D6.3: Methods for translating ethical analysis into instruments for the ethical development and deployment of emerging technologies

[WP6 – Generalizing project methods, and exploitation measures]

Contributors	**Chapter 1**: Philip Brey and Brandt Dainow, University of Twente
	Chapter 2: Yasemin J. Erden and Philip Brey, University of Twente; Amal Matar, Uppsala University
	Chapter 3: Brandt Dainow, Philip Brey, Philip Jansen, University of Twente; Amal Matar, Uppsala University
	Chapter 4: Rowena Rodrigues, Nicole Santiago, Anais Resseguier, Trilateral Research
	Chapter 5: Lisa Tambornino, Dirk Lanzerath, Philipp Hoevel, EUREC Office gUG
Reviewers	Alexei Grinbaum, CEA-Saclay/Larsim
	Marcelo de Araujo, Federal University of Rio de Janeiro
Contact	Philip Brey, University of Twente, p.a.e.brey@utwente.nl Lisa Tambornino, EUREC Office gUG, tambornino@eurecnet.eu

Due date	31 March 2021
Delivery date	29 March 2021
Type	Report
Dissemination level	PU

Keywords	general methods, emerging technologies, codes, operational guidelines, guidance documents, ethics by design, policy guidance, research ethics committees

The SIENNA project - *Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact* - has received funding under the European Union's H2020 research and innovation programme under grant agreement No 741716.

Abstract

In this SIENNA deliverable we present five general methods for translating ethical analysis into frameworks and methods for the ethical guidance of new emerging technologies. These are:

- a multistakeholder, coevolutionary strategy for ethically responsible development, deployment and use of new technology,
- a step-by-step method for the development of ethics guidelines and ways in which guidelines can be operationalized,
- a general approach for Ethics by Design, that works for all technology fields,
- suggestions for ethics and human rights projects on new and emerging technologies for engaging with policy-maker, and finally
- a method on how research ethics committees can support ethics in new emerging technology research.

Document history

Version	Date	Description	Reason for change	Distribution
V1.0	10 March 2021	Review version	Development of report	Quality assurance reviewers
V2.0	29 March 2021	Final version for submission	Finalisation	European Commission

Information in this report that may influence other SIENNA tasks

Linked task	Points of relevance
Tasks 2.7, 3.7, 4.7	The generalized methods developed for this report are based on the methods that were used for specific technologies in these tasks.
Tasks 5.1	
Tasks 5.2, 5.3, 5.4	

Table of contents

Executive summary

In this SIENNA deliverable we present general methods for translating ethical analysis into frameworks and methods for the ethical guidance of new emerging technologies. The generalized methods are based on the methods that were used for specific technologies in other tasks of the SIENNA project. We developed the following five methods:

- a multistakeholder, coevolutionary strategy for ethically responsible development, deployment and use of new technology,
- a step-by-step method for the development of ethics guidelines and ways in which guidelines can be operationalized,
- a general approach for Ethics by Design, that works for all technology fields,
- suggestions for ethics and human rights projects on new and emerging technologies for engaging with policy-maker, and finally
- a method on how RECs can support ethics in new emerging technology research.

Section 1: Emerging technology and ethics: a multistakeholder coevolutionary strategy

In section 1 we propose a multistakeholder, coevolutionary strategy for the ethically responsible development, deployment and use of new emerging technologies. The approach we present aims to help technology actors in finding and strengthening their role in responsible and ethical innovation practices. The basis of our strategy is the Ethics and Technology Coevolutionary Model (ETCOM), a normative model for the coevolution of technological innovation and ethics.

Section 2: The development and operationalisation of ethics guidelines

In this section we offer a step-by-step method for the development of ethics guidelines and ways in which guidelines can be operationalised. We differentiate between ethics guidelines and ethics codes, and between general and practice-specific ethics guidelines. We present a proposal for (general) ethics guideline development, which comprises eleven steps towards guideline development, with brief instructions and recommendations for each step. We then proceeded to discuss special considerations in the development of practice-specific guidelines and end with ideas and proposals for the operationalisation of ethics guidelines.

Section 3: Ethics by Design: A General Approach

In this section we present a general approach for Ethics by Design, which works for all technology fields.

At first, we outline the objectives and core assumptions of Ethics by Design. Thereafter, we explain that technological designs, whether of a product, system or process, are not neutral but have consequences or effects and even values “embedded” in them. We describe how this realization led to creation of the *value-sensitive design* approach and a broader family of approaches called *design for values*, which try to include human values in the design process. Finally, we made the case for the *ethics by design* approach, which aims for the systematic inclusion of ethical values, principles, requirements and procedures into design and development processes.

Section 4: Engaging policy-makers in projects on ethical and human-rights aspects of new and emerging technologies: tips for successful engagement

In section 4 we offer tips for ethics and human rights projects on new and emerging technologies for engaging with policy-makers. Alongside drawing from the above research, we examined some key documents to draw out general tips, tips on 'how' to engage and key performance indicators to measure policy impact.

Section 5: General approach on how RECs can support ethics in new emerging technology research

For section 5 we examined what role Research Ethics Committees (RECs) can play when it comes to ethically sound research on and with new emerging technologies. RECs are well established to review health-related research, and they also exist in the humanities, social and behavioural sciences. However, in some research fields it is difficult for researchers to find a REC that fits well to review their research projects. Therefore, either new RECs need to be established or the old ones need to be extended. We make suggestions for the composition and guidance of RECs.

List of figures

List of tables

List of acronyms/abbreviations

Abbreviation	Explanation
AI	Artificial Intelligence
AI&R	Artificial Intelligence and Robotics
AI HLEG	High-Level Expert Group on Artificial Intelligence
APA	American Psychology Association
ATE	Anticipatory Technology Ethics
CAD	Computer Aided Design
CoE	Code of Ethics
CPA	Canadian Psychological Association
CRISP-DM	CRoss Industry Standard Process for Data Mining
CSR	Corporate Social Responsibility
EbD	Ethics by Design
EIA	Ethical Impact Assessment
ETCOM	Ethics and Technology Coevolutionary Model

HG	Human Genomics
HE	Human Enhancement
HET	Human Enhancement Technology
IEEE	Institute of Electrical and Electronics Engineers
NASMM	National Association of Senior Move Manager
R&D	Research & Development
R&I	Research & Innovation
REC	Research Ethics Committee
SIS	Smart Information Systems
TLC	Technology Life Cycle
TRL	Technology Readiness Levels
VSD	Value-sensitive Design
WP	Work Package

Table 1: List of acronyms/abbreviations

Glossary of terms

Term	**Explanation**
Artefact	Within this report used as general term for any individual technological device.
Autonomy	The value of a person's ability to decide and act on her own authentic desires and preferences, without being unduly influenced, coerced or manipulated by others.
Emerging technologies	Emerging technologies are innovative, new, and still in development. They are innovative in the sense that they promise new and potentially superior solutions to problems. They are new in the sense that they employ new concepts, methods and techniques and cannot be subsumed under existing technologies. They are still in development in that they are still, to some extent, a promise: few, if any, products and applications have resulted from them, and few, if any, are marketed and used on a large scale.
Ethical analysis	Ethical analysis is the process by which ethical issues associated with a situation, action, process or thing are studied in a systematic manner.

Table 2: Glossary of terms

Introduction

In this report we present general methods for translating ethical analysis into frameworks and methods for the ethical guidance of emerging technologies. The generalized methods are based on the methods that were used for specific technologies in other SIENNA tasks, most important are the results from tasks 2.7, 3.7, 4.7 and 5.1 – 5.4.

In the first section, we aim to develop a multistakeholder, coevolutionary strategy for ethically responsible development, deployment and use of new technology. We will propose an ideal (or normative) model for the coevolution of technological innovation and ethics which we call the 'Ethics and Technology Coevolutionary Model' (ETCOM). We model the manner in which ethical assessments and practices would accompany the evolution of an emerging technology at different stages of its emergence so as to best support the ethical development and deployment of the new technology. In section 2, we present a step-by-step method for the development of ethics guidelines and ways in which such guidelines can be operationalised and implemented. In section 3, we present a general approach for Ethics by Design, which works for all technology fields. This document generalizes from the Ethics by Design approach we developed for Artificial Intelligence and robotics.[1] We have considered various other technologies, and studied the design processes in them, and have on this basis generalized the approach to apply to, in principle, any technology. In section 4, we present suggestions for ethics and human rights projects on new and emerging technologies for engaging with policy-makers. And finally, in section 5 we present a general approach on how Research Ethics Committees (RECs) can support ethics in new emerging technology research.

Although this report has been written as a balanced whole, the reader may also skip entire chapters and/or read individual chapters independently of the others.

1. Emerging technology and ethics: a multistakeholder coevolutionary strategy

1.1 Introduction

In this chapter, we aim to develop a multistakeholder, coevolutionary strategy for ethically responsible development, deployment and use of new technology. Our strategy is multistakeholder in that it emphasizes a multistakeholder view of the governance of emerging technologies; successful governance of emerging technologies, including successful consideration and mitigation of ethical issues, requires the cooperation and shared responsibility of multiple actors across society. Our strategy is coevolutionary in that it models technological innovation and ethical practice as processes

---

[1] Brey, Philip and Brandt Dainow, "Ethics by Design and Ethics of Use in AI and Robotics", Annex 2 of Resseguier, Anaïs, Philip Brey, Brandt Dainow, Anna Drozdzewska, Nicole Santiago, David Wright, SIENNA D5.4, Multi-stakeholder Strategy and Practical Tools for Ethical AI and Robotics, 2021. https://www.sienna-project.eu/publications/deliverable-reports/. This is an improved version of Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the development and use of SIS, 2019, Deliverable D3.2 of the SHERPA project. https://doi.org/10.21253/DMU.11316833.

that should coevolve in a symbiotic and interactive manner. It should be added that no procedure for inclusion of ethical considerations guarantees ethical outcomes, but it is hoped that our strategy makes such outcomes considerable more likely.

We will propose an ideal (or normative) model for the coevolution of technological innovation and ethics that we call the 'Ethics and Technology Coevolutionary Model' (ETCOM). We model the manner in which ethical assessments and practices would accompany the evolution of an emerging technology at different stages of its emergence so as to best support the ethical development and deployment of the new technology. This is a multi-actor model in that different actors have differing responsibilities for making this coevolution of ethics and technology come about. Our initial focus is not on individual actors, but on the coevolutionary process at a macro level. We want to understand the general steps which must be taken to include ethical considerations in the technology development process itself.

After laying out the ETCOM model, we go down one level of description and focus on the technology actors who are involved in the development, deployment and use of new technologies, the methods and instruments these actors have (or might have) available to address ethical issues and engage in ethical practices, and finally the roles and responsibilities that they should adopt. Thus, we define the normative model not at the macro-level, but at the meso- and microlevel of organisations and individual actors. We describe the key technology actors typically present; we propose methods, tools and instruments which they need to confront ethical issues and promote ethical practices, and finally we describe the roles and responsibilities of the technology actors at different points in time, in accordance with the different coevolutionary stages of the ETCOM model. Our approach is intended to stimulate new ways of thinking about responsible research and innovation with particular attention to the role of ethics. We are confident that a normative, multistakeholder, coevolutionary approach, such as the one we present here, could greatly assist technology actors in finding and strengthening their role regarding responsible and ethical innovation practices.

The chapter is structured as follows. First, we start building the coevolutionary model by describing evolutionary life cycle of technology, defining what an emerging technology is within this life cycle and distinguishing different stages in the evolution of an emerging technology. We then use this evolutionary model to evaluate at which point interventions are needed by technology actors in order to address ethical issues and to develop ethical practices. This gives us the ETCOM model. We then move to lower levels of analysis and focus on the technology actors, the methods we recommend they should use, and their roles and responsibilities in the context of the ETCOM model. This gives us a better understanding of the actions which actors need to undertake at different points in time in order to ensure responsible and ethical practice in technology development, deployment and use – and the means and methods these actors require so as to bring them about.

1.2 Emerging technology and its stages

In this section, we will define what an emerging technology is and will distinguish different stages in its evolutionary path. Let us start by defining the term 'technology.' There are many ways of looking at technologies. Technology is a complex topic and philosophy has debated the nature of technologies for centuries without coming to agreement on a single definition. This lack of agreed definition is actually useful because it allows us to define technology according to the needs of the moment. If we are concerned with issues related to ownership and use rights, defining a technology in terms of *artefacts*—that is, as individual technological devices[2]—is more useful than thinking about the scientific theories which made them possible. On the other hand, if we are concerned with issues related to, for instance, education or economic advancement, we need to understand the relationship between scientific research and innovation, and so defining the technology as a set of ideas and procedures that become embodied in physical objects is more useful. One popular perspective is exemplified by Charles Schwab, who discussed the changes created by the new technologies of the current era under the term 'the fourth industrial revolution.' He defined modern technologies by the material composition of their artefacts – organising them into the physical, biological and digital.[3] Conversely, technology is often defined as a socio-technical system.[4] This definition allows for consideration of factors outside the artefact, such as the nature of the organisations creating or using a technology, or political, economic and social factors, all of which can significantly affect a technology's development and impact. This range of approaches allows one to define the technology in a manner that best suits the concerns of the project and how the technology will be analysed. SIENNA used the model provided by Anticipatory Technology Ethics (ATE)[5] to deal with this range of possible definitions of technology in an inclusive manner.

Under the ATE model, different aspects of a technology emerge at different phases of its development. The earlier a characteristic emerges, the more pervasive it will be across all instances of that technology. Should that characteristic have ethical issues, then those issues are likely to be more difficult to resolve because they are more fundamental to the technology itself. On the other hand, a particular product or artefact may implement a technology in ethically problematic ways while other artefacts deliver the same functionality[6] without generating ethical concerns. Addressing an ethical concern in such a situation is less complex because the issue is less universal and simply requires design changes in that particular artefact. On the other hand, unproblematic artefacts may be used in unethical ways. Here the resolution requires changes in practices, such as through legislation or cultural change, leaving the artefacts unchanged.

It is therefore important to organise the technology so as to distinguish between fundamental and universal characteristics, individual artefacts and patterns of use. ATE arranges the technology and its artefacts into three layers. The *technology level* is the most wide-ranging level of description, covering the technology in general, its subfields, and its basic techniques and approaches. The *product* level defines the artefacts and processes that are being developed for practical use. The *application* level

[2] We use 'artefact' as the general term for any individual technological device. This does not imply anything about its material composition. All technologies work by changing something in the world. 'Artefact' is simply the term for the whatever it is which makes that change. In commercial domains, an artefact usually takes the form of a product (e.g., machinery, chemicals, software applications and genetically engineered bacteria), but at a more theoretical level, even institutions and laws can be conceived of as 'artefacts'. See for instance Burazin, L., Himma, K. E. and Roversi, C. (eds) (2018) *Law as an Artifact*. Oxford: Oxford University Press.

[3] Schwab, Klaus, The Fourth Industrial Revolution, World Economic Forum, Geneva, Switzerland, 2016.

[4] Dainow, Brandt, "Threats to Autonomy from Emerging ICTs", Australasian Journal of Information Systems, Vol. 21, No. 0, 2017; Stahl, Bernd Carsten, Richard Heersmink, Philippe Goujon, Catherine Flick, Jeroen van den Hoven, Kutoma Wakunuma, Veikko Ikonen, and Michael Rader, "Identifying the Ethics of Emerging Information and Communication Technologies: An Essay on Issues, Concepts and Method", International Journal of Technoethics (IJT), Vol. 1, No. 4, 2010, pp. 20–38.

[5] Brey, Philip, "Anticipatory Ethics for Emerging Technologies", *Nanoethics* 6, 1–13 (2012).

[6] Here, 'functionality' refers to what the artefact does. When technologists design a new artefact, they are trying to find ways to create something capable of delivering that desired functionality. Functionality can also be unexpected. For example, email was originally invented to enable computer network administrators to send technical messages to each other while trying to get two networks to connect. However, when people began using email to send personal messages to each other, they added functionality which was never intended by the inventors.

discusses how the technology is applied, defining particular uses of the products, in particular contexts and domains, by particular users. The benefit of stratifying technology into these levels is that it explicitly identifies both the social and technical dimensions of a technology and how they interact.

The technology level summarises the state of science and engineering regarding the technology and identifies those characteristics that are universal across the technology, including key terms and concepts, important subfields, techniques and methods. At this level, one is concerned with "collections of techniques that are related to each other because of a common purpose, domain, or formal or functional features."[7] A technology can be defined by its purpose or its capabilities[8] at this level. For example, automotive technology could be defined as the collection of technologies for the purposes of moving goods and people with vehicles that contain their own propulsion power and fuel supply. Conversely, space technology could be defined as moving people and goods into outer space, performing operations while there, and returning them to earth. Unfortunately, it is not always obvious how to define a technology, thus it may require considerable research.

The product level identifies the products or artefacts a technology generates. Every technology, regardless of how it is defined, achieves its effects through artefacts. The creation of these artefacts requires unique processes designed for the task. Identification of artefacts at the product level is usually fairly straightforward, however it is important to note that the product level also includes *processes*. Sometimes, relevant processes are those used to create the artefacts, but processes can also be artefacts. For example, automotive technology underwent profound changes during the 1970s with the introduction of robotic production processes to the assembly line.[9]

The application level describes the ways in which the artefacts of the product level are actually being used. Artefacts usually make some activities within a technology easier while making others more difficult. Often this is the result of design choices rather than unavoidable necessity.[10] Most technologies offer a range of functions and different artefacts will focus on different subsets of functionality. For example, smaller automotive vehicles, such as motorbikes, are more manoeuvrable and easier to park, but can carry less goods. Furthermore, some artefacts (and processes) are easier to 'subvert'. Designers of technology must anticipate how their artefacts will be used in order to determine what functionality is required for people to be able to use them. However, their vision may not match what users see, so that people extend the functionality by adding their own ways of using the artefacts.[11]

The Technology Life Cycle

Having defined important aspects of the term 'technology', let us now turn to the Technology Life Cycle. Technologies are understood as being created, developed, and then used, with patterns of usage going through a series of stages from initial adoption to maturity. Technologies, then, remain in use

[7] Brey, op. cit., p. 7.
[8] When considering the future evolution of a technology, it is often important to consider not just what it can do now (functionality), but what it might be able to do in the future or how that functionality might support new usages (capability). For example, the functionality of a car is concerned with moving people and goods. However, because a car is a large heavy metal object which can move at speed, it is also capable of being used as a murder weapon by being driven into a crowd.
[9] Pardi, Tommaso, "Fourth Industrial Revolution Concepts in the Automotive Sector: Performativity, Work and Employment", Journal of Industrial and Business Economics, Vol. 46, No. 3, September 2019. pp. 379–389.
[10] Hutchby, Ian, "Technologies, Texts and Affordances", Sociology, Vol. 35, No. 2, 2001. Pp.441-456.
[11] Norman, Donald A., The Design of Everyday Things, Basic Books, New York, 2002.

until they are superseded by others, though not all technologies are eventually superseded.[12] The rate of technological innovation underwent a rapid increase with the rise of the industrial revolution and has been increasing since then, especially since invention of enabling technologies such as electricity and then computing.[13,14]

There are various ways of describing the TLC, but the most common is the S-Curve model.[15] This is a component of a cyclic model of technological evolution. The cyclic model states that a technology begins with a discontinuity in the form of a breakthrough innovation affecting either processes or artifacts. This is followed by a period of instability during which competition between variations of the original breakthrough eventually lead to the selection of a single dominant configuration, which then becomes the industry standard. Following the emergence of the dominant design, the technology evolves more slowly through incremental innovations until a new breakthrough triggers a new discontinuity and the technology becomes superseded.[16] The S-Curve models this process within a single technology cycle. There are a variety of metrics by which to assess a technology's progress, such as diffusion through society (e.g., the percentage of households owning an artefact), or performance characteristics, such as fuel efficiency. This allows the S-Curve model to chart the best metrics for the aspect of interest.[17] For example, economic assessment of technological maturity could focus on diffusion, while an engineering assessment could focus on internal aspects such as power consumption.

To count as an emerging technology, a technology must be sufficiently developed, in such a way that we can understand its general features, even if it has not yet achieved maturity in all relevant aspects. An emerging technology has little market penetration and may still be in prototype. As a result, its final place and role in the market is yet to become evident. Its position on the S-Curve will be prior to, or just entering, the phase of competition and rapid innovation (see figure 1). Because it has not completed the phase of competitive rapid innovation, its final set of features are unlikely to be completely determined. Similarly, the business models under which it will function will be are likely to be nascent and may also undergo substantial change. Due to its lack of market penetration users will

[12] Anderson, Philip, and Michael L Tushman, "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change", Administrative Science Quarterly, 1990, pp. 604–633.

[13] Schwab, Klaus, The Fourth Industrial Revolution, World Economic Forum, Geneva, Switzerland, 2016.

[14] The Technology Life Cycle (TLC) is sometimes confused with the industry life cycle and the product life cycle because the concepts are somewhat interrelated. The product life cycle is concerned with the introduction, growth, maturity and decline of a product in the market and is usually assessed in terms of sales volume or revenue. Sometimes other metrics can be used so as to assess diffusion of products through society. The industry life cycle tracks the growth and development of the manufacturing processes by which the products are produced and distributed. Both industry and product life cycles represent some aspect of the technology but fail to consider the knowledge-base which underpins it.

[15] Taylor, Margaret, and Andrew Taylor, "The Technology Life Cycle: Conceptualization and Managerial Implications", International Journal of Production Economics, Vol. 140, No. 1, 2012, pp. 541–553; Ryu, Jiyeon, and Soon Cheon Byeon, "Technology Level Evaluation Methodology Based on the Technology Growth Curve", Technological Forecasting and Social Change, Vol. 78, No. 6, 2011, pp. 1049–1059.

[16] Anderson, Philip, and Michael L Tushman, "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change", Administrative Science Quarterly, 1990, pp. 604–633.

