RockEU
Robotics Coordination Action for Europe

Grant Agreement Number: 611247

17.01.2013 – 16.07.2016

Instrument: Coordination and Support Action

ELS issues in robotics and steps to consider them

Part 3: Ethics

Roadmap

Pr. Alan Winfield[1]

Deliverable D3.4.1 – part 3

Lead contractor for this deliverable:	CEA
Due date of deliverable:	June 30, 2016
Actual submission date:	June 30, 2016
Dissemination level:	Public
Revision:	1.0

[1] Director of the Science Communication Unit, Bristol Robotics Laboratory, University of the West of England, Bristol, UK

Executive Summary

This document presents a robot ethics roadmap. The roadmap links a number of elements, including standards, regulation, responsible research & innovation and public engagement, which together contribute to robot ethics. The roadmap is offered as a new framework for ethical governance for robots, which in turn could support the process of building public confidence and trust in robotic systems.

This report has been produced for the Partnership for Robotics SPARC in Europe via RockEU: Robotics Coordination Action for Europe, a Coordination and Support Action (CSA) project that has received funding by under the EU FP7 research and innovation programme (grant agreement 611247).

Content

1. Introduction

The aim of this document is to set out a robot ethics roadmap. The value of a roadmap is twofold, firstly it connects and maps the different elements that each contribute to robot ethics, and secondly it provides us with a framework for ethical governance for robots.

This work continues and builds upon previous work on ethical legal and societal issues in robotics [2][3][4].

This document is structured as follows: in part 1 we build the roadmap. Part 2 then provides a commentary on the roadmap, linking the roadmap with existing related work toward ethics, standards and regulation. Part 3 suggests some recommendations arising from the roadmap.

2. Part 1: Building the Roadmap

The core of our roadmap connects ethics, standards and regulation. Standards formalise ethical principles into a structure which could be used either to evaluate the level of compliance or, more usefully perhaps for ethical standards, to provide guidelines for designers on how to conduct an ethical risk assessment for a given robot and mitigate the risks so identified. Ethics therefore underpin standards. But standards also sometimes need teeth, i.e. regulation which mandates that systems are certified as compliant with standards, or parts of standards. Thus ethics (or ethical principles) are linked to standards, which are in turn linked to regulation as shown in Fig 1 below. (Noting that not all ethical principles become codified into standards, and not all standards are mandatory.)

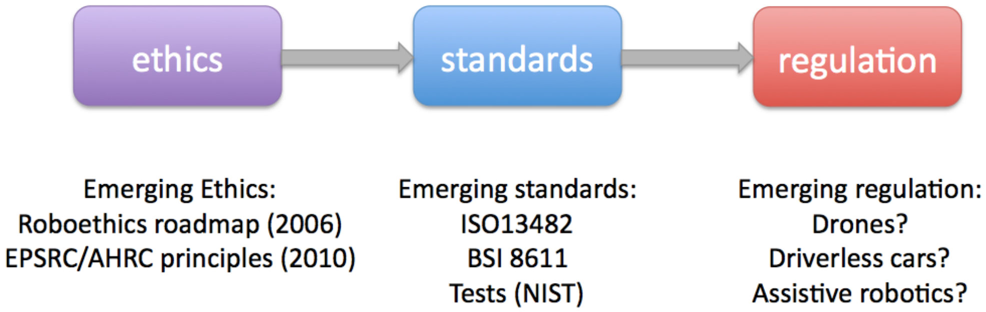

Figure 1: Standards and Regulation build on a foundation of Ethics

---

[2] Veruggio G (2006), EURON Roboethics Roadmap:

[3] C Leroux, R Labruto, euRobotics CA, D3.2.1 Ethical Legal and Societal issues in robotics, Grant Agreement Number: 248552, public report, Dec 2012

[4] C Leroux, R Labruto, euRobotics CA, A green paper on legal issues in robotics, Grant Agreement Number: 248552, public report, Dec 2012

Figure 1 shows some indicative ethical frameworks, including the 2006 Euron Roboethics Roadmap [1], and the 2010 EPSRC Principles of Robotics [2]. Also shown are emerging standards such as ISO 13482 [3] and BS 8611:2016 [4].

