Embedding Ethics in Human Factors Design and Evaluation Methodologies

Joan Cahill(✉)

Trinity College Dublin, Dublin, Ireland
cahilljo@tcd.ie

Abstract. Methodologies are required to enable the active translation of ethical issues pertaining to the human and social dimensions of new technologies, in a manner that considers the diversity of practices across research and innovation and commercial research projects. This paper presents a new methodology for embedding ethics assessment in human machine interaction (HMI)/human factors (HF) design and evaluation activities.

Keywords: Human factors · Ethics · Impact assessment · Responsible research & innovation · Societal impact · Emerging technologies

1 Introduction

Human activity should not compromise the long-term balance between the economic, environmental and social pillars [1]. The evaluation of impact is a necessary part of all research and innovation (R&I) and attempts to improve the relationship between science and society. This follows a 'perspective oriented to humane and social values' [2] and recommendations from the European Union (EU) in relation to undertaking 'responsible research and innovation (RRI)' [3] and the involvement of societal actors in R&I [4]. New technologies have the potential to deliver benefits. However, such technologies are inherently uncertain. As stated by Capurro (2009), technology designers must examine the ethical implications of things which may not yet exist, or things which may have impacts we cannot predict [5]. In so doing, they must deal with uncertainty. This includes the 'uncertainty of future products, uses and consequences, and associated ethical issues that will result from an emerging technology' [6].

In asking what technology is and how it might be designed, we ask questions about who we are (identity) and what it means to be human [7]. As stated by Heidegger [1977], we examine the nature of existence and human autonomy [7]. Such ideas have led to the concept of 'ontological design' which addresses how the design of technology changes our human and social reality [8]. As such, we are designed by our designing and by that which we have designed [9].

Design/technology teams exercise choice in relation to what is valued and advancing technology that improves the human condition (and not worsens it). As researchers we need methods to assess and practice ethics, to ensure that new technologies positively contribute to human wellbeing and have positive impacts across the triple bottom

V. G. Duffy (Ed.): HCII 2020, LNCS 12199, pp. 217–227, 2020.
https://doi.org/10.1007/978-3-030-49907-5_15

line. In an ideal world, R&I teams are multi-disciplinary and include ethicists. Further, stakeholder evaluation underpins the generation of an evidence map and proposed solutions. However, this is not always the case. Methodologies are required to enable the active translation of ethical issues pertaining to the human and social dimensions of new technologies, in a manner that considers the diversity of practices across R&I and commercial research projects. To this end, this paper presents a new methodology for embedding ethics assessment in human machine interaction (HMI)/human factors (HF) design and evaluation activities.

2 Background

2.1 Underlying Concepts

Human Factors refers to 'the practice of designing products, systems, or processes to take proper account of the interaction between them and the people who use them' [10]. Ethics concerns the moral principles that govern a person's behavior or how an activity is conducted. As researchers, we must distinguish research ethics (i.e. the normative aspects of engaging in scientific research) and the ethics of technological innovation and its impacts at different levels. 'Digital ethics' or information ethics deals with the impact of digital information and communication technologies (ICT) on society and the environment. Data ethics is defined as a branch of ethics that evaluates data practices with the potential to adversely impact on people and society [11].

2.2 The Practice of Ethics in R&I

A recent systematic review indicates that the practice of ethics in R&I is a relatively new topic [12]. While academic discussion on specific practices commenced in the 1990s, this research has gained considerable momentum in the last ten years [12]. According to Reijers et al. (2017), health technologies is the most represented in the literature, followed by the fields of information systems research and computer science [12].