[17] Taylor, Margaret, and Andrew Taylor, "The Technology Life Cycle: Conceptualization and Managerial Implications", International Journal of Production Economics, Vol. 140, No. 1, 2012, pp. 541–553; Ryu, Jiyeon, and Soon Cheon Byeon, "Technology Level Evaluation Methodology Based on the Technology Growth Curve", Technological Forecasting and Social Change, Vol. 78, No. 6, 2011, pp. 1049–1059.

still be working out patterns of usage and the regulatory framework will most probably be undeveloped.[18]

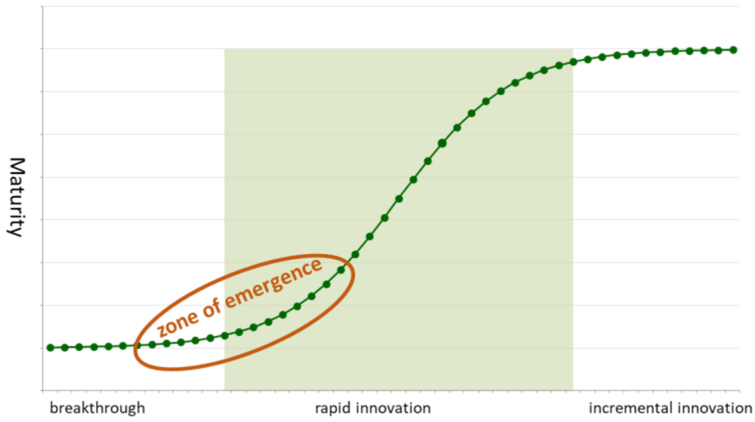

Figure 1: *The Technology Life Cycle S-Curve, showing where emergent technologies are situated*

Another popular life cycle model, which we will use for the ETCOM model, emphasizes the full lifecycle of the technology, not just the innovation path. It distinguishes four stages: R&D, ascent, maturity and decline.[19] While still in the R&D stage, technology is pre-market. During the ascent stage, market parties start observing that the technology yields products that have added value, and uptake begins, as innovation continues. During the maturity stage, the technology attains optimal competitive viability, and during the decline stage, it becomes increasingly obsolete. A technology can be called 'emerging' during the later R&D and ascent stages.

Technology Readiness Levels (TRL)

Technology Life Cycle models position an emerging technology on a trajectory between pure R&D and widespread use. Different technologies move along this path at different paces. It is therefore necessary to assess what stage in its evolution a technology has reached in order to determine whether it is now emerging. However, Life Cycle models are too gross a system for such detailed analysis. The Technology Readiness Levels (TRLs), which are presented below (table 3), have become a common method for determining where exactly a technology sits on the path between pure research and becoming available for general use. TRLs, thus, examine the S-Curve's breakthrough phase (or alternatively, the R&D phase) in more detail. Originally pioneered by NASA to determine when a technology was safe for use in space flight, they have become adapted for general assessment of any technology.

[18] Dainow, Brandt, "Threats to Autonomy from Emerging ICTs", Australasian Journal of Information Systems, Vol. 21, No. 0, 2017.
[19] Haupt, R., M. Kloyer and M. Lange, "Patent indicators for the technology life cycle development," Research Policy 36(3):387-398, 2007.

TECHNOLOGY READINESS LEVEL	DESCRIPTION
TRL 1 – basic principles observed and reported.	Scientific research begins to be translated into applied research and development, such as peer-reviewed papers regarding a technology's properties.
TRL 2 – technology concept and/or application formulated.	The first practical applications of the principles are invented. Applications are speculative, and there may be no proof or detailed analysis to support the assumptions.
TRL 3 – experimental proof of concept.	Active research and development commences. This includes laboratory studies to physically validate the predictions made during the earlier stages, such as the development of components that are not yet integrated or representative.
TRL 4 – aspects of the technology validated in the laboratory.	Basic technological components are integrated to establish that they will work together. At this stage, the technology is unlikely to resemble the final mature stage. Examples include breadboards and other ad hoc integration of components in a laboratory.
TRL 5 – technology validated in simulated environment.	Fidelity of breadboard or laboratory technology increases significantly. The basic technological components are integrated with realistic supporting elements so they can be tested in a simulated environment.
TRL 6 – early prototype technology demonstrated in laboratory environment.	Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment, such as a simulated operational environment.
TRL 7 – mature prototype demonstrated in operational environment.	Prototype is near, or at, that of the planned operational system. This represents a major step up from TRL 6 by requiring demonstration of an actual working prototype in the actual operational environment.
TRL 8 – system complete and qualified.	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of technology development. Examples include final evaluation of the system in its intended environment to determine if specific artefacts meet their design specifications
TRL 9 – actual system deployed and proven in operational environment.	Actual deployment of the technology in its final form and under real-world conditions.

Table 3: *Technology Readiness Levels*[20]

[20] Ibid., 12.

TRLs are now the *de facto* standard for technology assessment in many industries, from power systems to consumer electronics.[21] For example, since 2014 the EU has used TRLs to determine suitability for different types of technology research funding under the Horizon Europe program. Different funding programs are only available for technology at specific TRLs, according to the degree of pure research versus active deployment the funding program is intended for. TRLs have become a widely accepted standard of assessment of technology at its early stages and the TRL scheme is now a government-approved method for grading the development of a technology in the EU[22] and USA.[23]

It should be noted that there is no single TRL that indicates a technology is emerging. As indicated above, to count as 'emerging', a technology must be sufficiently developed, in such a way that we can understand its general features. This means it must be beyond TRL 3. If work is still being done to create stable and successful prototypes it may be said to be in "early emergence" and will occupy TRLs 4 or 5. TRLs 6 – 8 track the development of the prototypes, and are usually the levels at which manufacturing processes are developed. TRL 9 is achieved once actual products start to be deployed. Thus, technology emergence is not a single event, but a process which spans TRLs 4 – 9.

Additionally, it should be noted that a technology's trajectory through the TRLs involves some degree of uncertainty. The earlier the TRL, the less we can predict a technology's final form. The chance of a technology changing becomes less as it progresses from theory to final product. It is important to bear in mind that technologies do not always move at a regular pace or even in a consistent direction. In particular, many technologies hit a block in their development moving from TRL 6 (testing in a laboratory) to TRL 7 (testing in the real world). For example, self-driving cars moved at a regular and fairly rapid pace from TRL 1 to TRL 6. They demonstrated high levels of performance on test tracks, but encountered significant problems once they were tested amongst real traffic, primarily because human drivers cannot be relied upon to always follow the rules of the road. Real world testing also revealed that self-driving cars are extremely poor at detecting bicycles, to such a degree real-world testing was stopped prematurely in the Netherlands due to the presence of large numbers of bicycles on the roads. As a result, self-driving car technology has retreated from TRL 7 to TRL 4.[24] However, this does not constitute evidence self-driving cars will never reach TRL 9, merely that getting the technology to work in the real world is more complex than anticipated and so maturity will take longer to occur.

1.3 ETCOM: A normative model for the coevolution of emerging technology and ethics

Having described the typical emergence process of a new technology through the models of the Technology Life Cycle (TLC) and the Technology Readiness Levels (TRLs), let us now consider which steps should be taken at different phases in the development of emerging technologies in order to include ethical considerations. Under the ETCOM model for the coevolution of ethics and emerging technology there are five stages in the evolution of ethical practice which map onto stages in the

[21] Olechowski, Alison, Steven D Eppinger, and Nitin Joglekar, "Technology Readiness Levels at 40: A Study of State-of-the-Art Use, Challenges, and Opportunities", 2015 Portland International Conference on Management of Engineering and Technology (PICMET), IEEE, 2015, pp. 2084–2094.

[22] European Commission, Horizon 2020 – Work Programme 2016-2017 General Annexes, Publications Office of the European Union, 2016.

[23] US Government Accountability Office, "Technology Readiness Assessment Guide", 2020.

[24] Stilgoe, Jack, "Self-Driving Cars Will Take a While to Get Right", Nature Machine Intelligence, Vol. 1, No. 5, May 2019, pp. 202–203.

evolution of emerging technologies. These five stages move from early awareness of ethical issues to full institutionalization of ethical considerations within the development, deployment and use of the new technology.

The stages are as follows (figure 2):

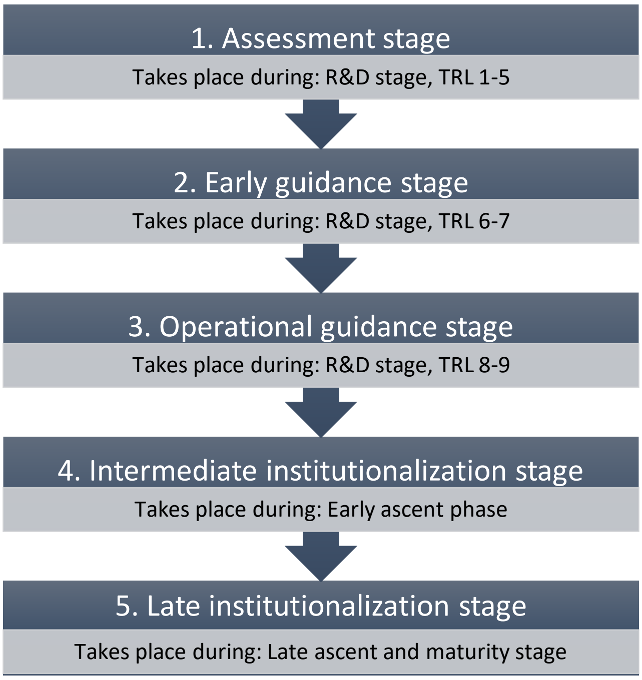

Figure 2: *Structure of the ETCOM model: Five stages in the evolution of ethical practice mapped onto the different stages in the evolution of emerging technologies.*

This is a model that applies to emerging technologies that are expected to have major economic and social impacts and raise significant ethical issues. Full institutionalization of ethical practices can only be achieved at a cost – research into ethical aspects must be carried out, guidelines developed, organizational units and committees formed, and so forth. This cost is warranted if it is determined that major ethical issues will arise with the technology if left unchecked, and if not addressing them results in a greater cost to society than pre-emptively addressing them. On the other hand, some emerging technologies may not raise serious ethical issues and therefore not warrant full institutionalization of ethics. Here minor ethical issues may be sufficiently addressed at the early or operational guidance stage with minimal institutionalization.

In practice, we find that for some such technologies, institutionalization of ethics is never achieved, and even ethical guidance may never be forthcoming. Even when ethical guidance is attained, it will often be later than the ideal model prescribes. For example, the guidance stages may only be reached in the early ascent stage of technological innovation, while intermediate ethics institutionalization may only be reached somewhere during the maturity stage of the technology.

In addition, the proposed mapping of ethics stages to stages of technological emergence is indicative only, as the ideal mapping depends on a number of factors. These include the types of ethical issues that are involved, and whether they already play out during the R&D phase or later, and to what extent they can be mitigated at different stages at emergence.

Let us now discuss the five stages of ETCOM in detail.

Assessment stage

During this stage, ethical issues relating to the new technology are identified and initial assessment is performed. Early identification of ethical issues can be performed by scientists, ethicists, journalists and other actors. Assessment of these issues is usually carried out by trained ethicists, often in collaboration with other stakeholders [see also SIENNA D6.1[25]]. Identification and assessment of ethical issues serves to create awareness of them amongst technology developers, policy makers and the general public. Ethical assessment can take place at different stages of emergence. At early stages of R&D, ethical assessment has its limitations. At these stages it is only possible to address general ethical issues which are not dependent on the precise products or applications expected to emerge but yet undetermined. Ethical issues are therefore speculative due to this uncertainty regarding future products, applications and impacts. Nevertheless, early assessment can serve to alert actors to ethical issues and can serve to initiate mitigation processes. Ethics assessment should not occur only in the assessment stage, but should also proceed beyond it – all the way to the maturity stage. As the technology evolves, new ethical issues may emerge. In addition, more information may become available relevant to the assessment of ethical issues already identified.

Identification and assessment of ethical issues is a first step towards ethical guidance of R&D. At what point in time such ethical guidance is needed in the R&D process depends on two factors. Firstly, whether the ethical issues are present already within the R&D process, or whether they are issues associated with final products in the future. Within human embryonic stem cell research, for example, significant ethical issues apply to the research process itself, since the destruction of human embryos that occurs in such research is controversial. In contrast, for R&D in biometric technologies, the main ethical issues play out during deployment and use of these technologies, after R&D has been completed, since it is then that privacy violations and unfair treatment may result. Secondly, for ethical issues that play out during deployment and use rather than during R&D, at what stage during the R&D can they be addressed and mitigated? For many ethical issues, mitigation is not possible at the stage of basic research (TRL 1) or applied research (TRL 2), but will have to take place at later stages, for example, at the stage of product design. Thirdly, are the ethical issues of a general nature (e.g., risks of health or environmental damage) or are they more specific – relating to specific envisioned products or applications? More general and fundamental ethical issues normally need to be addressed at an earlier stage in R&D than more specific, localized ones.

These three factors determine when ethical guidance will be needed for R&D. In many cases, ethical guidance will not be needed for the stages of basic and applied research, because there are no major ethical issues that play out during research or they can be mitigated in research. Ethical guidance may only be needed at later TRL stages, when R&D moves closer to product design. However, this depends on the precise circumstances and some ethical guidance may already be required earlier.

Early guidance stage

As it becomes progressively clear that an emerging technology will soon be introduced into society and will raise major ethical issues, ethical guidance should be put in place to guide R&D, and then deployment and use. Ethical evaluation at the assessment stage determines what the ethical problems are. This creates an awareness of them but does not yet provide directions for action. What is needed, therefore, are guidelines for action for different actors. In the early guidance stage, it is particularly

[25] Brey et al. *D6.1 Generalised Methodology for Ethical Assessment of Emerging Technologies*, SIENNA Project, 2021. Forthcoming at https://www.sienna-project.eu/publications/deliverable-reports/.

important to have ethics guidelines for R&D actors and also to start developing more general, multi-actor ethics guidelines. General, multi-actor ethics guidelines are guidelines that either apply to society as a whole, for technologies that have broad usage and impacts in different societal sectors (for example, artificial intelligence), or to a particular societal sector or application domain to which it is limited, for example, the medical domain (for medical technologies) or the transportation domain (for transportation technologies).

As the social and ethical implications of an emerging technology become clearer, it is increasingly important to have both specific R&D ethics guidelines which anticipate and help mitigate negative implications, and for general guidelines to be developed which imply joint responsibilities for the different actors that affect the development, deployment and use of the technology (the technology actors). In the early guidance stage, therefore, research ethics guidelines will be developed and implemented (covering both research and development), possibly also professional ethics codes for R&D actors will be developed or amended, and general ethics guidelines will be developed and published.

There is no required order in which these processes should take place. Research ethics guidelines will be the most pressing, since R&D activity is already underway. However, during later stages of the R&D process, other actors (including prospective deployers, vendors, users and regulators) will already be gearing up to assume their future roles. It is therefore possible to develop general ethics guidelines first, and research ethics guidelines second. These can then be based on the general ethics guidelines. Arguably, this is what is now happening with artificial intelligence – in recent years, several international organisations have produced general ethics guidelines for AI, while, conversely, a tradition of research ethics for AI has not yet been established.

During the early guidance stage, the first steps towards institutionalization of ethics for the emerging technology are taken. Guidelines are a type of institutionalization. For them to be developed and implemented, one needs guideline working groups, research ethics committees and other organizational and institutional structures.

Operational guidance stage

The focus during the operational guidance stage is on operationalization and implementation of the ethics guidelines and their application towards different products, application domains and technology actors. Technology actors also start negotiating their roles and responsibilities with respect to ethical issues during this stage.

Operationalization involves making ethics guidelines usable in everyday practices. Ethics guidelines are often quite general and abstract therefore requiring further interpretation for them to be applicable by different actors. Implementation involves the application of ethics guidelines to specific practices. Implementation requires operationalization because it facilitates implementation. Part of the operationalization process may include further specification of the guidelines towards different products, application domains and technology actors. As particular products come into view in an emerging technology, and different application domains shape up, it may be necessary to have ethics guidelines and accompanying ethical assessments which apply to these products and domains. For example, in addition to having general ethics guidelines for artificial intelligence, it may be necessary to add ethics guidelines for specific AI-based products, such as driverless cars and natural language processing systems. It may also be useful to have ethics guidelines for particular application domains of AI, such as healthcare or defence. In addition, it may be useful to have ethics guidelines for particular practices, such research, deployment and use, and ethics codes for different professional roles.

The implementation of ethics guidelines and considerations in specific practices can also be achieved through specialized implementation instruments. In technology development, for example, Ethics by Design is an approach for incorporating ethical requirements in design and development processes. It implements ethics guidelines and considerations in existing design methodologies. In the SHERPA project, we have developed a similar instrument for the implementation of ethics guidelines in the deployment and use of AI systems.[26]

During this stage, it is also important for technology actors to discuss their responsibilities for mitigation of ethical issues, as well as the collaborative relations they need to establish in order to take joint responsibility. For example, technology actors need to discuss how to what degree developers are responsible for the mitigation of particular ethical issues versus how much rests with deployers and end-users. They will also need to consider if mitigation is best achieved through self-policing by research institutes and industry, or whether regulation is needed. In addition, they need to consider how they will collaborate. For example, will they undertake joint policy development with respect to ethical issues, or jointly recognize an independent body for such policy development? Will they inform each other of their mitigation strategies and actions, and will deliberate and act jointly with respect to some issues? These deliberations on the distribution of responsibility will result in definitions of roles and responsibilities per technology actor, in addition to new collaborative practices and institutional support structures. Shared ethics guidelines are an important instrument for facilitating these collaborations and assignments of responsibility because they help ensure agreement on end-goals.

Intermediate institutionalization stage

During the early and the operational guidance stages, the first steps towards institutionalization of ethical practices for an emerging technology domain are taken. In the intermediate institutionalization stage, which we place at the early ascent stage of the emerging technology, further institutionalization occurs. This stage takes place at a point where products and applications of the technology have entered the market and are starting to effect society. At this stage, regulatory issues are being considered, attention goes towards the development of technical standards and various types of organisations and institutional structures are being created to support diffusion of products and applications in society. This is also a stage at which significant institutionalization of ethical practices may take place.

During this stage, major regulation deemed necessary to mitigate ethical issues is agreed upon, and ethics-related practices and institutional arrangements become the norm for relevant technology actors. For example, it may become a norm for universities and industry to implement certain ethics guidelines. At this time new units and institutions are also formed to support ethical practices in deployment and use as well as R&D. Technology actors may start education and training programmes to support future ethical practices, standards organisations may develop technical standards that incorporate ethical guidelines and considerations, and actors may move into the market for certification and audits for ethics compliance.

Late institutionalization stage

Finally, during the late institutionalization stage institutional support structures for ethical practice become fully formed. Ethics becomes an integrated part of normal practice. This normalization

[26] Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the development and use of SIS, 2019, Deliverable D3.2 of the SHERPA project.

happens in three ways. Firstly, mitigation actions at the earlier stages will have resulted in technology development, deployment and use practices which avoid the manifestation of harmful ethical issues to a significant extent. Secondly, ethical issues which cannot entirely be avoided are addressed by integrating mitigation actions in normal practices as much as possible. For example, many ethical guidelines will be incorporated in regular design methodologies so that values like privacy, fairness and responsibility function as design constraints, similar to issues like safety, cost and effectiveness. Thirdly, for those ethical issues which require more reflection, ethics assessment procedures will be in place to ensure that such reflection takes place. These may include research ethics assessment procedures, certification procedures and audits (either specialized versions for ethics assessment or general procedures in which ethical consideration are part of the certification or audit process), periodic awareness raising events, and others.

Moreover, at this stage the acquisition of ethics competencies by technology actors of the emerging technology will be fully integrated into relevant higher education curricula, as well as in education and training programmes for industry and other types of organisations. Students and employees will acquire relevant competencies through regular training programmes in which these competencies will be both integrated in regular courses and considered separately in dedicated ethics courses.

At the late institutionalization stage an optimum is reached in mitigation of ethical issues, both in development, deployment and use. All technology actors play their part, both wittingly and unwittingly, in bringing about desired ethical outcomes. However, because both technology and society keep evolving, this stage is never fully completed. Continuous calibrations will be needed during the maturity stage of the technology to adapt to this evolution.

1.4 Responsibilities and instruments for technology actors

The ETCOM coevolution model for ethics of emerging technologies defines at a macro-level a strategy for ethical practice for emerging technology fields. It describes the overall steps which need to be taken so as to establish an outcome in which new technology is developed, deployed and used in an ethically responsible way. In this section we will develop this strategy further by going one level below the macro-level and focusing on technology actors, their roles and responsibilities, and methods, instruments and competencies that they need to carry out these responsibilities. This section generalizes the multistakeholder strategy that we developed earlier for AI and robotics[27], and makes adaptations to ensure a good fit with other emerging technologies.

A detailed strategy for the ethical development, deployment and use of emerging technology should:

- Identify all the *technology actors* (stakeholders that partake in or influence development, deployment and use).
- Identify *methods* these stakeholders require to contribute to ethical emerging technologies (by a method, we mean a means or instrument of some sort, like a cognitive tool, organisation form, set of competencies or procedure), and ways of producing them if they are not in existence yet.

[27] Brey, Philip, "Research Ethics Guidelines for the Engineering Sciences and Computer and Information Sciences", in Kelly Laas, Elisabeth Hildt, and Michael Davis (eds.), Codes of Ethics and Ethical Guidelines: Emerging Technologies, Changing Fields, Springer, Dordrecht, Netherlands, 2021.

- Analyse how technology actors can assume relevant *roles and responsibilities* and *acquire and start using the methods* they need for ethical practice.

The technology actors will be different for different technologies, as will some of the methods, roles and responsibilities. At a more general level of description, however, we can distinguish different actors, method and role categories that are likely to be similar for different emerging technologies. We will map and discuss them in this section, while also paying attention to how they may play out differently for different emerging technologies.

Technology actors

In our strategic model, we distinguish the following technology actors:

1. Technology developers
2. Technology development support organisations
3. Organisations that deploy, sell and service technology
4. Organisations that use the technology
5. Governance and standards organisations
6. Educational and media organisations
7. Civil society organisations and the general public
8. Organisations and units working on ethics and social impacts

We will now discuss each in turn.

