CIHS White Paper: The Specification of a 'Human Factors and Ethics' Canvas for Socio-technical Systems

Author

Dr Joan Cahill, Centre for Innovative Human Systems (CIHS), School of Psychology, Trinity College Dublin, 2019.

Version **Version 2 (January 2020)**

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Table of Contents:

1. Introduction

Personalised technology and intelligent/automated systems have now become part of our personal and working lives. In healthcare, technology is becoming more pervasive. This is evidenced by the emergence of health monitoring/nudging technologies, promoting behaviour change centred around desirable physical, social and mental health outcomes. Such technologies involve a range of sensors (wearable, environmental and implantable) and novel artificial intelligence (AI) and machine learning (ML) components.

According to Sollie (2007), the primary problem for the ethics of emerging technology is the problem of uncertainty. That is, how to deal with the uncertainty of future products, uses and consequences, and associated ethical issues that will result from an emerging technology (Sollie, 2007).

Future technology is shaping (and will shape) our political, social and moral existence. The use and potential impact of these technologies raise many interrelated questions pertaining to ethics, law and social acceptability. In relation to ethics, these new technologies (including innovative AI & ML components) raise macro ethics questions concerning (1) the intended use and purpose of technology, (2) the role of the person, (3) the impact of these technologies on our behaviour and activities (including potentially negative consequences), and (4) societal values.

As highlighted by Brobst (2018), the beneficent uses of technology have a way of being co-opted for other purposes. The application of automation and advanced machine learning technologies changes the role of the person in the system. Often there can be a gap between the intended use of a system and its use in terms of what is implemented. The social effects of certain technologies are not always apparent. For example, the ways AI could affect social relationships and connections. Technologies both embed and promote values at a societal level. In this way, new technologies pose question in term of human identity (i.e. who we are and who we want to be). Design decisions are normative. As stated by Fry (2021), 'we are designed by our designing and by that which we have designed'.

Ethics issues are now being formally explored in commercial and research projects (European Commission, 2013). In parallel, both ethics canvases (Adapt, 2018) and data ethics canvases (ODI, 2019) are being used in technology projects. Much of the focus is on ethical issues related to data privacy and data quality (O' Keefe & O' Brein, 2018). Further, there is some focus on the psycho-social dimensions of new technologies, their impact on behaviour and activities, and risks and safeguards in relation to the use of AI and ML technologies which make decisions impacting on human wellbeing and rights (Adapt, 2018). However, there is limited integration with human factors themes and methodologies (for example, the collection of evidence about stakeholder goals, needs and benefits using stakeholder evaluation approaches).

Overall, we need methodologies to support the production and documentation of evidence in relation to addressing the human and ethical dimensions of future technologies. The responsibilities of designers and questions concerning the moral quality of technology belong to the field of Applied Ethics. However, they also belong to the field of Human Factors. To this end, this white paper presents a 'Human Factors and Ethics (HFAE) Canvas', which enables the active translation of ethical issues pertaining to the human and social dimensions of new technologies into ethically responsible solutions. The HFAE Canvas was developed across three human factors projects. This includes project pertaining to

- Developing an ethically responsible driving assistance solution for older adults.

- Addressing ethical issues in the development of assisted living technologies promoting wellbeing, independence and social participation.
- Advancing ethically responsible patient and caregiver monitoring technologies for use in acute and aged care settings

At the Centre of Innovative Human Systems (CIHS) we would like to promote engagement with issues around ethics and user acceptability. There is a need for honest conversations about the purpose of new technologies and their impact (and specifically, known and unknown consequences). 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 properly considered.

Designers, consumers and the community have responsibilities for the technologies that are created (Adamson, 2013). Ideally, new technologies should (a) enhance lived experience (b) protect human rights (for example, dignity, privacy, social interaction), (c) ensure human benefit and (d) prioritize human wellbeing. The objective of the 'human factors and ethics' canvas, is to create an evidence map in relation to the specification of an ethically responsible technology solution that properly addresses relevant human and ethical issues.

