

From Robot Ethics to Ethical Robots

Alan FT Winfield
Bristol Robotics Laboratory
alanwinfield.blogspot.com

British Computer Society
Bristol Branch
3 May 2017

Outline

- Part 1 **robot ethics**
 - Why and how *roboticists* must be ethical
- Part 2 **ethical robots**
 - How *robots* could be ethical



Robot Ethics

The societal/ethical problem of robots that displace jobs



Jobs in assembly plants

...warehouses



Care or health work



Taxi driving?

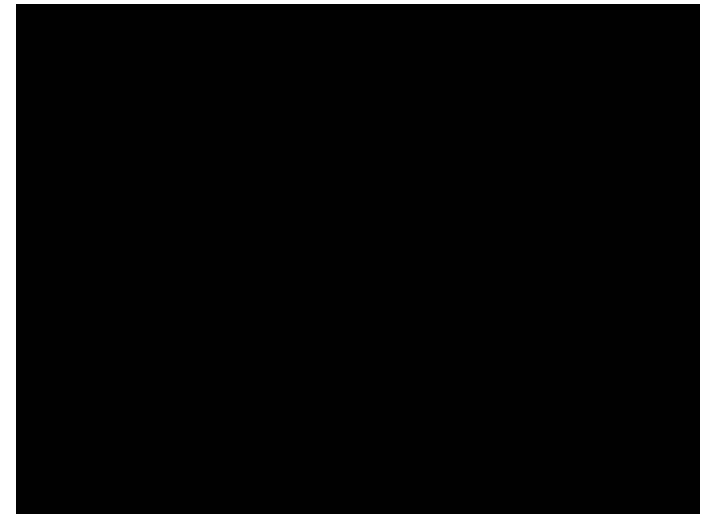
The societal/ethical problem of robots that pull the trigger



US Reaper drone

From man-in-the-loop to man-on-the-loop

TALON SWORDS



The ethical problem of humanoid robots that appear to be intelligent, but are not

We can build amazing android robot bodies, but we cannot (yet) build their brains



Geminoid-dk modelled on Henrik Schärfe



The brain-body mismatch problem

The ethical problem of robots that induce an emotional reaction, or dependency

Does our instinct to anthropomorphise make us vulnerable?