Technology developers

This is the class of actors that harness and integrate the knowledge and resources by which new technology is made possible. Within this broad category we can make some further distinctions. At the organisational level, developers include firms that develop emerging technologies and research institutes (universities and other research performing organisations) which engage in research and innovation for emerging technologies. At the intra-organisational level there are various units within these institutions that are involved in the planning, support and carrying out of research and innovation activities. At the individual level there are also professionals in various roles (e.g., IT project manager, IT director, hardware technician, professor of robotics) who are stakeholders in emerging technology development.

Technology development support organisations

Technology development support organisations are organisations that support R&I activities of emerging technology firms and research institutes. These include business and industry associations (also known as trade organisations) – organisations that support companies in a certain sector, chambers of commerce, research funding organisations, investment banks and other investors and funders, associations of universities and research institutes, science academies and associations of science academies, professional organisations for the emerging technology fields, advisory and consultancy firms for companies and research institutes.

Organisations that deploy, sell and service technology

Organisations that deploy, sell and service technology are private and public organisations that make the technology available to clients, ensure a good fit between client and technology, or provide services for continuous good operation of the technology to clients. Examples are a company that installs and maintains software suites for businesses, or a medical professional fitting prosthetic devices to patients. For the latter, deploying technology may not be their main business, but they nevertheless function as a technology actor in this context. Note that many organisations are simultaneously developers and deployers of technology.

Organisations that use technology

Organisations that use technology are private and public organisations that use technology to further the organisation's objectives. Usage can be intended to improve or support various organisational functions, such as operations, finance, marketing, human resources, customer service, regulation, etc. Within these organisations, one can also define various units and professional roles associated with the deployment and use of emerging technology systems within or by the organisation, such as information technology managers, database administrators and development operations engineers. Note that some organisations are simultaneously developers and users of technology, or deployers and users of technology. For example, technology companies like Apple and Google develop technologies and deploy and use them within their own organisation.

Governance and standards organisations

Governance and standards organisations are organisations involved in developing, implementing or enforcing policies, standards and guidelines – specifically those regarding the development, deployment and use of technology. Organisations also make policies and guidelines for themselves. These are not our concern here. This category refers to organisations that develop or implement guidelines, policies, regulations and standards for others. This includes, first of all, national, local and supranational governments, and government-instituted or -supported advisory and regulatory bodies. They also include intergovernmental organisations such as the United Nations, the Council of Europe and the World Health Organization WHO). Also included in this category are national and international standards bodies (e.g., ISO, IEEE), certification, quality assurance, accreditation and auditing organisations. Policies, standards and guidelines may also be issued by many of the emerging technology development support organisations discussed earlier.

Educational and media organisations

Educational institutes and media organisations both have a significant role, albeit a quite different one, in shaping people's knowledge and understanding of emerging technologies, the ethical issues associated with them, and the ways in which these ethical issues can be addressed. Educational organisations, from elementary school to postgraduate education, provide the major vehicle by which individuals acquire knowledge, skills and insights regarding emerging technologies, their impacts on society, their ethical aspects, and ways to address ethical issues in their profession. Of course, it is not only educational organisations that provide education and training. Companies may, for example, organise their own in-house training as well. Media organisations have a large role in generating public awareness and understanding of emerging technologies and the ethical issues raised by them and therefore should also be recognized as actors with respect to emerging technologies.

Civil society organisations and the general public

Civil Society Organisations (CSOs) are non-governmental, not-for-profit organisations that represent the interests and will of groups of individuals. They may be based on cultural, political, ethical, scientific, economic, religious or philanthropic concerns. They include civic groups, cultural, groups, consumer organisations, environmental organisations, religious organisations, political parties, trade unions, professional organisations, non-governmental policy institutes, activist groups, and several other kinds. Many CSOs want, and should, have a role in public policy or influence the way that organisations function in which they have an interest. For some of them the development and use of emerging technologies is a concern. As a result, these organisations may function as agents with respect to public policy and the actions of relevant other organisations. The general public, finally, can also perform as a stakeholder and should be considered as such by policy makers and other actors involved in the development, deployment and use of emerging technologies. The general public can be consulted through public opinion surveys and studies and studied through voting patterns, consumer purchases, and use or non-use of emerging technology products and services.

Organisations and units working on ethics and social impacts

Finally, it is important to mention organisations and units working on ethics and social impacts. These may be part of the various kinds of organisations and units listed above. These include ethics research units, ethics policy units, ethics officers, research ethics committees, integrity offices and officers, corporate social responsibility teams and officers, technology impact assessors, ethics educational programmes, ethics advisory bodies, and national and international ethics committees. Although all of these stakeholders have a role in ensuring ethical standards and practices, ethics organisations and units have a particular responsibility in this regard. This category also includes research institutes working on the ethics and social sciences of technology, especially emerging technologies. These are needed to follow technological developments closely and their short, medium and long terms impacts on the society. Considering the novelty and complexity of emerging technologies, it is necessary to conduct in-depth studies on ethical and social impacts of these technologies on the society and identify transformations that may remain invisible without the tools of ethics and social sciences. There still remain many unknowns and uncertainties regarding the ethical and social impacts of these technologies. The resources of ethics and social sciences are much needed to lift these and come to a better and understanding of the long-term impact on society in general and on particular groups and to mitigate negative implications.

1.5 Methods

In the context of this report, methods are any type of means that stakeholders can use to take into account ethical considerations, implement ethical guidelines and engage in ethical practices. We propose eight sets of methods for the ethical development and use of emerging technologies:[28]

[28] Points 1, 3-6 are taken from the SHERPA development and use guidelines: Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the development and use of SIS, 2019. Point 2 is an added point.

(1) Methods for incorporating ethics into research and development of technology (aimed at technology developers and support organisations).

(2) Methods to incorporate ethics into the deployment, selling and servicing of technology (aimed at organisations that deploy, sell and service the technology).

(3) Methods to incorporate ethics into the use of emerging technologies by organisations (aimed at organisations that use the technology in question).

(4) Corporate responsibility policies and cultures that support ethical development and use of the technology (aimed at both developers, deployers/users and support organizations).

(5) National and international guidelines, standards and certification for ethical technology (aimed at governance and standards organisations, indirectly affecting developers, deployers/users and support organizations).

(6) Education, training and awareness raising for the ethical and social aspects of emerging technologies (aimed at all stakeholders).

(7) Policy and regulation to support ethical practices in technology (aimed at governance and standards organisations; indirectly affecting developers and deployers/users).

(8) Methods for in-depth ethical and social analysis of emerging technologies and their impacts.

We next discuss these sets of methods in more detail.

Methods for incorporating ethics into research and development

These are methods to make ethical considerations, principles, guidelines, analyses or reflections a part of research and development processes. They apply to the first actor category identified above – technology developers. Four main classes of methods fall into this category:

- Research ethics guidelines
- Ethical impact assessment methodologies
- Ethics by design approaches
- Codes of professional conduct for researchers and developers

We now discuss each in turn.

Research ethics guidelines

Research ethics guidelines for technology are ethics guidelines and procedures by which researchers, developers, research ethics committees and ethics officers can ethically assess research and innovation (R&I) proposals and ongoing R&I practices. We can differentiate between ethics guidance documents for research ethics committees, and ethical checklists, assessments or guidance documents for developers. These guidelines can be used to improve R&I plans and practices in order to make them more ethical. The SIENNA project developed its own proposals for guidelines for research ethics committees[29] and a research ethics protocol specifically focused on AI projects (Annex 4 of D5.4).[30]

Ethical impact assessment methodologies

Ethical impact assessment (EIA) methodologies are methods for assessing present and potential future impacts of emerging technologies, including specific products and applications, and identifying ethical issues associated with these impacts. EIA, in short, is an approach for assessing not only present but also potential future ethical issues in relation to a technology. EIA, in its current form, was developed within the EU-funded FP7 SATORI project.[31] It has also been developed into a CEN pre-standard.[32] See also the SIENNA report *Generalised methodology for ethical assessment of emerging technologies*, which also contains an advanced EIA methodology.[33] EIA is not just a method for emerging technology developers, but can also be used, amongst others, by government agencies and bodies to support technology policy, and by research funding organisations to help set priorities in research funding.

Ethics by design approaches

Ethics by design methodologies are methods for incorporating ethical guidelines, recommendations and considerations into design and development processes. They fill a gap in current research ethics approaches, which is that it is often not clear for developers how to implement ethical guidelines and recommendations, which are often of a quite general and abstract nature. Ethics by design methodologies identify how, at different stages in the development process, ethical considerations can be included in development. It does this by finding ways to translate and operationalize ethical guidelines into concrete design practices. Ethics by design approaches have existed in computer science and engineering since the early 1990s, initially under the name 'Value-sensitive Design' (VSD)[34] and later also under the label of 'Design for Values.'[35] During 2020, the term "ethics by design" came into vogue. An extensive ethics by design approach for AI was published as part of the EU Horizon 2020-funded SHERPA project.[36] To the best of our knowledge, as of writing this report, no other formal ethics by design approaches had been published for emerging technologies, although the IEEE is working on one. The SIENNA project builds on the SHERPA report to present an extended approach for

---

[29] Tambornino, Lisa, Dirk Lanzerath, Philipp Hoevel, Tom Lindemann, D5.1: Report Documenting Elements to Open and Complement Operational Guidelines for Research Ethics Committees, SIENNA Project, 2021.

[30] See also Brey, Philip, "Research Ethics Guidelines for the Engineering Sciences and Computer and Information Sciences" for a review of research ethics guidelines for the engineering sciences and computer sciences. Brey, Philip, "Research Ethics Guidelines for the Engineering Sciences and Computer and Information Sciences", in Kelly Laas, Elisabeth Hildt, and Michael Davis (eds.), *Codes of Ethics and Ethical Guidelines: Emerging Technologies, Changing Fields*, Springer, Dordrecht, Netherlands, 2021.

[31] https://satoriproject.eu

[32] CEN, Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework. CEN workshop agreement, CWA 17145-2, 2017.

[33] Brey et al., *D6.1 Generalised Methodology for Ethical Assessment of Emerging Technologies*, SIENNA Project. Forthcoming at https://www.sienna-project.eu/publications/deliverable-reports/.

[34] Friedman, Batya, Peter Kahn and Alan Borning, "Value Sensitive Design and Information Systems", in P. Zhang and D. Galletta (eds.), Human-Computer Interaction in Management Information Systems: Foundations, Armonk, NY: M.E. Sharpe, 2006, pp. 348-372.

[35] Hoven, Jeroen van den, Pieter E. Vermaas, and Ibo van de Poel, eds., Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains, Springer Netherlands, 2015.

[36] Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the development and use of SIS, 2019, Deliverable D3.2 of the SHERPA project.

ethics by design that has wider applicability than the one proposed in that report. Annex 2 of the present report presents the ethics by design guidelines developed in SIENNA.

Codes of professional conduct for researchers and developers of emerging technologies

Codes of professional conduct, also called codes of ethics, are codified personal and corporate standards of behaviour that are expected in a certain profession or field. These codes are often set by professional organisations. Codes of conduct exist for various professions in science, technology, and engineering. However, these are often broader codes that do not address the specific challenges and responsibilities, which emerge when someone specialises in a new technology field and this may necessitate the development of new codes specifically for professions within the scope of emerging technologies.

Interdisciplinary research and innovation

This is perhaps not so much a method as a form which R&I can take by making R&I processes interdisciplinary, involving not only science and engineering fields, but also social sciences and humanities, including ethics. Such research activities allow for a better incorporation of social and ethical concerns into engineering practice, and are therefore highly advisable at different stages in the R&D process.

Methods to incorporate ethics into the deployment, sale and servicing of technology and methods to incorporate ethics into the use of technology by organisations

We discuss these methods together because deployment and use are often closely related practices. After the development of technology systems, services and solutions, they are deployed and used.[37] The deployment and use of new technologies often require their own ethical guidelines and solutions, which are, to some extent, different from those that apply to their development. Ethical questions that are typically asked in relation to deployment and use include questions like: Is it ethical to deploy a system that is intended to do X/ is capable of doing X/ can be used to do X? How can unethical uses of the system be monitored and prevented? What is the responsibility of different stakeholders in preventing or mitigating unethical use? What policies to prevent unethical use should be put in place and how can they be implemented effectively?

Deployment and use scenarios come in various forms, but the following are the most typical:

(1) Deploying emerging technology to enhance organisational processes. An organisation acquires technology and uses them to improve its organisational processes, such as manufacturing, logistics, and marketing. End-users are IT specialists or other employees.

(2) Embedding emerging technology in products and services. An organisation acquires emerging technologies and incorporates them into products or services that it offers to customers. This is a different application of emerging technology than its application in the development, manufacturing and marketing of products and services. For example, emerging technologies can be used to better design, manufacture or market automobiles that do not themselves contain these emerging technologies.

Emerging technologies can be embedded in products and services for different purposes:

[37] Of course, deployment and use cycles are often followed by repeated redevelopment of systems.

- To enhance the value of a product or service for customers by offering enhanced functionality or usability. For example, by powering an online dating service with AI algorithms, or by enhancing an automobile with a self-drive mode.
- To enhance the value of a product or service through intelligent monitoring, self-repair, communications with customer service, or data collection for future upgrades.
- To further the interests of the organisation or of third parties, for example, by collecting data for marketing purposes or allowing for targeted messaging.

It is not always clear which party is the end-user of the emerging technology in these three scenarios, since the end-user of an emerging technology embedded in a product or service may be different from the end-user of that product or service, and there may also be multiple end-users (e.g., Uber drivers and customers using the same AI algorithms).

In addition, in relation to an organisation that starts using technology, deployment actors can be in-house, external or a mixture of the two. These configurations result in different responsibilities for the user-organisation.

Taking these scenarios into consideration, the following five methods can contribute to ethical deployment and use of emerging technologies:

- Operational ethics guidelines and protocols for the deployment and use of the technology for the enhancement of organisational processes
- Operational ethics guidelines and protocols for the deployment and use of the technology in products and services
- Codes of professional ethics for managers, technical support specialists and other management, engineering staff responsible for the deployment and use of the technology in an organisation or its embedding in products and services
- End-user guidelines for ethical usage of (products and services that include) emerging technologies
- Operational ethics guidelines, protocols and codes for servicing and maintenance of the technology.

Corporate responsibility policies and cultures

Ethics guidelines and professional ethical codes, even when fully operationalized for particular practices, will have little impact if they are not supported by organisational structures, policies and cultures of responsibility. Therefore, corporate social responsibility (CSR) is itself a major method for securing ethics practices. In truth, it is, of course, not a single method, but is manifested in informal codes and cultures, policies, units, communications, actions and events. CSR needs support from an organisation's board of directors, who should support ethical practice throughout the company. They should meet regularly to discuss ethical and corporate responsibility issues and best practices within the organisation, and carry responsibility for their ethics and CSR policies. In addition, management strategies should be in place that place a high value on ethics, including ethical practice in relation to the technologies developed, deployed, or used by, and in, the organisation. There should be policies, structures and compliance mechanisms in place to support ethical practice.

National and international guidelines, standards and certification

In this report, we distinguish between *practice-specific ethics guidelines*, which are guidelines developed for specific practices by specific stakeholders (e.g., research practices, deployment practices), and *general ethics guidelines*, which are statements of ethical principles and general guidelines that apply to a broad range of stakeholders and practices. While it is possible to develop practice-specific guidelines without general guidelines, it is often beneficial to have shared general guidelines on the basis of which practice-specific guidelines are developed. These general guidelines would ideally be supported by national governments and intergovernmental organisations.

Standards, developed by recognised national and international standards organisations or by particular (associations of) companies or organisations, are different from ethics guidelines in two ways. Firstly, they apply to specific products, services, processes or methods, while ethics guidelines apply to any action, thing or event that has ethical implications. Secondly, they define specific norms or requirements to which the phenomenon to which the standard applies must adhere. Standards are intended to leave limited room for subjectivity and interpretation and are intended to define intersubjective requirements that different stakeholders can apply, identify or assess. Standards sometimes aim to codify ethical requirements, procedures or methods.[38] Examples are ISO 26000[39], which is an international standard for corporate social responsibility, CEN CWA 17145-1[40], which is a standard for ethics assessment by research ethics committees, and CEN CWA 17145-2[41], which is a standard for the method of ethical impact assessment for R&I. Standards can also include ethical requirements, procedures or methods, while not themselves having ethics as a focus. For example, ethics is discussed in the context of the ISO 9000 and 9001 standards[42] for quality management. For AI & robotics, a number of ethical standards are currently being developed by IEEE[43] and by ISO.

Certification is the process by which an external third party (typically a certifying body) verifies that an object, person or organisation is in possession of certain characteristics or qualities. Amongst others, certification can be applied to persons, in professional certification, to products, to determine if it meets minimum standards, and to organizations or organizational processes, through external audits, to verify that they meet certain standards. Certification can be a means to verify and validate the quality of ethics processes and procedures in organisations. In relation to standards, in particular, certification can be a means of ensuring conformity to the requirements of the standard. For example, IEEE is currently developing its own certification programme to certify adherence to the ethics standards it is developing for AI and robotics. ISO does not carry out certification itself, but third-party certification organisations could assess compliance to ISO ethics-related standards in the future.

---

[38] ISO is the International Organisation for Standardisation which develops and publishes international standards: https://www.iso.org/home.html. CEN is the European Committee for Standardization. It brings together the National Standardization Bodies of 34 European countries to develop and define standards at the European level.

[39] https://www.iso.org/iso-26000-social-responsibility.html

[40] https://satoriproject.eu/media/CWA_part_1.pdf

[41] https://satoriproject.eu/media/CWA17145-23d2017.pdf

[42] https://www.iso.org/iso-9001-quality-management.html

[43] IEEE, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition, IEEE, 2019.

Education, training and awareness raising

Education is a powerful method for stimulating ethical behaviour in relation to emerging technologies. In professional and academic education, specifically, education that concerns ethical and social issues in emerging technologies would benefit future professionals, especially those working in the technology field, those in other fields who may deploy and use these technologies in the future, and more generally, any individuals to make informed decisions about the technology in question.

Most active professionals who currently develop, deploy and use technology have not received ethics education in relation to technology in their professional education. For them, continuing education programmes in ethics (of technology and engineering) would be valuable. Such training programmes could even be accompanied by professional certification, for example, certification in ethics by design methodology, preparing for ethics review, or all-round ethical knowledge and skills in relation to the relevant technology. In addition to external organisations setting up such training and education programmes, organisations could also organise their own in-house training in ethics for emerging technologies.

Turning now from educational institutions to the media, we should acknowledge that media organisations and journalists (including independent ones) have a large role in generating public awareness and understanding of emerging technologies, including the ethical issues raised by them. These are often complicated technologies that are difficult to understand for the general public. Since it is expected that many of these technologies will have a significant impact on people's lives, a proper understanding of them and their ethical issues could be important. A certain degree of awareness of the technologies and their social and ethical impacts is also essential to ensure proper public oversight. Media companies and journalists are an important type of organisation that can provide such an understanding to the general public. Therefore, relevant media stories on emerging technologies and its social and ethical dimensions, whether in print, podcast, television or other formats, are important.[44] While media organisations and journalists have a major responsibility here, technology developers also have a responsibility to be transparent and communicate with the public about these issues, as do governments in ensuring that sufficient information is provided.

Policy, regulation and governance

While policy can be made by any kind of organisation, our concern is with public policy, as made by governments, as well as the laws and regulations created by them. The key question here is: what policies, laws and regulations should governments develop, if any, to stimulate the ethical development, deployment and use of emerging technologies? Policies, laws and regulations can relate to ethical criteria in three ways – they can explicitly institute, promote or require ethics guidelines, procedures, or bodies, they can have a focus on upholding certain moral values or principles without explicitly identifying them as ethical (e.g., well-being, privacy, fairness, sustainability, civil rights), and they either explicitly or implicitly take on board ethical considerations in broader social and economic policies.

Some decisions that governments should be making when confronted with emerging technologies that raise significant ethical issues include the following:

[44] See in particular the media analysis conducted as part of SIENNA Deliverable D4.4 and the public perception studies conducted in SIENNA in D4.5 and D4.6: Jansen, P., et al, op. cit., 2020. Hamlyn, R., op. cit., 2020. Kantar (Public Division), op. cit., 2019.

- Whether or not to issue, or support the issuing of, ethical guidelines for the emerging technology.
- Whether or not to put any ethical guidelines for the emerging technology into law.
- Whether or not to revise existing institutional structures to better account for ethical issues or to create new governmental bodies or unites for ethical and social issues relating to the emerging technology.
- Whether or not to mandate ethics standards, certification, education, training, ethical impact assessments or ethics by design methods in relation to the emerging technology.
- How to include ethical considerations concerning the emerging technology in policies, laws and regulations, including the creation of new policies, laws and regulations specifically for the technology and updating existing ones to account for the technology, such as in the areas of consumer protection, data protection, medical law, environmental law, criminal law, non-discrimination provisions, civil liability and accountability.
- What financial support and funding to provide, if any, for ethics research, ethics education, ethics dialogue, ethics awareness raising and other ethics initiatives in relation to the technology.
- How to regulate the government's own use of the technology so as to ensure ethical conduct.

Methods for in-depth ethical and social analysis of emerging technologies and their impacts

The last category of methods that we wish to highlight to ensure ethical considerations are taken into account in the development and use of emerging technologies concerns methods for the ethical and social analysis of emerging technologies. Ethics is in need of advanced methods to analyse emerging technologies.[45] Several methods have been developed recently, however, including the previously mentioned methodologies for ethical impact assessment. Methods from the social sciences, including technology assessment, social impact assessment, science and technology studies and human rights impact assessment, are important for acquiring a better understanding of consequences and impacts of emerging technologies that we often do not fully comprehend nor are able to mitigate properly.