2. Background: Terms & Definitions

Progress

Progress is typically defined in relation to concepts of advancement and improvement. As stated by the Organization for Economic Co-Operation and Development's (OECD) 'Being able to measure people's quality of life is fundamental when assessing the progress of societies' (2007).

Lived Experience

In qualitative phenomenological research, lived experience refers to a representation of the experiences and choices of a given person and the knowledge that they gain from these experiences and choices (Van Manen, 1990). Experience includes both objective and subjective (i.e. first-person point of view) elements.

Values

Values are defined as "an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence" (Rockeach, 1973). Values are not solely morally defined and can depend on cultural traditions. Communities or members of a culture share values about what is good or desirable.

Responsibility & Duty

Responsibility refers to the state of being accountable for something or someone. Concepts of responsibility link to concepts of moral obligation and duty. As stated, 'Making judgments about whether a person is morally responsible for her behaviour and/or holding others and ourselves responsible for actions and the consequences of actions, is a fundamental and familiar part of our moral practices and our interpersonal relationships' (Standford, 2019).

Ethics & Applied Ethics

Overall, ethics concerns the moral principles that govern a person's behaviour or how an activity is conducted (OED, 2019). A key distinction in ethics is the distinction between that which is unethical

and that which is undesirable. Primarily, moral principles apply to a person. However, moral code can also be ascribed to the behaviour of automated or intelligent systems (A/IS) or artificial moral agents (AMA).

The discipline of Ethics has three branches (1) metaethics (whether morality exists), (2) normative ethics (sometimes referred to as ethical theory), and (3) applied ethical (practical application of moral considerations in the area of private and public life, health, technology law and so forth).

Critically, the field of applied ethics concerns real world actions and their moral considerations. It involves the application of ethical theory for the purpose of choosing an ethical action in each issue.

Digital Ethics/Information Ethics

'Digital ethics' or information ethics deals with the impact of digital Information and Communication Technologies (ICT) on our societies and the environment at large [19]. As defined by Capurro (2009), it addresses the ethical implications of things which may not yet exist, or things which may have impacts we cannot predict.

Data Ethics

The Open Data Institute defines data ethics as: 'A branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use'.

Rights

Philosophers distinguish between negative and positive rights. Negative and positive rights are rights that oblige either action or inaction. A negative right is a right not to be subjected to an action of another person or group. A positive right is a right to be subjected to an action or another person or group.

The Universal Declaration of human rights (1948) enshrines all persons with human rights [16]. This includes rights pertaining to dignity (Article 1), autonomy (Article 3), privacy (Article 12), and safety (Article 29). In terms of information technology, the right to privacy has been established and defined in relation to specific rules around data collection, access, storage, sharing and use.

Some would argue that rights also apply to technology and artificial agents. These are referred to as 'transhuman right's'. To this end, the field of roboethics has emerged. Specifically, roboethics is concerned with the moral behavior of humans as they design, construct, use and treat artificially intelligent beings.

3. Human Factors, Ontological Design & Evidenced Based Design

New Technologies

Automation and Intelligent Systems

Automation is the 'technique, method, or system of operating or controlling a process by highly automatic means, as by electronic devices, reducing human intervention to a minimum' (Dictionary.com).

Emerging Technologies

As defined by Brey (2017), emerging technologies are technologies that are new, innovative, and still in development. Importantly, emerging technologies are still technologies in the making (Brey,

2017). As such, they are not fully developed and entrenched in society (Brey, 2017). Further, these technologies are expected to have significant socioeconomic impact across different sectors including education, entertainment, healthcare, transportation and the retail industry.

Human Factors & Socio-technical Systems Design

As defined in ISO 6385, human factors (HF) concerns 'the practice of designing products, systems, or processes to take proper account of the interaction between them and the people who use them' (2016). Human factors research applies theory, principles, data and methods to design to optimize human well-being and overall system performance.