Ethics and Standards both fit within a wider framework of **Responsible Research and Innovation** (RRI). In a sense RRI provides a *scaffold* for ethics and standards, as shown below in Fig 2. Responsible Innovation typically requires that research is conducted ethically, so **ethical governance** connects RRI with ethics. RRI also connects directly with ethics through principles of, for instance, public engagement, open science and inclusivity. Another key principle of RRI is the ability to systematically and transparently measure and compare system capabilities, typically with standardised tests or **benchmarks**.

A further key element of RRI, especially when systems move into real world application, is the need for **verification and validation**, to provide assurance both of safety and fitness for purpose. Verification and validation might be undertaken against published standards, and – for safety critical systems – conformance with those standards may be a legal requirement without which the system would not be certified. Hence verification and validation links to both standards and regulation.

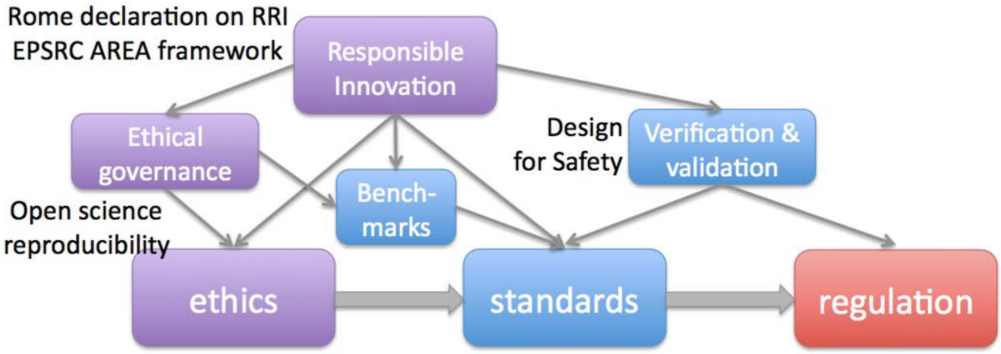

Figure 2 Responsible Research and Innovation provides the Scaffold

In general technology is **trusted** if it brings benefits while also being safe, well-regulated and, when accidents happen, subject to robust investigation. One of the reasons we trust airliners is that we know they are part of a highly regulated industry with an excellent safety record. The reason commercial aircraft are so safe is not just good design, it is also the tough **safety certification** processes and, when things do go wrong, robust processes of air **accident investigation**. Should driverless cars, for instance, be regulated through a body similar to the Civil Aviation Authority (CAA), with a driverless car equivalent of the Air Accident Investigation Branch?

Regulation requires regulatory bodies, linked with public engagement to provide transparency and confidence in the robustness of regulatory processes. All of which supports the process of building public trust, as shown below in Fig 3.

