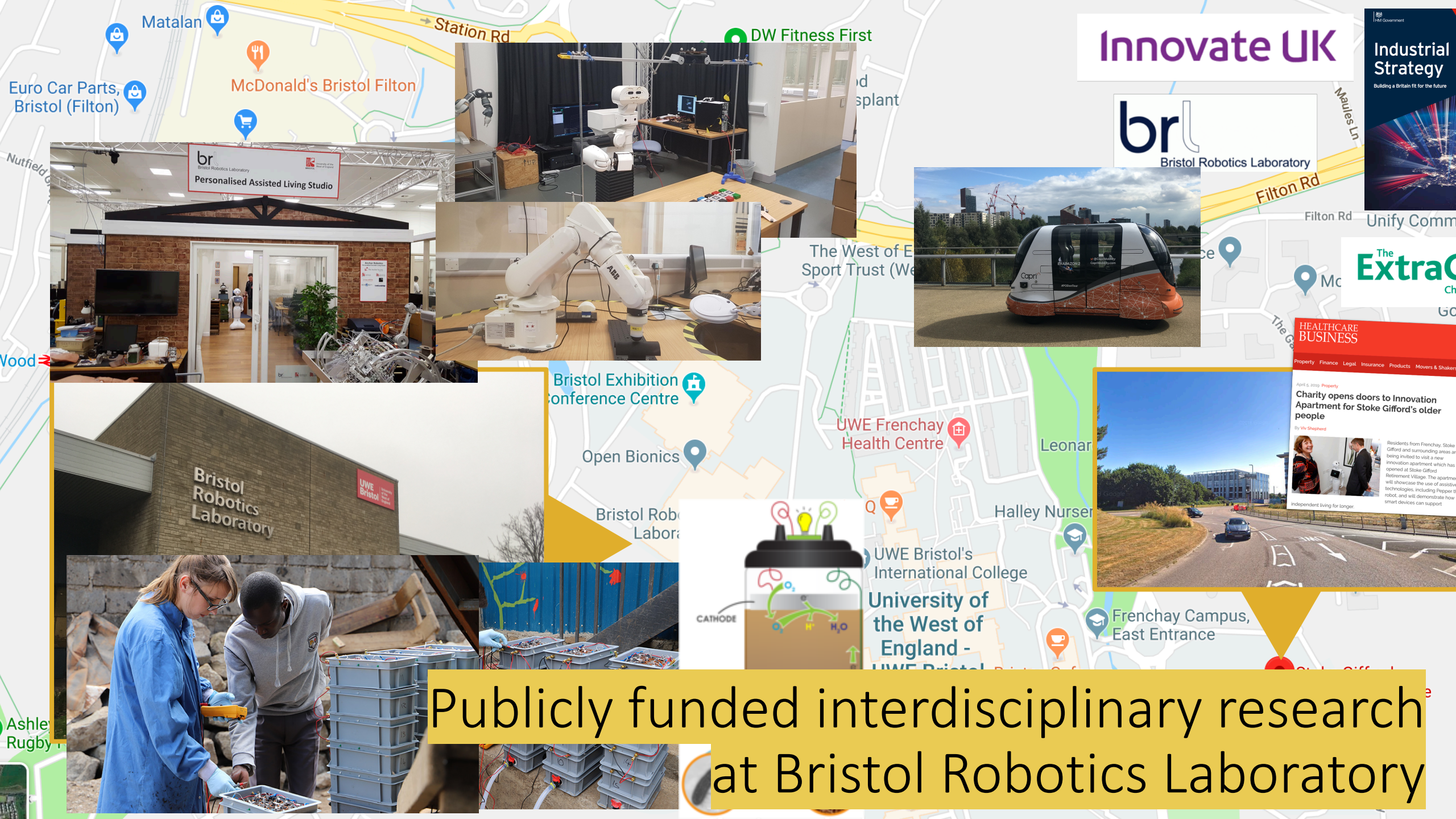
How to use AI for good | #WebSci20 | Southampton | July 7th 2020