HF reflects a principle-based approach:

- **Role of person/person comes first** – intended use of technology system and human role, lived experience, wellbeing, value placed on autonomy/dignity/privacy
- Consider broader **socio-technical systems (systems approach)**
- Consider relationship between technology, behaviour change and impact (at individual and societal level)

Overall, the goal is to engage in stakeholder evaluation research (Cousins, 2015), to build an evidence map in relation to stakeholder need, requirements, acceptability and benefits, and to design and implement a solution which properly addresses this. In this way, we use human factors methodologies to collect evidence to justify or validate a technology design which reflects an appropriate balance (i.e. stakeholder benefit and harm) and a suitable implementation approach. Briefly stated, these methods include ethnography, process mapping, participatory design and evaluation, implementation planning & design etc

Importantly, human factors researchers investigate the functions and benefit of technology from the perspective of all relevant stakeholders.

Sociotechnical systems design is an approach to organizational work design that recognizes the interaction between people/behaviour, technology/tools, work processes, workplace environments and work culture (Baxter & Sommerville, 2011). As highlighted by Waterson (2009), human factors research focuses on system interactions and follows a 'systems approach'. Performance results from the interaction of a sociotechnical system comprising many elements (person, task design, technology, internal and external environment, culture etc.)

Ontological Design

The concept of 'ontological design' addresses the normative dimensions of design and design practice. Specifically, it focuses on the 'the relation between human beings and the design of lifeworld's' (Winnograd and Flores, 1986). As argued by Winnograd and Flores (1986), new technology does not simply change the task, it changes what it means to be human. Put simply, we are designed by our designing and by that which we have designed (Fry, 2012). The concept of ontological design follows from precepts around the design of human existence as defined by Heidegger. In 'The Question Concerning Technology', the philosopher Heidegger suggests that in asking what technology is, we ask questions about who we are (1977). In so doing, we examine the nature of existence and human autonomy (Heidegger, 1977).

Evidence Based Design

Evidence-based design approaches have become quite popular in relation to the design of healthcare technology (Ulrich et al. 2008). Here the focus is on making credible cases for new technology and associated design decisions (Pawson, 2006). Central to this is the idea of using evidence-based design decisions based on the best available information (Webster and Steinke 2009). This includes information from research, project evaluations and evidence gathered from client operations (Webster and Steinke 2009). In terms of software development practices, the production of evidence maps and argumentation to validate software design decisions is becoming more popular (John 2005). An evidence map includes the research questions, the answers to the research questions and the source of evidence.

4. Addressing Ethical Issues in Technology Development & the Emergence of an 'Ethics Canvas'

Introduction

In 1972, the IEEE Standards Association established the forerunner to the Society on Social Implications of Technology (SSIT). In 2000, the IEEE proposed the phrase "Advancing Technology for Humanity." Today, SSIT has five areas of focus: (1) technology ethics, (2) development technology, (3) technology sustainability, (4) access to technology, and (5) the impact of emerging technologies.

The IEEE P7000 family of standards (2019) has been described as a 'process model by which engineers and technologists can address ethical consideration throughout the various stages of system initiation, analysis and design'. Expected process requirements include management and engineering view of new IT product development, computer ethics and IT system design, value-sensitive design, and, stakeholder involvement in ethical IT system design (IEEE, 2019).

Further, the IEEE Standards Association has recently articulated a desire to create technology that improves the human condition and prioritizes wellbeing. Specifically, the 'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems' have defined a set of core ethical principles for autonomous and intelligent systems (A/IS). As stated in 'Ethically Aligned Design (EAD1e), A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems' 'for extended intelligence and automation to provably advance a specific benefit for humanity, there needs to be clear indicators of that benefit'. Further, the IEEE Global Initiative argue that 'the world's top metric of value (Gross Domestic Product) must move beyond GDP, to holistically measure how intelligent and autonomous systems can hinder or improve human well-being'.

In parallel, the Information Technology (IT) sector has taken many leaps in relation to addressing these questions. Currently, there is a large focus on issues pertaining to wellbeing, data privacy and cybersecurity. In 2016, Amazon, Google, Facebook, IBM, and Microsoft have established a non-profit partnership (i.e. the Partnership on Artificial Intelligence to Benefit People and Society) to formulate best practices on artificial intelligence technologies.

