Ten Problems in Industrial AI Ethics

Joanna J. Bryson

Artificial Models of Natural Intelligence

University of Bath, United Kingdom

Mannheimer Zentrum für Europäische Sozialforschung

MZES

Outline

- Non-Problems

- Problems

UK EPSRC's Principles of Robotics (2011)

1. **Robots are multi-use tools.** Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.

2. **Humans, not robots, are responsible agents.** Robots should be designed & operated as far as is practicable to comply with existing laws & fundamental rights & freedoms, including privacy.

3. **Robots are products.** They should be designed using processes which assure their safety and security. (of 5...)

UK EPSRC's Principles of Robotics (2011)

4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.

5. The person with legal responsibility for a robot should be attributed. [like automobile titles]

Outline

- Non-Problems

- Problems
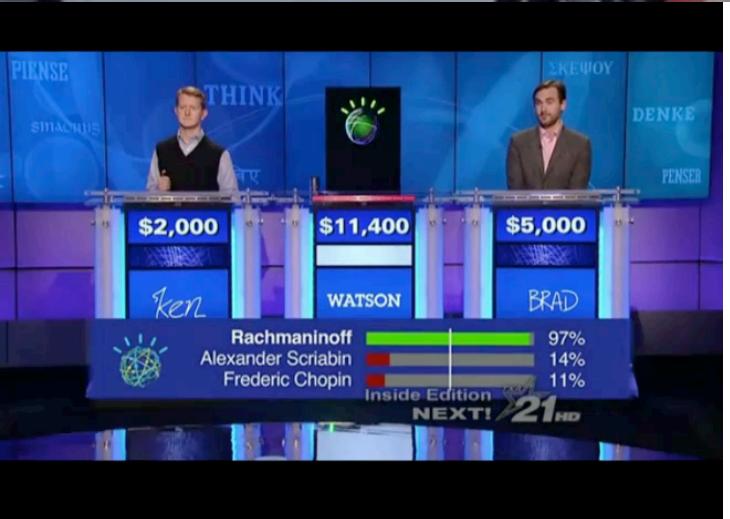
-3 Intelligence

big dog
Watson
driverless cars

Big data + Bayesian-based machine learning has successfully mined human intelligence in a broad range of domains.

-2 Reliability

Industrial robots are capable of thousands of operations a day, for years at a time.

-1 Moral Patiency
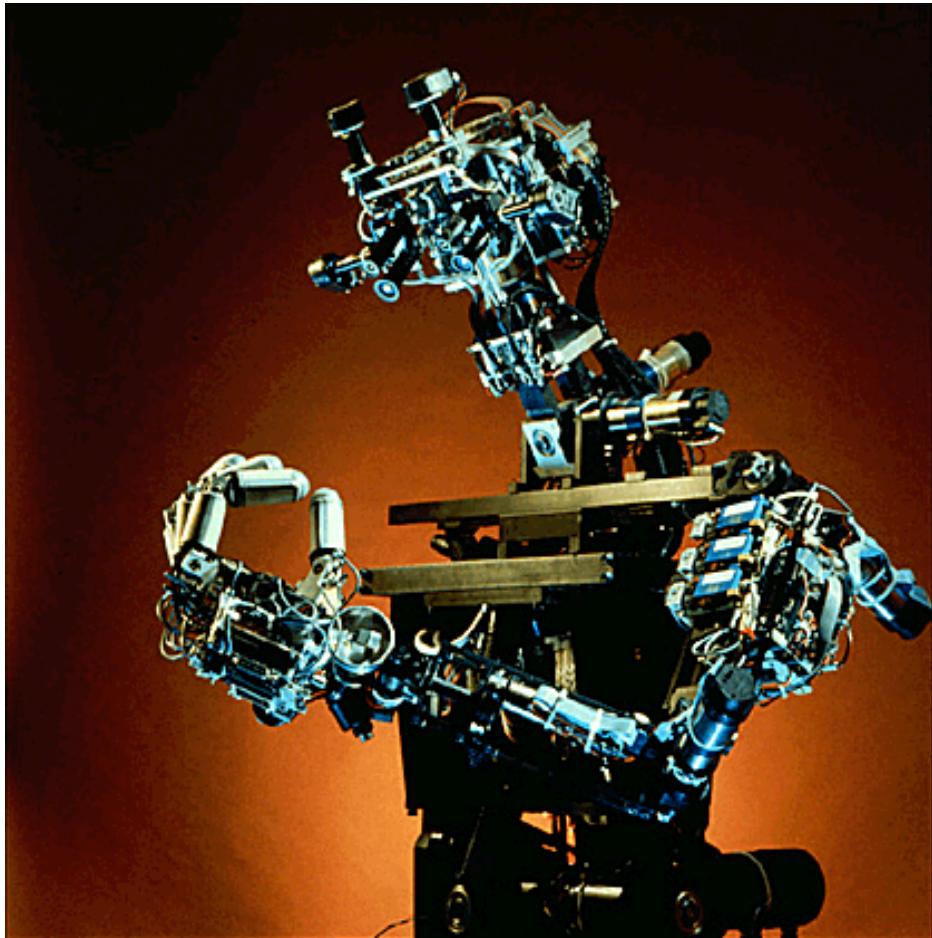
Both robots and ethical systems are artefacts of our culture.