SCALINGS SCALING UP CO-CREATION

STS UCL STS

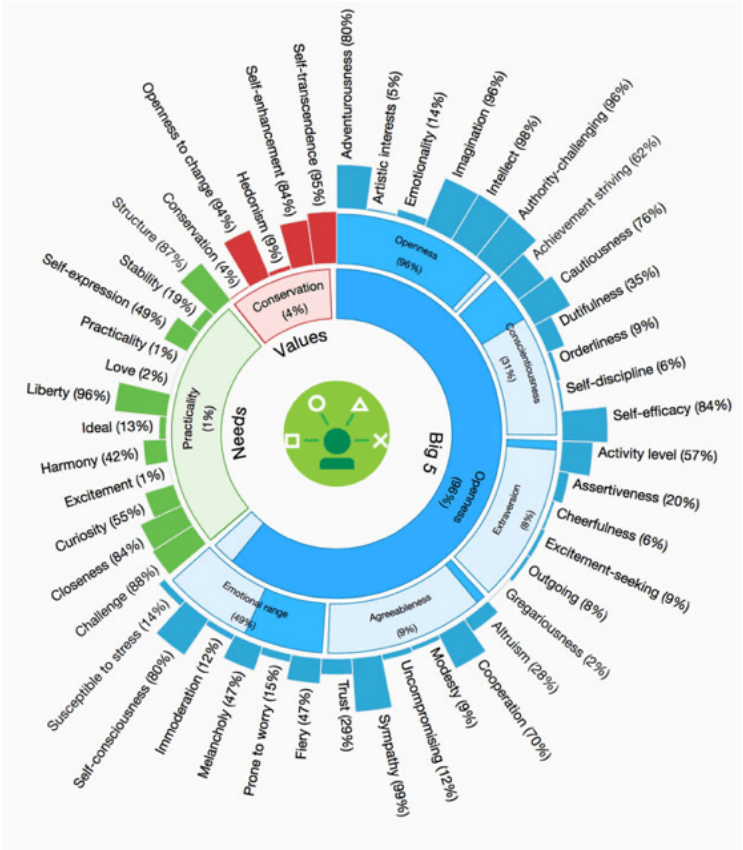
UCL

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 788359



Publicly funded interdisciplinary research at Bristol Robotics Laboratory

Interdisciplinary AI research – a public good?



1. The problem – the ethics of interdisciplinary research

- Interdisciplinary research is suggested as an answer for **societal challenges**
- But what KIND of interdisciplinary research
- And how do we assess and evaluate it
- How do researchers (and others) make the difficult decisions identified by Will
  - Going beyond ethics frameworks

Alan Winfield's Web Log

Mostly, but not exclusively, about robots

Home | Media | Robotics Q&A | Robotics: a very short introduction | The Ethical Roboticist | Stuff I've Invented

Thursday, April 18, 2019

An Updated Round Up of Ethical Principles of Robotics and AI

This blogpost is an updated round up of the various sets of ethical principles of robotics and AI that have been proposed to date, ordered by date of first publication.

I previously listed principles published before December 2017 here; this blogpost appends those principles drafted since January 2018 (plus one in October 2017 I had missed). The principles are listed here (in full or abridged) with links, notes and references but without critique.

Scroll down to the next horizontal line for the updates.

If there any (prominent) ones I've missed please let me know.

Asimov's three laws of Robotics (1950)

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

I have included these to explicitly acknowledge, firstly, that Asimov undoubtedly established the principle that robots (and by extension AIs) should be governed by principles, and secondly that many subsequent principles have been drafted as a direct response. The three laws first appeared in Asimov's short story Runaround [1]. This wikipedia article provides a very good account of the three laws and their many (fictional) extensions.

Murphy and Wood's three laws of Responsible Robotics (2009)

@alan_winfield

Tweets by @alan_winfield

Blog Archive

- ► 2020 (2)
- ▼ 2019 (9)
  - ► September (1)
  - ► July (2)
  - ► June (1)
  - ► May (1)
  - ▼ April (1)
    - An Updated Round Up of Ethical Principles of Robot...
  - ► March (1)
  - ► February (1)
  - ► January (1)
- ► 2018 (9)
- ► 2017 (10)
- ► 2016 (12)
- ► 2015 (9)
- ► 2014 (15)

n > 20

O'Donovan, C. (2020) *Explicitly ethical standards for robotics*. Brighton, UK. Available at: https://www.sussex.ac.uk/webteam/gateway/file.php?name=odonovan-2019-robotics-standards-210210.pdf&site=25.