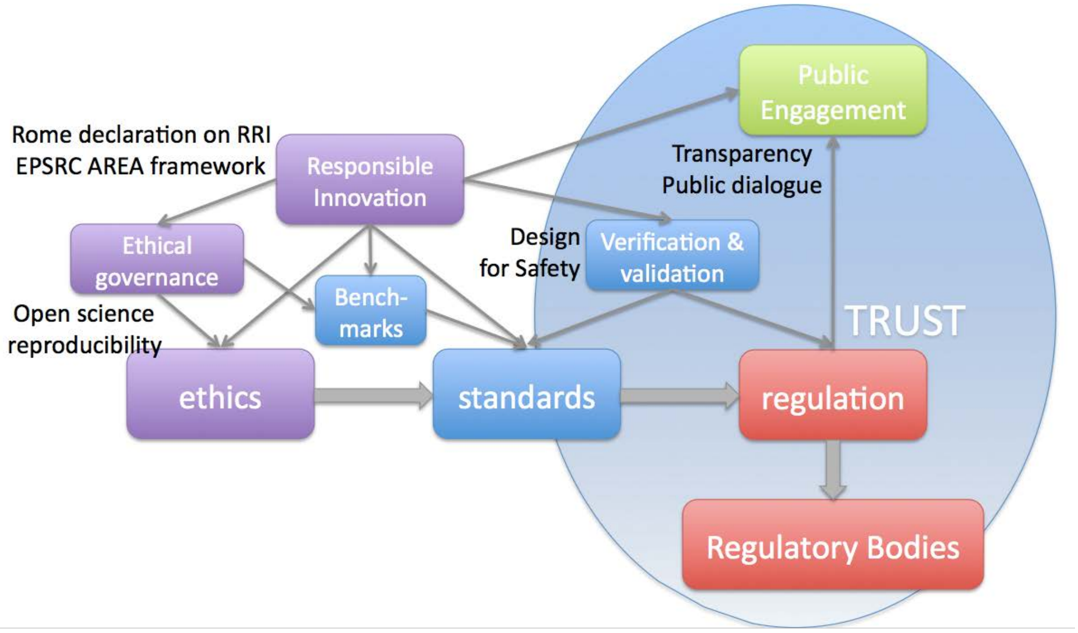

Figure 3 Building Public Trust

3. Part 2: commentary on the Roboethics roadmap

Public attitudes. It is well understood that there are public fears around robotics and artificial intelligence. Many of these fears are undoubtedly misplaced, fuelled perhaps by press and media hype, but some are grounded in genuine worries over how the technology might impact, for instance, jobs or privacy. The most recent Eurobarometer survey on autonomous systems showed that the proportion of respondents with an overall positive attitude has declined from 70% in the 2012 survey [5] to 64% in 2014 [6]. Notably the 2014 survey showed that the more personal experience people have with robots, the more favourably they tend to think of them; 82% of respondents have a positive view of robots if they have experience with them, where as only 60% of respondents have a positive view if they lack robot experience. Also important is that a significant majority (89%) believe that autonomous systems are a form of technology that requires careful management.

Ethics are the foundation of trust, and underpin good practice. Principles of good practice can be found in **Responsible Research and Innovation** (RRI). Examples include the 2014 Rome Declaration on RRI [7]; the six pillars of the Rome declaration on RRI are: Engagement, Gender equality, Education, Ethics, Open Access and Governance. The UK EPSRC framework for responsible innovation [8] incorporates the AREA (Anticipate, Reflect, Engage and Act) approach.

The first European work to articulate ethical considerations for robotics was the EURON Roboethics Roadmap [1].

In 2010 a joint AHRC/EPSRC workshop drafted and published the **Principles of Robotics** for designers, builders and users of robots [2]. The principles are:

(i) 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;

(ii) 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.

(iii) Robots are products. They should be designed using processes which assure their safety and security.

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

(v) The person with legal responsibility for a robot should be attributed.

Work by the British Standards Institute technical subcommittee on Robots and Robotic Devices led to publication – in April 2016 – of **BS 8611**: *Guide to the ethical design and application of robots and robotic systems* [4]. BS8611 is not a code of practice; instead it gives "guidance on the identification of potential ethical harm and provides guidelines on safe design, protective measures and information for the design and application of robots". BS8611 articulates a broad range of ethical hazards and their mitigation, including societal, application, commercial/financial and environment risks, and provides designers with guidance on how to assess and then reduce the risks associated with these ethical hazards. The societal hazards include, for example, loss of trust, deception, privacy & confidentiality, addiction and employment.

Significant recent work towards **regulation** was undertaken by the EU project RoboLaw. The primary output of that project is a comprehensive report entitled *Guidelines on Regulating Robotics* [9]. That report reviews both ethical and legal aspects; the legal analysis covers rights, liability & insurance, privacy and legal capacity. The report focuses on driverless cars, surgical robots, robot prostheses and care robots and concludes by stating: "The field of robotics is too broad, and the range of legislative domains affected by robotics too wide, to be able to say that robotics by and large can be accommodated within existing legal frameworks or rather require a *lex robotica*. For some types of applications and some regulatory domains, it might be useful to consider creating new, fine-grained rules that are specifically tailored to the robotics at issue, while for types of robotics, and for many regulatory fields, robotics can likely be regulated well by smart adaptation of existing laws".

The primary focus of the paragraphs in this section 2 is robotics and autonomous systems, and not **software artificial intelligence**. This reflects the fact that most work toward ethics and regulation has focused on robotics. Because robots are physical artefacts (which embody AI) they are undoubtedly more readily defined and hence regulated than distributed or cloud-based AIs. This and the already pervasive applications of AI (in search engines, machine translation systems or intelligent personal assistant AIs, for example) strongly suggest that greater urgency needs to be directed toward considering the societal and ethical impact of AI, including the **governance and regulation of AI**.

AI systems raise serious questions over **trust** and **transparency**:

- How can we trust the decisions made by AI systems, and – more generally – how can the public have confidence in the use of AI systems in decision making?