There's no predetermined place for robots in our ethical system.

We are responsible for robots as their creators, making ethical obligation to them would be unethical.

Outline

- Non-Problems

- Problems

1 Privacy
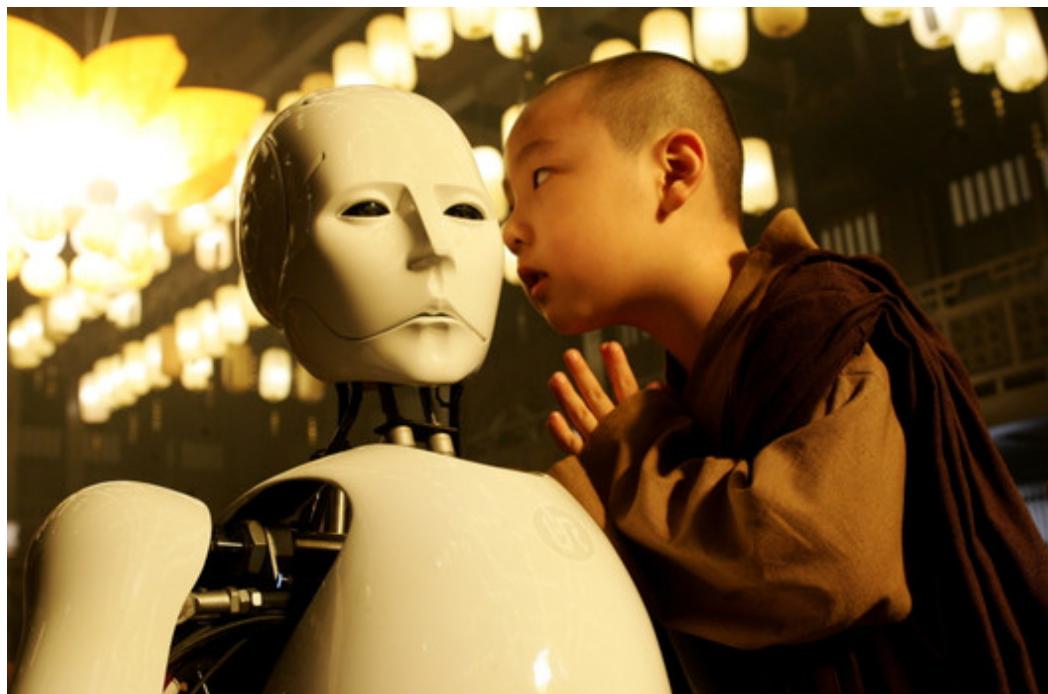
Doomsday Book

Networked gadgets with cameras, microphones, accelerometers that can learn your daily and exceptional patterns.