Roles, responsibilities and use of the methods

In our strategic model, emerging technology is developed, deployed and used in an ethically responsible way if, at the different steps of the ETCOM model, the technology actors assume appropriate roles and responsibilities, and use the methods that are needed to bring about responsible, ethical practices of development, deployment and use. Let us review what should happen at these steps.

Assessment stage

Key technology actors with a responsibility for ethical issues at this point are technology developers, technology development support organisations, particularly research funding organisations, ethicists,

[45] Brey, Philip, "Ethics of Emerging Technologies", in Sven Ove Hansson (ed.), The Ethics of Technology: Methods and Approaches, Philosophy, Technology and Society, Rowman & Littlefield International, 2017, pp. 175–192.

social scientists and media actors, particularly journalists and news organisations. Technology developers have a responsibility to do early signalling of ethical issues in their emerging technology field, to the extent that they are able discern them. No particular methods are needed for this, just a general sensitivity to ethical issues, which could be strengthened by ethics education which they would have received as part of their education in science and engineering. Research funding organisations should ethically assess project proposals or require that funded projects undergo ethics review locally, and should make funding available for ethics research into emerging technologies. Research funding organisations and research ethics committees may not have specialised ethics guidelines for the emerging technology at this time, but could rely on existing frameworks and competencies to spot and assess ethical issues. Ethicists should do ethical analysis of emerging technologies at this stage already, and social scientists should start doing foresight studies and impact assessments. Journalists and news organisations, finally, should report on ethical issues that have been spotted by technology developers and ethicists, and stimulate public debate.

Early guidance stage

Key technology actors at this stage are technology support organisations (particularly professional organisations for scientists and engineers, science academies, and industry associations), governmental and intergovernmental organisations, and organisations and units working in ethics, particularly (associations of) research ethics committees and national ethics committees. All these organisations have a potential role in developing ethics guidelines, particularly general ethics guidelines and research ethics guidelines for the technology in question. It is not determined in advance which of them must issue ethics guidelines, but it is important that at least one of them takes this responsibility if the emerging technology raises significant ethical issues, including ones that can be mitigated in R&D.

Operational guidance stage

Virtually all technology actors need to come into action during this stage, if ethical issues are to be addressed properly, since products and applications from the emerging technology field are about to hit the market, and technology actors have to start negotiating their responsibilities to ensure that the deployment and use process will take ethical considerations into account. Moreover, as the technology gradually materializes, and the understanding of ethical issues and social impacts grows, technology development actors have an enhanced responsibility to address ethical issues in the ultimate design of products and applications. This may be a period in which Ethics by Design methods for the technology are developed and put into place, further ethical impact assessment is carried out, and ethics guidelines for deployment and use are formulated. Governmental actors will have to start thinking about regulation and technology deployers will have to start considering ethical issues and ways to address them.

Intermediate institutionalisation stage

Governmental and standards organisations are key actors at this stage, since this is a time during which new legislation and regulations should be put in place, some of it possibly as a way to address ethical issues. For example, it may be deemed that new regulations need to be put in place to better protect fundamental rights of citizens which may be harmed due to the new technology, to level playing fields, or to avoid harms. At this point, technical standards, some of which could address ethical issues, may also be developed. Additionally, all technology actors, particularly deployers and user organisations, should consider how to properly implement and (where possible) formalize strategies for ethical practice. This includes proper implementation of ethics guidelines for deployment and use.

Late institutionalisation stage

Key actors include educational actors, technology support organisations (including certification agencies), as well as all other technology actors. During this stage, further institutionalisation occurs, and ethical practice is, in large part, normalized and made part of routine professional practice. All technology actors have a responsibility in making this happen. Educational actors, particularly universities and professional schools, may start including ethics education for the emerging technology field in their curricula, and certification organisations may offer ethics certification for ethical development, deployment or use of the new technology, or may include it in regular certification and quality assurance frameworks.

We cannot provide a precise account of the roles and responsibilities of different actors because roles and responsibilities are, to a significant extent, negotiated between the parties. Such negotiation can be explicit, as when actors reach formal agreement on roles and responsibilities, but it is also often implicit, as when actors start assuming roles and others hold them to these roles. Similarly, we cannot provide a precise account of the methods that technology actors should use or the timing that this involves, for them to properly address ethical issues and move towards ethical practices. This depends, first of all, on the negotiated responsibilities, but it also depends on preferences - the methods are instruments and the same result can sometimes be reached by means of different instruments.

More studies are needed to reveal which methods and distributions of roles and responsibilities are most effective in supporting ethical development, deployment and use practices for emerging technologies. Much can be gained from such studies. In addition, more insight is needed regarding what motivates technology actors to invest in ethical practices, and what keeps them from doing so. A lack of investment in ethics and ethical practices is not necessarily the result of a lack of interest. It could be the result of a lack of efforts of the different parties to establish roles and responsibilities in communication with each other, a lack of usable methods that are available to actors for addressing ethical issues or a lack of knowledge and skills needed to address them.

If technology actors fail to step up, leaving a responsibility vacuum in society, governments are often seen as the responsible stakeholder to step in and enact policies, laws and regulations that help fill this vacuum. While there are some limitations, governments, after all, have a particular responsibility to promote the public good, protect human rights, and support fair socioeconomic conditions, and have the powers to stimulate and compel other stakeholders to act responsibly and in the public interest. However, it is our belief that much can be gained by the further development of useful, operational methods for addressing ethical issues in technology development, deployment and use, by facilitating dialogue between technology actors regarding their moral and social responsibilities. In addition, it needs to be advocated that ethics and success often go hand in hand – if technology does not conform to ethical standards, then it is likely to meet resistance from users and other stakeholders, thereby making the technology less successful. This is something that would not benefit any technology actor.

1.6 Conclusion

In this chapter, we proposed a multistakeholder, coevolutionary strategy for the ethically responsible development, deployment and use of new technology. We started out with the presentation of the Ethics and Technology Coevolutionary Model (ETCOM), a normative model for the coevolution of technological innovation and ethics. ETCOM proposes five stages of ethical practice for emerging technologies that correspond to different stages of emergence. They are the assessment stage, early guidance stage, operational guidance stage, intermediate institutionalisation stage, and late institutionalisation stage. The first three follow the technology as it is still largely in the R&D stage, and the fourth and fifth follow the technology as it hits the market and starts its ascent.

Following this macro-level normative model, we moved to a lower level of analysis to discuss roles and responsibilities of technology actors in relation to ethical issues, and the methods that they would benefit from using to address these issues. We did so in the context of the ETCOM model. We identified and described eight main types of technology actors, with subtypes in each category. We then focused on methods for addressing ethical issues and engaging in ethical practices. We described a large number of such methods, coupling them to the different classes of technology actors. Some of these methods are relatively new, and in general, better development of these methods should be high on the agenda for the technology actors. Other chapters in this report are, in fact, focused on the development and improvement of methods for strengthening ethical practice. Finally, we discussed the roles and responsibilities that actors should assume, the actions that they should perform, and the methods they should use at different stages in the coevolutionary model to contribute properly to ethical development, deployment and use of emerging technology.

We are convinced that a normative, multistakeholder, coevolutionary approach such as the one we present here could greatly benefit technology actors in finding and strengthening their role in responsible and ethical innovation practices. It will also help them in engaging in fruitful dialogue with other actors regarding their collective responsibilities, and in prioritizing the methods that need to be developed and used to strengthen ethical practice. We also pointed out, however, that more studies are needed to assess which methods and which distributions of roles and responsibilities for actors are most effective in supporting ethical development, deployment and use practices for emerging technologies, and what motivates technology actors to invest or refrain from investing in ethical practices. Our approach will be helpful in identifying the kinds of investigations that are needed here.

2. The development and operationalisation of ethics guidelines

2.1 Introduction

In this chapter, the objective is to present a step-by-step method for the development of ethics guidelines. In addition, we will discuss ways in which such guidelines can be operationalised and implemented. Ethics guidelines, codes, rules of conduct, and standards exist in many forms, for many purposes. To establish the scope of this chapter, we will therefore first define our terms.

Centrally, we make a distinction between ethics guidelines and ethics codes.[46] We define ethics guidelines as ethics standards for a particular set of practices. For example, ethics guidelines for human experimentation are ethical standards to which experiments involving human subjects should adhere. Ethics guidelines for the social sciences are ethics standards for conducting research in the social sciences. Note that ethics guidelines are prescriptive for practices, and not necessarily for individual conduct. Practices typically involve multiple actors, and are typically also dependent on instruments, infrastructure, and other conditions. Ethics guidelines prescribe how this total configuration of humans and nonhumans should operate and does not necessarily prescribe particular individual actions or behaviours.

Ethics codes, as defined here, are guidelines for ethical behaviour by individual professionals in various professional fields.[47] They aim to regulate professional conduct so as to ensure it exhibits high ethical standards, professional quality, and trustworthiness. Often, a distinction is made between two types of codes: codes of ethics and codes of conduct. The former describes ethical values or principles that an organisation or a profession aspires to embrace. The main criticism of such codes is they are too broad or vague and do not provide necessary guidance on how users should act. Thus, a code of conduct can deliver a more practical document to users which details accepted behaviour in the form of do and do-nots. However, some critics argued that such codes usually have a narrow focus and do not address all anticipated situations and can inadvertently let users furlough ethical judgement and individual responsibility. As a result, the favoured direction now is to create documents that combine both “integrity-based” and “compliant-based” aspects (codes of ethics and professional conduct).

This chapter is concerned with the development and operationalisation of ethics guidelines, not ethics codes. However, many elements of our proposed approach may also apply to ethics codes. In addition, there is a significant body of literature on the development and implementation of ethics codes, whereas less has been written on the development of ethics guidelines. Accordingly, our proposals build on literature primarily concerned about ethics codes.

Within the family of ethics guidelines, a further distinction can be made between general and practice-specific ethics guidelines. General ethics guidelines are guidelines that apply all relevant practices relating to a phenomenon (like a technology), and to all actors who have a role in these practices. They can be used in different contexts and settings, by different actors. General ethics guidelines define

[46] Our proposed terminological distinction mirrors how these terms tend to be used in practice. There are exceptions. For instance, the Nuremberg Code is, using our terminology, a set of ethics guidelines for research, not a code.

[47] Martin, Clancy, Vaught, Wayne and Solomon, Robert C. (eds.), Ethics across the Professions: A Reader for Professional Ethics, New York: Oxford University Press, 2010.

desirable outcomes for society and direct different actors to engage in practices that contribute to these outcomes. They contain general ethical principles and statements like “Informed consent must be guaranteed for all genomic therapies”, “AI systems should allow for meaningful human control” and “technology actors should take responsibility for contributing to sustainability goals in the development of new building technologies”.

Practice-specific ethics guidelines relate to one type of practice only, typically engaged in by a limited number of technology actors, and provide guidance for this practice. For emerging technologies, research ethics guidelines that aim to guide research and development practices are particularly important. But ethics guidelines can also be developed for other technology-related practices, such as deployment and use. Since multiple practices are typically associated with a new technology (e.g., development, deployment, regulation, use, certification), it is conceivable that general ethics guidelines are developed for it, and, in addition, several sets of practice-specific guidelines that apply to different practices and thereby, indirectly, to the actors associated with these practices.

For example, for AI, general ethics guidelines have been proposed by various organisations (e.g., The Organisation for Economic Co-operation and Development (OECD), and the High-Level expert Group on AI, instituted by the European Commission), which are clearly stated to apply to all technology actors and practices. Next to these, practice-specific guidelines have also been proposed, such as the research ethics guidelines that were developed in the EU-funded SIENNA project[48] and the deployment and use guidelines that were developed in the EU-funded SHERPA project.[49] In addition, it is conceivable that ethics guidelines are developed that are restricted to the practices in particular domains (e.g., ethics guidelines for AI in medicine, or AI in defence), or for practices relating to particular products associated with the technology (e.g., ethics guidelines for autonomous vehicles, or for decision support systems).

[48] Brey, Philip, “Research Ethics Guidelines for Artificial Intelligence”, Annex 4 to Resseguier, Anaïs, Philip Brey, Brandt Dainow, Anna Drozdzewska, Nicole Santiago and David Wright, SIENNA D5.4, Multi-Stakeholder Strategy and Tools for Ethical AI and Robotics, forthcoming.

[49] Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the Ethical Use of AI and Big Data Systems, SHERPA project, July 2019.

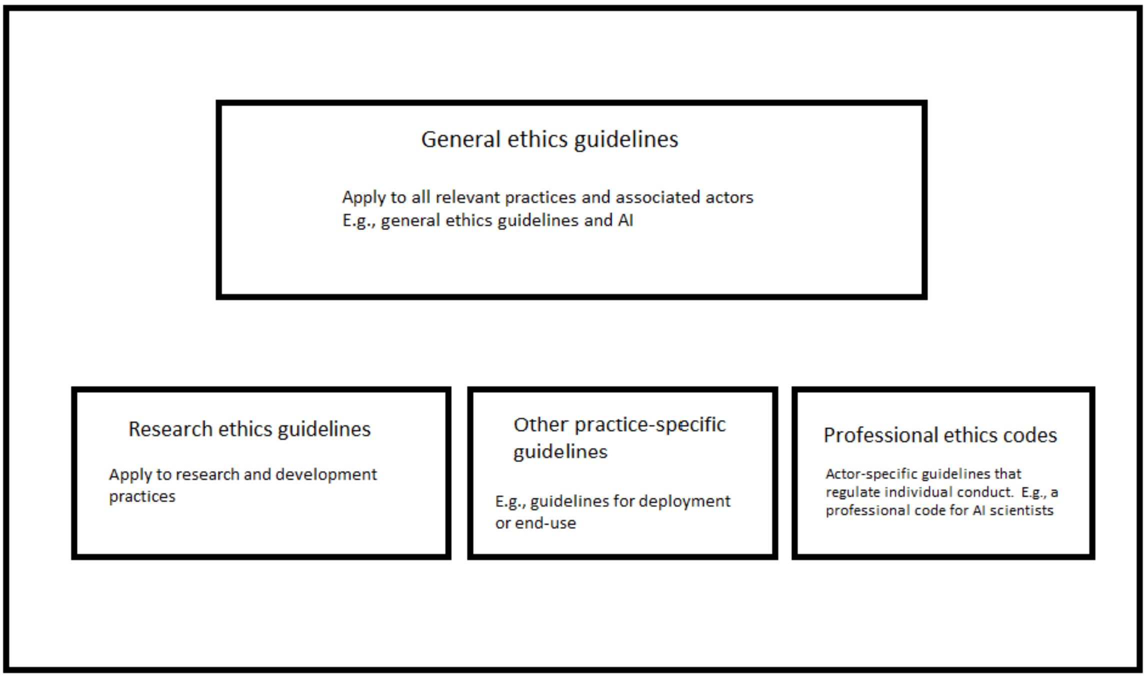

Figure 3: *Types of guidelines and codes*

In the remainder of this chapter, we will proceed as follows. First, in section 2.2, we will review existing literature on the development of ethics guidelines and codes and will present several cases of guideline and code development. Next, in section 2.3, we will propose a step-by-step approach for the development of (general) ethics guidelines. In section 2.4, we will discuss the development of practice-specific guidelines. In section 2.5, we will discuss how ethics guidelines can be operationalised. Finally, in the conclusion, we summarise our findings and discuss limitations and future studies.

2.2 Review of literature and cases

Very little literature exists that addresses procedures for developing ethics guidelines. Much more literature exists about the process of developing codes of ethics and codes of conduct.[50] Reasoning that the literature on code development can be instructive for guideline development as well, we start by reviewing some of this literature. We then move to the presentation of cases, both of the development of codes as well as of guidelines. The literature review and the cases will help us construct our step-by-step approach for the development of ethics guidelines.

2.2.1 Literature review

The white paper 'Developing a Code of Conduct: A Step-By-Step Guide', produced by Lighthouse Services,[51] notes a number of key factors to consider and steps to take in the development of a code.

[50] Rodrigues, Rowena, Stearns Broadhead, Philip Brey, Zuzanna Warso, Tim Hanson, Lisa Tambornino, and Dirk Lanzerath. SIENNA D1.1: The Consortium's Methodological Handbook (Version V0.6), 2018. https://doi.org/10.5281/zenodo.4247383

[51] Lighthouse, "Developing a code of conduct: A step by step guide", 2013. https://www.lighthouse-services.com/documents/Developing%20a%20Code%20of%20Conduct%20[A%20Step-by-Step%20Guide].pdf

These include giving sufficient thought to a title for the code, for instance, to encourage ownership. In addition, the code itself should avoid an overly formal or legal tone and include key sections such as (1) an introduction from someone with legitimate authority, (2) an outline of purposes and benefits, (3) an account of key mission statements and values (4) an outline of key definitions, scope, provisions and examples (5) a framework for decision making, including guidance for the assessment of ethical dilemmas (6) resources related to compliance, non-compliance, and how to approach grievances.

The white paper also details a number of steps for the development of such codes, which can be summarised as follows:

Step one: Information and idea gathering, which takes into account the specificity of the context within which a code will be applied and makes use of stakeholder input, e.g. through focus groups.

- *To be considered:* (1) any existing codes and values, whether formal and informal, from in or outside the organisation, (2) laws and other regulatory instruments to include, (3) pertinent (previous, current, potential) ethical dilemmas, (4) any relevant specifics such as conflicts of interest or regarding integrity, (5) information about the kinds of professional relationships affected or governed by the code.

Step two: Creating a usable draft, which is at its core, "a positive, values-based document that serves as a guideline for appropriate behaviour instead of merely a list of rules and regulations that must be obeyed at all costs." [52] To these ends, they recommend clear and concise language, avoiding approaches that are highly formal or legal, thereby making a document impractical, or which are threatening or intimidating, thereby alienating users.

- *To be considered:* (1) identification of key topics and definitions, (2) clear statement of intent and core principles, (3) guidelines and procedures with examples.

Step three: Review, which checks for compliance, as well as for usability. To ensure the latter, it is recommended that the code is tested by stakeholders.

- *To be considered:* (1) selection of key stakeholders who are not otherwise involved with the development of the code, (2) involvement of senior leaders,[53] (3) oversight to ensure legal compliance.

Step four: Adoption of the code, which includes formal presentations, approvals, and ensuring adoption by key stakeholders.

- *To be considered:* (1) how to frame the process of adoption to ensure ethical outcomes, (2) who will have responsibility for the final adoption of the code.

Step five: Introduction of the code in a way that ensures successful implementation. They propose that senior leaders should be responsible for this step, and that they should also ensure there are sufficient resources for successful execution. Whether buy-in could be more successfully achieved with a bottom-up approach is not discussed.

- *To be considered:* (1) methods of introduction, for instance at unveiling events or kick-off meetings, (2) methods for distribution of written versions, and decisions about how and when,

[52] Ibid, p. 5.

[53] The authors offer no guidance here on how to manage potential conflicts, including of interest, between the values of stakeholders and those of senior leaders. We suggest this would also be an important factor to consider, not least in terms of practical steps for the resolution of such conflicts.

e.g. added to newsletters and emails, given to new staff on arrival, promoted in posters (3) methods for training, e.g. presentations, scenario workshops, e-learning (4) methods to promote buy in for training, e.g. certificates and scope for regular or annual retraining, or on an ad hoc or case-by-case basis.

Step six: Enforcement, to ensure effective compliance.

- *To be considered:* (1) who has responsibility for enforcement, e.g. an ethics or compliance officer, (2) methods for staff engagement, including an outline for how staff can report possible violations anonymously and without risk of retaliation, i.e. as outlined in a policy, (3) methods for enforcement, e.g. monitoring processes and procedures and an outline for investigation processes and disciplinary actions, (4) methods for review and amendment, including an outline that can be temporal, i.e. on a periodic basis or context driven, i.e. as a result of significant change (internal or external, expected or unpredicted), as well as processes to solicit and evaluate feedback from stakeholders in advance of and/or during review.

Lawton[54] offers similar guidance for the implementation of codes of ethics, which includes: (1) that buy-in needs support from, and demonstrated commitment by, those in leadership roles, (2) that codes should be positive, e.g. aspirational as well as regulatory, (3) embedding codes by using effective consultation processes and wide dissemination practices, (4) use of regular and consistent processes of application, and ensuring access to the codes to encourage recognition and understanding, (5) that benefits are positively and regularly demonstrated, (6) that methods are devised by which the code can be adapted in response to internal or external challenges, not least as arise from changes in leadership or policy. Lawton recommends that successful implementation of a code requires additional skills of those who undertake the task, including of persistence and communication. Furthermore, he suggests that such skills should complement existing methods of compliance and adoption, such as achieved through training in ethics, as well as by creating processes for measuring and enforcing accountability.

We can add here a number of additional criteria for processes of rational regulation. These criteria include those devised in order to set the standards against which clauses or rules within a code are framed. Some of these criteria overlap with those we have listed above, and include comprehensibility, capacity for conflict resolution, equality, effectiveness, acceptability, feasibility, enforceability, and lastly the capacity to integrate a code. Below we explain in more detail some of the reasoning for these criteria:[55]

1. ***Comprehensibility***: as outlined above, a regulatory document should be understandable and clear. Rules should be unambiguous to users. The purpose the document serves should be explicit; the rules should also outline whether such purposes are legal, or to ensure best practice. The target audience should also be clear.

2. ***Conflict resolution***: conflicts, especially in the interpretation and application of rules, and as they occur between users, are likely to arise, especially in practice. The code of conduct/ethics should describe a functional system to resolve them. This can be in the form of an agency within academic institutions or via court procedures. Furthermore, the document should uphold principles of equality before review boards or the law or ethics committees.