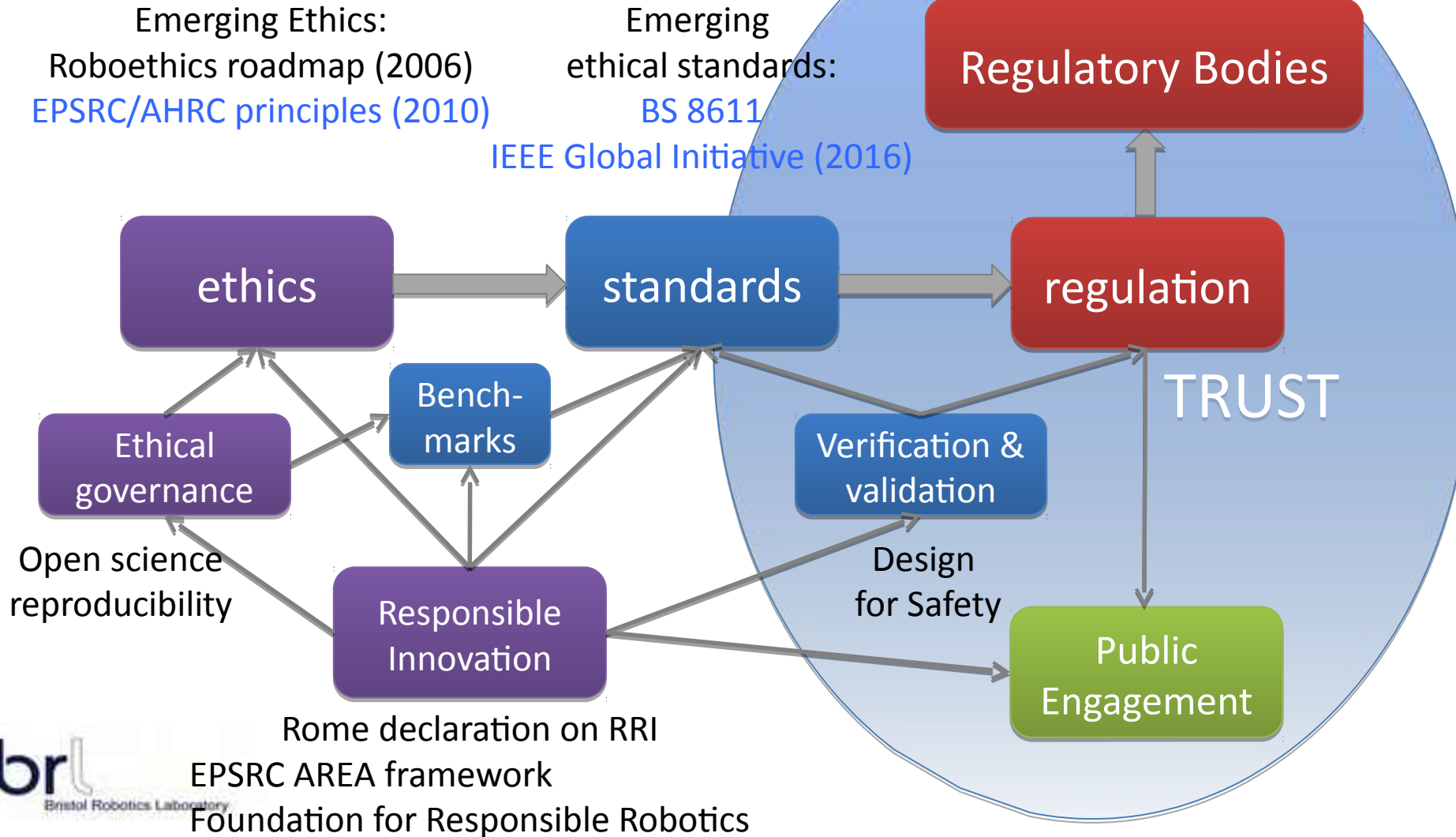
Paro robot baby harp seal

The problem of dependency

EPSRC Principles of Robotics

1. **Robots are multi-use tools.** Robots should not be designed solely or primarily to kill or harm humans.
2. **Humans, not robots, are responsible agents.** Robots should be designed 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.
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.**

Building Public Trust*





BSI Standards Publication

Robots and robotic devices

Guide to the ethical design and application of robots and robotic systems

Ethical Risk Assessment

- BS8611 sets out 20 distinct *ethical hazards and risks*, grouped under four categories:
 - *societal*,
 - *application*,
 - *commercial/financial*, and
 - *environmental*.
- Advice on measures to mitigate the impact of each risk is given, along with suggestions on how such measures might be verified or validated.

Find Standards

Develop Standards

Get Involved

News & Events

About Us

Buy Standards

eTools

The Global Initiative for Ethical Considerations in the Design of Autonomous Systems

An incubation space for new standards and solutions, certifications and codes of conduct, and consensus building for ethical implementation of intelligent technologies



INDUSTRY CONNECTIONS

The Global Initiative for Ethical Considerations in the Design of Autonomous Systems

ICAID 

NEWS AND EVENTS

Register Now! [IEEE Symposium on Ethics of Autonomous Systems \(SEAS Europe\) Sunday, 28 August, 2016 - Monday, 29](#)

ABOUT

The purpose of this Initiative is to ensure every technologist is educated, trained, and empowered to prioritize ethical considerations in the design and development of autonomous and intelligent systems.