- o If an AI system makes a decision that turns out to be disastrously wrong, how do we investigate the logic by which the decision was made?

Of course much depends of the consequences of those decisions. Consider decisions that have real consequences to human safety or well being, such as those made by medical diagnosis AIs or driverless car autopilots. Systems that make such decisions are *critical* systems.

Existing critical software systems are not AI systems, nor do they incorporate AI systems. The reason is that AI systems (and more generally machine learning systems) are generally regarded as impossible to verify for safety critical applications - the reasons for this need to be understood.

- o First is the problem of **verification of systems that learn**. Current verification approaches typically assume that the system being verified will never change its behaviour, but a system that learns does – by definition – change its behaviour, so any verification is likely to be rendered invalid after the system has learned.
- o Second is **the black box problem**. Modern AI systems, and especially the ones receiving the greatest attention, so called Deep Learning systems, are based on Artificial Neural Networks (ANNs). A characteristic of ANNs is that after the ANN has been trained with data sets (which may be very large, so called “big data” sets – which itself poses another problem for verification), any attempt to examine the internal structure of the ANN in order to understand why and how the ANN makes a particular decision is impossible. The decision making process of an ANN is not transparent.

The problem of **verification and validation** of systems that learn may not be intractable, but is the subject of current research, see for example work on verification and validation of autonomous systems and [10]. The black box problem may be intractable for ANNs, but could be avoided by using algorithmic approaches to AI (i.e. that do not use ANNs).

The IEEE has recently launched a global initiative on Ethical Considerations in the Design of Autonomous Systems, to encompass all intelligent technologies including robotics, AI, computational intelligence and deep learning [11].

4. Part 3: Recommendations

It is vital that we address public fears around robotics and artificial intelligence, through **renewed public engagement and consultation**.

Work is required to first identify and then **develop new standards** for intelligent autonomous robots, together with the **benchmark tests** and **verification & validation processes** that would assure compliance against those standards.

Work is required to **identify the kind of governance framework(s)** and regulatory bodies needed to support Robotics and Artificial Intelligence in Europe.

5. References:

[1] Veruggio G (2006), EURON Roboethics Roadmap: http://www.roboethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf

[2] Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B. and Winfield, A. Principles of Robotics, EPSRC, May 2011.

https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

[3] International Standards Organisation 13482:2013 Robots and robotic devices – Safety requirements for personal care robots. http://www.iso.org/iso/catalogue_detail.htm?csnumber=53820

[4] British Standards Institute BS8611:2016 Robots and robotic devices: guide to the ethical design and application of robots and robotic systems, ISBN 978 0 580 89530 2. http://shop.bsigroup.com/ProductDetail?pid=000000000030320089

[5] Special Eurobarometer 427, Autonomous Systems, June 2015. http://ec.europa.eu/public_opinion/archives/ebs/ebs_427_en.pdf See also analysis summary http://robohub.org/study-shows-public-perception-of-robotics-generally-positive-in-eu-but-declining/

[6] Special Eurobarometer 382, Public Attitudes towards Robots, Sept 2012.

http://ec.europa.eu/COMMFrontOffice/PublicOpinion/index.cfm/ResultDoc/download/DocumentKy/56814

[7] The Rome Declaration on Responsible Research and Innovation (2014): http://www.science-and-you.com/en/sis-rri-conference-recommendations-rome-declaration-responsible-research-and-innovation and http://ec.europa.eu/research/science-society/document_library/pdf_06/responsible-research-and-innovation-leaflet_en.pdf

[8] The EPSRC framework for responsible Innovation: https://www.epsrc.ac.uk/research/framework/

[9] Palmerini, E., Azzarri, F., Battaglia, A., Bertolini, A., Carnevale, A., Carpaneto, J., Cavallo, F., Di Carlo, A., Cempini, M., Controzzi, M., Koops, B.J., Lucivero, F., Mukerji, N., Nocco, L., Pirni, A., Shah, H., Salvini, P., Schellekens, M. and Warwick, K. D6.2 – Guidelines on regulating robotics. Pisa, Italy: Robolaw project, 2014. http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf

[10] EPSRC project Verifiable Autonomy: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/L024845/1 and the EPSRC Network on the Verification and Validation of Autonomous Systems:

http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/M027309/1

[11] The global initiative on Ethical Considerations in the Design of Autonomous Systems http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Note: Winfield co-chairs the General Principles committee.