[54] Lawton, Alan. "Developing and implementing codes of ethics.", *Viešoji Politika ir Administratyvimas*, Vol. 3, No. 7, 2004, pp. 94–101.
[55] Eriksson, S, 'Rationality of rules', 2021. (unpublished)

3. ***Effectiveness:*** procedures should be in place to examine the effectiveness of the code. Effectiveness is broadly defined as achieving the objectives of the code, though additional criteria can of course be added.
4. ***Acceptability:*** to fulfil the criterion of acceptability, the rules should be fair and relevant. The target users of the documents should be well-described, and the roles clearly defined in order to avoid that users deflect or even abscond from responsibility.
5. ***Feasibility and Enforceability:*** for the code to be enforceable, it has to be feasible and users will need to agree that it is relevant to their operations. Furthermore, those administering the code should also have the means to implement, monitor and, where applicable, enforce it.
6. ***Capacity to integrate***: reflects the internal cohesiveness of the regulatory document and the relationship between the different rules that it contains. The rules should not conflict with each other. As far as possible, there should not be conceptual gaps. A successful code is one with a clear logical and systematic structure.

Support and enabling systems are frequently recommended for the successful development and implementation of the codes we outline above. These require a committed leadership who not only practice what they preach but are also seen to do so. In order to ensure the success of a code, there need to be robust systems in place to challenge and defend behaviours considered to be unacceptable. In addition, adequate ethics training is needed so as to ensure that authentic buy in as well as ownership of a code. There also needs to be means for users to seek advice in ethically challenging situations, and a system for reporting misconduct and rewarding good behaviour. Codes benefit from clear relations with legal foundations, and with guidance about whether the relation is direct or indirect. Whether a system of development is top-down or bottom-up can make a difference to both acceptance and compliance, and there can be good arguments for both positions, as we have briefly shown in this section. We have however found some consensus that a bottom-up approach to the development of the code offers a better approach for the following reasons: It allows a sense of ownership from users and ensures greater compliance, it creates a relevant code that addresses realistic and concrete situations as key stakeholders themselves experience them, and it contributes to building an ethical environment since users themselves partake in designing the code by which the environment is then governed.[56, 57]

2.2.2 Cases

In the following section, we discuss four individual cases of code and guideline development. The first two concern the development of codes specifically. Since codes are already covered in some detail in the literature section above, here we offer only very brief accounts, and primarily so as to identify and explicate some general methods from such endeavours. The cases we consider in this section are (i) the code of ethics (CoE) devised by the Canadian Psychological Association (CPA), and (ii) methods adopted by a commercial organisation called Newton to draft its CoE. After this we provide more detailed cases regarding the development of ethics guidelines. We hope that by offering more detail in these cases we can redress the imbalance caused by the dearth of literature on the topic of guideline development. Accordingly, in the second half of this section we cover (iii) The development of 'The

[56] Lawton, ibid.

[57] Gilman, SC, "Ethics codes and codes of conduct as tools for promoting an ethical and professional public service: Comparative successes and lessons", Prepared for the PREM, the World Bank, 2005.

Ethics Guidelines for Trustworthy Artificial Intelligence' for the European Commission, (iv) The SIENNA approach that was taken for the preparation of ethics guidelines for human enhancement.

Canadian Psychological Association (CPA) code of ethics

Here we describe the method that the Canadian Psychological Association (CPA) employed to create its CoE. The assigned taskforce turned to the existing literature first. They critically reviewed the CoE of the American Psychology Association (APA) and examined it in relation to a wider literature of ethical codes. From this they determined theoretical foundations for their method, which relied on a theory developed by Kohlberg (1969). He proposed that moral thinking was "a cognitive, stage-related, developmental process."[58] The last stage of this three-level course enabled individuals to apply ethical principles in solving ethical dilemmas. In addition, the task force was able to conceptualise the main objectives of CPA's CoE and specify the suitable method of constructing it.[59]

The method consisted in distributing a questionnaire on a number of hypothetical ethical scenarios to a randomly selected sample of CPA members. The text included ethical dilemmas that psychologists encountered during their practice, teaching, or research. The written feedback was subjected to content analysis that generated a number of statements that were later arranged into categories. Each group was abstracted into an overarching ethical value. The resulting document delineated major core values and ethical principles as well as standards of behaviour. These were further supplemented with provisions from other countries' CoEs and relevant literature on ethical guidelines. The final document contained high level ethical values as well as more detailed standards of behaviours.[60] Finally, the code was tested for validity. This was carried out by sending the CoE to internal (CPA psychologists) and external (regulatory bodies, lawyers, academics) stakeholders for review and amendment. The feedback resulted in minimal changes.[61]

Newton code of ethics

Here we review the method that was adopted by a commercial organisation to draft its CoC. The concept of ethical validity, as developed by Newton,[62] founded a three-stage framework that was used to develop a CoE for the National Association of Senior Move Manager (NASMM). The NASMM is a company that assists families, the elderly and other individuals in need of support to move to "assisted living facilities."[63] The method the organisation adopted included "participation, content validity and authenticity of leadership."[64] Members of the organisation or company would take part in the process of developing the code by, for example,

1. Identifying the core values of the organisation/company

[58] Kohlberg, L, "Stage and sequence: The cognitive-developmental approach to socialization", Handbook of socialization theory and research, Vol. 347, 1969, pp. 347-480.
[59] Sinclair, C, Poizner, S, Gilmour-Barrett, K, Randall, D, "The development of a code of ethics for canadian psychologists", Canadian psychology = Psychologie canadienne, Vol. 28, No. 1, 1987, pp. 1-8.
[60] Ibid.
[61] Ibid.
[62] Newton, LH, Hoffman, W, Frederick, R, 'The many faces of the corporate code', Taking Sides: Clashing Views on Controversial Issues in Business Ethics and Society, 3rd Edition (Dushkin Publishing Group, Guildford, CT), 1994, pp. 81-88.
[63] National Association of Senior and Specialty Move Managers, Nasmm, 2021. https://www.nasmm.org/.
[64] Messikomer, CM, Cirka, CC, "Constructing a code of ethics: An experiential case of a national professional organization", Journal of Business Ethics, Vol. 95, No. 1, 2010, pp. 55-71.

2. deliberating on ethical responses to common ethical dilemmas users commonly encounter and
3. reviewing the final product.

Wide participation was achieved through a series of workshops involving all members[65]of NASMM, where they responded to commonly encountered ethical dilemmas. The workshops resulted in a draft of the CoE that fulfilled "content validity" where rules were relevant and functional for users.

The ultimate goal of the code was to foster ethical behaviour within an organisation rather than a more mundane objective of improving "company image". This was evidenced by the organisation's leaders showing commitment and actively supporting the creation of the code. As such, the third dimension of "authenticity" of the high-level management proposed by Newton was accomplished.[66]

Ethics Guidelines for Trustworthy AI (European Commission, 2019)

The Ethics Guidelines for Trustworthy Artificial Intelligence (AI) are general ethics guidelines for AI that were prepared by an independent expert group, the 'High-Level Expert Group on Artificial Intelligence' (AI HLEG), set up by the European Commission in June 2018.[67] The guidelines were one of their two assignments; the other was to prepare policy and investment recommendations for AI. The AI HLEG guidelines consist of a set of fundamental ethical principles (chapter 1), followed by specific ethics requirements on AI and methods for satisfying these requirements (chapter 2), and a pilot-version assessment list to operationalise the requirements (chapter 3). The AI HLEG published a later separate report which contained a final assessment list, after extensive stakeholder consultation.[68] The report also contains a brief chapter with examples and cases.

We now reconstruct the steps taken in the process of developing these guidelines. Our reconstruction is based on public and non-public documents from the AI HLEG that were available to us, including from our own participation in AI HLEG meetings and interaction with AI HLEG members. A first step was the EC's decision that European ethics guidelines were needed for AI, and that they should be initiated by the EC. This was followed by the decision to task the development of the guidelines to a commission of independent experts, supported by EC staff. High-level expert groups are already a familiar entity within EC policy, and many were instituted before to provide advice and recommendations to the EC, so instituting a HLEG for AI was a natural choice. HLEGs consist of independent experts that are recruited through a selection procedure that is open to all inhabitants of the EU. They typically contain a mix of academics, industry representatives, civil society representative, and others. The EC tends to make sure that major stakeholder constituencies are adequately represented. For example, in the AI HLEG, representatives were present from the IT industry in the EU. The group consisted of 51 members.

The EC provided the AI HLEG with a written mandate, rules of procedure, and a draft plan with timetable for development of the ethics guidelines. As far as we can see, this draft plan was then modified and approved by the HLEG. The HLEG subsequently met a limited number of times in plenary sessions to work on the guidelines (about eight, by our estimate), and also had working groups to develop specific parts. In some of the plenary meetings, stakeholders and experts from outside the

[65] Ibid.
[66] Ibid.
[67] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI", 2019. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines
[68] High-Level Expert Group on Artificial Intelligence, "The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment", 2020. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

HLEG were asked to make presentations and provide input. The AI HLEG also maintained online dialogue with stakeholders through the European AI Alliance, an open, multi-stakeholder online forum engaged in wide-ranging and open discussion of all aspects of AI development and its impact on the economy and society. In December 2019, the AI HLEG presented draft ethics guidelines for public consultation. More than five hundred individuals and organisations responded to the request. The HLEG issued its final report in April 2019.

Our investigations have shown that although a final report was eventually issued, the AI HLEG had great difficulty reaching consensus on the AI HLEG ethics guidelines. This is, in our analysis, due to several factors. First is the compressed schedule in which the guidelines had to be developed. The HLEG did so in nine months, with only intermittent meetings, during which time they also worked on another report, and they received little compensation, meaning that little paid time was available to do the work. The very diverse backgrounds and interests of the group also made consensus difficult. Most members had little knowledge of ethics, ethical principles, or guideline construction, and those who did, the handful of ethicists on the group, often championed different approaches and principles. In addition, there were conflicting interests, with industry representatives advocating a more limited set of guidelines, and others advocating for more stringent ones. Regardless, the guidelines were issued, and gained much publicity, and are now one of the instruments of EC policy. However, the assessment list issued by the group, which was tested by a large number of R&D actors, was criticised for being too long and cumbersome, putting a cloud over its uptake by these actors.

The SIENNA approach to preparing ethics guidelines for human enhancement

Background, scope and limitations

In SIENNA we have developed a number of guidelines, but in this section, we focus on the process of developing ethics guidelines for human enhancement technologies (HET). These guidelines are centred on research in and development and application of human enhancement technologies and procedures, so it is important to note that different factors may apply for guidance for other areas or types of technology, and that such differences may also require different steps. For instance, it is important to note that human enhancement (HE) does not refer to a specific technology or application, but to a wide range of interventions and technologies that aim at improving human beings beyond what is considered typical, or as sometimes problematically referred to as 'normal.' The ethics guidelines we produced are therefore similarly broad, which means that they cover a greater number of technologies and involve a larger number of actors and stakeholders than some other technologies with more limited scope of research or application may require. In what follows we describe how these guidelines were developed, and some of the particular challenges that arose in undertaking this work. In a later section we describe ways in which we sought to operationalise the guidelines, i.e. as practical and actionable. In what follows we have summarised these actions in terms of steps.

Step one: mapping relevant fields

This step fits that identified in the literature review above, namely assessment of whether there is a demand, need, or requirement for ethics guidelines in a particular domain or field, and whether there should be ethics guidelines at this point in time. The guidelines we developed build on extensive prior studies of human enhancement in the SIENNA project (D3.1, D3.4, and D3.7), in which the ethical implications of human enhancement were examined, and the lack of substantial policy and guidelines were discussed (D3.4), and then some methods for promoting ethics were considered, evaluated, and proposed, including the proposals guidelines (D3.7). Bias can be difficult to avoid in the gathering of data, and in the normative analyses, given values can guide actions and decisions. Those who develop guidelines should therefore be transparent about any specific, foundational values that guide their research, and on which any later guidelines rely. Taking all this into account, this step requires a number of further sub-steps that include:

i. *An extensive literature review*, using scholarly material from central and related disciplines. For instance, human enhancement covers many fields in the sciences, social sciences, as well as in the humanities. Each offers different assessments and analyses, and it cannot be assumed which will prove most fruitful. For instance, scientific literature may be focussed on technical possibility, social sciences may cover social and political aspects, and humanities may focus on ethical insight and analysis.

ii. *Mapping of the technological field* takes into account the different kinds of technologies at stake. For human enhancement this means differentiating between those with enhancement as a primary aim, those where enhancement is secondary, those where enhancement is not expected but possible, and those where there is potential, but this is not identified by the researchers or developers. In the latter case, recognising scope for 'dual use' is therefore essential in the mapping process.

iii. *Mapping of legal and governance fields.* Proposals for guidelines should complement existing ethical principles where possible, and adhere to relevant laws, regulations and other legal oversight. Mapping of these areas is essential to ensure there is no contradiction with existing legal and regulatory instruments such as charters, as well as to ensure broad support for the guidelines by aligning them with familiar and accepted existing instruments.

Step two: stakeholder engagement

Where the literature and mapping step indicates a need or requirement for ethics guidelines, a proposal needs to be developed to which stakeholders can provide input and for which their support can be assessed. To achieve this, the following sub-steps have been identified:

i. *Establishing which stakeholders* should be involved. This may involve further mapping of the field. Stakeholders should be involved so as to help ensure that ethical issues are not overlooked, to help arrive at ethical propositions, and to find solutions to ethical issues. They are defined in SIENNA as those who "(1) might be affected by the project; (2) have the potential to implement the project's results and findings; (3) have a stated interest in the project fields; and/or, (4) have the knowledge and expertise to propose strategies and solutions."[69]

ii. *Preparation of proposals*. Consultation can be most fruitful where people are presented with ideas on which to reflect and to prompt discussion. For the SIENNA guidelines, we offered a number of options to stakeholders in advance of a workshop in which methods for ethical guidance were discussed. We list these options below, though further details can be found in the forthcoming deliverable:[70]

Option 1: No ethics guidelines

[69] D1.1, ibid, p. 23.
[70] Erden, Yasemin J. and Philip Brey, D5.3: Methods for promoting ethics for human enhancement, forthcoming.

Option 2: General ethics guidelines

Option 2a: Stand-alone general guidelines

Option 2b: Incorporation of general guidelines into existing guidelines

Option 3: Field-specific guidelines

Option 3a: Stand-alone field-specific guidelines for enhancement

Option 3b: Incorporation into existing field-specific guidelines

Option 4: Dual strategy (general plus field-specific guidelines)

iii. *Consultation*, which can involve participation in meetings, workshops, or in written feedback, with scope for a public engagement stage of consultation to ensure that feedback is not limited only to actors within the same sphere of those who are preparing guidelines. There are a number of challenges with this process of stakeholder involvement, including the need to take into account competing and disparate views. A related challenge concerns how to take into account that those who may be highly organised or present themselves as representatives may not in fact speak for the interests of all stakeholders. Similarly, it is important to avoid presenting groups of people as having uniform views. Binary presentations of positions can obscure the fact that many stakeholders may hold a range of views on different technologies, and that these can also depend on context.

iv. *Discursive evaluation* with agreement sought wherever possible. A challenge here for human enhancement comes from the fact it covers a broad range field, as already noted. Individual stakeholders may have a similarly broad range of expertise, or they may be focussed on narrow areas in HE. The task was therefore to balance judgments about key ethical issues and the enhancement potential of various technologies and techniques, while taking into account the varieties of expertise, interests, and competencies. Development of such methods in HE (and potentially in other new and emerging fields) is particularly hampered by, on the one hand, the low volume of research that is officially characterised as HE research, and on the other by the fragmented nature of HE research, which involves many different techniques, applications and domains.

Step three: gathering consensus

Once some agreement was reached about what guidelines could be produced, and ideas were refined, the next step was to circulate a draft for further consultation as well as to test the likelihood of consensus on key areas. This step required a delicate balance of skills and planning, including ensuring that there was transparency about the foundations for the text that was produced, and maintaining a process that was highly discursive with a range of stakeholders. On the one hand there was a need to be flexible and responsive to input received, while yet not being swayed by minority views that might claim to speak for many. This process included many steps of consecutive proposals where stakeholder input was collected, the text redrafted, and stakeholder input sought again, with further redrafts. This step can continue in the same fashion over many iterations, until a final version achieves sufficient support from target stakeholder groups. We found that discussion helped to tease out when consensus could be reached, and we also found that stakeholders were often keen to offer practical advice and suggestions to improve the guidelines. These amendments greatly helped with gathering further consensus once the next iteration was circulated.

Step four: public consultation

While similar in some ways to the previous steps of stakeholder engagement and consensus gathering, public consultation brings additional challenges. First, participation means that a person identifies themselves as a stakeholder, and this is achieved by virtue of their interest and willing to participate, over and above any particular vested interest, expertise, or likely impact. In these respects, then, engaging the public requires only the third element of stakeholder identity as noted by SIENNA, namely that by the *act* of participation a person expresses their interest in a topic. Whether they are or are not affected by the topic, or have the potential to implement results and findings, or even whether they have knowledge and expertise is only relevant after the fact of this particular action. For the HE consultation, we were interested to know a little more about the people who participated, and so we asked about participant's 'field of expertise', and about their affiliation, yet this information was used to help us gauge the reach of the consultation itself, and if participants chose not to answer fully or truthfully, the system would not reject their submission.

Public consultation, especially for human enhancement, will be impacted by people's understanding and knowledge, their preferences, fears and hopes, as well as by what is being said about a technology at the time of consultation. The ubiquity of a technology in the media can count for more in terms of opinion setting than concrete impacts of such technologies on the people.[71] This means that people may have strong opinions about technologies that don't yet impact them, and which may not have any impact during their lifetime. For instance, what, and how much, human enhancement potential there is in neural prostheses will range depending on who you ask. Companies that promote such technology, such as Elon Musk's Neuralink[72] may be inclined towards hype, and this can feed public consciousness, as well as give hope, or strike fear, among other emotions. Yet neuroscientists with whom we discussed such technologies were far more conservative in their estimates regarding what the technology is likely to achieve. We can expect the public who respond to consultation to have a similarly broad range of views, though we may not always be able to authenticate the rational and evidential foundations for those views, nor even their legitimacy. Uninformed members of the public may be inexplicably afraid of technologies, and such fears may be more or less legitimate or misplaced. Hype from researchers, developers, industry, and marketing, and as plays out in the media, all play a role in the kinds of outcomes that a public consultation can achieve, and human enhancement with its science fiction connotations is particularly susceptible to these kinds of outcomes. Finally, it has to be expected that pressure groups or lobbyists will seek to use public consultations to their own ends, and though we were fortunate to have avoided these outcomes in our public consultation process, this is not a dependable result.

Step five: buy-in

In this step, it is crucial to identify those stakeholders who most concretely answer to points (2) and (4) in SIENNA's account of stakeholders, namely, that they have the potential to implement the guidelines and that they have the knowledge and expertise to propose strategies and solutions. While those in (1), i.e. who might be affected by the project can also be consulted at this stage, it needs to be recognised that such consultation needs to be with the aim of creating the possibility to engage

[71] The same can be said about political decisions that generate very strong views even among people who stand to be affected the least.
[72] https://neuralink.com

with those concrete others, for instance, as a process of networking, and as a way to build momentum and awareness.

Buy-in will only be successful if the first four steps outlined above have been fully and transparently implemented. If stakeholders are unconvinced by the proposals, or by the consultation process and the process of review, or if they have doubts about the legitimacy of stakeholder engagement, for instance, because they think that you have been too narrow in your focus or given too much ground to only one kind of position, then buy-in will be unlikely. For HE, we focussed on achieving maximum practical impact by exploring a range of implementation options. We liaised with EU and international organisations and discussed ways in which the guidelines we had developed could be used. We met with a range of stakeholders including from industry, from The Organisation for Economic Co-operation and Development (OECD), as well as from The Institute of Electrical and Electronics Engineers (IEEE), and the European Commission. We promoted the guidelines to professional organisations and within academic and research networks, including in webinars and conferences, and by offering training within the EC.

2.3 The development of general ethics guidelines

Based on the literature on guideline development, the cases we presented, and our own experiences, we now propose a step-by-step approach for the development of general ethics guidelines. This approach can be used by any organisation or group with an interest in producing general ethics guidelines (henceforth called “the actor”). It is particularly intended for the development of general ethics guidelines for new technologies, but is likely to have broader applicability.

We envision an eleven–step process:

(1) *Establishing rationale*

This is the process of determining whether ethics guidelines are needed for a practice or set of practices, for instance a set of practices related to a new technology.

(2) *Establishing support*

In the first step, the initiating actor has established rationale, but has not yet established that this rationale is supported by key stakeholders. Therefore, during this step, the initiating actor discusses the rationale with key stakeholder and gauges whether there is a need for, and support for, guidelines. This also could include gathering provisional commitments from key stakeholders to be involved in the development process and/or to use the guidelines when completed. During this step, the initial rationale can change, based on stakeholder input, and further constraints and requirements for the guidelines can be determined. The actor could also choose to expand during this step and include additional stakeholders in the team.

(3) *Making a development plan*

During this step, an initial plan is made for developing the guidelines. This plan would define the objective, constraints and requirements in realising this objective, resources needed, an outline of organisational structures and procedures, and timetables with different steps to be taken in the process and points in time when they should be realised. The actor could opt to discuss this plan with stakeholders before it takes its final form.

We recommend the inclusion of a timetable of 9 to 18 months for guideline development in the development plan. Steps 1 to 3 may have already taken several months, meaning that a typical guideline development process takes one to two years.

(4) *Securing resources and doing base organisation*

The development plan contains an assessment of the resources needed for the guideline development process. If those resources have not yet been secured, the next step is to secure them, especially the financial resources. Next, base organisation is undertaken. The plan may specify that there should be a secretariat, or other management structures. It may also specify that there should be a development team in place that is not in existence yet and therefore needs to be formed, or perhaps a stakeholder board, or whatever organisation is chosen for the development process. Stakeholders and experts may need to be recruited at this point. A mandate for the development group may need to be established, as well as rules of procedure. Meetings need to be planned, and an adapted version of the timetable may need to be established.

It is recommended that the in the development team (which could also consist of a core team and stakeholder board), the major stakeholders are adequately represented, and needed expertise is also represented. For emerging technologies, this typically means that key actors are present from industry, academia, civil society, and policy organisations. From industry, both technology developers, deployers and users should be represented. It should also be taken into account that the interests of vulnerable populations are represented. In terms of expertise, it is recommended that sufficient expertise is represented in the team in ethics (including expertise in the development of ethics guidelines), law, social sciences, and policy.