- [View specifics regarding the Mission and deliverables for the Initiative.](#)
- [See a list of The Initiative's Executive and other Committees.](#)

Exploration of this activity will include:

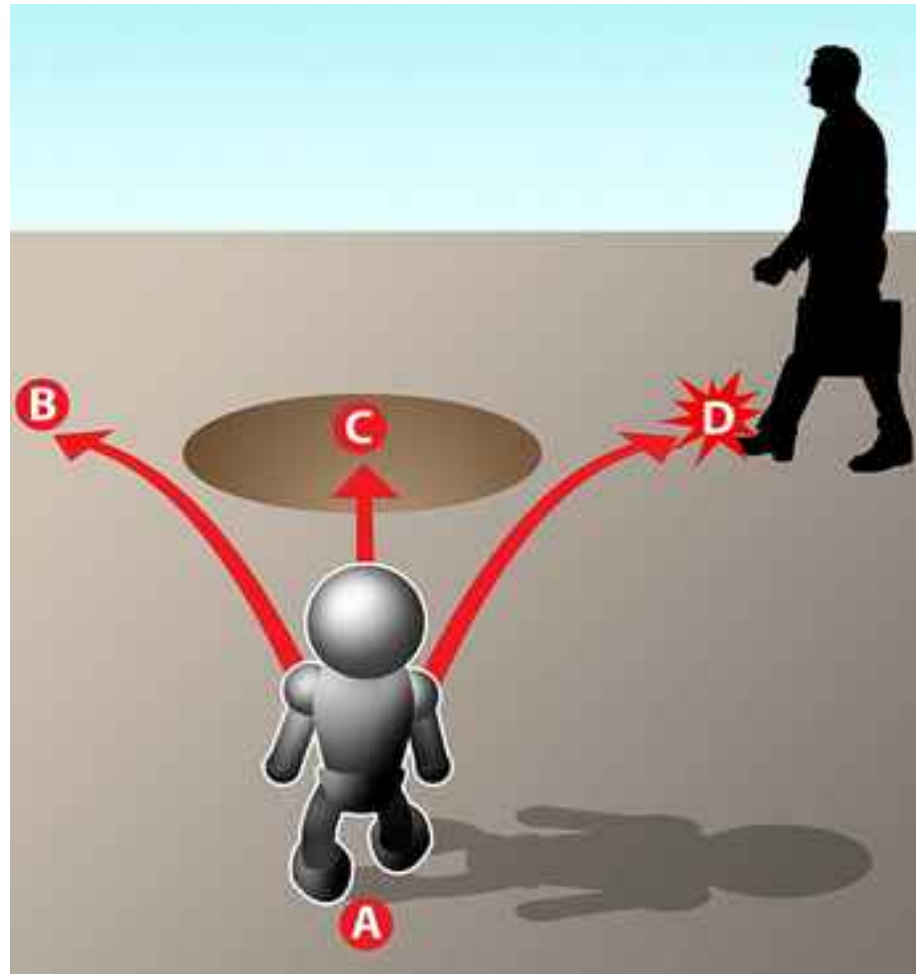
- Proposals for standardization projects, certifications, and codes of conduct
- Comprehensive roadmaps, white papers and educational materials that may include Massively Open Online Courses (MOOCs)
- Certification materials and/or programs
- An efficient, economical environment for building consensus and producing shared results
- Language for codes of conduct or principles regarding ethics and autonomous systems to be iterated in open collaborative groups
- Proposals for conferences and summits
- Cooperation with additional industry alliances focused in this field

Ethical Robots



After





An Ethical Rule

IF for all robot actions, the human is equally safe

THEN (* default safe action *)

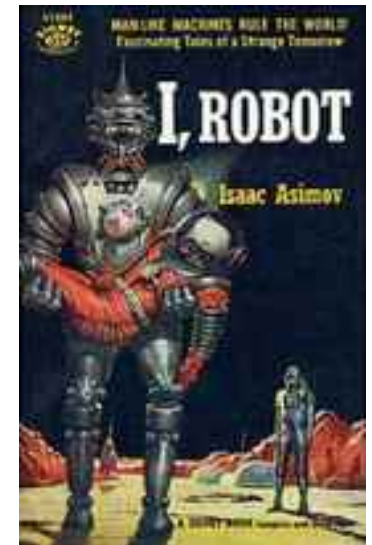
output safe robot actions

ELSE (* ethical action *)

