Informing the Design of HRI Systems through Use of the Ethics Canvas

Conor McGinn
School of Engineering
Trinity College Dublin
Dublin, Ireland
mcginnc@tcd.ie

***Abstract*—If used in the right way, robotics technology has the potential to improve lives and address many problems faced by vulnerable members of society. However, if used for the wrong reasons or without due consideration for the consequences, it may threaten the privacy and safety of citizens. HRI practitioners have a responsibility to ensure that potential benefits of the technology are weighed against the ethical cost. The aim of this research was to investigate if the Ethical Canvas, a practical tool derived from the Business Model Canvas (BMC) to foster ethically informed technology design, could be useful in a HRI context. To evaluate the suitability of the tool in a robotics application, a case study was undertaken which involved applying the tool to a practical use-case. The use-case concerned a robot being deployed to mediate a group-based activity (like a game of 'Bingo') in a retirement community. Although on first appearance this use-case did not seem ethically complex, using the canvas provoked consideration of a wide range of issues and potential impacts that were not initially apparent. As a result of insights gained from this pilot study, it would seem that the Ethical Canvas has great potential for use in HRI research, and it is suggested that researchers conducting studies involving human subjects might consider using the tool to inform experimental design and help establish a standard best practise.**

***Index Terms*—ethics canvas; business model canvas; roboethics; eldercare robotics**

I. Introduction

As the impact of technology on everyday life continues to grow, it is becoming increasingly important to reflect on the ethical impact of their use and on-going development. This is especially true for disruptive new technologies, such as artificial intelligence and robotics, which have the potential to cause significant adverse effects to user privacy and/or safety if used in the wrong way. However, detailed consideration of ethics during product development is far from straightforward and ethics has traditionally been slow to catch up with technological developments [1].

In Universities and other academic research institutions, ethics is typically governed by professional codes of conduct as well as institutional review boards (IRBs). Commercial entities concerned with conducting research and innovation activities are less internally regulated, and ethics tends to be taken into account primarily through compliance with legal requirements (GDPR, HIPAA, etc.) which also tend to lag technological developments [2]. This has current ethics practise to be a top-down process that is bureaucratic, focuses on compliance rather than critical reasoning based on specific use-case(s), can differ greatly between jurisdictions and institutions, and lacks the adaptable to meet the needs of exploratory research that contains high degrees of experimental uncertainty (such as long-term product pilot studies). Furthermore, since these approaches are typically carried out in a 'snapshot' fashion, before deployment of the technology, they fail to incorporate valuable insights that might be generated from on-going critical evaluation or from reflection which incorporates the lived experience of the people effected by the technology [3]. Separating ethics from the underlying technology and its use fails to recognize that 'ethics' is not a not some sort of a separate field, but is "intertwined within the fabric of technology" [4].

Within the field of robotics, there has been considerable research activity in the area of ethics, driven by the large potential impact robot technology may have on the world [5]. According to Veruggio and Operto, "roboticists cannot avoid engaging in a critical analysis of the social implications of their researches". Research into robot ethics has motivated a new discipline within the robotics research community, known as 'Roboethics', which explores how humans relate to these machines in both the design and use phase of their operation. Roboethics[1] can be considered as an offshoot to computer ethics that pays special attention to the alterations that need to be made to computer ethics when we give the computer mobility and a means to interact directly in the human environment [6].

The creators of robot systems have a critical ethical role to play, especially considering "designers intentions do not always correspond with the users practice" [7]. According to Sullins, "ethics has now become something that the designers of robots must take into careful consideration at some point during each project" [8]. However, despite this important role, it has been observed that HRI researchers frequently choose not to directly address ethical issues due to a cited lack of experience in ethics and a failure to recognize significant ethical considerations in their work [9].

[1]Roboethics is disntict from 'machine ethics' or 'machine morality' which is concerned with describing how machines could behave ethically towards humans [6].