(5) *Collecting information*

After the development team is in place, we recommend that strike initial agreements on procedure, and that they then move on to an information collection stage, in which relevant information is collected for the guidelines. This could consist of previous guideline documents, documents on fundamental ethics principles, constitutional and human rights law documents, relevant ethics cases and ethical analyses, viewpoints and opinions of relevant stakeholders, information on methods and procedures for guideline development, and information on the context in which the guidelines are to operate. The members of the development should read (summaries of) this literature and discuss it with one another. They should also get familiar with each other's backgrounds, viewpoints, and expertise.

(6) *Establishing basic principles and general constraints*

As a next step, it is recommended that the development team reaches agreement on the moral foundations of the guidelines. These could be fundamental ethical principles, including principles enshrined in constitutional and human rights law. The team should establish which ethical principles are relevant to the guidelines and have the support of stakeholders. In addition, it is recommended that the development team establishes additional requirements and constraints for the guidelines. For example, regarding the level of specificity, whether an operationalisation of the guidelines will be included, whether only general guidelines are pursued or also guidelines for specific sectors or domains, or for specific topics, whether methods for conflict resolution will be included, and so forth.

(7) *Interactive drafting of guidelines and stakeholder consultation*

What follows next is an iterative process in which guidelines are drafted, discussed with the team (or working groups within the team), redrafted and then discussed again. During this process, consultations can also be done with external stakeholders, if so desired. This is especially important if the development team is relatively small and does not contain full representation of all relevant stakeholder groups. A workshop format with invited stakeholders is then a good way to proceed. It is desirable that consensus is reached for the guidelines, if not in the entire team, then at least amongst the key stakeholder constituencies. Specific points of attention in the drafting process are that guidelines should be clear and unambiguous, that they have a sufficient degree of specificity, that they have broad stakeholder support, that they are supported by the moral foundations that were established earlier, and that they are amenable to being operationalised so as to provide concrete guidance for key actors.

(8) *Public consultation and final drafting*

Once the development has drafted guidelines that have the support of the team (and of any external stakeholders that were consulted), it is recommended that the draft guidelines are offered for public consultation. Public consultation processes should be open, so that any organisation or individual can participate. They should allow participants to provide evaluations of (different parts of) the guidelines and to offer detailed advice for revisions. The public should be given sufficient time to respond. The team should then review the commentaries, and propose and discuss revisions to the draft, eventually settling on a final version.

(9) *Establishing a mechanism for ownership and revision*

Earlier on in the process, decisions may have already been made on issues like ownership and copyright. If not, this should be done at this stage. It is recommended that guidelines are not subjected to heavy copyright restrictions, so that they can be easily disseminated. At this point, if not done earlier, it is also important to establish revision procedures. Guidelines may require periodic updates. How will this update process proceed, who will be responsible, and how will resources be secured for this process? To the extent possible, this will be established by this point in time.

(10) *Dissemination and implementation*

A strong dissemination effort is recommended after completion of the ethics guidelines. All stakeholders should be informed. It is moreover recommended that the development team provides instructions for using the guidelines. These can be part of the guideline document or be included in a separate document. Members of the development team can also provide or facilitate trainings on using the guidelines. An option is that cases are provided in which the guidelines are applied. Another option is that the development team provides operationalisations of the guidelines, in which the guidelines are translated into more detailed, actor-specific guidelines or instruments (like assessment lists and checklists) to make them more directly usable by specific actors.

(11) *Enforcement*

Ethics guidelines are an instrument of soft law, meaning that they do not have a legally binding force, but nevertheless tend to have the ability to regulate conduct. This ability may be the result of the intrinsic commitment of actors, or of enforcement by other actors with the authority to do so (e.g., employers, funders, governments). Any enforcement of the guidelines will likely not be a responsibility of the development team. The decision likely goes to key stakeholders (e.g., governments, industry associations, companies, universities, research funding organisations) whether they want to set up enforcement mechanisms for the guidelines. Implementation of the guidelines by stakeholders is often implicitly paired with enforcement. For example, if a company, university or research funding organisation implements the ethics guidelines as part of its ethics review procedure, it thereby stimulates that the guidelines are followed by the research and development teams that are subjected to ethics review. Governments can also choose to enshrine important ethics guidelines into law, if they believe that adherence to them is so vital that soft-law provisions are insufficient. Another option is to have policies to make adherence to guidelines attractive to key actors. For example, governments may make it a requirement of public-private partnership that the private partner adheres to ethics guidelines.

2.4 The development of practice-specific ethics guidelines

The development of practice-specific ethics guidelines will, by and large, follow the same step-by-step procedure as for general ethics guidelines. A significant difference, however, is that the group of actors for whom the guidelines are relevant will be smaller, as the guidelines will only pertain to one type of practice, along with the actors associated with it. Possibly, however, the group of stakeholders is equally large, since if a stakeholder is likely to be affected by a technology, he or she is also likely to be affected by a particular practice relating to that technology. For example, if someone is a stakeholder with respect to AI, he or she is also likely to be a stakeholder with respect to the development of AI, because choices in development may well affect him or her.

A different development procedure can be followed if general ethics guidelines are already in existence. In that case, it is to be expected that the practice-specific guidelines will be in line with the general ethics guidelines. For example, if a general ethics guideline states that cognitive enhancement technologies should not impair the potential and capacity for human rationality and independent thought, then it would seem to follow that R&D in cognitive enhancement technologies should not be directed at applications that can impair the potential and capacity for human rationality and independent thought. The correlation between general and practice-specific guidelines will not always be this linear, but one can expect there to be systematic correspondences between them.

We propose that general ethics guidelines for a new technology can provide a good foundation for the development of practice-specific guidelines. For each general ethics guideline, an effort can be made to translate it into a practice-specific guideline. Such a translation requires an understanding of how the specific practice relates to the general guideline. For example, if the guideline states that the privacy of individuals should be respected, or that benefits of the technology should be distributed fairly over society, it needs to be interpreted what this implies for the specific practice to which the practice-specific guidelines pertain. In research ethics guidelines, for example, which relate to R&D practices, the relevant implications could for example be that R&D practice should not include actions in which the technology in question is made capable of violating privacy if used in a normal way, and that the technology should be designed according to principles of universal design, so that different user groups can benefit from it.

There is no reliable formula for this interpretation process, and it is also dependent on one's conception of the responsibilities of the actors engaging in the practice. If, for example, parties were to hold that privacy protection is not a responsibility of technology developers, but of deployers and users, then naturally, it would follow that research ethics for developers need not require guidelines for privacy. Note, in addition, that a mere translation of general into practice-specific guidelines may not be sufficient for the establishment of practice-specific guidelines. Additional guidelines may be needed that are entirely specific to the practice in question. For example, research ethics guidelines for a particular technology may need to include guidelines for human subject research and informed consent, guidelines that have no mirror image in the general ethics guidelines for that technology.

2.5 Operationalising ethics guidelines

In this section we offer a brief statement on the process of operationalising ethics guidelines. By operationalisation, we mean the process of making them more concrete, specific, measurable, verifiable, and action-guiding for specific actors. A frequent critique of ethics guidelines is that they are vague and abstract, and do not provide actors with concrete guidance for action. Therefore, operationalisation can be an important means to ensure practical effect. It should be noted that operationalisation can also come as a cost, as it could result in very detailed instructions that leave limited flexibility for actors. It is therefore desirable to strike a balance as regards the level and degree of operationalisation that is required for specific actors and situations.

We distinguish the following types of operationalisations of ethics guidelines:

(1) *Definition*. This is the process of defining values and other normative concepts. E.g., "Autonomy is self-governance or self-determination. It is the ability to have one's own thoughts and to construct one's own goals and values, and the freedom to make one's own decisions and to perform actions based on them." Definition is a way of translating vague and abstract terms into terms that may be more concrete and familiar.

(2) *Decomposition*. This is the process of going from more general values or principles to more specific ones that can be seen as components or subdivisions. For example, a principle of liberty can be further divided into principles of positive or negative liberty, or into a list of basic liberties (e.g., freedom of speech, freedom of assembly, etc.).

(3) *Means specification.* This is the process of specifying the most important means through which a guideline can be satisfied. For example, the High-Level Expert Group on AI proposes that its principle of diversity and fairness for AI can be satisfied through avoidance of unfair bias, accessibility and universal design, and stakeholder participation.

(4) *Action, actor and practice introduction*. Ethics guidelines may merely specify conditions that are to apply, or events that are to occur, without specifying any actions to be taken or referring to actors who should take these actions. E.g., "There should be adequate privacy protections in place for data subjects" or "Human enhancement requires informed consent". Action introduction is the process by which actions are attached to such guidelines, even if the actors in question remain unspecified. E.g., the first principle can become "Actions should be taken so that adequate privacy protections are in place for data subjects" or "Human enhancements should not be applied without the consent of the subject". Actor introduction is the process by which explicit reference is made to actors in the guidelines. For example, the last mentioned guideline could also be translated into: "Professionals who apply human enhancements to persons should not do so without their informed consent". Practice introduction is the process by which a guideline is related to a concrete practice or set of practices. E.g., the first mentioned principle could be translated to: "Adequate privacy protections should be put in place for data subjects both during development, deployment and use of data systems".

(5) *Actor, actor and practice specification.* This is the process of moving from guidelines that apply to multiple actors to those that apply to a single actor, and from guidelines that apply to broad or multiple practices or action, to a narrower or singular practice action. For example, the guideline that "All AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to AI systems that includes risks to privacy, security, safety and bias" can be translated into a guideline that "developers should implement

risk management procedures during the development process that ensure adequate assessment of privacy, security, safety and bias risks of the developed system, and implement mitigation actions at different stages in the design process". An action or practice can be broken down into component actions or practices. This is what happens, e.g., in our Ethics by Design approach. General guidelines for design practices, like designing systems without algorithmic bias, or with adequate privacy protection, are broken down into specific guidelines for specific steps in the design process.

(6) *Application to cases.* A final good way to make ethical guidelines more operational is to present cases in which they are applied. Seeing them being applied will help actors in applying the guidelines themselves. Inclusion of commentary regarding the application process in the case presentation will further facilitate understanding.

Operationalisation of guidelines is a topic that has been under-investigated, and would benefit from more studies, and efforts towards developing methodologies for operationalisation, and studies of the effectiveness of ethics guidelines in relation to their degree of operationalisation.

2.6 Conclusion

In this chapter, we endeavoured to present a step-by-step method for the development of ethics guidelines. We also discussed ways in which guidelines can be operationalised. We started the chapter with making relevant distinctions, notably between ethics guidelines and ethics codes, and between general and practice-specific ethics guidelines. We then proceeded to review the literature on guideline development and presented several cases. We observed that literature on guideline development is hardly in existence, and therefore proceeded to review development methods in the adjacent area of ethics code. We then presented several instructive cases of code development and guideline development.

Next, we presented a proposal for (general) ethics guideline development. This is a proposal we developed based on the literature review and case analysis, as well as our personal experience with guideline development processes. The proposal comprises eleven steps towards guideline development, with brief instructions and recommendations for each step. We then proceeded to discuss special considerations in the development of practice-specific guidelines and ended with ideas and proposals for the operationalisation of ethics guidelines.

At different points, we observed that methodology for the development and operationalisation of ethics guidelines is in its infancy, and that more studies are needed to bring these processes to a higher level. We present this study as a first attempt to do so and hope that more studies will be undertaken to further advance these topics.

3. Ethics by Design: A General Approach

3.1 Introduction

Ethics by Design is the systematic inclusion of ethical values, principles, requirements and procedures in design and development processes. Traditionally, ethical issues with new technological products have been discovered after their design and introduction to market. Or, when they are being addressed during the development process, they are addressed in a way that developers find difficult to feed into their developed practices. The Ethics by Design (EbD) approach is intended to ensure ethical problems with new products are not generated in the first place, and that ethics can be applied successfully in design and development. This requires specific ethically focused activities at each stage of the design, development and deployment phases of a project. The Ethics by Design approach details how these activities can gain shape.

This document generalizes from the Ethics by Design approach we developed for Artificial Intelligence and robotics.[73] We have considered various other technologies and studied their design processes in them, as well as ethical issues relating to them, and have on this basis generalized the approach to apply to, in principle, any technology.

In section 3.2, we will explain in more detail what an Ethics by Design approach is, and what its benefits are. By means of illustration, we will also give a brief sketch of the Ethics by Design approach we developed for AI. In section 3.3, we then proceed to develop a general EbD approach. We aim to provide a set of step-by-step instructions for developing a dedicated EbD approach for a particular technology field. Each technology field is different, both in its techniques, product and applications, the structure of design and development processes, and the ethical issues that are at play. Section 3.3 provides, in effect, a template for an EbD approach, using generic steps, and a generic model of design, and contains instructions for filling that template for a particular technology. The result of following these instructions is therefore an EbD approach for a technology field, that can be used for the anchoring of ethical considerations to a variety of design processes in that field, for a variety of products. In the concluding section, finally, we summarize our findings, and discuss limitations of this study and suggest further research.

3.2 The Ethics by Design approach

3.2.1 Objectives and core assumptions

Ethics by design rests on the idea that technological designs, whether of a product, system or process, are not neutral but have consequences or effects and even of values "embedded" in them. Designs have consequences built into them in the sense that particular consequences manifest themselves in all of the main uses of a technological product based on it. For example, an ordinary gasoline-powered

---

[73] Brey, Philip and Brandt Dainow, "Ethics by Design and Ethics of Use in AI and Robotics", Annex 2 of Resseguier, Anaïs, Philip Brey, Brandt Dainow, Anna Drozdzewska, Nicole Santiago, David Wright, SIENNA D5.4, Multi-stakeholder Strategy and Practical Tools for Ethical AI and Robotics, 2021. https://www.sienna-project.eu/publications/deliverable-reports/. This is an improved version of Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the development and use of SIS, 2019, Deliverable D3.2 of the SHERPA project. https://doi.org/10.21253/DMU.11316833.

car can be used in a variety of different ways: for leisure driving, for commuter traffic, for cargo transportation, for drive-by shootings, as a shelter for bad weather, as an intentional barricade, etc. While there is no single consequence that results from all of these uses, there are several consequences that stem from many of these uses: in all but the last two uses, gasoline is consumed, greenhouse gases and other pollutants are being emitted, noise is being produced, and at least one person is being transported. This example goes to show that designs and technological products are not neutral in terms of their consequences. Most technological products have only a limited range of (reasonable) uses, and across most of these uses, there are recurrent consequences, which can be both positive (beneficial) and negative (harmful).

Additionally, designs can have *values* embedded in them. Just as technological products can have positive or negative consequences for persons, the economy, or the environment, they can also be beneficial or harmful in terms of the realisation of values. That is, they can have positive or negative effects in terms of the extent to which events and states-of-affairs that are shaped or brought into being accord with specific values, such as freedom and equality. Design, through technological products, can have systematic tendencies to promote or benefit values, as well as tendencies to harm or detract from them. For example, a technological product like a surveillance camera has a systematic tendency, across its different uses, to bring about harms to privacy and benefits to security. The approach to technology that associates technological products and their designs with values embedded in them is called the embedded values approach.[74]

The idea that technology is not neutral and that values can be embedded in design, is at the heart of various approaches to design that have been developed in recent decades. Although there were predecessors who advocated the idea, Batya Friedman and her associates were arguably the first to propose a design approach based on the concept of embedded values.[75] They developed the *value-sensitive design* approach, which tries to comprehensively account for and include human values in the design process. Value-sensitive design advocates studies into values, designs, contexts of use and stakeholders, with the aim of designing products that embody and balance the values of different stakeholders. The term *design for values* is sometimes used to refer to a broader family of design approaches that utilise the concept of value embeddedness.[76] Key activities include (1) the identification of stakeholders and the benefits and harms for these stakeholders as a result of the product that is to be designed, (2) the mapping of benefits and harms onto corresponding values, conceptual investigations of key values, and the identification of possible value conflicts and the suggestion of solutions for them, and (3) investigations of how properties of the product to be designed may promote or hinder human values and how the product may be designed proactively in order to support particular values that have been found important during the conceptual investigations.

---

[74] Nissenbaum, Helen, "Values in the Design of Computer Systems", Computers and Society 28 (1), 1998, 38–39.
[75] Friedman, Batya, and David G. Hendry, Value Sensitive Design: Shaping Technology with Moral Imagination, Cambridge, MA: The MIT Press, 2019; Friedman, Batya, David G. Hendry, and Alan Borning, A Survey of Value Sensitive Design Methods, Boston Delft: Now Publishers Inc, 2018; Friedman, Batya, Peter Kahn and Alan Borning, "Value Sensitive Design and Information Systems", in P. Zhang and D. Galletta (eds.), Human-Computer Interaction in Management Information Systems: Foundations, Armonk, NY: M.E. Sharpe, 2006, pp. 348-372.
[76] Hoven, Jeroen van den, Pieter E. Vermaas, and Ibo van de Poel, "Design for Values: An Introduction", in Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel (eds.), Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains, Dordrecht: Springer Netherlands, 2015b, pp. 1–7.

Most design for values approaches are not constrained to strictly moral values and include aesthetic, economic and social values, amongst others. When they are focused on the moral values, we can use the term *ethics by design*. An *ethics by design* approach aims for the systematic inclusion of ethical values, principles, requirements and procedures into design and development processes.

The reason why we need ethics by design is that ethical issues with technological products are traditionally only discovered after these products have been put on the market and start to cause harm. Ethics by Design tries to ensure that ethical problems are not created in the first place. This requires specific ethics-focused activities at each stage of the design, development and deployment phases of a project. An overview of a general framework for such activities is provided in the next subsection and tested in subsection 3.2.3.

3.2.2 Overview of the approach

The general Ethics by Design approach advocated by SIENNA can be described in a five-layer model. This model is similar to many others in computer science in that higher levels are more abstract, with increasing levels of specificity going down the levels. The layers are as follows.

1. **Ethics by Design Values**. These are the primary ethical values by which we want to guide the ethical status of an AI or robotics system. Where a system violates these values, it may be considered unethical. Values are to be upheld and enhanced. Privacy and fairness are examples of such values.
2. **Ethical Requisites**. Ethical requisites are the conditions that a solution or application must meet in order to achieve its goals ethically. In Ethics by Design, ethical requisites are instantiations of values within AI and robotics systems. Values may be instantiated in many ways; through functionality, in data structures, in the process by which the system is constructed, and so forth. For example, one way the value of fairness can be instantiated as an ethical requisite is to require that a system does not exhibit racial bias. Asimov's Three Laws of Robotics are an example of ethical requisites.
3. **Ethics by Design Guidelines.** Whereas ethical requisites were concerned with the system, guidelines are concerned with the steps by which it is created. Ethics by Design works on the basis that there are steps in the development process which are common to all design methodologies. The Ethics by Design approach offers a generic description of these phases in the development process and maps the ethical requisites onto these phases. This yields specific guidelines (usually formulated as tasks) at each phase which ensure that the final system instantiates the ethical requisites and therefore does not violate any ethical values. For example, the guidelines state that during the data gathering stage, data should be screened for fairness and any discriminatory biases that are found should be corrected.
4. **Methodologies.** There are a variety of design methodologies used in any technology field. They are, at least partially, distinguished by the manner in which they organise the development process. Each methodology offers its own steps and sequence. Here Ethics by Design maps its principles onto the components of each individual methodology. If a project is using a different methodology, the researcher should return to the generic model (see section 3.3). By mapping the steps in the generic development process to their own methodology, they can allocate each guideline to the appropriate steps in their methodology.
5. **Tools & Methods**. The Tools and Methods layer accommodates specific programmatic artefacts and processes deployed within the development process to undertake Ethics by Design. It is possible some could be specific to a particular methodology and inapplicable to others, but at this stage, those which have emerged in the development community are tuned to ethical requisites

and useable under any methodology. For example, Datasheets for Datasets[77] are employed to interrogate the ethical characteristics of data, and so can be used at any stage which works with that data and for any norm relating to data. They can thus be deployed at multiple stages of the development process and are methodology-neutral.

3.2.3 Example: Artificial intelligence

This subsection details the application of the model presented in subsection 3.2.2 to the field of AI. To begin, for step 1, the primary ethical values by which we want to guide the ethical status of an AI or robotics system were established as:

- Human Agency;
- Privacy and Data Governance;
- Fairness;
- Well-being;
- Accountability and Oversight; and
- Transparency.

These ethical values were arrived at using the *Ethics Guidelines for Trustworthy AI* of the EU High-Level Expert Group on AI and other guidelines as a foundation.

Then, for step 2, we developed "ethical requisites" for AI systems on the basis of these values. Ethical requisites are the conditions that an application must meet in order to achieve its goals ethically. For example, for the value of *fairness*, these include:

- *Avoidance of algorithmic bias:* AI systems should be designed to avoid bias in both input data and algorithm design.
- *Universal accessibility:* AI systems should be designed so that they are usable by different types of end-users with different abilities.
- *Fair impacts:* Applications should demonstrate that possible social impact on relevant groups has been considered and what, if any, steps will be taken to ensure the system does not cause them to be discriminated against or stigmatized, or otherwise have their interests affected in a negative way.

For step 3, we developed from these ethical requisites sets of ethical guidelines to be followed at different stages of the design, development and deployment of the system. For this, we first established a model of what the typical design process of an AI system looks like.

The six main tasks in this generic model are:

1. **Specification of objectives.** The determination of what the system is for and what it should be capable of doing.

[77] Gebru, T., J. Morgenstern, B. Vecchione, J.W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford, "Datasheets for Datasets", ArXiv:1803.09010 [Cs], March 19, 2020.

2. **Specification of requirements.** Development of technical and non-technical requirements by which to build the system, including initial determination of required resources, together with an initial risk assessment and cost-benefit analysis, resulting in a design plan.
3. **High-level design.** Development of a high-level architecture. This is sometimes preceded by the development of a conceptual model.
4. **Data collection and preparation.** Data must be collected, verified, cleaned and integrated.
5. **Detailed design and development.** The actual construction of a full working system.
6. **Testing and evaluation.** Testing and evaluation of the system.