Note: already a problem with phones.
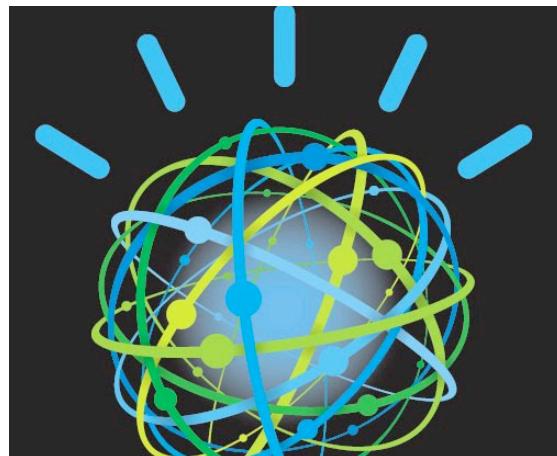
2 Redundancy

Not just robot office cleaners: AI scientists & lecturers

Note: obsolescence already a problem, millions of people banned from work

3 Liability

Probably less of a problem than expected – either operator or manufacturer.

Note: EPSRC Principles deliberately reduce problem to the same as for cars, guns, heavy machinery.

4 Autocracy

Say we can drop deaths per year by 90%. But it will be a different 3,500 Americans who die next year. And we can't go the wrong way on one way streets ever.

Note: Already a problem with automated ticket machines.

5 Isolation / Asociality

AI may be more entertaining than anyone in your village, or at least entertaining enough and less likely to laugh.

Note: Already a problem with books, plays, movies, TV… but may alter behaviour earlier. But may be fun.

6 Sustainability

Sell lots of power-hungry plastic, metal, rare-earth stuff.

Note: Already a problem with computers & phones.

7 Hackers

The flip side of privacy: actions not just sensing.

Note: Actually a new problem?

As people who know the value of morphology / embodiment in intelligent behaviour, we could argue that every machine, every prosthetic memory, has been part of our construction of AI.

12,000 years of AI

AI enhances human power. The real problem is governing AI-enhanced humans. But we are obliged to help make our work governable.

$N=2$

$N=1$

$N=4$

F_e

F_r

Pulley $IMA = N$

World Population

Population (millions)

10,000

1,000

100

10

1

-10,000   -6000   -2000   0   2000
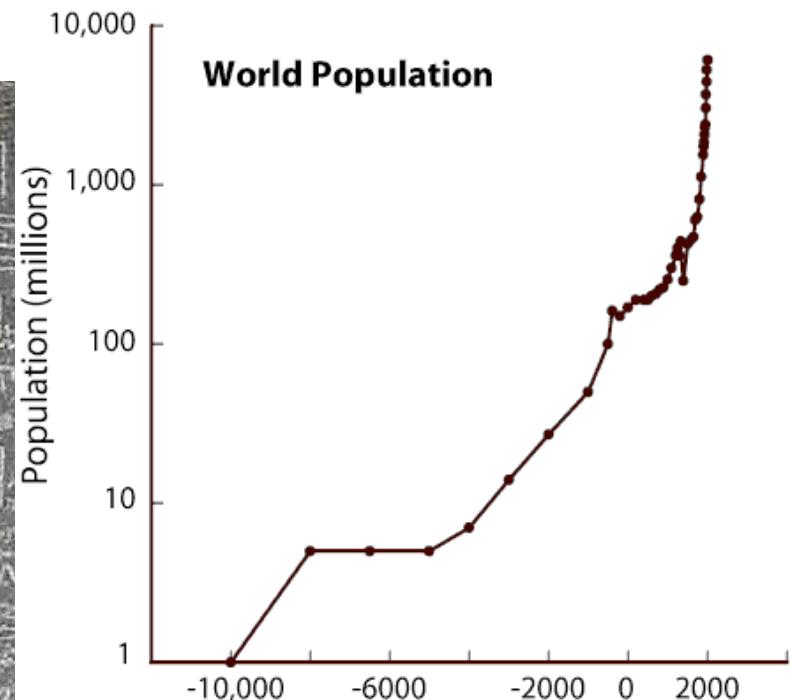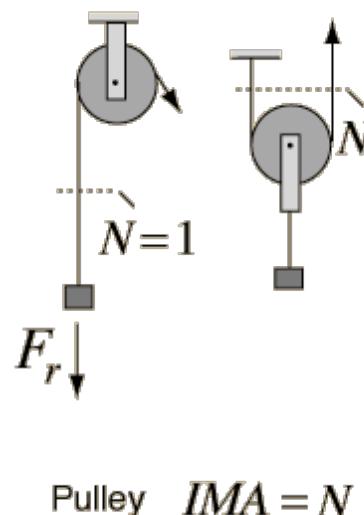
Ten Problems in Industrial AI Ethics