2. What's at stake in this problem

- **Technology as treatment**
  - neglects the voices, values and interests of a wide range of people and communities
- Innovation policy tends to narrowly focus on **inducing acceptance**
- But **technology can be steered** in different directions
  - and the benefits distributed to a variety of people and communities (Weber and Rohracher 2012; Stirling 2009).

Image: https://www.idiap.ch/workshop/iros2018/images/participants-1.jpg/@@images/1980d26f-41c7-415f-a5de-620a52331dc1.jpeg

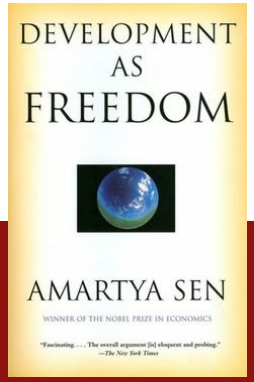
3.1 From *ethics* to *politics* of interdisciplinary research

- Robots from BRL shape society – society also shapes robots at BRL (co-production)
  - How this happens is a matter of politics
- Three ways we can think about the politics of interdisciplinary research for robotics and ai
  - **Interdisciplinarity reach** (wide vs narrow) Kelly 1996
  - **Discourses and logics of interdisciplinarity** Barry et al. 2008
  - **Capabilities for research** O'Donovan et al. forthcoming

O'Donovan, C., Michalec, O. and Moon, J. R. (forthcoming) *Capabilities for transdisciplinary research. An evaluation framework and lessons from the ESRC Nexus Network+*. SPRU Working Paper Series, Brighton, UK.

Michalec, O., Sobhani, M. and O'Donovan, C. (forthcoming) 'What is robotics made of? The politics of interdisciplinary robotics research'.

3.2 Capabilities for interdisciplinary research



- The capability to *do* research and *be* a researcher (focus on human agency/well-being)
- Capabilities required / cultivated during interdisciplinary research. For example:
  - Cognitive capabilities such as expert disciplinary knowledge
  - To manage a research team
  - for pluralism
  - To collaborate
  - To be humble
  - To be reflexive
  - To build democratic struggle (?!?!)

O'Donovan, C., Michalec, O. and Moon, J. R. (forthcoming) *Capabilities for transdisciplinary research. An evaluation framework and lessons from the ESRC Nexus Network+*. SPRU Working Paper Series, Brighton, UK.

4. Locating capabilities at BRL

Bibliometric research profile illustrating key disciplines represented in BRL publications indexed in Web of Science.

(485 publications; 2004-2020)

Bibliometric map yielded the following "disciplinary" clusters:

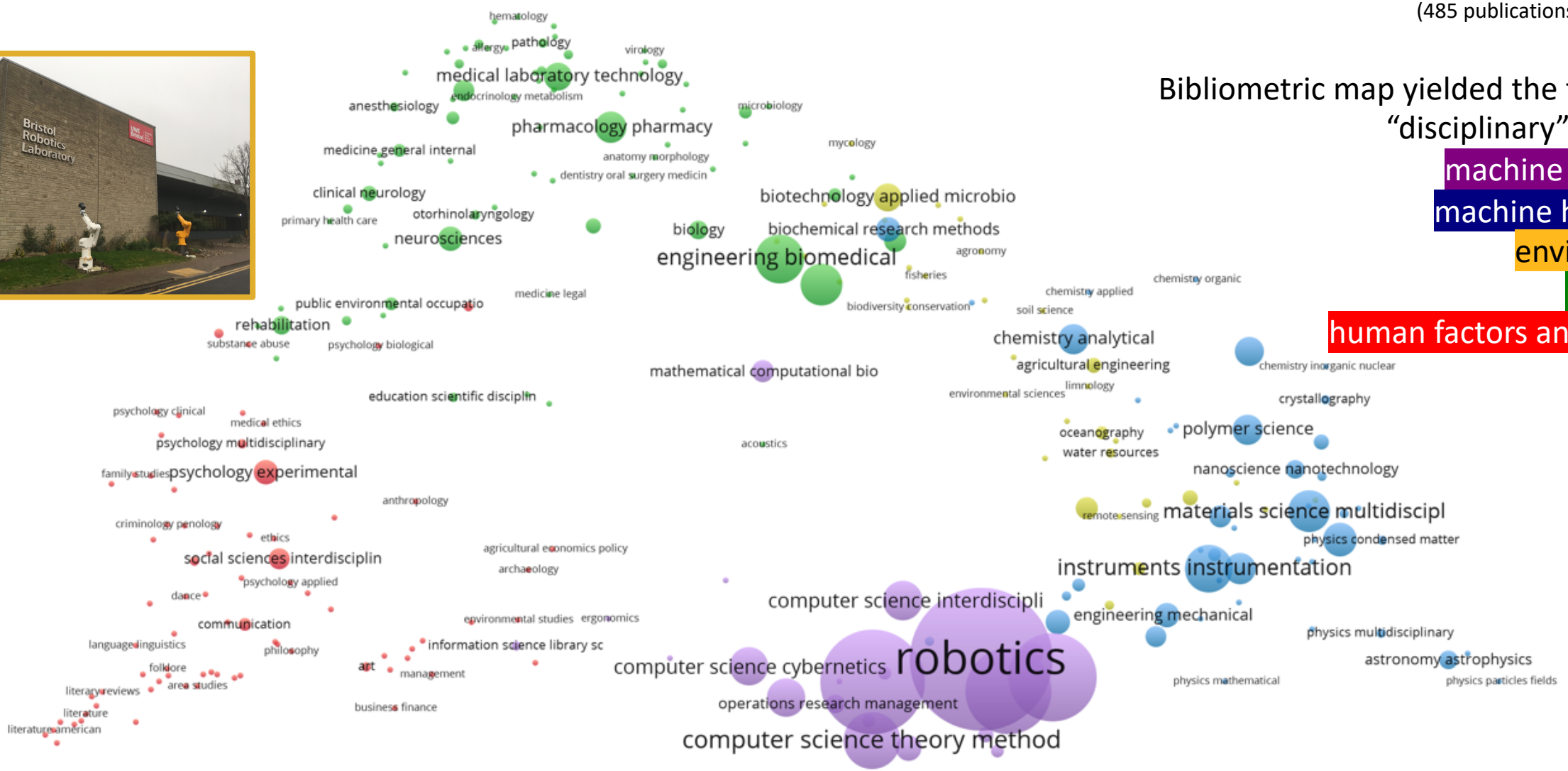
machine software

machine hardware

environment

medicine

human factors and society



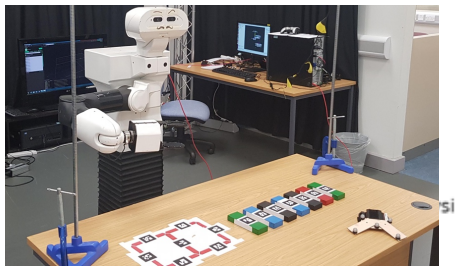
Michalec, O., Sobhani, M. and O'Donovan, C. (forthcoming) 'What is robotics made of? The politics of interdisciplinary robotics research'.

4. Locating capabilities at BRL

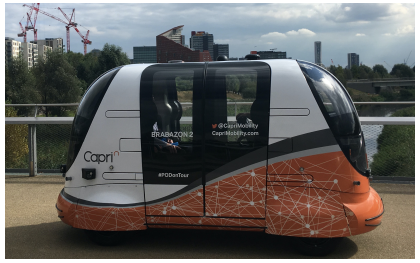
Bibliometric research profile illustrating most prominent key words represented in BRL publications indexed in Scopus