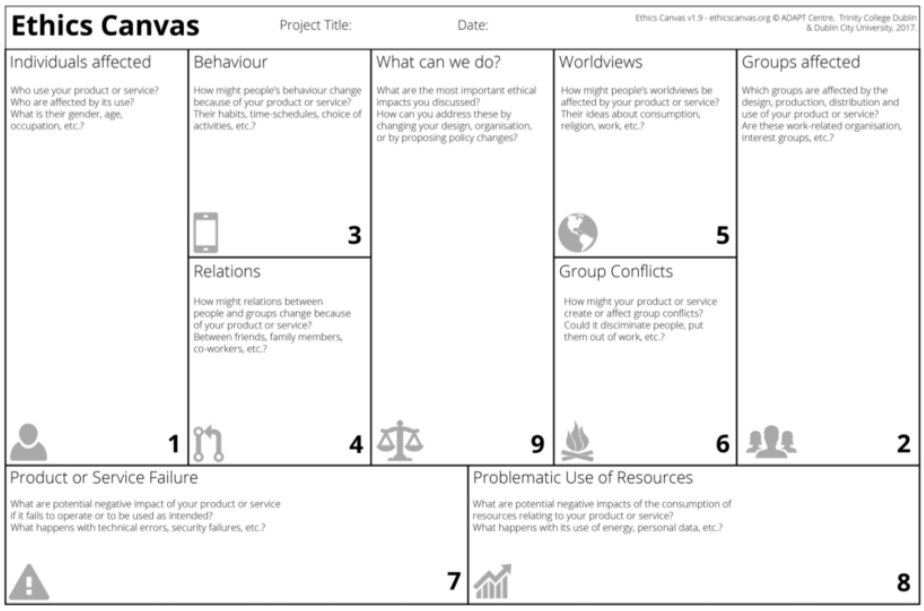


Fig. 1. Ethical Canvas template - extracted from [10].

In recent years, there has been significant research activity involving to the development of ethics guidance resources for HRI practitioners. For example, a code of ethics has been suggested by Ingram et al. which outlines seven principles that robotics engineers should follow in their work [11]. A more comprehensive code of ethics has been proposed by Riek et al. based around the overall principle of "respect for human persons, including respect for human autonomy, respect for human bodily and mental integrity, and the affordance of all rights and protections ordinarily assumed in human-human interactions" [12]. These codes of ethics, while useful in establishing an overarching ethical framework, provide limited practical support/guidance to researchers involved with explicit design, planning and execution of robots and robot-based studies.

There has been recent progress in the development of standards concerning the ethical design of robot systems, notably the British standard BS 8611-2016 (Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems) [13]. This standard presents a taxonomoy, information about how to conduct an ethical risk assessment, and outlines a range of guidelines and ethics-related system design recommendations. While there is yet to be a published international standard governing Roboethics and design, the IEEE P7007 working group (for Ontological Standard for Ethically Driven Robotics and Automation Systems) are actively on developing a standard in this area.

The lack of harmonized standards and methods for the ethical design of robots presents major problem, and motivates the development and utilization of new techniques/methods that promote increased levels of awareness and engagement on ethical issues by HRI researchers. To be maximally effective in practise, these tools must be accessible to people who may not have deep theoretical ethics knowledge, support continuous use (i.e. can be modified and used continuously throughout the project), encourage and support reflection, and consider the holistic impact of the research being conducted on a wide range of effected stakeholders. It is proposed that the Ethics Canvas [14], a recently proposed tool for identifying ethical impacts of research and innovation may provide a useful aid to helping researchers to identify important ethical issues with their work, recognise relevant stakeholders, and may potentially trigger reconsideration of the robot or experiment/application design.

II. Ethics Canvas

The Ethics Canvas is a collaborative brainstorming tool with the overall aim to foster ethically informed technology design by improving the engagement of research and innovation practitioners with the ethical impacts of their activities [14]. The method was initially conceived by the Adapt research centre based in Trinity College Dublin, and was inspired by a widely used a business-modelling tool known as the Business Model Canvas (BMC) [15]. The BMC is a visual-linguistic tool that can be used in a collaborative process in which participants generate ideas by offering and discussing certain narratives that are related to a series of text boxes displayed on a one page canvas. The canvas is divided into nine thematic building blocks, which enable key elements of a business model to be described in a holistic manner. Two key benefits of the BMC is that it it is "highly accessible and understandable to people without specific knowledge of the field" and it also supports interdisciplinary use within teams since it "relies on the collaborative generation of participant narratives" [14]. The tool is widely used within "lean" and "agile" business practises, since it is quick to complete and can be easily refined over time.

