Institution: University of the West of England, Bristol

Unit of Assessment: 12

Title of case study: Robot ethics to ethical robots: informing standards, policy and practice

Period when the underpinning research was undertaken: 2009 - 2019

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Alan Winfield	Professor of Robot Ethics	December 1991 – present
Wenguo Liu	Research Fellow	April 2008 – January 2015
Dieter Vanderelst	Research Fellow	April 2015 – August 2016
Paul Bremner	Research Fellow	June 2009 – present

Period when the claimed impact occurred: 01.08.2013 – 2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

Professor Alan Winfield's research and engagement activities have contributed materially to robot and Artificial Intelligence (AI) ethics, building on academic discourse to inform and impact on intense public and policy debate, both nationally and internationally. The research, conducted at the University of the West of England, has:

- informed the development of new national and international standards for the ethical design and application of robots and robotic systems;
- influenced the development of new organisational standards within the robotics industry, and defined best practices;
- enhanced wider public understanding and informed public debate on robot ethics;
- guided the work of the UK government, and UK and EU parliaments, and informed policy debate in the House of Commons;
- helped shape the NHS strategy for preparing the healthcare workforce to deliver digital healthcare technologies in the future.

2. Underpinning research

Considering the ethical impacts of robots and work on standards

UWE research into robot ethics and the related field of ethical robots, emerged from Professor Winfield's 2006-2009 EPSRC grant *Walking with Robots* (**G1**) and 2009-2013 EPSRC Senior Media Fellowship, *Intelligent Robots in Science and Society* (**G2**). *Walking with Robots*, which won the Royal Academy of Engineering Rooke Medal in 2010, considered the ethical implications of robotics research, and *Intelligent Robots in Science and Society* looked at the ethical impact of robotics on society. Winfield was asked to present findings from these projects to the EPSRC Societal Impact Panel, which led to him being asked to co-organise a joint EPSRC/AHRC workshop, which in turn resulted in the publication of the EPSRC/AHRC Principles of Robotics (**R1**).

These principles (**R1**) directly influenced subsequent ethical principles in robotics and AI. They also prompted the formation of a working group on robot ethics, which led directly to the development of British Standard BS 8611 – the world's first published ethical standard in robotics (see Section 4). Winfield's engagement with the EPSRC principles and the British Standards Institute (BSI) in turn led to his invitation to join the IEEE Standards Association's global ethics initiative. Winfield's contributions to the work of this initiative have been the development of new general ethical principles for autonomous and intelligent systems, and new IEEE ethical standards.

Introducing the ‘ethical black box’ approach and pillars of ethical governance
An ongoing collaboration with Professor Marina Jirotka (University of Oxford) on *ethical governance*, resulted in a proposal in 2017 that all robots and AI should be equipped with the equivalent of an aircraft flight-data recorder to support robot accident investigation (**R2**). Winfield and Jirotka are currently being supported by a five-year EPSRC grant (**G3**) to develop this ethical black box. The work on **R2** led to a paper (**R3**) that developed the framework linking ethical principles to standards and regulations. The paper argued that ethical governance was essential to building public trust in robotics and AI, and proposed four ‘pillars’ of good ethical governance for companies and organisations:

1. Publish an ethical code of conduct;
2. Provide ethics and responsible research and innovation training for all members/staff;
3. Practice responsible innovation, including the engagement of wider stakeholders within a framework of anticipatory governance;
4. Be transparent about ethical governance.

Using simulation-based architectures for ethical robots and introducing verifiability
Winfield’s work with Professor Michael Fisher (University of Liverpool), brought work on formal methods from computer science together for the first time with robotics, to develop completely new approaches to the validation of robot systems and robot safety. Also, in parallel, Winfield’s EPSRC project *The Emergence of Artificial Culture in Robot Societies* (**G4**), led to the idea of robots with simulation-based internal models – that is, a simulation of the robot itself, other robots, and its environment inside itself. Building on this work, the EPSRC project *Verifiable Autonomy* (**G5**) focused on explicitly ethical robots, i.e. robots that can take ethical considerations into account when deciding how to behave. The inclusion of models of other agents was a novel feature of this research, enabling the robot to predict the consequences of both its and another agent’s actions. UWE researchers conducted a series of successful experimental trials and demonstrated the world’s first transparent and verifiable ethical robot. Outputs from this work include **R4**, **R5** and **R6**. **R6** appeared in a special issue on machine ethics of the *Proceedings of the IEEE*, co-edited by Winfield. This special issue represents the most comprehensive survey of the emerging field of practical machine ethics to date, with papers on both the engineering and governance of ethical machines.