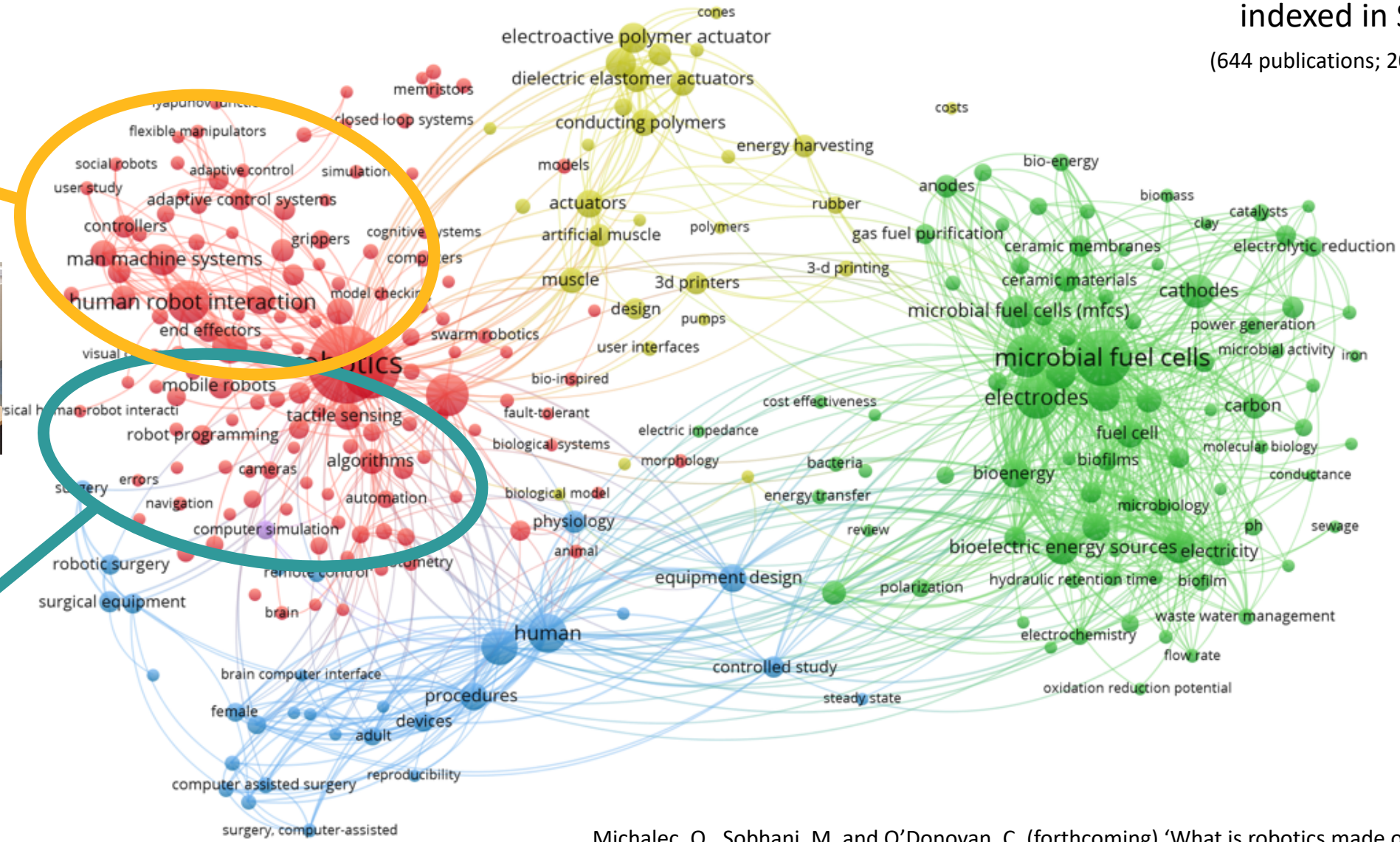
(644 publications; 2004-2020)

Assistive living robotics



Self-driving vehicles

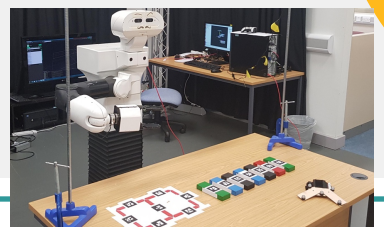




Michalec, O., Sobhani, M. and O'Donovan, C. (forthcoming) 'What is robotics made of? The politics of interdisciplinary robotics research'.

5. Assessing capabilities at BRL

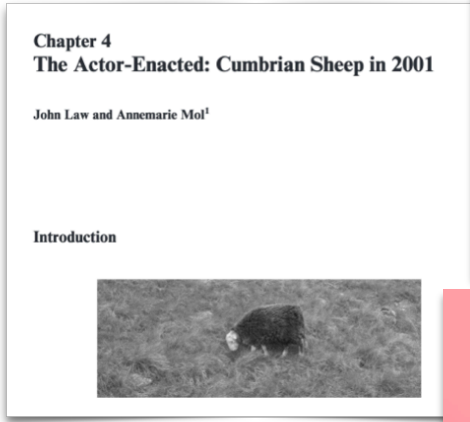
	Assisted Living robotics	Driverless vehicles (CCAV projects)
Societal challenge	Maintaining care for more elderly Current and future labour shortages	Mobility for the future Road safety
Scope of interdisciplinary research	**Wide** - Expert researchers, civil society organizations, users may sometimes ask research questions	**Medium** - Multi rather than inter-disciplinary; Users are typically subjects of research
Logics of interdisciplinary research	Innovation as **crisis response**; **Market** creation; Interdisciplinarity as **accountability**	**Testing** innovation (governance of emergent technology); **Market** growth
Capabilities noted	Capabilities to **broaden participation** in research; **Infrastructure** maintenance Interdisciplinary **network building**; Capabilities to build a diverse network of **regional stakeholders**	**Accelerate** innovation (steer innovation); Capabilities to work with **large consortia**; Capabilities to **draw in internal resource** Capabilities to **win CCAV funds** (Large consortia are blunt instruments – more precise steering required) (**Reflexivity; accountability; democracy**)