Using this framework, the Ethics Canvas was conceived to offer guidance for thinking about ethical impacts of a technology in a holistic way. For this purpose, it has retained the nine 'building blocks' of the BMC and reorientated to help answer three basic questions: (1) Who might be affected by the technology? [blocks 1,2] (2) What are the potential ethical impacts for these people and groups [blocks 3-8] and (3) how can we address these ethical impacts? [block 9] [16]. The Ethics canvas template (Fig. 1) is available free for use through a Creative Commons Attribution-Share Alike 3.0 unported license. It is available for download online, or can be completed online using a dedicated web-interface [10].

III. Case Study

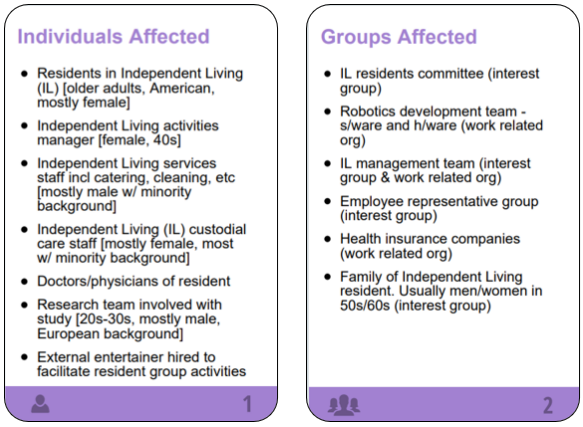
A case study was undertaken to explore the potential usefulness of the Ethical Canvas for addressing ethical issues in the design of a HRI experiment. The experiment that was chosen involved a pilot of robot technology at a Continued Care Retirement Community (CCRC) located on the east coast of the US. The specific use-case involved the Stevie robot (Fig. 2(a)) mediating a group-based game, such as Bingo, with residents housed within the Independent Living (IL) wing of the facility. Group-based game playing is a popular activity in many retirement communities [17], helping to increase cognitive stimulation as well as reduce boredom/loneliness. Prior to the completion of the Ethical Canvas, the researcher had gained a detailed understanding of the facility from a series of needfinding exercises including focus groups, interviews and observation sessions with residents and staff.

The application of robot technology to care-related tasks involving older adults remains an ethically complex domain [12]; on the one hand, the technology has tangible care benefits, however their use also raises significant concerns regarding: (1) the potential reduction in the amount of human contact; (2) an increase in the feelings of objectification and loss of control; (3) a loss of privacy; (4) a loss of personal liberty; (5) deception and infantilisation; (6) the circumstances in which elderly people should be allowed to control robots [18]. While the robot was not intended to physically interact with participants during the game, this scenario presents several 'hidden' dangers due to factors including physical/cognitive impairment of some of the older adults in the test group (MCI, reduced mobility, etc.), the fact that the robot was an experimental hardware platform, and that certain aspects of the robots behaviour was autonomous.

A. Stage 1: Identify Stakeholders

The first stage of the Ethical Canvas is concerned with identifying key individual (block 1) and group stakeholders (block 2). The purpose of this stage is to better understand the people and groups who are likely to be most affected by the the introduction of the technology. This part of the analysis promoted a greater understanding of primary users (such as identifying that most residents and care staff were female). It also revealed that many of the front-line workers were of minority background, and may be an especially vulnerable group within the workforce. Of the groups affected, there was a combination of interest groups (resident committees, family/friends of residents, etc.) as well as commercial and work-related organisations (i.e. design team, insurance companies). Since the retirement community in this study is part of an not-for-profit organisation, it can be considered as both interest group and work organisation. A summary of the responses provided for stage 1 is presented in figure 2.