3. References to the research

R1 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. (2017) Principles of robotics: Regulating robots in the real world. *Connection Science*, 29 (20), pp. 124–129. https://doi.org/10.1080/09540091.2016.1271400
R2 Winfield, A. and Jirotka, M. (2017) The case for an ethical black box. In: Gao, Y., ed. (2017) *Towards Autonomous Robot Systems*. Springer, pp. 1-12. https://uwe-repository.worktribe.com/output/904084
R3 Winfield, A. and Jirotka, M. (2018) Ethical governance is essential to building trust in robotics and AI systems. *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, 376 (2133). ISSN 1364-503X. https://doi.org/10.1098/rsta.2018.0085
R4 Winfield, A., Blum, C. and Liu, W. (2014) Towards an ethical robot: Internal models,

consequences and ethical action selection. In: Mistry, M., Leonardis, Aleš, Witkowski, M. and Melhuish, C., eds. *Advances in Autonomous Robotics Systems: Proceedings of the 15th Annual Conference*, TAROS 2014, Birmingham, UK, 1-3 September 2014, pp. 85-96. http://dx.doi.org/10.1007/978-3-319-10401-0_8
R5 Vanderelst, D. and Winfield, A. (2018) An architecture for ethical robots inspired by the simulation theory of cognition*. Cognitive Systems Research*, 48. pp. 56-66. ISSN 1389-0417. https://doi.org/10.1016/j.cogsys.2017.04.002
R6 Bremner, P., Dennis, L., Fisher, M. and Winfield, A. (2019) On proactive, transparent and verifiable ethical reasoning for robots. *Proceedings of the IEEE*, 107 (3). pp. 541-561. ISSN 0018-9219. https://doi.org/10.1109/JPROC.2019.2898267

Evidence of the quality of the underpinning research
G1 Winfield, A. *Walking with Robots*, EPSRC, 2006 – 2009, £249,557.
G2 Winfield, A. *Intelligent Robots in Science and Society,* EPSRC, 2009 – 2012, £112,078.
G3 Winfield, A. *RoboTIPS: Developing Responsible Robotics for the Digital Economy,* EPSRC, 2019 – 2024, £428,068.
G4 Winfield, A. *The Emergence of Artificial Culture in Robot Societies*, EPSRC, 2007 – 2011, £735,507.
G5 Winfield, A. *Verifiable Autonomy*, EPSRC, 2014 – 2018, £340,338.

4. Details of the impact

The development of new national and international ethical standards

UWE research has informed the development of national and international standards for the ethical design and application of robots and robotic systems:

- The principles articulated in **R1** prompted the formation of the BSI working group that drafted BS 8611, the world’s first published standard (2016) in robot ethics (**S1**).
- As founding co-chair of the IEEE Standards Association ethics initiative General Principles Committee, Winfield brought the insights of his research to bear on the development of new general ethical principles for Intelligent and Autonomous Systems, a foundational part of IEEE framework ‘*Ethically Aligned Design*’ 2017 (**S2**). For example, **S2** (p30) makes use of **R3**.
- Drawing on **R2** and **R3**, Winfield led a proposal that one of the IEEE framework’s general ethical principles related to transparency should form the basis of a new IEEE standard (IEEE P7001). In 2017, that proposal was accepted, and he now chairs Working Group P7001 drafting a new standard on *Transparency in Autonomous Systems* (for an outline of P7001 see **S3**).

Ideas and practice within robotics industry

The robotics industry and affiliated unions are using Winfield’s research to inform their thinking and practice on robot ethics. His research has underpinned the development of organisational standards and helped define best practice:

- The UNI Global Union – an association of over 650 unions in 140 countries with over 20 million members – has developed a set of principles for ethical AI. Principle 2 (**S4**, p7) is *Equip AI Systems With an “Ethical Black Box”* – an idea first proposed in **R2**. UNI Global Union cites **R1** and an article in *Futurism,* which draws on **R2**, as providing inspiration and insight for the development of the principles (**S4**, p10).

- Ethical standards are becoming incorporated into industry practice; Sastra Robotics, for example, reference the IEEE initiative (which draws on **R3**) in their statement on *Rules and Ethical Considerations of Robotics Technology* (**S5**).

Informing public debate on robot ethics
UWE research has extensively informed public debate on robot ethics through Winfield's wide-ranging engagement with the public – drawing on the training and development received as a science communicator through his EPSRC Senior Media Fellowship award. From August 2013, Winfield contributed to over 50 public lectures and panel debates on ethical challenges in robotics and AI. For example, he gave the Campaign for Science and Engineering (CaSE) 2018 annual lecture alongside Dame Wendy Hall and Jim Al-Khalili at the Institute of Physics, and he debated AI with Professor Brian Cox and Robin Ince on BBC's popular radio programme *The Infinite Monkey Cage* in January 2016. Winfield is frequently called upon by the press and media to comment on topical issues in robot ethics; notably he was a guest on BBC R4 programme *The Life Scientific* in February 2017, and interviewed for BBC News *HARDtalk* in October 2017 (**S6**). Winfield's blog, which deals mainly with issues relating to robot ethics and ethical robots, has been visited over 500,000 times since August 2013.