Michalec, O., Sobhani, M. and O'Donovan, C. (forthcoming) 'What is robotics made of? The politics of interdisciplinary robotics research'.

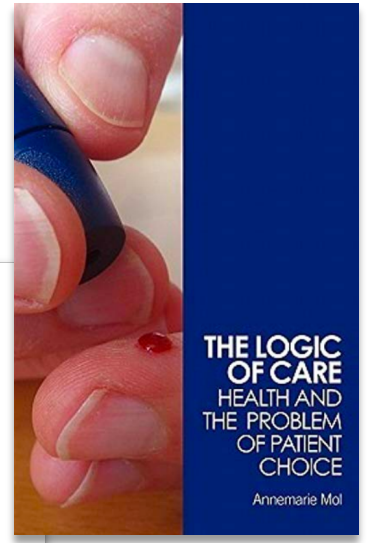
6. Some implications: from societal challenges to societal collaboration

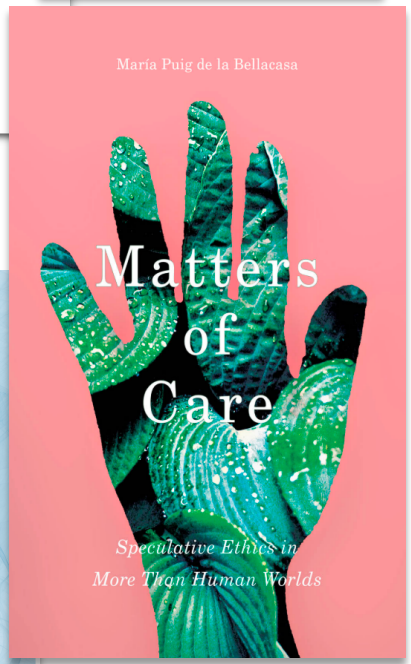
- Putting practices before principles of AI research
  - Applicable for doctoral training, curricula design, research mgmt
- Steering the direction of research
  - funders, government, civil society, communities
- Capability mapping: framework for ethics/values from ground up
- Emphasizing **care (not treatment) in capabilities for grand challenges**:
  - Attention to neglected things and devalued doings (e.g. carers and cleaners)
  - Both means and ends of methods matter
  - Knowledge AND values
  - Recognise the inherent uncertainty technology introduces


Chapter 4
The Actor-Enacted: Cumbrian Sheep in 2001

John Law and Annemarie Mol[1]

Introduction





Thanks for staying with it

Cian O'Donovan
c.o'donovan@ucl.ac.uk. | @cian

www.scalings.eu

Slides at cianodonovan.com

SCALINGS SCALING UP CO-CREATION

STS UCL STS

UCL

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 788359

Post script: cultivating human capabilities for care

- Many possible configurations of people, knowledge and things
- Capabilities for care help locate, assess and address tensions and choices between:
  - Strong emergence of new public management ideal which comes with a logic of **clear cut choices** VS the need to allow more open-ended processes which follow a **logic of care**
  - Between the ideal of **focused and controlled futures** VS the fact that we need to understand the future as more open for exploration involving a **variety of visions**
  - Between the focus on **invited participation**
    - **Ready made technoscientific futures** VS the reality of context dependent, **ever changing futures-in-the-making**
  - Between the strong imaginary of a community of **homogenous values** VS a reality of a growing **diversity of values**, each asking for a voice
  - Between normatively pre-established assessment structures VS more open valuing processes

Care

- Attention to neglected things and devalued doings (e.g. carers and cleaners)
- Means and ends of methods considered
- Knowledge AND values considered
- Recognise the inherent uncertainty technology introduces