Ethical Theories & New Technologies

Many ethical theories have been used to examine the ethical issues pertaining to new technologies. Two main classical theories – namely, Consequentialism and Deontology are often used in relation to the analysis of technology innovation from the perspective of potential benefit versus harm. Consequentialism is the view that normative properties depend only on consequences. Put simply, the consequences of one's behaviour/action are the ultimate basis for any judgment about the rightness or wrongness of that behaviour/action. On the other hand, in Deontologists argue that the morality of a given behaviour/action should be based on whether that action itself is right or wrong under a series of rules, rather than based on the consequences of the action. As highlighted by Moor (1999), a combination of two is required.

Principlism is an approach to the examination of moral dilemmas based upon the application of certain ethical principles. Principlism provides for a way of 'ethical reasoning that is based on a core set of ethical principles that are not limited by any particular case or by theoretical constraints'(Beever & Brightman, 2015).

Example Impact & Ethics Canvas

As identified by Reijers et al (2017), academic discussion on practising ethics in R&I commenced in the 1990s. There has been considerable progress in this field in the last 10 years.

Several tools/canvases have emerged in relation to examining ethical issues pertaining to new technologies. The existing available Ethics Canvas follow from the paradigm developed in Osterwalder's Business Model Canvas (Osterwalder & Pigneur, 2010). The BMC features a visual chart with elements describing a firm's or product's value proposition, infrastructure, customers, and finances.

The 'Research Impacts Canvas' (RIC) is a project tool addressing societal impacts and value (Fecher & Kobsda, 2019). Although it is not focused on specific ethical issues, it attempts to focus attention on improving the relationship between science and society. Please see: https://elephantinthelab.org/meet-the-research-impact-canvas-a-structured-guide-for-planning-your-science-communication-activities/

The Online ETHICS Canvas (developed by the ADAPT Centre for Digital Content Technologies provides a framework for brainstorming about ethical issues and representing the outcomes in a structured template (https://ethicscanvas.org/).

The Data Ethics Canvas is a tool for anyone who collects, shares or uses data. It helps identify and manage ethical issues – at the start of a project that uses data, and throughout (https://theodi.org/article/data-ethics-canvas/).

Other tools include 'The Digital Product Ethics Canvas and Impacts Canvas' (Threebility, 2020), the Humans & Machines Ethics Canvas's (Vaish, 2020) and the Ethical Matrix (Forsberg, 2004).

Currently, there is little information about how industry is using these tools. However, they represent significant progress.

5. Evidence, HF Methodologies & the Production of a 'Human Factors & Ethics' Canvas

The objective of the human factors and ethics canvas is to create an evidence map in relation to the specification of an ethically responsible technology solution, that properly addresses human and ethical issues.

The HFEC allows non-ethicists such as Designers, Human Factors Researchers, Engineers, and Computer Scientists to engage in ethical issues pertaining to the emerging technology product.

The ethics canvas reflects an integration of ethics and HF methods, particularly around the collection of evidence using stakeholder evaluation methods (Cousins et al, 2013; Wenger 1998). Specifically, the ethics canvas makes use of the data gathered and analysed in from literature and stakeholder evaluation methods (i.e. interviews, observations and participatory design/evaluation sessions).

A key dimension of the human factors and ethics canvas is the application of personae-based design (Pruitt & Grudin, 2003) and scenario-based design (Carroll, 1995) approaches. The application of a personae/scenario-based design approach and integration within an ethics canvas allows us to consider the human and ethical dimensions of these technologies.

Key performance indicators (KPIs) relevant to the potential success of this technology once it is introduced and used by the public (including psychosocial dimensions) are specified in the ethics canvas. Importantly, the translation of system objectives into set of objectives spanning key themes (for example, wellbeing, human benefit, social interaction and relationships and societal values and norms) and associated metrics, ensures that wellbeing, human benefit and values are both a reference point and a design outcome. Further, potential failures, potential negative impacts and unknowns are also defined.

Questions pertaining to data ethics and ethics in data analytics must be asked.

Further, ethics happens and is addressed in implementation. As such, we must locate the technology from a 'socio-technical' perspective. Technology is one of many system dimensions to be accounted for. The HFAE Canvas calls for a holistic solution to ethical issues (i.e. technology, task design, environment, process, culture and so forth).