Informing the work of the UK government, and UK and EU parliaments
Winfield's research and profile within robot ethics has also led to prolific engagement with UK government departments and parliament, and has informed policy debate in multiple Parliamentary Committees and similar bodies:

- An invitation by the Foreign Office to brief the G8 non-proliferation committee on the risks of robotics and AI in October 2013 (co-presenting with Lord Martin Rees) (**S7**, p1).
- Invitations by the UK Chief Scientific Advisor, Sir Mark Walport, to attend expert round table briefings in January 2015 and January 2016; the latter on the '*risks and opportunities in the use of artificial intelligence in government decision-making*' (**S7**, p2, p3).
- Winfield was invited to submit written evidence to the House of Commons Select Committee on Science and Technology inquiry on Robotics and AI, and is cited in connection with the need to be able to investigate the logic by which AI decisions are made and the implications of this for public confidence (**S8**, *Robots and artificial intelligence*, p18, p22, cf. **R2**).
- Winfield was invited to give oral evidence to the 2017 House of Lords Select Committee on Artificial Intelligence's session on AI ethics. This oral evidence is cited in the Committee's 2018 report in connection with technical transparency (**S8,** *AI in the UK*, p38).
- Both the work of the House of Commons Select Committee on Science and Technology on Robotics and AI, and the concept of a 'logging mechanism' to give a step-by-step account of processes involved in decision making, (first proposed in **R2**) informed a debate held in the House of Commons on 17 January 2018 where the House considered ethics and AI (**S9**, Column 353WH).
- Winfield has attended several meetings of the All-Party Parliamentary Groups (APPGs) on AI and on Data Analytics. Winfield and Jirotka's work on transparency

(**R2**) and ethical governance (**R3**) was cited in the 2019 APPG Data Analytics report on *Trust, Transparency and Technology* (**S10**, p23, p46).

- In 2020, the European Parliament Panel for the Future of Science and Technology published a report on *The ethics of artificial intelligence*, which cites **R3** (**S11**, p32, p34).

Informing NHS strategy

Finally, UWE research has helped shape the NHS strategy for preparing the healthcare workforce to deliver digital healthcare technologies in the future. Winfield was invited by Health Education England (HEE) to join the Topol Review as the Robotics and AI ethics advisor, contributing to the final report published in February 2019 (**S12**); Winfield co-drafted two sections of that report: *Ethical Considerations* and *AI & Robotics.* The NHS' *Interim people plan* (June 2019) identified implementing the recommendations of the Topol report as a major objective for the NHS (**S13**, p54, p71).

5. Sources to corroborate the impact

S1 *Robots and Robotic Devices - Guide to the Ethical Design and Application of Robots and Robotic Systems* BS 8611:2016 (British Standards Institution, 2016)

S2 IEEE Standards Association (2017) *IEEE Global Initiative on the Ethics of Autonomous and Intelligent Systems - Ethically Aligned Design*

S3 Winfield, A. F. (2019) Ethical standards in Robotics and AI. *Nature Electronics*, 2. pp. 46-48. ISSN 2520-1131

S4 UNI Global Union *Top 10 Principles for Ethical Artificial Intelligence*

S5 Sastra Robotics (2017) *Rules and ethical considerations of robotics technology*

S6 BBC R4, The Infinite Monkey Cage: Artificial Intelligence, first broadcast 19.01.2016; https://www.bbc.co.uk/programmes/b06wcsng; BBC R4, The Life Scientific: Alan Winfield on Robot Ethics, first broadcast 27.02.2017 https://www.bbc.co.uk/programmes/b08ffv2l; BBC World News, HARDtalk, interview by Stephen Sackur, first broadcast 31.10.2017 https://www.bbc.co.uk/programmes/n3ct2km5

S7 Invitations from the Foreign Office and the UK Chief Scientific Advisor

S8 House of Commons (2016) – *Robotics and artificial intelligence*, Select Committee on Science and Technology, Fifth Report of Session 2016-2017; House of Lords (2018) *AI in the UK: ready, willing and able?* Select Committee on Artificial Intelligence, Report of Session 2017–19

S9 Hansard *Ethics and Artificial Intelligence* Volume 634: debated on 17.01.2018

S10 All Party Parliamentary Group on Data Analytics (2019) *Trust, Transparency and Technology: Building Data Policies for the Public Good*

S11 European Parliament Panel for the Future of Science and Technology (STOA) *The ethics of artificial intelligence: Issues and initiatives*

S12 *The Topol Review* Preparing the healthcare workforce to deliver the digital future: an independent report on behalf of the Secretary of State for Health and Social Care, NHS Health Education England, Feb 2019

S13 NHS Interim People Plan June 2019