Specific ethics approaches in R&I can take many forms. Reijers et al. (2017) categorize the different methods in relation to their application in the technology development lifecycle – distinguishing (1) ex ante methods, dealing with emerging technologies (2) intra methods, dealing with technology design and (3) ex post methods, dealing with ethical analysis of existing technologies [12]. Research evidence can include information from horizon scanning and participatory foresight activities, literature reviews, and field research with stakeholders [13]. Specific stakeholder evaluation research (i.e. empirical research) may take different formulations. Stakeholders may engage directly or indirectly with R&I teams. Researchers and stakeholders may engage with ethical challenges in a collaborative workshop. Or, research may be undertaken with stakeholders and later examined by research and design teams in a structured format. This format may follow specific conceptual frameworks and assessment approaches. Several key frameworks for ethical assessment have emerged. This includes (but is not limited to): ontological design [9], anticipatory technology ethics/foresight approach [13], value sensitive design [14], ethical impact assessment [15], the ETICA approach

[16], and the techno-ethical scenarios approach [17]. Brey (2017) classifies five sets of ethical impact assessment approaches. This includes generic approaches, anticipatory/foresight approaches, risk assessment approaches, experimental approaches and participatory/deliberative ethics approaches [13]. Increasing, researchers are combining approaches. For example, Cotton (2014) combines participatory/deliberative ethics approaches and stakeholder approaches [18].

2.3 Ethics Canvases

Ethics canvases or visual tools which support the ethics assessment approach are not being used in commercial and research projects. In principle, these canvases allow non-ethicists such as Designers, Human Factors Researchers, Engineers, and Computer Scientists to engage in ethical issues pertaining to the emerging technology product. Examples of such canvases include the 'Research Impacts Canvas' (RIC) [19], The Ethical Matrix [20], The Digital Product Ethics Canvas and Impacts Canvas [21], The Humans & Machines Ethics Canvas's [22], The Online Ethics Canvas [23], and the Data Ethics Canvas [24]. Some canvases focus on ethics and impact in a general sense, while others address specific themes. For example, the Online Ethics Canvas addresses the impact of new technology on human behavior and activity at individual and societal levels [23]. The Data Ethics Canvas considers ethical issues related to data privacy, data use and data quality [24].

3 Human Factors and Ethics Canvas

3.1 Rationale

Critically, human factors and ethical issues must be explored in an integrated way. Although valuable, the existing ethics canvases require further emphasis on framing the problem, specifying the psychosocial dimensions and impacts of new technologies and addressing specific stakeholder/end user requirements and impacts. Further, ethical issues need to be managed in terms of design decisions. These decisions need to be agreed and documented.

The 'Human Factors & Ethics Canvas' introduced by Cahill (2019) [25] reflects an integration of ethics and HF methods, particularly around the collection of evidence using stakeholder evaluation methods [26, 27] personae-based design [28], scenario-based design approaches [29]. Further, it makes use of ethical theories/perspectives that are used in relation to the analysis of technology innovation in relation to the analysis of benefit versus harm including Consequentialism, Deontology & Principlism [30].

3.2 Procedure

The HFEC can be used at any stage of the design process. As such, it spans the classification of methods proposed by Reijers et al. [12]. Overall, it combines anticipatory/foresight approaches and participatory/deliberative ethics approaches. In line with stakeholder evaluation approaches, the canvas can be evaluated using the 'community of

practice' [27]. That is, using internal stakeholders (project team) and external stakeholders (relevant ends users/stakeholders and legitimate other parties who may be impacted by the technology). At a minimum, core internal stakeholders/core team members (including an ethicist {if available}, the HF lead, the design lead and the product owner/manager) are involved in completing the canvas. If the project team includes an ethicist, then they should take the role of the 'HFEC' coordinator - recording relevant information in the HFEC. Otherwise, this can be done by the HF lead or another designated member of the project team.