It should be noted that the human factors and ethics canvas 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 (i.e. Consequentialism, Deontology & Principlism). Principles need to be articulated and embedded in the design concept. Human factors methods are useful here in relation to considering the needs/perspectives of different stakeholders and adjudicating between conflicting goals/principles. In this way, the solution needs to carefully balance goals and issues pertaining to human benefit for different stakeholders.

The HFEC does not simply record issues to be addressed. It also records decision decisions and actions. Critically, three of the deep dives record decision decisions pertaining to the relevant theme. This includes:

- Personae & Scenario (Section 4)
- Data Ethics (Section 5)
- Implementation (Section 6)

6. 'Human Factors and Ethics' Canvas: Overview & Stages & Procedure

Introduction

There are different stages in the HFAE Canvas. All stages are important. It is not possible to get to the last stage (i.e. the human factors & ethics summary), without progressing through the other stages (or at least a subset of the other stages).

At the beginning of a project, it is likely that the first three stages may be completed, along with stage 7 (summary information).

Depending on project scope and timing, you may not have scope for some of the 'deep-dives'. Deep dives are grouped thematically

- Benefits, outcomes and impact (Section 3)
- Personae & Scenario (Section 4)
- Data Ethics (Section 5)
- Implementation (Section 6)

In line with stakeholder evaluation approaches, the canvas can be evaluated using the 'community of practice' – i.e. 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.

Stage 0

Stage 0 records relevant project information including who is responsible for co-ordinating the HFEC inputs. Critically, it captures the time/point in the project when the HFEC (i.e. stage in the research and innovation process) was documented. Ideally, there are several iterations of the HFEC as the project progresses. Key sources of research/evidence on which the HFEC inputs are based are also documented, along with research ethics (i.e. methodologies) information.

Stage 1

Stage 1 is all about framing the problem. Values and human and ethical issues are built into how we frame the problem (and set the design brief for new technologies).

- Are you thinking about the problem in the right way and posing the right question?
- To do this, you need to understand the problem correctly - behaviour/symptoms, contributory factors and outcomes/consequences

Stage 2:

Stage 2 involves understanding how the technology fits to the problem, and specifically what we know about stakeholder goals and needs. A key component of this is the specification of expected benefits for different stakeholders.

Stage 3

Stage 3 is a deep dive into benefits, outcomes and impact.

Stage 4.

Stage 4 involves a deep dive in relation to personae and scenario.

Stage 5.

Stage 5 focuses on data ethics. The specific questions posed reflect a simplification of what is covered in the 'Data Ethics Canvas' (ODI, 2019).

Stage 6.

Stage 6 concerns technology implementation. As noted previously, ethics both happens and is addressed in implementation. The questions posed follow from a 'system approach' to human factors. Technology is one of many system dimensions to be accounted for. The HFEC calls for a holistic solution to ethical issues (i.e. technology, task design, environment, process, culture and so forth).

Stage 7.

The final stage presents the outcomes of the preceding analysis.

7. Example HFEC

0: PROJECT INFORMATION & RESEARCH SUMMARY	
Date	
Project Name	
Product Owner	
HF & Ethics Coordinator	
HF & Ethics Canvas Version No.	
Research & Innovation Phase	
Summary of Research Completed & Key Sources of Information/Evidence	
Research Ethics Approval & Date	
1: FORMULATING THE PROBLEM & FRAMING QUESTION	
What is the problem that the proposed technology will address?	
Who is it a problem for? Key stakeholders? Who effect (directly and indirectly?)	
Setting & Environment?	
Causes of the problem?	
Ethical codes that apply in this setting?	
Ethics embedded in the problem definition?	.
Ethics & Impact of Problem. Individual Level. Societal level. Ethics of acting/not acting?	
Summary of ethical issues to be addressed.	
Summary of relevant ethics principles and frameworks.	
Ethics & Key KPI	
2: UNDERSTANDING TECHNOLOGY & FIT TO PROBLEM/STAKEHOLDER NEEDS & EXPECTED BENEFITS	
What is the technology? How does the proposed technology address the problem?	