output robot action for least unsafe human outcome

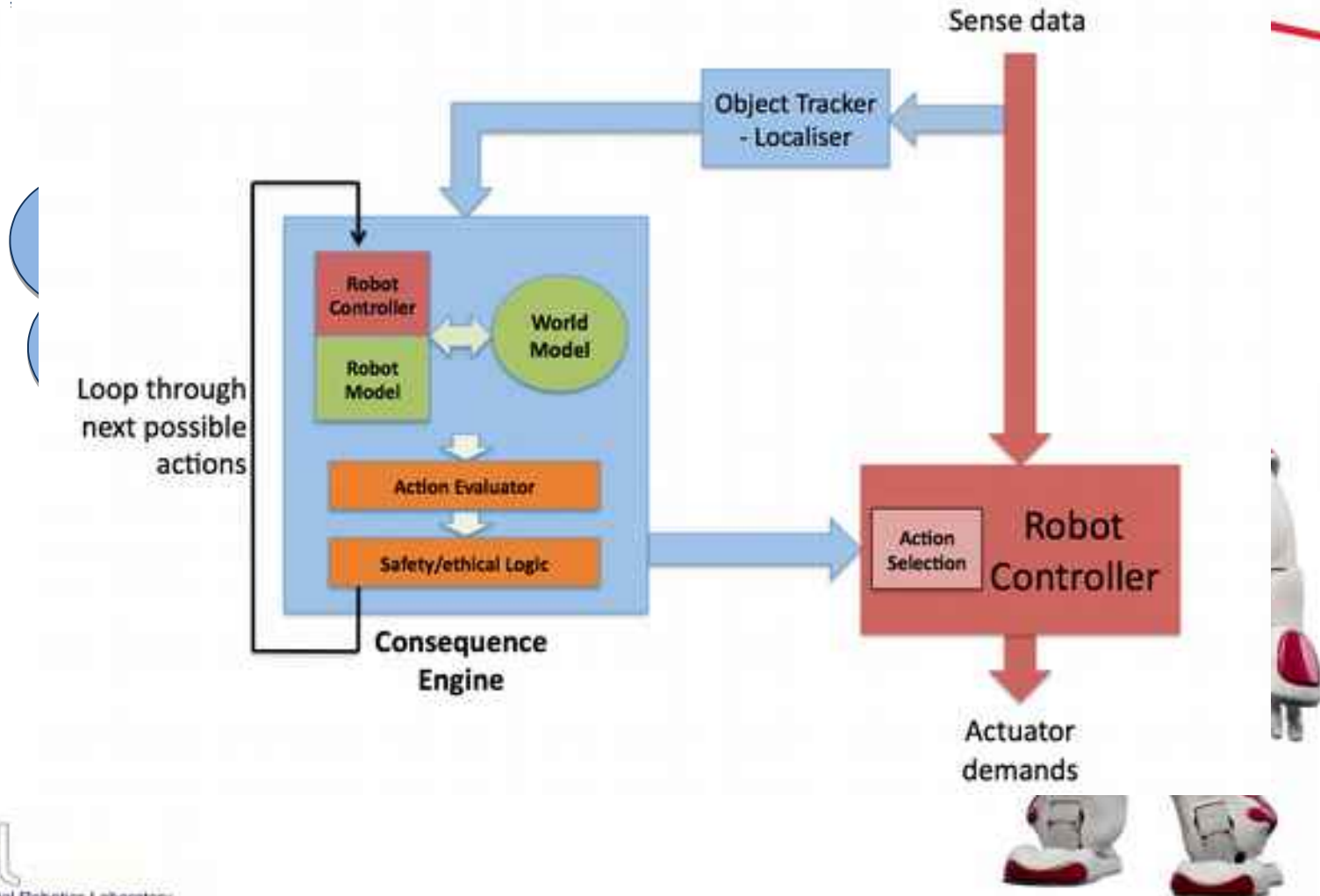
Asimov's three laws of robotics

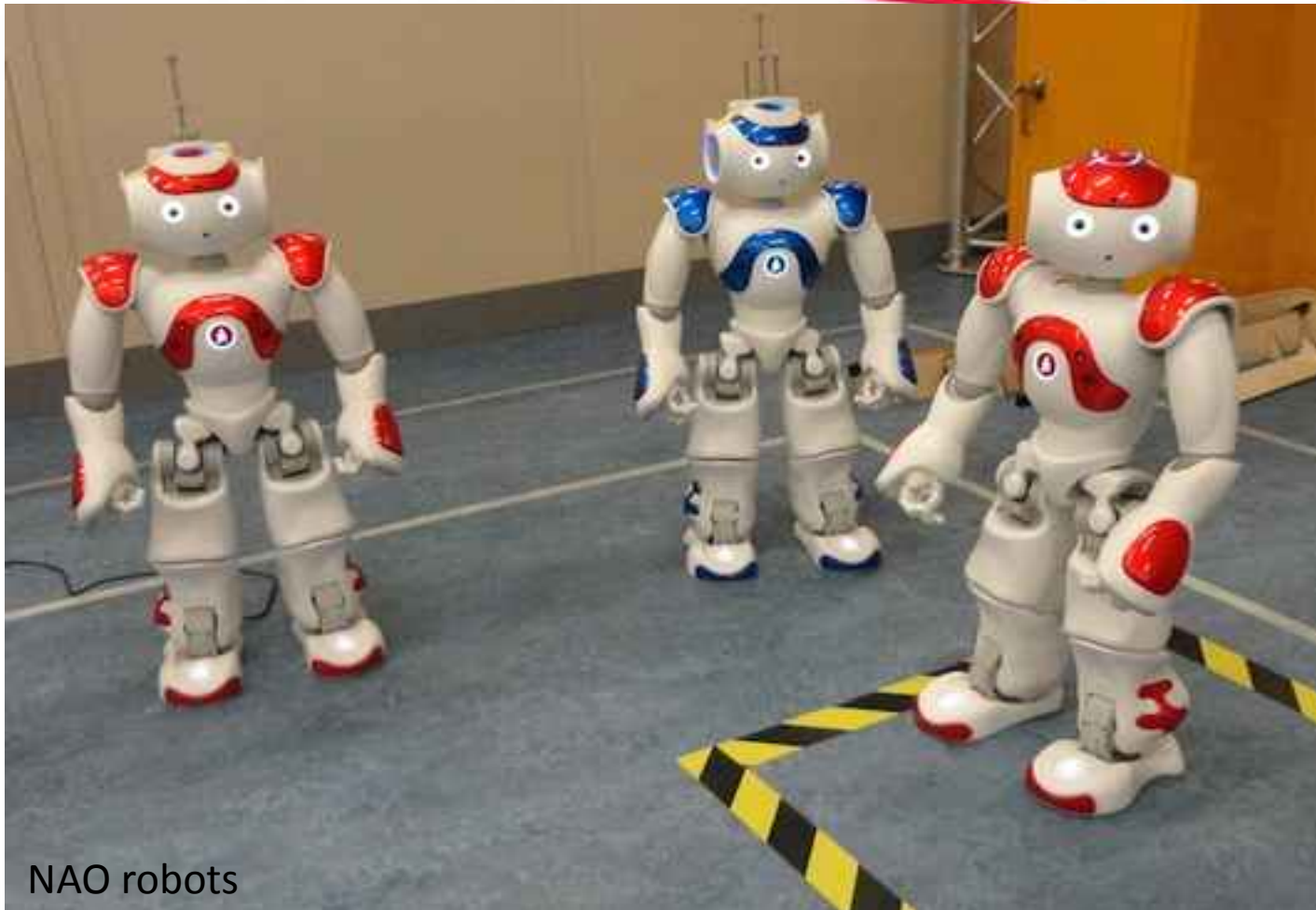
1. a robot *may not injure a human* being or, through inaction, allow a human being to come to harm;
2. a robot *must obey* any orders given to it by human beings, except where such orders would conflict with the first Law, and
3. a robot *must protect its own existence* as long as such protection does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