As indicated in Fig. 1, the HFEC is divided into seven stages or sections. For more, please see Appendix A. Stage 0 records project information. Stage 1 is all about framing the problem. Stage 2 involves understanding how the technology fits to the problem, defining stakeholder goals and needs and the specification of expected benefits for different stakeholders. This is followed by several more detailed examinations of core themes. These are: benefits, outcomes and impact (stage 3), personae and scenario (stage 4), data ethics (stage 5) and implementation (stage 6). The final stage (stage 7) presents the outcomes of the preceding analysis. An analysis of literature review data and information from team problem solving sessions can be used to populate the HFEC. However, it is best to complete Stage 3 and 4 either using stakeholder evaluation approaches (either direct engagement of stakeholders in ethical assessment or following the analysis of field research with stakeholders). In addition, Stage 6 can only be completed following implementation and evaluation of the proposed technologies. Ideally, this might occur in a field setting. However, information from simulation studies can also be used.

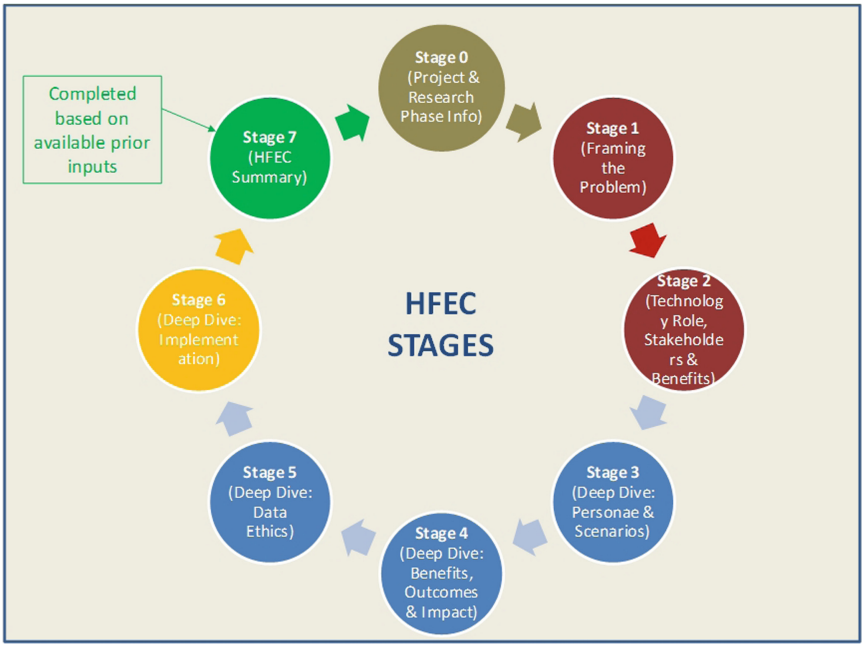

Fig. 1. Stages in Human Factors & Ethics Canvas

4 Discussion

As illustrated in the ethics canvas, there is much convergence between the analysis of new technology both from an ethics and human factors perspective (for example, addressing stakeholder need, expected benefits and outcomes, and impact [intended and unintended] – both at an individual and societal level). Ethical principles need to be both articulated and then embedded in the design concept. Personae/scenarios are useful in relation to considering and documenting the needs/perspectives of different stakeholders and adjudicating between conflicting goals/principles. Moreover, the translation of system objectives in relation to wellbeing and human benefit objectives (and associated metrics) ensures that wellbeing and human benefit are both a reference point and a design outcome.

As highlighted by Brey (2017), the ethics of emerging technologies 'harbors the promise of early intervention when a technology is still malleable and there is still much room for choice in its development and social embedding' [13]. However, researchers have a limited range of empirical data to use. As the technologies are not in use, there are 'significant uncertainties regarding future developments and impacts' (Brey 2017). Some theorists present philosophical objections to speculation about future impacts [31, 32]. For example, Nordmann (2007) contends that speculation about the future should be rejected as researchers cannot gain sufficient knowledge about the future to stipulate procedures for action or guidance in R&I processes [32]. Others argue that the available theories and methods do not provide adequate theoretical grounding in terms of how values might be embedded in design solutions [33]. In addition, VSD and related approaches must address the difference between designer's intentions and user practice [34].