(a)



(b) (c)

Fig. 2. (a) Image of Stevie robot interacting with residents during a focus group session, (b) Outcomes from stage 1 of the Ethical Canvas.

B. Stage 2: Identify Ethical Impacts

During the second stage, which consideres blocks 3-8, potential ethical impacts for the different stakeholders are captured. This involves firstly identifying the 'micro' impacts that influence the everyday lives of people using and living with the robot. This part captured how peoples behaviour may change because of the robot (block 3) and also how relations between people/groups may change. The analysis of the use-case suggested that the introduction of the robot may result in several beneficial behavioural and relation changes, such as: increased levels of engagement from socially isolated residents, pressure alleviation for front line staff and the possibility that the availability of the robot might spawn additional group activities such as training to use the robot. However, the canvas also revealed several threats including more missed/late appointments, reduction in contact time between care staff

and residents, and the possibility that residents would feel objectified by the presence of the robot and possibly perceive it as a safety risk. The responses provided for blocks 3 and 4 are shown in figure 3(a)-3(b).

Blocks 5 and 6 deal with potential 'macro' impacts. These impacts extend beyond the level of an individuals everyday life and pertain to broader social structures. They consider how the use of the robot may influence people's worldviews (block 5) as well as inter-group conflict that may arise as a direct or indirect result of the technology. The canvas suggests that successful introduction of the robot would help validate the use of social robots in a care setting and may help combat damaging stereotypes pertaining to technology usage among older adults. However, it may also demonstrate that even jobs that require high levels of social interaction may soon be automated by advanced technology. It was also identified that deploying the robot for a prolonged period may cause conflicts among existing staff who may need to take on new roles and responsibilities, residents who feel overlooked and objectified by the presence of the robot, family/friends of residents who feel there loved ones are being put at unnecessary risk of harm or deprived of human interaction, and dedicated activities staff who feel that they are being replaced by a piece of technology. The responses provided for blocks 5 and 6 are shown in figure 3(c)-3(d).

Finally, the remaining two blocks in stage 2 concern the negative side-effects of the design and deployment of the robot. In block 7, potential negative impacts of the robot failing to operate or to be used as intended are suggested. In block 8, negative impacts arising from the consumption of resources are detailed. This part of the analysis revealed several distinct failure modes and their potential consequences. Four distinct outcomes of potential failures emerged during this part of the analysis: (i) failure resulting in physical harm. (ii) failure resulting in compromise of personal data, and (iii) failure resulting in harm to the mental well-being of the resident, and (iv) failure resulting in distrust of the robot technology. This part of the study also revealed ways in which deploying the robot may comprise important resources. For example, it is possible that while the robot may alleviate some of the pressures of care workers, managing the robot may present new responsibilities. Additionally, it identified that the benefits of the technology may not be balanced by the overall cost[2] to deploy the technology. Outcomes for blocks 7 and 8 are shown in figure 3(e)-3(f).