We subsequently described these phases and devised ethical guidelines to be followed during each of them. These guidelines are concrete tasks which must be performed in order to achieve the ethical requisites that were arrived at in the previous step. For example, the guidelines state that during the data collection and preparation stage, data should be screened for fairness and any discriminatory biases that are found should be corrected.

In step 4, the Ethics by Design principles need to be mapped onto the components of specific design methodology used in a given AI project. Three of the most important methodologies include Agile, CRISP-DM and V-model.

Finally, in step 5, the Tools and Methods layer accommodates specific programmatic artefacts and processes deployed within the development process to undertake Ethics by Design. For example, in AI projects, Datasheets for Datasets[78] may be employed to interrogate the ethical characteristics of data, and so can be used at any stage which works with that data and for any norm relating to data.

3.3 Developing Ethics by Design for a technology field

Before an Ethics by Design approach can be used in a specific technology field, it first must be developed. In principle, this is a one-time process, after which the approach can be used in a multiple of design processes. An Ethics by Design approach is normally defined in relation to an established design methodology. For example, an Ethics by Design approach for the design of big data systems is defined in relation to a particular methodology for their design, such as CRISP-DM or Agile. It may be possible to define a so-called generic method for the design of big data systems, a high-level description of the design process that fits multiple design methodologies that are in existence for the development of these systems. And subsequently, an Ethics by Design approach may then be defined in relation to such a generic method. However, this approach to developing Ethics by Design will lack some of the specificity that designers will need when they use a particular design methodology like Agile or CRISP-DM.

Therefore, in this section, we describe a five-step process for the development of an Ethics by Design approach that is defined in relation to a particular technology and in relation to a particular design methodology in the field. This process can be applied to a wide variety of technological fields (e.g., biomedical engineering, computer engineering, chemical engineering) and create approaches that are applicable to all kinds of design objects (e.g., techniques, products, processes, socio-technical systems, components).

[78] Gebru, T., J. Morgenstern, B. Vecchione, J.W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford, "Datasheets for Datasets", ArXiv:1803.09010 [Cs], March 19, 2020.

In subsection 3.3.1, we present an overview of our five-step approach. In subsection 3.3.2, we detail steps 1 and 2 of the approach (Defining values and ethical requisites). In subsection 3.3.3, we expand on step 3 (Choosing and describing a design methodology). In subsection 3.3.4, we offer details on step 4 (Developing operational guidelines). And, finally, in subsection 3.3.5, we flesh out step 5 of the approach (Developing additional tools and methods and special topics).

3.3.1 Developing an Ethics by Design approach for a specific technology field

The development of an Ethics by Design approach that is defined in relation to a particular technology field and in relation to a particular design methodology used in that field, is a five-step process that goes as follows:

Step 1: Reach consensus on the key moral values and principles that apply to the technology field.

These are the primary moral values by which we want to guide the ethical status of a technological product. Privacy and fairness are examples of such values. Where a product violates these values, it may be considered unethical. To uncover which moral values are important in any given technological field, one should study the existing ethics and popular media literature on the field, discuss with ethics experts in the field, utilize ethics checklists, and consult their own intuitions. In addition, one should apply stakeholder methods, e.g., inquiring with stakeholders which values they consider of prime importance in relation to the technological field. The moral values should be conceptualised and described in terms of their most important aspects in relation to the technological field.[79]

Step 2: Derive ethical requisites (or norms) from these values.

Ethical requisites are the conditions that a solution or application must meet in order to achieve its goals ethically. For each moral value a set of ethical requisites needs to be devised that fully ensures realisation of that value in the context of the technology in question. There is no single clear-cut method to arrive at these ethical requisites, but one can decompose the values into their constituent parts and specify the means through which they can be satisfied. In addition, one can simply brainstorm about ways in which the technology and its uses can strengthen or undermine adherence to the moral values.

Step 3: Choose and describe an established design methodology for the development of technology in the technology field.

One should describe the established methodology for the development of technological products in the field, and the resulting description should distinguish different steps or phases in the design process. One can describe methodologies such as V-model, Agile or CRISP-DM, or variants thereof, depending on what is being used most. In describing the methodology, one can draw from our description of the generic engineering design model containing six phases, which may roughly apply to most engineering projects (see subsection 3.3.3).

Step 4: Develop operational ethics guidelines that involve a translation of the ethical requisites to actionable methodological guidelines.

[79] Note that it will not be easy to attain consensus over what are the relevant values and principles. The development of general ethics guidelines for a technology field is normally an elaborate, multistakeholder process. See chapter 2 for an approach to the development of ethics guidelines for technology fields. Note, also, that values may change over time, not least due to technological change, so that ethics guidelines need to be updated regularly.

For each of the different steps of the design methodology described in the previous step, operational ethics guidelines should be created, guiding the ethical process of these steps. Creating such guidelines involves a translation of the ethical requisites of step 2 to actionable methodological guidelines. General guidelines that are applicable to almost any engineering project are presented in subsection 3.3.4, and can be copied into the specific methodology one is using. In addition to these general guidelines, it is necessary to create more concrete guidelines for this task that are specific to the technological domain of the engineering project. These guidelines ensure adherence to the general ethical requisites devised for the technological domain.

Step 5: Develop tools, methods and special topics.

Additional tools and methods may need to be developed, or tuned to the ethical requites, to undertake Ethics by design. For example, methods and tools need to be in place for proper stakeholder engagement, managing resources Ethics by Design, and training in using the approach.

Furthermore, “special topics” guidelines should be created that deal with the most important technologies, techniques, artefacts within the technological field that raise specific ethical issues. These guidelines would form an addition to the general guidelines for the technology based on the ethical requisites. We propose that they also include guidelines for the development of products for particular application domains that are prominently associated with the technological field.

3.3.2 Defining values and ethical requisites (step 1 and 2)

We can reliably uncover the primary moral values by which we want to guide the ethical status of a technological product through a variety of methods. An important source can be the existing ethics literature on the technology, including existing ethical guidelines, in which many moral values may already be identified. Also, one can rely their own ethical analysis, mostly based on their moral intuitions and mainstream methods of applied ethics. Furthermore, one can also consult outside ethics experts to help identify moral values that might otherwise be missed.

In addition, bibliometrics can be used to identify and examine relevant debates that were taking place in different national, geographic, and linguistic communities. Careful searches of popular media in carefully chosen locales can uncover concerns and associated moral values that may not be in the mainstream international discussion or in the awareness of the researchers operating in the field.

Another method is the systematic consideration of checklists of standard ethical issues or moral values. This is a method prescribed by the ATE approach to ethical analysis of emerging technologies.

Finally, stakeholder consultations and workshops with ethicists, technologists, futurists, and other members of the public may uncover unnoticed moral values. Consultations with members of the public regarding technology acceptance and uptake may also provide insights. For this step, less heavily structured interactions, especially with opportunities for creative input and interaction, may be helpful.

Once the primary moral values have been identified, it is time to derive the ethical requisites (or ethical requirements) from these values. Ethical requisites are the conditions that a solution or application must meet in order to achieve its goals ethically. For each moral value a set of ethical requisites needs to be devised that fully ensures realisation of that value in the context of the technology in question.

There are no proven method to arrive at these ethical requisites, but one can apply methods of decomposition and means specification to the moral values. Decomposition is the process of going from more general values to more specific ones that can be seen as components. For example, a principle of freedom can be subdivided into principles of positive or negative freedom, or into a list of basic freedoms (e.g., freedom of speech, freedom of movement, freedom of assembly).

Means specification is the process of specifying the most important means through which a value can be satisfied. For instance, the High-Level Expert Group on AI proposes that its principle of diversity and fairness for AI can be satisfied through avoidance of unfair bias, accessibility and universal design, and stakeholder participation.

Additionally, one can consider the range of potential products resulting from the technology, as well as their use context, and brainstorm about ways in which the technology and its uses can strengthen or undermine adherence to the moral values.

3.3.3 Choosing and describing a design methodology (step 3)

Ethics by Design is premised on the idea that development processes for an engineering product or system can be described by means of a design methodology that is used in the field. This can be a specific methodology such as V-model, Agile or CRISP-DM. Presented below, however, is generic engineering design model containing six phases, which may roughly apply to any engineering project. One can draw from this model in describing their own methodology. While the six phases are presented in a list format, this is not necessarily a sequential process.

The six tasks of the generic model are:

1. ***Product planning*** (or *specification of objectives*). The determination of what the product or system is for and what it should be capable of doing.

2. ***Specification of requirements*** (or *task clarification*). Development of technical and non-technical requirements by which to build the product or system, including initial determination of required resources, together with an initial risk assessment and cost-benefit analysis, resulting in a design plan.
3. ***Conceptual design****.* The development of one or more conceptual designs for the product or system.
4. ***High-level design*** (or *embodiment design* or *preliminary design*)*.* The development of a high-level design for the chosen conceptual design.
5. ***Detailed design and development****.* The actual construction of a full working product or system.
6. ***Testing and evaluation****.* Testing and evaluation of the constructed product or system.

There may be some variation between different technology fields in terms of their adherence to this generic model. For example, data-intensive information technology sometimes omits the conceptual design task (or combines it with high-level design), and usually has a "data collection and preparation" task following the high-level design task, during which data must be collected, verified, cleaned and integrated.

In some cases, a feasibility study may be conducted during the specification of requirements stage (and/or concept design stage). This can be a discrete task within the design process. In such a study, an evaluation and analysis is performed of the potential for the proposed project to reach a successful conclusion.

In some engineering design models, where significant amounts of time and resources are put into the production of prototypes, a separate prototyping task is added after the concept design stage.

Let us now further describe each of the six main phases in the generic model for design.

Product planning

During the product planning stage, an idea for a new product is developed to meet on a (real or perceived) need which has been expressed by a client or is thought to exist in the market. This idea for a new product leads to a task description for an engineering department to develop the product. This task description provides the product's objectives – what the product is for and what it should be capable of doing. Oftentimes, product planning is not done by designers, but by clients and/or the product planning department and/or marketing departments of companies.

Specification of requirements

Requirement specification is the process of determining the kind of product that is needed and important resources for creating it in terms of requirements and constraints. The primary functions include (1) arriving at a development plan that includes design specifications for the system, (2) designing the development infrastructure, (3) determining staff resources required, (4) setting milestones and other deadlines and so forth. Product planning and specification are often integrated processes in which there is some movement back and forth between the two.

Determining design requirements and performing requirement analysis is one of the most important activities in the design process This task is often conducted at the same time as a feasibility analysis (if one is conducted). The requirements guide the design of the product or system throughout the various phases of its development. These include functions, attributes and specifications, which are (in part) gleaned from the objectives and user needs which were established at the product planning stage. Design requirements often include parameters in terms of hardware, software, maintainability, availability, testability and so forth.

Conceptual design

Conceptual design is the process of finding a solution at a conceptual level to the problem outlined by the specification of requirements. Conceptual design involves identifying essential problems through abstraction, establishing structures in which top-level functions are broken up into subfunctions, and searching for appropriate working principles which can drive the subfunctions and combine them into working structures.[80] Solutions can be found through ideation, which is the mental process by which ideas are generated. Ideation techniques include (amongst others) *brainstorming*, *trigger words*, *morphological analysis*, and *synetics*.[81]

The result of this stage is called a "design concept" or "principle solution". The conceptual design stage can result in multiple concepts, of which one is usually chosen as the basis for further development at the high-level design phase. This choice is made by comparing the respective (expected) performance of each of the concepts in terms of how well they adhere to the specified requirements. Such comparative analysis can be done by (for example) plotting performance values for each of the concepts in a web-diagram.

Prototyping is often used to test concepts. Prototyping is the creation of inexpensive, scaled-down versions of a product (or specific features thereof), so that problem solutions generated at an earlier stage can be investigated.[82]

---

[80] Brey, Philip, "Engineering Design: Its Nature and Its Social, Political and Moral Dimensions", undated.

[81] Daly et al., "Comparing Ideation Techniques for Beginning Designers". Element Method, 2016.

[82] Brey, Philip, "Engineering Design: Its Nature and Its Social, Political and Moral Dimensions", undated.

High-level design

High-level design is a phase during which a design concept is turned into a definitive layout of the product or system proposed. It bridges a gap between design conception and detailed design, and involves developing a layout design in which the general configuration and spatial features (in terms of schematics, diagrams, etc.) of the product are determined. Additionally, it is during this phase that component shapes, materials and production processes are defined in broad terms.

Importantly, the high-level design phase involves technical and economic considerations and results in a design which can be comprehensively checked for its functionality, durability, production and assembly requirements, operation and cost, and its adherence to the design requirements. High-level design can involve several repeat design processes before a definitive design emerges.

During the detailed design and development phase, the parameters of the product or system being created may change, but the high-level design focuses on generating the general framework on which to build the engineering project.

Detailed design and development

In the detailed design and development phase, the high-level design of the previous phase is completed with final instructions before production. These concern shapes, forms, dimensions and properties of components, a definitive selection of materials, a final specification of production methods, operating procedures and costs, and the development of production documents which include component and assembly drawings and parts lists.[83] Oftentimes, *computer-aided design* (CAD) programs, such as solid modelling programs, are extensively used at this stage to create a highly detailed final design of a product or system. CAD programs can provide various benefits. For example, they can provide optimization to reduce a product's volume without diminishing its quality. They can also calculate stress and displacement by means of the finite element method[84] to evaluate stresses throughout the product.

In addition to providing a complete description of the physical aspects of the product or system itself, the detailed design phase may also involve the creation of assembly instructions, transportation documentation, quality control measures for the production department, and operating, maintenance or repair manuals for users.[85]

Testing and evaluation

Testing is the assessment of the performance, safety, quality or compliance with standards of a designed product or system, subsystem or component.[86] Testing can be done at earlier stages of the design process through prototyping, but is often done with a fully realized product, subsystem or component.[87] Testing can be done by the developers or by specialists, and may happen in dedicated testing environments.

Some engineering design models, such as V-model, place significant emphasis on testing and evaluation. These models may advocate testing the product or system at various levels (e.g., system

[83] Ibid.
[84] Reddy, J. N., "Introduction to the Finite Element Method", 2005.
[85] Ibid.
[86] Ibid.
[87] Ibid.

component testing, testing the system as a whole, testing against the system requirements) and providing multiple feedback loops to earlier design stages.

3.3.4 Developing operational guidelines (step 4)

For each of the different steps of the design methodology described in the previous step, operational ethics guidelines should be created, guiding the ethical process of these steps. Creating such guidelines involves a translation of the ethical requisites of step 2 into actionable methodological guidelines. General guidelines that are applicable to almost any engineering project are presented below, and can be copied into the specific methodology one is using. In addition to these general guidelines, it is necessary to create more concrete guidelines for this task that are specific to the technological domain of the engineering project. These guidelines ensure adherence to the general ethical requisites devised for the technological domain.

Product planning

While each project is unique, Ethics by Design lays down a set of standardised requirements which all engineering projects should meet. An important first step is to ethically assess the objectives of a development project against the ethical requisites. Sometimes, objectives are unethical, or even illegal.

The ethics guidelines for the product planning phase

- Assess whether the objectives for the design project will meet the relevant ethical requisites. It is recommended that a professional ethicist, if available, is enlisted to do the assessment of objectives, in collaboration with members of the development team. Potential violations differ in their degree of seriousness. Some violations may be only potential or less serious. Such concerns do not mean the objective should be abandoned, but that concrete steps will have to be taken to avoid the system becoming unethical.
- If your project has external stakeholders, such as researchers in other fields who will use the system, it is important you plan to include them in the specification of objectives and specification of requirements phases. In particular, stakeholders may be aware of wider ethical issues which could arise from the use of the system. Stakeholders should be consulted regarding what ethical issues they believe are at stake and how they should be dealt with. Stakeholders should be appropriately diverse (gender, age, ethnicity, etc.). In this way, an appropriately diverse range of ideas and preferences will inform design choices.
- When assessing objectives, consider the potential for intentional misuse. Where possible, modify the system's objectives to reduce such potential. If the potential misuse is significant, conduct a social risk assessment. This should outline the risks, the elements of the design needed to mitigate this and any procedures which may be required to reduce this risk once the system is deployed and operational.

Determining technology-specific guidelines

In addition to these general guidelines, one can devise more concrete guidelines for this task that are specific to the technological domain of the engineering project. These guidelines ensure adherence to the general ethical requisites devised for the technological domain.

One can create such technology-specific guidelines by determining which of the ethical requisites are important when ethically assessing a project's objectives. Guidelines should then be formulated so as to test the objectives for compliance with the relevant ethical requisites.

General ways to make to turn ethical requisites into concrete guidelines include dividing the ethical requisites into their component parts, specifying any *actions* to be taken in order to realise the ethical requisites, referencing *actors* who should take these actions, and relating the requisites to a concrete *practice* or set of practices. For example, "Adequate privacy protections should be put in place." could thus become "Action X should be taken by person X so that adequate privacy protections are in place for data subjects during development, deployment and use of the product."

Specification of requirements

The primary function of the requirement specification phase is to arrive at a development plan that includes design specifications for the system, design the development infrastructure, determine staff resources required, set milestones and other deadlines and so forth. Most organisations have a standardised set of development tools used for all projects. The organisational and management structures and procedures are usually tuned to these tools, as are the development methodologies. Changing these can be more challenging than building systems. Nevertheless, it cannot be assumed that any tool, process or organisational elements will support Ethics by Design. Some of the ethical requisites present new problems during development. For example, requirements for human oversight and audit may impose a need to document many internal processes to a greater degree than has previously been the case. It must therefore be recognised it is possible development methods, tools or even organisational structures used on previous projects will need modification. As a result, it must be recognised there is likely to be a need to adapt (or even replace) aspects of customary development systems so that they become capable of delivering the project's ethical requisites.

In some cases, it may not be technically possible to meet every ethical requisite due to lack of suitable tools. However, one should be extremely rigorous when searching for suitable tools. The degree to which a technical inability to meet the ethical requisites blocks a project also depends on the particular ethical requisite in question and the system's functionality.

General ethics guidelines for the specification of requirements phase

- Once a complete design plan has been produced, an ethical impact assessment may be justified in order to assess specific ethical risks in the development, deployment and use of the product (on the basis of proposed design specifications, constraints, selected resources and infrastructure). A professional AI ethicist, if available, should be able to perform such an assessment. Ethical impact assessment should be planned and budgeted for. This assessment should be scaled to the nature of the project, the severity of ethical risks and the overall budget of the development project. A standard for ethical impact assessment is available at: https://satoriproject.eu/media/CWA17145-23d2017.pdf.

- As a part of the ethical impact assessment, steps should be carefully planned to mitigate any unavoidable ethical risks. This risk assessment should be updated at later points in the development process as more information comes in.

- One should ensure that relevant ethical requisites are covered in the list of design specifications. For this purpose, consider creating an Ethical Requisites document. At the objectives phase this document will only cover ethical aspects of the overall system and the most obvious features of the development process. However, it can be refined and added to as the project proceeds.

- It is recommended that at this stage an *EbD implementation plan* is completed. This will specify future steps to be taken in order to incorporate EbD into the development process, the actors responsible for carrying them out and who is responsible for monitoring them. This implementation plan should incorporate the ethical risk and impact assessment, if one has been completed. We recommend, in addition, that it includes an ethical compliance architecture embedded into the development infrastructure and a set of organisational structures and procedures. The ethical compliance architecture will need to focus on tools and processes at the developer or engineer level, but may also need mechanisms for external communication from end-users (or other stakeholders) during testing and evaluation. Secondly, we recommend inclusion of an ethical governance model which includes organisational structures for governance of the EbD process, including, most likely, ethical review committees. The governance model needs to address the following issues: How will governance be exercised? What is the project's version of a supervisory authority to ensure the ethical requisites are met? What powers will it have? How will it be selected fairly and inclusively? What procedures will be used in cases of a conflict between the ethical governance authority and developers, engineers or clients? The governance mechanisms should formally detail the steps which have to be taken in order to incorporate EbD into the development process, the actors responsible for carrying out EbD-related tasks and those who monitoring this.

Conceptual design

Conceptual design is the process of finding a solution at a conceptual level to the problem outlined by the specification of requirements. At this stage, it is important that the ethical requisites are paid close attention to and that stakeholder involvement is considered.

Ethics guidelines for the conceptual design phase

- Ensure that relevant ethical requisites are taken into account during concept *development*. The ethical requisites should be treated just the same as any other requirements for the system.
- Ensure that relevant ethical requisites are taken into account during concept *selection*, and that while doing so, their importance is not marginalised in favour of the technical requirements.
- Since concept selection has an important impact on the design trajectory going forward, it is recommended that stakeholder involvement takes place during this phase. Stakeholders, especially those in vulnerable groups, can be invited to test prototypes and/or voice their opinions on some or all of the concepts and/or prototypes. Stakeholder input should be taken seriously in the final selection of a concept for further development.

Determining technology-specific guidelines

In addition to these general guidelines, one can devise more concrete guidelines for this task that are specific to the technological domain of the engineering project. These guidelines ensure adherence to the general ethical requisites devised for the technological domain.

On can create such technology-specific guidelines by determining which of the ethical requisites are important in ethically assessing a project's (1) conceptual design process and (2) its resulting concept(s). Guidelines should then be formulated so as to test these for compliance with the relevant ethical requisites.

General ways to make to turn ethical requisites into concrete guidelines include dividing the ethical requisites into their component parts, specifying any *actions* to be taken in order to realise the ethical requisites, referencing *actors* who should take these actions, and relating the requisites to a concrete *practice* or set of practices.

High-level design

High-level design is a phase during which a design concept is turned into a definitive layout of the product or system proposed. Like during the concept design phase, it is important that careful attention is paid to adherence to the ethical requisites during this phase.

Ethics guidelines for high-level design phase

- To integrate ethical requisites, ensure that ethical requisites are communicated to all engineers and/or designers at this stage.
- Ethical requisites should be treated just the same as any other requirements for the system. During this phase adherence of the high-level design to the ethical requisites should be regularly checked, especially when relevant decisions are made.