5 Conclusion

Assessing the ethical implications of things which may not yet exist, or things which may have impacts we cannot predict, is very difficult. However, this should not be barrier to posing important questions and ensuring that these questions are addressed as part of the design process. Thinking about both potential positive, negative consequences and unintended consequences enables designers to build in protections into the design concept. Overall, it is argued that the specification of an ethics canvas as part of a broader human factors design approach ensures that ethical issues are considered.

Appendix A: Human Factors and Ethics Canvas (HFEC)

Stage 0 (Project Information and Research Summary)

See Table 1.

Table 1. HFEC: Stage 0 (Project Information & Research Summary)

#	0: Project Information & Research Summary
1	Date
2	Project Name
3	Product Owner
4	HF & Ethics Coordinator
5	HF & Ethics Canvas Version No.
6	Prior HFEC Iterations
7	Research & Innovation Phase
8	Summary of Research Completed & Key Sources of Information/Evidence

Stage 1 (Formulating the Problem and Framing the Question)

See Table 2.

Table 2. HFEC: Stage 1 (Formulating the Problem & Framing the Question)

#	1: Formulating the problem and framing the question
1	What is the problem that the proposed technology will address?
2	Who is it a problem for? Key stakeholders? Who effect (directly and indirectly?)
3	Setting & Environment?
4	Causes of the problem?
5	Ethical codes that apply in this setting?
6	Ethics embedded in the problem definition?
7	Ethics & Impact of Problem. Individual Level. Societal level. Ethics of acting/not acting?
8	Summary of ethical issues to be addressed?
9	Summary of relevant ethics principles and frameworks?
10	Ethics & Key KPI?

Stage 2 (Understanding Technology and Fit to Problem/Stakeholder Needs & Expected Benefits)

See Table 3.

Table 3. HFEC: Stage 2 (Understanding Technology & Fit to Problem/Stakeholder Needs & Expected Benefits)

#	2: Understanding Technology & Fit to Problem/Stakeholder Needs & Expected Benefits
1	What is the technology? How does tech address the problem? What part of the problem does it address?
2	Who is it a problem for? Key stakeholders? Whom effect (directly and indirectly?)
3	What is the goal/objective? Intended purpose/function?
4	Setting & Environment?
5	Direct users of technology? Goals? Needs? Expected Benefits?
6	Other stakeholders impacted by technology? Goals? Needs? Expected Benefits?

Stage 3 (Deep Dive: Benefits, Outcomes and Impact)

See Table 4.

Table 4. HFEC: Stage 3 (Deep Dive: Benefits, Outcomes & Impact)

#	3: Deep Dive: Benefits, Outcomes & Impact			
1	Overall benefits and outcomes: key stakeholders? Expected positive impacts?			
2	Expected Impact for key stakeholders (psycho-social themes). Individual level? Societal Level?	(A) Human role in the system	(B) Human Identity	(C) Lived experience, wellbeing, quality of life
		(D) Social Interaction & Relationships	(E) Activity & Behavior	(F) Attitudes & Values
3	What could go wrong? Potential failures? Potential negative impacts? Psychosocial? Environmental?			
4	Unintended consequences			
5	Unknowns			

Stage 4 (Deep Dive: Personae and Scenarios)

See Table 5.

Table 5. HFEC: Stage 4 (Deep Dive: Personae & Scenarios)

#	4: Deep Dive: Personae & Scenarios
1	Example Scenario
2	Example Personae
3	How is it expected to work?
4	What does success look like? Benefits for whom? Expected positive outcomes and for whom?
5	What could go wrong? Potential failures? Potential negative impacts?
6	Unintended consequences?
7	Unknowns?
8	Design Decisions & Safeguards

Stage 5 (Deep Dive: Data Ethics)

See Table 6.