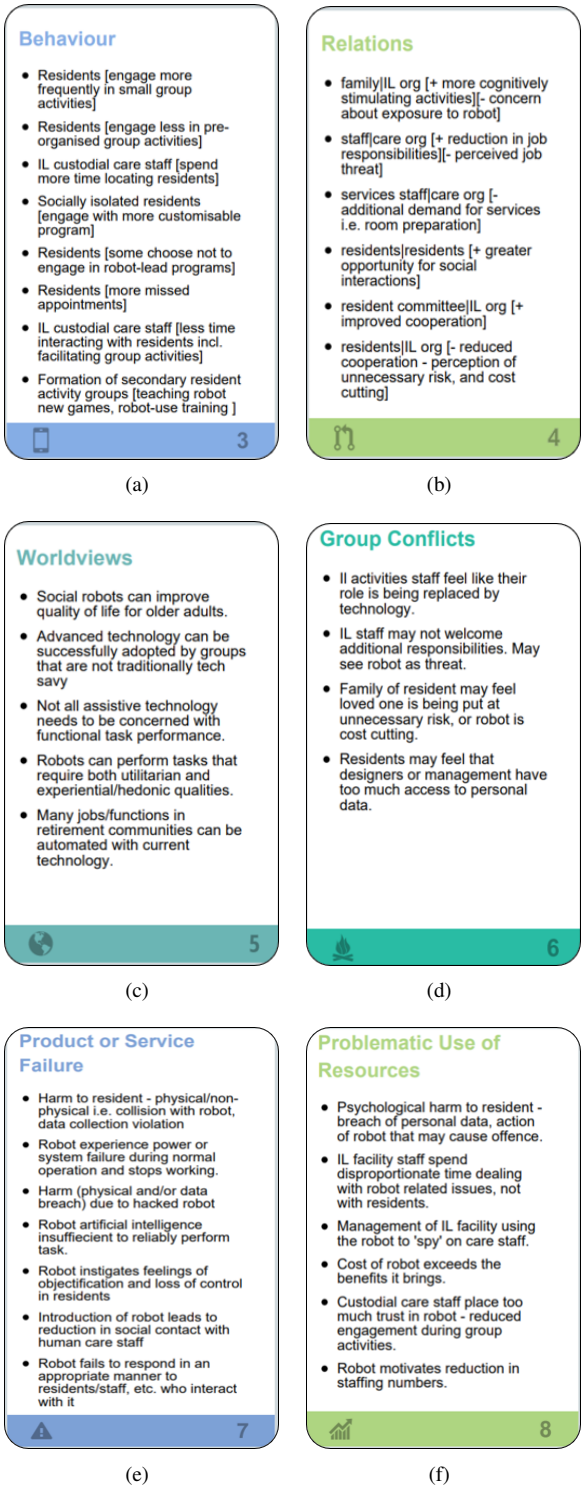


Fig. 3. Outcomes from stage 2 of the Ethical Canvas.

C. Stage 3: Address Ethical Impacts

After the potential ethical impacts have been identified, the final stage is concerned with outlining possible ways to address them. In block 9, responses to the most important ethical impacts are formulated; responses may involve adapting the design of the system, or making changes to how it is deployed/used in the application. Based on the ethical impacts identified, several suggestions are made to address issues that arose with his use case. Given the number of conflicts that arose in block 6, it is suggested that co-design activities are conducted involving a range of different stakeholders to determine exactly how the robot is introduced to the facility. It is also suggested that these workshops are complemented

[2]The overall 'cost' considers the direct financial cost to the organisation, the requirement to upgrade building infrastructure, increased insurance premiums, need to train staff, etc.

by frequent information sessions which can serve as a vehicle to disseminate outcomes of co-design exercises in addition to providing a clear stream of communication to those affected by the introduction of the technology. Given the large number of failure modes, and the potential consequences of these failures, it is suggested that risk of failure may be mitigated through the utilization of a semi-autonomous system. The response to this stage also advocates for a gradual roll-out of the technology to allow for on-going ethnographic evaluation. Risk of harm is likely to be further reduced through adherence with the relevant technical standards, including the performance of a formal ethical review in adherence with BS 8611-2016. Suggestions from stage 3 of the canvas are presented in figure 4.

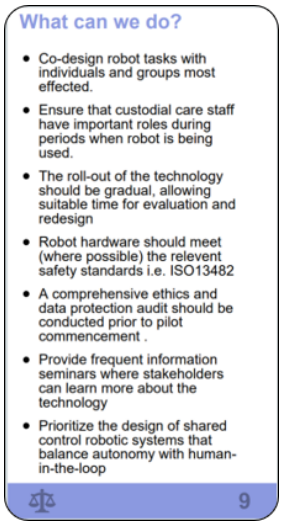


Fig. 4. Outcomes from stage 3 of the Ethical Canvas.