(+ 3 others)
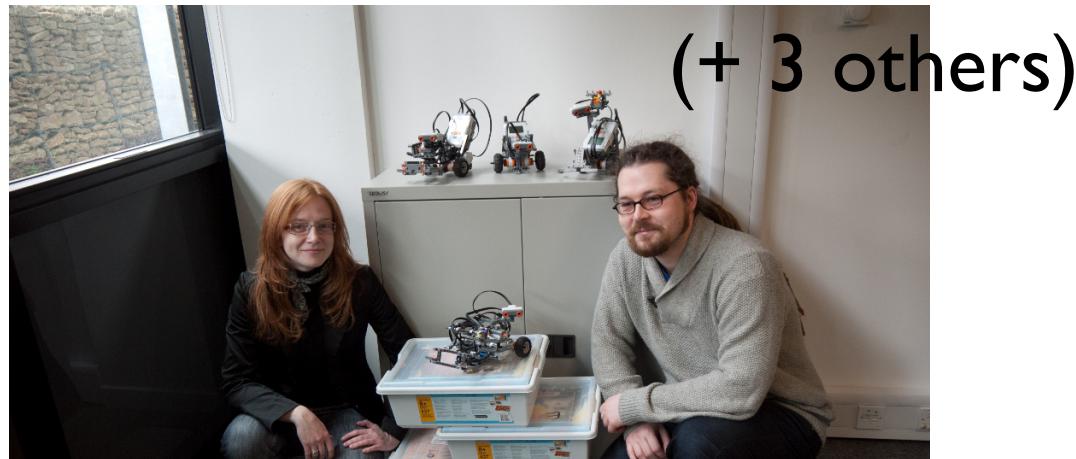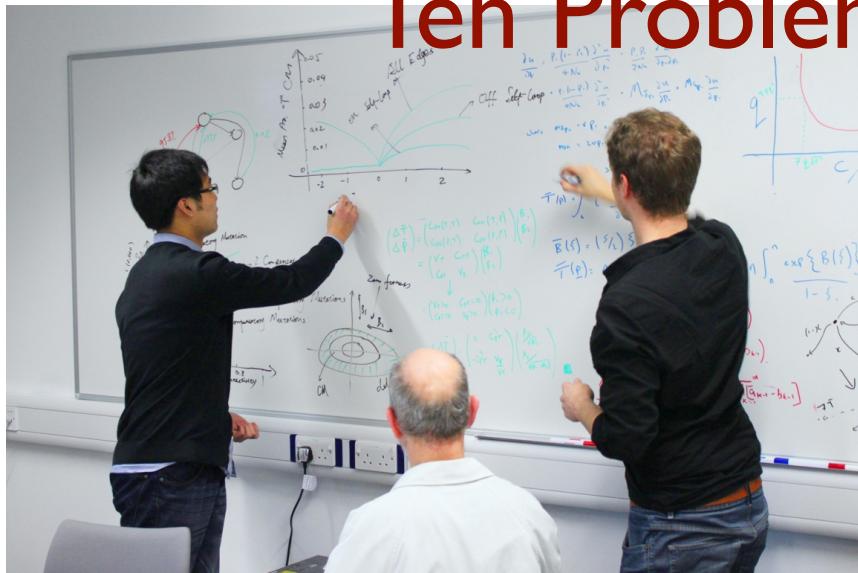
Joanna J. Bryson

@j2bryson
University of Bath, United Kingdom

Mannheimer Zentrum für Europäische Sozialforschung

Amonl
Artificial Models of Natural Intelligence

Extending the Human

We don't need a humanoid Rosie.

We need intelligent prosthetics that make humans more productive at their employment.