Webots simulation of NAO
robot soccer





NAO robots

Experimental work by Dr Dieter Vanderelst

brl

Bristol Robotics Laboratory



University of the West of England



University of BRISTOL



brl

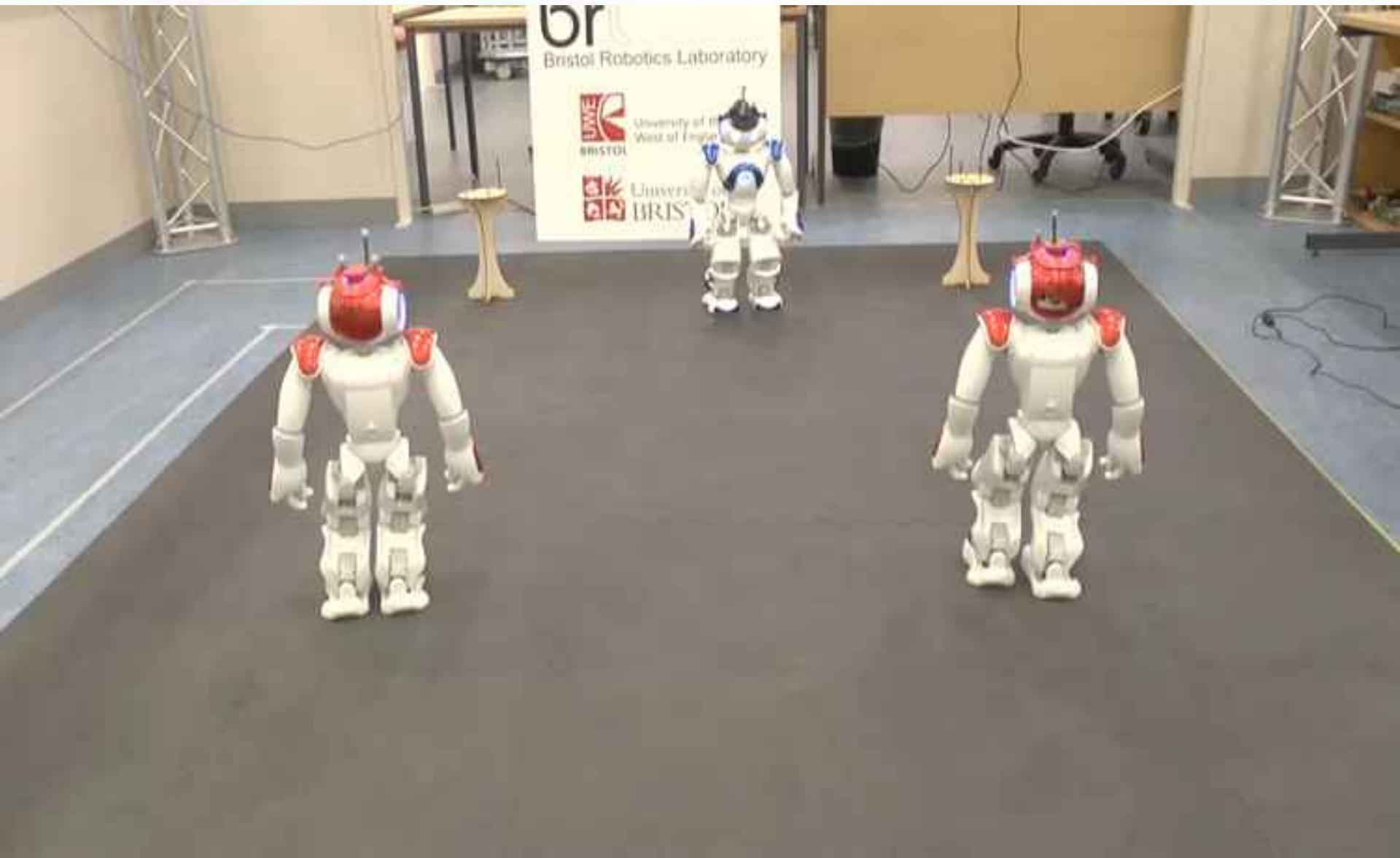
Bristol Robotics Laboratory



University of the
West of England



University of
BRISTOL





An ethical zombie



We don't need a major breakthrough in AI to build an ethical robot

In conclusion

- We need an open and informed debate about the societal, ethical and economic consequences of robotics
- We should also consider the ethics of ethical robots
- Above all we need **Responsible Research and Innovation**