IV. Conclusions

This study set out to investigate if the Ethics Canvas, a recently proposed tool for guiding ethically informed technology design, might be useful in a HRI context. To test this hypothesis, the Ethics Canvas was applied to better understand the ethical implications of a use-case concerning a social robot being deployed to facilitate group activities in a retirement community. It emerged that the Ethics Canvas provoked thorough consideration of the stakeholders affected, potential ethical impacts (at both micro and macro levels), potential consequences of robot failure and resource demand, as well as possible remediation measures. The method proved to be highly compact, with the completed table fitting on just one A4 page. Furthermore, the accessibility of the method to non-ethicists was evidenced by the fact that analysis was predominately undertaken by engineers with no formal ethics training. Using the Ethics Canvas permitted a holistic evaluation of the problem, and ultimately cumulated in a comprehensive representation of the ethical landscape along with a number of recommendations to address key ethical issues.

The findings from this research suggest that the Ethical Canvas may provide a very useful way for HRI practitioners to gain fundamental insights into the ethical issues associated with real-world applications involving robots. Furthermore, it is proposed that if made part of standard experimental reporting in the field (i.e. if published HRI studies involving human subjects were to include a Ethical Canvas in the paper/appendices), it would serve the dual purpose of informing ethical best practise as well as providing tangible evidence that ethical issues were considered throughout the performance of the study.

References

[1] P. Lin, K. Abney, and G. A. Bekey, *Robot ethics: the ethical and social implications of robotics*. The MIT Press, 2014.

[2] C. Holder, V. Khurana, F. Harrison, and L. Jacobs, "Robotics and law: Key legal and regulatory implications of the robotics age (part i of ii)," *Computer Law & Security Review*, vol. 32, no. 3, pp. 383–402, 2016.

[3] A. Lindseth and A. Norberg, "A phenomenological hermeneutical method for researching lived experience," *Scandinavian journal of caring sciences*, vol. 18, no. 2, pp. 145–153, 2004.

[4] B. Friedman and P. H. Kahn Jr, "Human values, ethics, and design," in *The human-computer interaction handbook*. CRC Press, 2007, pp. 1223–1248.

[5] I. R. Nourbakhsh, *Robot futures*. Mit Press, 2013.

[6] J. P. Sullins, "Introduction: Open questions in roboethics," *Philosophy & Technology*, vol. 24, no. 3, p. 233, 2011.

[7] A. Albrechtslund, "Ethics and technology design," *Ethics and information technology*, vol. 9, no. 1, pp. 63–72, 2007.

[8] J. P. Sullins, "Applied professional ethics for the reluctant roboticist," in *The Emerging Policy and Ethics of Human-Robot Interaction workshop at HRI*, 2015.

[9] K. Zawieska, "Is roboethics really optional?" in *An Alternative HRI Methodology: The Use of Ethnography to Identify and Address Ethical, Legal, Societal (ELS) Issues workshop at HRI 2018*, 2018.

[10] "Online Ethics Canvas," https://www.ethicscanvas.org/, accessed: 2018-01-15.

[11] B. Ingram, D. Jones, A. Lewis, M. Richards, C. Rich, and L. Schachterle, "A code of ethics for robotics engineers," in *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*. IEEE Press, 2010, pp. 103–104.

[12] "A code of ethics for the human-robot interaction profession," *Proceedings of We Robot*.

[13] BS 8611-2016 , "Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems," British Standards Institution, Standard, 2016.

[14] W. Reijers, K. Koidl, D. Lewis, H. J. Pandit, and B. Gordijn, "Discussing ethical impacts in research and innovation: The ethics canvas," in *IFIP International Conference on Human Choice and Computers*. Springer, 2018, pp. 299–313.

[15] A. Osterwalder and Y. Pigneur, *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons, 2010.

[16] D. Lewis, W. Reijers, H. Pandit, and W. Reijers, "Ethics canvas manual," 2017.

[17] W.-Y. G. Louie, J. Li, C. Mohamed, F. Despond, V. Lee, and G. Nejat, "Tangy the robot bingo facilitator: A performance review," *Journal of Medical Devices*, vol. 9, no. 2, p. 020936, 2015.

[18] A. Sharkey and N. Sharkey, "Granny and the robots: ethical issues in robot care for the elderly," *Ethics and information technology*, vol. 14, no. 1, pp. 27–40, 2012.