What part of the problem does it address?			
What is the goal/objective? Intended purpose/function of technology?			
Setting & Environment?			
Direct users of technology? Goals? Needs? Expected Benefits?			
Other stakeholders impacted by technology? Goals? Needs? Expected Benefits?			
Design Decisions & Safeguards			
3: DEEP DIVE: BENEFITS, OUTCOMES & IMPACT			
Overall benefits and outcomes: key stakeholders? Expected positive impacts?			
Expected Impact for key stakeholders (psycho-social themes). Individual level? Societal Level?	Human role in the system	Human Identity	Lived experience, wellbeing, quality of life
	Social Interaction & Relationships	Activity & Behaviour	Attitudes & Values
What could go wrong? Potential failures? Potential negative impacts? Psychosocial? Environmental?			
Unintended consequences.			
Unknowns			
Design Decisions & Safeguards			
4: DEEP DIVE: PERSONAE & SCENARIOS			
Example Scenario			
Example Personae			
How is it expected to work?	Scenario 1:	Scenario 2:	Scenario 3:
What does success look like? Benefits for whom? Expected	Scenario 1:	Scenario 2:	Scenario 3:

positive outcomes and for whom?					
What could go wrong? Potential failures? Potential negative impacts?	Scenario 1:		Scenario 2:		Scenario 3:
Unintended consequences.	Scenario 1:		Scenario 2:		Scenario 3:
Unknowns?	Scenario 1:		Scenario 2:		Scenario 3:
Design Decisions & Safeguards	Scenario 1:		Scenario 2:		Scenario 3:
PART 5: DEEP DIVE – DATA ETHICS					
Ethical issues relevant to data collection? What data? Why collecting? Potential for bias in data collection?					
Ethical issues relevant to data, model & algorithms? Potential for harm and risk?					
Ethical issues relevant to data use & predictions (i.e. application of model/algorithms)?					
Ethical issues relevant to data sharing?					
Design Decisions & Safeguards					
PART 6: DEEP DIVE – IMPLEMENTATION					
Implementation approach					
Implementation enablers					
Implementation barriers					
Systems Perspective: Addressing Ethics as part of Implementation.	People	Process	Technology	Culture	Training & Education
Design Decisions & Safeguards					
PART 7: HUMAN FACTORS & ETHICS SUMMARY					
Key stakeholders? Who is this technology designed for?					
What does success look like? Success for whom?					
Human/Societal Vision & Technology Role/Purpose.					

Summary of Key Ethical Issues to be Addressed?	
Ethical Principles Underlying Technology Design	
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?	
Data Ethics Summary.	
Implementation Summary	
Ethics & Key KPI	

8. Conclusion

The application of ethics to questions concerning technology development is not new.

Design/technology teams exercise choice in relation to what is valued and advancing technology that improves the human condition (and not worsens it). Technologies need to positively contribute to human wellbeing and our lived experience. Intelligent automation/technologies must put the human at the centre.

In relation to new technology design, much of the literature focuses on the what and the how. Psychosocial dimensions are addressed, but not always questions of purpose, values, and benefits. It is important to also look at the why. What behaviour are we seeking to change and why, what is going on at individual/societal level that we have this behaviour/problem, how will the proposed technology address the problem, how will it shape identity and behaviour (who are we, purpose and role of tech, acceptability of tech), what will it change (and is this a good idea), and how to design/implement new tech to create positive impacts and manage potential negative impacts. This goes beyond psychology...and stretches the discipline of human factors to include ethics.

The proposed ethics canvas – (1) integrates with HF methods and collection of evidence and (2) addresses key dimensions of data ethics. Specifically, it addresses the (1) the problem specification (i.e. why require technology, impact of problem and values/ethics underpinning the problem specification), (2) the fit to stakeholder needs, (3) future stakeholder use scenarios, behaviour change and ethical issues, (4) expected benefits and outcomes of the proposed technology, (5) examination of technology impacts (intended and unintended) – at an individual and societal level,

(6) key KPI from ethics perspective, and (7) adopting a systems perspective as part of technology implementation.