Table 6. HFEC: Stage 5 (Deep Dive: Data Ethics)

#	5: Deep Dive: Data Ethics
1	Ethical issues relevant to data collection? What data? Why collecting? Potential for bias in data collection?
2	Ethical issues relevant to data, model & algorithms? Potential for harm and risk?
3	Ethical issues relevant to data use & predictions (i.e. application of model/algorithms)?
4	Ethical issues relevant to data sharing?
5	Design Decisions & Safeguards

Stage 6 (Implementation)

See Table 7.

Table 7. HFEC: Stage 6 (Implementation)

#	6: Implementation
1	Implementation Approach
2	Implementation Enablers
3	Implementation Barriers
4	Systems Perspective: Addressing Ethics as part of Implementation. People. Process. Technology. Culture. Training & Education
5	Design Decisions & Safeguards

Stage 7 (Human Factors and Ethics Summary)

See Table 8.

Table 8. HFEC: Stage 7 (Human Factors & Ethics Summary)

#	7: Human Factors & Ethics Summary
1	Key stakeholders? Who is this technology designed for?
2	What does success look like? Success for whom?
3	Human/Societal Vision & Technology Role/Purpose
4	Summary of Key Ethical Issues to be Addressed?
5	Ethical Principles Underlying Technology Design
6	Design Approach: Balancing Benefits & Harm How managing ethics issues? How increasing potential positive impacts? How preventing risk/harm? How managing potential negative impacts and unintended consequences? How addressing unknowns?
7	Data Ethics Summary
8	Implementation Summary
9	Ethics & Key KPI

References

1. Elkington, J.: Cannibals with forks: the triple bottom line of 21st century business, ISBN 9780865713925. OCLC 963459936. Capstone, Oxford (1999)

2. Stephanidis, C.C., et al.: Seven HCI grand challenges. Int. J. Hum.– Comput. Interact. **35**(14), 1229–1269 (2019). https://doi.org/10.1080/10447318.2019.1619259
3. European Commission: Ethics for researchers. http://ec.europa.eu/research/science-society/document_library/pdf_06/ethics-forresearchers_en.pdf. Accessed 13 Feb 2020
4. Geoghegan-Quinn, M.: Responsible Research & Innovation. European Union Publications Office, Brussels (2014)
5. Capurro, R.: Digital ethics. In: The Academy of Korean Studies (ed.): Civilization and Peace. Academy of Korean Studies 2010, Korea, pp. 203–214 (2009)
6. Sollie, P.: Ethics, technology development and uncertainty: an outline for any future ethics of technology. J. Inf. Commun. Ethics Soc. **5**(4), 293–306 (2007)
7. Heidegger, M.: The Question Concerning Technology, and Other Essays, 6th edn. Harper & Row, New York (1977)
8. Winograd, T., Flores, F.: Understanding Computers and Cognition: A New Foundation for Design. Ablex Publishing Corporation, Norwood (1986)
9. Fry, T.: Becoming Human by Design. Berg Publishers, Oxford (2012)
10. International Standards Organisation (ISO): Standard 6385 (2020). https://www.iso.org/standard/63785.html
11. Open Data Institute (ODI): The Data Ethics Canvas (2020). https://theodi.org/article/data-ethics-canvas/
12. Reijers, W., et al.: Methods for practising ethics in research and innovation: a literature review, critical analysis and recommendations. Sci. Eng. Ethics **24**(5), 1437–1481 (2017). https://doi.org/10.1007/s11948-017-9961-8
13. Brey, P.: Ethics of emerging technologies. In: Hansson, S.O. (ed.) Methods for the Ethics of Technology. Rowman and Littlefield International, Lanham (2017)
14. Friedman, B., David, G.: Value Sensitive Design: Shaping Technology with Moral Imagination. MIT Press, Cambridge (2019)
15. Wright, D., Mordini, E.: Privacy and ethical impact assessment. In: Wright, D., De Hert, P. (eds.) Privacy Impact Assessment Law, Governance and Technology Series, vol. 6. Springer, Dordrecht (2012). https://doi.org/10.1007/978-94-007-2543-0_19
16. Stahl, B., Heersmink, R., Goujon, P., Flick, C., Van den Hoven, J., Wakunuma, K.: Identifying the ethics of emerging information and communication technologies: an essay on issues, concepts and method. Int. J. Technoeth. **1**(4), 20–38 (2010)
17. Boenink, M., Swierstra, T., Stemerding, D.: Anticipating the interaction between technology and morality: a scenario study of experimenting with humans in bionanotechnology. Stud. Ethics Law Technol. **4**(2), 1–38 (2010)
18. Cotton, M.: Ethics and Technology Assessment: A Participatory Approach. Springer, Berlin (2014). https://doi.org/10.1007/978-3-642-45088-4
19. Fecher, B., Kobsda, C.: Research Impacts Canvas (RIC). https://elephantinthelab.org/meet-the-research-impact-canvas-a-structured-guide-for-planning-your-science-communication-activities/. Accessed 13 Feb 2020
20. Forsberg, E.M.: The ethical matrix—a tool for ethical assessments of biotechnology. Glob. Bioeth. **17**(1), 167–172 (2004). https://doi.org/10.1080/11287462.2004.10800856
21. Gerlach, R. The Digital Product Ethics Canvas. https://www.threebility.com/post/the-digital-product-ethics-canvas. Accessed 13 Feb 2020
22. Vaish, P.: Humans & machines ethics canvas (2020). https://adataanalyst.com/wp-content/uploads/2016/09/ETHICS_Canvas_2.pdf
23. Adapt Centre for Digital Content Technologies. The Data Ethics Canvas. https://ethicscanvas.org/. Accessed 13 Feb 2020
24. Cahill, J.: Human factors & ethics canvas: a white paper. https://www.tcd.ie/cihs/projects/hfaecanvas.php. Accessed 13 Feb 2020