Principles need to be both articulated and then embedded in design concepts. Certain core concept must be addressed in a human and ethics canvas. This includes: (1) human role, (2) human benefit, (3) rights, (4) progress and (5) wellbeing. These concepts provide structuring principles to guide the design of new systems. Any proposed solution will emerge from an analysis of these principles in relation to the goals and needs of specific stakeholders (i.e. personae) in different situations (i.e. scenarios).

The HFAE Canvas calls for a holistic solution to ethical issues (i.e. systems approach - technology, task design, environment, process, culture and so forth).

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. 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). Typically, the human factors discipline is concerned with issues around intended use, user interface design and technology acceptability. Arguably, human factors and human machine interaction design research must extend its remit and 'go beyond the user interface'. Specifically, it should address issues pertaining to the psycho-social impact of technology, and how wellbeing, rights and human value/benefit are considered in terms of the design solution.

It is argued that the application of a personae/scenario-based design approach and integration within an ethics canvas allows us to consider the ethical dimension of these technologies. Further, 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 'holds 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'. 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).

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. Critically, thinking about both potential positive and negative consequences enables designers to build both enablers and protections into the design concept.

9. References

Albrechtslund, A. (2007). Ethics and technology design. Ethics and Information Technology, 9(1), 63–72. doi:10.1007/s10676-006-9129-8.

Adapt (2019). https://ethicscanvas.org/

Baxter, Gordon & Sommerville, Ian. (2011). Socio-technical systems: From design methods to systems engineering, Interacting with Computers, Volume 23, Issue 1, January 2011, Pages 4–17, https://doi.org/10.1016/j.intcom.2010.07.003

Beever, J., & Brightman, A. O. (2015). Reflexive Principlism as an Effective Approach for Developing Ethical Reasoning in Engineering. Science and Engineering Ethics. http://doi.org/10.1007/s11948-015-9633-5

Brey, P. (2017). Ethics of Emerging Technologies. In S. O. Hansson (Ed.), Methods for the Ethics of Technology. Rowman and Littlefield International.

Capurro, R. Digital Ethics. (2009). The Academy of Korean Studies (ed.): 2009 Civilization and Peace, Korea: Academy of Korean Studies 2010, pp. 203-214.

Carroll JM (ed) (1995) Scenario-based design: envisioning work and technology in system development. Wiley, New York

Cinoğlu, H.; Arıkan, Y. Self, identity and identity formation: From the perspectives of three major theories. *Int. J. Hum. Sci.* **2012**, *9*, 1114–1131.

Cousins, J.B.; Whitmore, E.; Shulha, L. Arguments for a common set of principles for collaborative inquiry in evaluation. Am. J. Eval. 2013, 34, 7–22.

Dodd J, Stern-Gillet S (1995) The is/ought gap, the fact/value distinction and the naturalistic fallacy. Dialogue 34(4):727–746

European Commission. (2013). Ethics for researchers. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/ethics-forresearchers_en.pdf

Fecher & Kobsda, 2019. Research Impacts Canvas (RIC). Available from: https://elephantinthelab.org/meet-the-research-impact-canvas-a-structured-guide-for-planning-your-science-communication-activities/

Forsberg, E.M. (2004) The Ethical Matrix—A Tool for Ethical Assessments of Biotechnol-ogy, Global Bioethics, 17:1, 167-172, DOI: 10.1080/11287462.2004.10800856

Fry, T (2012). Becoming human by design. Berg Publishers.

Hoven, J. van den. (2008). Moral methodology and information technology. In K. E. Himma & H. T. Tavani (Eds.), The Handbook of Information Systems Research. John Wiley & Sons, Inc. http://doi.org/10.4018/978-1-59140-144-5

Heidegger, M. (1977). The question concerning technology, and other essays. Sixth Edition. New York: Harper & Row.