25. Cousins, J.B., Whitmore, E., Shulha, L.: Arguments for a common set of principles for collaborative inquiry in evaluation. Am. J. Eval. **34**, 7–22 (2013)
26. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, Cambridge (1998)
27. Pruitt, J., Grudin, J.: Personas: practice and theory. In: Proceedings of the 2003 Conference on Designing for User Experiences (DUX 2003), pp. 1–15. ACM, New York (2003). https://doi.org/10.1145/997078.997089
28. Carroll, J.M.: Scenario-Based Design: Envisioning Work and Technology in System Development. Wiley, New York (1995)
29. Beever, J., Brightman, Andrew O.: Reflexive principlism as an effective approach for developing ethical reasoning in engineering. Sci. Eng. Ethics **22**(1), 275–291 (2015). https://doi.org/10.1007/s11948-015-9633-5
30. Markus, M.L., Mentzer, K.: Foresight for a responsible future with ICT. Inf. Syst. Front. **16**, 353–368 (2014). https://doi.org/10.1007/s10796-013-9479-9
31. Nordmann, A.: If and then: a critique of speculative nanoethics. NanoEthics **1**(1), 31–46 (2007). https://doi.org/10.1007/s11569-007-0007-6
32. Poel, I.: Translating values into design requirements. In: Michelfelder, Diane P., McCarthy, N., Goldberg, David E. (eds.) Philosophy and Engineering: Reflections on Practice, Principles and Process. PET, vol. 15, pp. 253–266. Springer, Dordrecht (2013). https://doi.org/10.1007/978-94-007-7762-0_20
33. Van den Hoven, J., Manders-Huits, N.: Value-sensitive design. In: Kyrre, J., Olsen, B., Hendricks, V.F. (eds.) A Companion to the Philosophy of Technology. Blackwell Publishing, Malden (2009). https://doi.org/10.1002/9781444310795.ch1
34. Albrechtslund, A.: Ethics and technology design. Ethics Inf. Technol. **9**(1), 63–72 (2007). https://doi.org/10.1007/s10676-006-9129-8