IEEE. Ethically Aligned Design, A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. First Edition (EAD1e) i. 2018. Available from: https://ethicsinaction.ieee.org/

IEEE. (2019). https://site.ieee.org/sagroups-7000/

John BE (2005) Evidence-based practice in human–computer interaction and evidence maps. ACM SIGSOFT Softw Eng Notes 30:1–5

LaFollette, Hugh (2002). Ethics in Practice (2nd Edition). Blackwell Publishing.

https://www.standardsuniversity.org/e-magazine/march-2017/ethics-and-technology/

Miake-Lye, I.M., Hempel, S., Shanman, R. et al. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. Syst Rev 5, 28 (2016) doi:10.1186/s13643-016-0204-x

Moor, J. H. (1999). Just consequentialism and computing. Ethics and Information Technology, 1(1), 61–65. http://doi.org/10.1023/A:1010078828842

ODI (2019). https://theodi.org/article/data-ethics-canvas/

O Keefe, C & O' Brein, D (2018). Ethical Data and Information Management: Concepts, Tools and Methods.

Fecher, B & Kobsda, C. (2019). https://elephantinthelab.org/meet-the-research-impact-canvas-a-structured-guide-for-planning-your-science-communication-activities/

Osterwalder, Alexander; Pigneur, Yves; Clark, Tim (2010). Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Strategyzer series. Hoboken, NJ: John Wiley & Sons. ISBN 9780470876411. OCLC 648031756.

Oxford English Dictionary. Definition of Ethics. 2019. Available from: https://en.oxforddictionaries.com/definition/ethics

Organization for Economic Co-Operation and Development OECD 2 Nd World Forum – Istanbul 2007, Measuring and Fostering the Progress of Societies. Available from: http://www.oecd.org/site/0,3407,en_21571361_31938349_1_1_1_1_1,00.html

Partnership on Artificial Intelligence to Benefit People and Society to formulate best practices on artificial intelligence technologies. (2016). Available at: https://www.partnershiponai.org/about/

Pawson R (2006) Evidence-based policy: a realist perspective. Sage, Newcastle upon Tyne

Pruitt, J.; Grudin, J. Personas: Practice and Theory. In Proceedings of the 2003 conference on Designing for user experiences (DUX '03). 2003. ACM, New York, NY, USA, 1-15. DOI=http://dx.doi.org/10.1145/997078.997089

Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., Gordijn, B. (2017). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. Science and Engineering Ethics.

Reijers W., Koidl K., Lewis D., Pandit H.J., Gordijn B. (2018) Discussing Ethical Impacts in Research and Innovation: The Ethics Canvas. In: Kreps D., Ess C., Leenen L., Kimppa K. (eds) This Changes Everything – ICT and Climate Change: What Can We Do?. HCC13 2018. IFIP Advances in Information and Communication Technology, vol 537. Springer, Cham

Rickman, H.P. (1979). Wilhelm Dilthey: Pioneer of the Human Studies, Berkeley: University of California Press.

Rockeach, M. (1973). The nature of human values. Free Press. P.5

Sollie, P. (2007). Ethics, Technology Development and Uncertainty: An Outline for any Future Ethics of Technology. Journal of Information, Communications & Ethics in Society, 5(4), 293–306.

Standford Encyclopaedia of Philosophy (2019). Available from: https://plato.stanford.edu/entries/moral-responsibility/

United Nations. Universal Declaration of human rights. 1948. Available from: https://www.un.org/en/universal-declaration-human-rights/index.html.

Ulrich R, Zimring C, Xuemei Z, DuBose J, Seo H, Choi Y, Quan Xiaobo, Joseph Anjali (2008) A review of the research literature on evidence-based healthcare design. HERD 1(3):61–12

Values (2017) In Businessdictionary.com. Available from http ://www. business dictionary.com/definition/values.html

Van Manen, M. (1990) Researching the lived experience Human sciences for an action sensitive pedagogy. The Althouse Press, London, 78-135.

Waterson PE. A critical review of the systems approach within patient safety research. Ergonomics. 2009;52:1185–1195.

Webster L, Steinke C (2009) Evidence-based design: a new direction for health care. Design Quarterly, Winter 2009

Wenger, E. (1998). Communities of Practice: Learning, Meaning, and Identity; Cambridge University Press: Cambridge, UK, 1998.

Winograd, T. & Flores, F. Understanding Computers and Cognition: A New Foundation for Design Norwood (New Jersey). 1986. Ablex Publishing Corporation

10. Further Information

For more information on the ethics canvas, please contact Dr Joan Cahill (cahilljo@tcd.ie).