Determining technology-specific guidelines

In addition to these general guidelines, one can devise more concrete guidelines for this task that are specific to the technological domain of the engineering project. These guidelines ensure adherence to the general ethical requisites devised for the technological domain.

One can create such technology-specific guidelines by determining which of the ethical requisites are important in ethically assessing a project's (1) high-level design process and (2) its resulting high-level design. Guidelines should then be formulated so as to test these for compliance with the relevant ethical requisites. The guidelines can recommend a more detailed testing for compliance with respect to certain ethical requisites (than during previous stages), since important design decisions are being made as the design is taking shape.

Detailed design and development

In the detailed design and development phase, the high-level design resulting from the previous phase is completed with final instructions before production. As with the high-level design phase, it is important that proper attention is paid to adherence to the ethical requisites during this phase.

Ethics guidelines for this design phase are the following:

- To integrate ethical requisites, ensure that ethical requisites are communicated to all engineers and/or designers at this stage.
- Ethical requisites should be treated just the same as any other requirements for the system, such as reliability. During this phase adherence of the high-level design to the ethical requisites should be regularly checked, especially when relevant decisions are made.
- Quality control measures for the production department should be designed in a way that ensures that all products leaving quality control adhere to the ethical requisites.
- Operating, maintenance and repair manuals for users should be designed in a way that maximises the probability that products maintain their adherence to the ethical requisites during operation and after maintenance and repair. Explicit warnings about unethical consequences of certain improper procedures can be made in operating, maintenance and repair manuals.

Determining technology-specific guidelines

As was the case with high-level design, one can devise more concrete guidelines for this phase that are specific to the technological domain of the project. These guidelines ensure adherence to the ethical requisites devised for the technological domain.

By determining which of the ethical requisites are important in ethically assessing a project’s detailed design and development process one can create more specific guidelines. These guidelines should be formulated so as to test for compliance with the ethical requisites. This may call for in-depth testing for ethical compliance now that the design is being finalised in great detail.

Testing and evaluation

Testing is the assessment of the performance, safety, quality or compliance with standards of a designed product or system, subsystem or component. It may be that the system achieves its functional requirements but not all ethical requisites. If this is the case, the system cannot be considered to have been successfully completed. However, the whole point of Ethics by Design is to avoid such an outcome. If rigorously applied, the Ethics by Design approach should prevent ethical issues at this stage of the development process.

Ethics guidelines for the testing and evaluation phase

- As part of the testing and evaluation phase, you should use the project’s ethical requisites document to design a testing regime which can check the system’s ethical compliance. It is highly unlikely any standard testing regime will consider all of the system’s ethical requisites, so the choice of testing methodology is important here. Implement this testing to determine whether the system meets all of its ethical requisites.
- Treat departures from the system’s desired ethical characteristics just as seriously as a reliability flaw and undertake remedial work until the product meets its ethical requisites.
- It is highly recommended that stakeholder involvement takes place during this phase. Stakeholders, especially those in vulnerable groups, can be invited to take part in the testing (insofar as this can expected to be done safely) and/or give their opinions on the product. Stakeholder input should be taken very seriously in any potential remedial work on the product.

Determining technology-specific guidelines

One should devise more concrete testing guidelines that are specific to the project. These guidelines ensure the finished product meets the ethical requisites devised for it.

Tests for ethical compliance can be developed by determining which of the ethical requisites are most important. Guidelines should then be formulated so as to test these for compliance with the relevant ethical requisites. The guidelines can specifically recommend a more in-depth testing with regard to the final product’s effects on users and other stakeholders, especially those in vulnerable groups.

3.3.5 Developing additional tools and methods and special topics (step 5)

Additional tools and methods may need to be developed, or tuned to the ethical requites, to undertake Ethics by design. For example, methods and tools need to be in place for proper stakeholder engagement, managing resources Ethics by Design, and training in using the approach.

Furthermore, “special topics” guidelines should be created that deal with the most important technologies, techniques, artefacts within the technological field that raise specific ethical issues. These guidelines would form an addition to the general guidelines for the technology based on the ethical requisites. We propose that they also guidelines for the development of products for particular application domains that are prominently associated with the technological field (e.g., healthcare, defence, law enforcement, entertainment). For new technologies, it is often known that they will be used in particular application domains, or at least it can be foreseen that they will be. Ethics guidelines for technologies in relation to application domains guide development choices that for the mitigation of ethical issues that can occur in particular domains.

The recommended use special topics guidelines is that designers standardly apply the general guidelines as devised for the field in their development practices, but that they also determine whether their practices also includes one or more techniques and artefacts listed in the special topics section. If so, then these technology-specific ethics guidelines should be applied as well.

To find out which technologies, techniques and artefacts within the technological field raise specific ethical issues, one can cross-reference lists of such technologies, techniques and artefacts with the lists of ethical requisites for the technological field. Once the important technologies, techniques and artefacts have been identified, one can create guidelines for them. One can place the ethical requisites in the context of the technologies, techniques and artefacts, and specify any actions to be taken in order to realise the ethical requisites, reference actors who should take these actions, and relate the requisites to a concrete practice or set of practices.

3.4 Conclusion

In this chapter, we presented a general approach for Ethics by Design, that works for all technology fields. We developed this approach in two sections. In section 3.2, we laid out the objectives and core assumptions of Ethics by Design and gave an overview of the approach. We explained that technological designs, whether of a product, system or process, are not neutral but have consequences or effects and even values "embedded" in them. We detailed how this realization led to creation of the *value-sensitive design* approach and a broader family of approaches called *design for values*, which try to include human values in the design process. We then made the case for the *Ethics by Design* approach, which aims for the systematic inclusion of ethical values, principles, requirements and procedures into design and development processes. Subsequently, in section 3.2.3, we presented an overview of the approach and used AI technology as an example to apply the approach. We explained how Ethics by Design approach advocated by SIENNA can be described in a five-layer model. The first layer is about establishing the primary ethical values by which we want to guide the ethical status of a technological product. The second layer is about creating ethical requisites, which are the conditions that a solution or application must meet in order to achieve its goals ethically. The third layer involves operationalising the ethical requisites into ethical guidelines at each phase of the design process. The fourth layer involves mapping the Ethics by Design principles onto the components of any particular design methodology. And the fifth and final layer accommodates for methods and tools deployed within the development process to undertake Ethics by Design.

In section 3.3, we described a five-step process for the development of an Ethics by Design approach that is defined in relation to a particular technology and in relation to a particular design methodology in the field. We established that this process can be applied to a wide variety of technological fields (e.g., biomedical engineering, computer engineering, chemical engineering) and create Ethics by Design approaches that are applicable to all kinds of design objects (e.g., techniques, products, processes, socio-technical systems, components). The first step in the process, it was argued, is to define the key moral values that apply to the technology field. Consulting existing ethics and popular media literature, ethics experts, checklists, and one's own intuitions are useful in uncovering these values. The second step is to use these values to derive ethical requisites for the technological field. One could do so by decomposing the values into their constituent parts and specifying the means through which they can be satisfied. In addition, one could simply brainstorm about ways in which the technology and its uses can strengthen or undermine adherence to the moral values.

The third step is to choose and describe an established design methodology for the development of technology in the technology field. One can describe methodologies such as V-model, Agile or CRISP-DM, or variants thereof, depending on what is being used most. The fourth step is to develop operational ethics guidelines that involve a translation of the ethical requisites to actionable methodological guidelines for each step in the design process. General ways to do this include dividing the ethical requisites into their component parts, specifying any *actions* to be taken in order to realise the ethical requisites, referencing *actors* who should take these actions, and relating the requisites to a concrete *practice* or set of practices. And finally, the fifth step is to develop additional tools and methods to undertake Ethics by design, including for stakeholder engagement, managing resources Ethics by Design, and training in using the approach. In addition, in this final step, "special topics" guidelines should be created that deal with the most important technologies, techniques, artefacts within the technological field that raise specific ethical issues.

This is a preliminary proposal for a general EbD approach. While we have considered the fit of our approach to several technology fields, including AI, robotics, software engineering, human genomics, biochemical engineering and biomechatronics, we realize that more studies and more testing is needed to ensure that our general approach works optimally for different technology fields. We will eventually have to say more about differences between technology fields, and how relevant features of specific technologies will be accounted for in our approach. In addition, we will need to develop instructions for the implementation of EbD approaches in concrete design projects. We hope that next to ourselves, others will take these ideas forward and will develop this promising approach further.

4. Engaging policy-makers in projects on ethical and human-rights aspects of new and emerging technologies: tips for successful engagement

4.1 How EU ethics and/or new technology research projects engage with policy makers

Ethics and human rights research projects on new and emerging technologies can help policy-makers further conceptualise, clarify[88], draw insights, and form opinions on issues and impacts related to such technologies. SIENNA carried out a brief review[89] of selected EU research projects to identify how projects have, and are engaging with policy-makers. The objective of this review was to determine, mainly, the target policy-makers, how the projects engaged with them, the strengths of their approach and the challenges and limitations faced. Projects were selected based on their focus on ethics, human rights and new and emerging technologies. The projects were: EASI-GENOMICS[90] (H2020), EGAIS (FP7-SIS)[91], ETICA (FP-7-SIS)[92], IRISS (FP7)[93], PANELFIT (H2020)[94], SHERPA (H2020)[95], SIENNA (H2020)[96], and SATORI (FP7).[97] The projects were selected from the SIENNA partner networks, SwafS[98] cluster and using a search on CORDIS with key words 'ethics/human rights'+policy+technology. Insights drawn are presented below.

Target policy-makers

In the projects studied, target policy-makers are, in some cases, broadly defined and can range from international, EU-level to national (depending on the topic). In most cases, and given the European nature of the projects, targeted policy-makers are from the EU-level, e.g., European Parliament, Council, European Commission (e.g., DG Research and Innovation, DG Connect), and other European Union Agencies and bodies (e.g., European Agency for Fundamental Rights[99], AI HLEG[100]). National-level targeting is also evident through members of the project consortia, where projects carry out

[88] Gauttier, Stéphanie, Robert-Jan Geerts, Sinan Senel, Michael Nagenborg, "Building capacities for policy and industry outreach", Report, 4TU, November 2017.
[89] The review of projects was carried out in early 2021 using desktop research. Some requests for information were also sent out. SIENNA acknowledges inputs received from Prof. Dr Bernd Stahl, De Montfort University (ETICA).
[90] https://www.easi-genomics.eu/home
[91] https://cordis.europa.eu/project/id/230291/reporting
[92] https://cordis.europa.eu/project/id/230318
[93] http://irissproject.eu
[94] http://panelfit.eu
[95] https://www.project-sherpa.eu
[96] https://www.sienna-project.eu
[97] https://satoriproject.eu
[98] Science with and for Society, Horizon 2020.
[99] https://fra.europa.eu/en/about-fra
[100] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

national studies, and/or good connections exist, for example with the national embassies or attachés or national policy-makers. Projects also engage with data protection authorities and other regulatory agencies (e.g., law enforcement agencies, surveillance commissioners, or human rights institutions). Engaging with national level policy makers is harder when the partners involved are not embedded or do not have direct connections and prior expertise in engagement in the policy and/or regulatory sector. Based on our review, we find a lack of in-depth focus on specific policy-makers to be targeted – which policy-makers should a project primarily focus on to bring about the greatest change and impact based on its results and findings? Greater attention should be paid to this aspect.

Forms and tools of engagement

The projects analysed adopted different strategies and tools to engage policy-makers – sometimes structured, more often ad-hoc, needs-based or opportunistic. Some examples[101] include:

- Policy watch/monitoring via a dedicated task/work package[102];
- Inputs to existing bodies or committees, e.g., The European Group on Ethics in Science and New Technologies (EGE), European Parliament;[103]
- Inputs to open public consultations;[104]
- Letters/emails to Members of Parliament (MEPs);
- Newsletters;[105]
- Participation in policy-making activities at the request of policy-maker[106];
- Policy recommendations and policy briefs[107]
- Public debates;
- Training, mobilisation and mutual learning activities;
- White papers; and
- Workshops and events (targeted at policy-makers, hosted by policy-makers[108] or where policy-makers participate[109]).

Strengths identified

---

[101] Publicly available references are listed for some of the examples.
[102] E.g., https://satoriproject.eu/work_packages/policy-watch-and-policy-recommendations/
[103] See IRISS coordinator statement at the hearing of the Committee on Civil Liberties, Justice and Home Affairs of the European Parliament on "Electronic Mass Surveillance of EU-Citizens". https://www.youtube.com/watch?v=qaiVupd1mbM
[104] See e.g., https://www.project-sherpa.eu/commentary-on-the-european-parliament-committee-on-legal-affairs-draft-report-with-recommendations-to-the-commission-on-a-framework-of-ethical-aspects-of-artificial-intelligence-robotics-and-related/; https://www.sienna-project.eu/news/news-item/?tarContentId=902450 ; https://www.sienna-project.eu/news/news-item/?tarContentId=902358;
[105] See e.g., https://satoriproject.eu/media/D9.2_The-consortium-newsletter.pdf
[106] See PANELFIT. https://cordis.europa.eu/project/id/788039/reporting
[107] https://satoriproject.eu/publication_type/policy-briefs/; https://researchportal.unamur.be/en/publications/european-policy-brief-the-ethical-governance-of-emerging-technolo/activities/
[108] E.g., European Parliament STOA events. See ETICA. https://cordis.europa.eu/project/id/230318/reporting
[109] E.g., Computers, Privacy, Data Protection Conference. https://www.cpdpconferences.org

The studied projects benefitted from having **strong and/or well-developed links** with policy-makers via the European Commission as the research funding body. This has enabled, for example, feeding research results into policy discussions on new technologies, regulatory developments and legislative amendments, or changes in ethics processes. Other strengths include having a **work package focussed on policy monitoring**, including task on **advocacy, including policy-makers in project** consortia and/or Advisory Boards, **setting up community platforms** for information exchange, **systematically identifying policy-makers** as part of stakeholder identification and analysis activities. Another strength was the **presence of MEPs on a project's Stakeholder Board**, which enabled the project to directly connect with the European Parliament and ensure MEPs interest in and knowledge of the work of the project. Having partners in the consortium with **members well-versed in policy engagement, capable of rising to the task and willing to engage with policy-makers** is a bonus.

Challenges and limitations

The studied research projects have faced and/or continue to face many challenges in engaging with policy-makers. Our review identified the following major challenges:

- Lack of structure and ad-hoc nature of policy engagement activities;
- Late start of policy activities (back-ended) which affects resource allocation;
- Lack of experience within research teams in carrying out advocacy activities;
- No topical interest, political priorities and developments focussed elsewhere;
- Unequal interest of policy-makers;
- Lack of/limited resources to carry out policy activities efficiently;
- Policy engagement as an after-thought and not pre-planned activity;
- Inadequate monitoring of relevant policy activities; and
- Inability to control the timing of policy inputs (i.e., not always aligned with research timelines)[110]
- Inability to translate research results into concrete policy recommendations.

In addition to this, as highlighted by the SATORI project report, "It is difficult to make an impact at the highest level of policy-making due to some barriers that exist, e.g., policymakers are not used to endorsing or using research that is carried out under a participatory approach."[111] This is something that should also be considered when understanding and addressing the challenges of engaging with policy-makers on the ethics of new and emerging technologies – which itself calls for a highly participatory research and collaboration approach to identify and address societal impacts of such technologies.

[110] Noted also Gauttier, Stéphanie, Robert-Jan Geerts, Sinan Senel, Michael Nagenborg, "Building capacities for policy and industry outreach", Report, 4TU, November 2017.

[111] Shelley-Egan, Clare, David Wright, Rok Benčin, Jelica Šumič Riha, Gregor Strle, Daniela Ovadia, Adelina Pastor Cañedo, Christine Angeli, Menelaos Sotiriou, "SATORI Deliverable D2.1 Report (handbook) of participatory processes", July 2014, p.32. (in relation to the GAP2 project) https://satoriproject.eu/media/D2.1_Report-handbook-of-participatory-processes_FINAL1.pdf

4.2 Tips for successful engagement with policy makers

This section presents tips for ethics and human rights projects on new and emerging technologies for engaging with policy-makers[112]. Alongside drawing from the above research, we examined some key documents to draw out (a) general tips, (b) tips on 'how' to engage and (c) key performance indicators to measure policy impact. The documents examined and from which this section deeply draws are the UKRI IGF's 'How to engage with policy makers: A guide for academics in the arts and humanities"[113], the SATORI Report (handbook) of participatory processes[114], Phoenix, et. al.'s *Creating and communicating social research for policymakers in government*[115] and Oliver and Cairney's *The do's and don'ts of influencing policy*[116]. We have adapted the tips to fit the ethics and human rights context in new and emerging technologies. These tips should be read and appropriately applied in specific contexts.

(A) GENERAL TIPS

- Do **high-quality** research.
- Understand the **full range of individuals, groups, key actors** and different **processes** involved in policy making on ethics and human rights – who are the key players and who do they talk to?
- Be **aware of the political context** – how does the research fit in with current thinking on the issue? How is it relevant?
- Communicate in ways that policy makers find useful – consider your audience and **make practical recommendations**.
- **Develop and maintain networks** – become visible and make connections with people who share a policy interest in ethics and human rights, both in person and online.
- As the expert, be prepared to **share your general knowledge** of the subject and your **specific research**.
- Know **how to express** what ethics and human rights has to offer to aid decision-making and problem-solving (along with identifying the ethical and human rights issues)
- Adopt a **long-term perspective** – be open-minded and patient to engage successfully.

[112] Note the tips presented here are also more broadly applicable in a variety of contexts.
[113] UKRI, Institute for Government, "How to engage with policy makers: A guide for academics in the arts and humanities", 2020. https://www.instituteforgovernment.org.uk/publications/how-engage-policy-makers https://www.instituteforgovernment.org.uk/publications/how-engage-policy-makers
[114] Shelley-Egan, op. cit., 2014.
[115] Phoenix, J.H., L.G. Atkinson, and H. Baker, "Creating and communicating social research for policymakers in government", *Palgrave Commun* 5, 98, 2019. https://doi.org/10.1057/s41599-019-0310-1
[116] Oliver, K., P. Cairney, "The dos and don'ts of influencing policy: a systematic review of advice to academics", *Palgrave Commun* 5, 21, 2019. https://doi.org/10.1057/s41599-019-0232-y

- **Engage with influencers**, e.g., civil society, think tanks, lobby and interest groups, charities and non-governmental organisations (NGOs), industry bodies, media – conventional and social, constituents, political parties, academia, international organisations, national human rights institutions (NHRIs), governments.
- **Time it right –** identify, monitor and make the most of opportunities presented.

(B) HOW TO ENGAGE

- Map out all the people and organisations with influence in the relevant policy areas, weighing up the potential for influence.
- Analyse the current political landscape around the issues in your research. Consider the political cycle and current preoccupations of policy makers – what could make your ethical/human rights issue get public attention?
- Engage potential collaborators and policy influencers at project design stage and keep them interested by regularly sharing results/exchanging information on ethical and human rights issues. Engage multiple times and not just once for greater impact.
- Include policy makers to facilitate required adaptations of policies, revisions of regulations, expediting the standardisation process etc in project work and activities such as workshops, Advisory Boards, Networks.
- Be ‘accessible’ and invest resources (time and personnel) to develop and maintain relationships with policymakers
- Build relationships and be persistent in finding ways to contribute – find out what is of interest to policy-makers and relate the research to their interests (which will then open further avenues to contribute)
- Stay connected with people and organisations who may provide policy ideas when the ethical/human rights issue rises to prominence
- Utilise ‘windows of opportunity’, e.g., open public consultations on policies, regulations, standards, strategic policy interest in a topic.
- Make research accessible outside academic paywalls e.g., deposit it in an open access repository, write accessible publications.
- Write up research in clear, easily understandable formats that are useful to the people you want to engage with. Adapt the message, presentation format and communication styles to the target policy-maker.
- Ensure recommendations are practical and realistic – considering resource constraints, government priorities and current political narratives.
- When meeting people, consider what they will be most interested in and tailor conversations accordingly.

- ➢ Attend events where you will find people with a possible interest in your research. Think broadly about who might be useful to engage with – many groups have influence that are not directly involved in policy making themselves.
- ➢ Set up or join cross-disciplinary working groups to get different perspectives on common research problems and access a wider range of contacts.
- ➢ Offer to run trainings for policy-makers or implementers
- ➢ Develop policy relationships before applying for funding and include collaborative activities in your proposal.
- ➢ Use social media such as Twitter as a newsfeed and interact on topics that interest you. Tweet about your own work and that of your contacts where relevant.
- ➢ Write articles for blogs and websites.
- ➢ Look for current events that the research gives a unique perspective and could be used as a hook to share results and recommendations.
- ➢ Put a description of your research activities online on your project website in an easily format that is easily understandable to non-academics, clearly stating areas of expertise.
- ➢ Seek support from your organisation's engagement or communications teams for valuable advice and resources.
- ➢ Where possible, be generous with your time and open-minded about what might produce impact in the long-term.
- ➢ Be prepared to explain how you can help – several times if necessary.
- ➢ Seek out colleagues with an engagement role or strong policy engagement experience, to ask for advice or assistance.
- ➢ Consider recruiting an "engagement fellow"/impact officer with experience in policy and advocacy to focus purely on impact.
- ➢ Reflect on your engagement activities and follow-up with people to see what has happened as a result.

(C) KEY PERFORMANCE INDICATORS (KPIs) TO MEASURE POLICY IMPACT

Below is a list of qualitative and quantitative measures drawn from studied literature[117] and project team insights that will help assess the policy impacts of research projects on ethics of new and emerging technologies. Measuring policy impact has its own challenges, but is useful to carry out nonetheless, to ensure that policy activities are on the right track, meeting their objectives and making desired impact.

[117] E.g., Gauttier et al, op. cit., 2017.

- Number and type of policy makers on the project contact list and/or Stakeholder Board
- Number of direct contacts made
- Number of materials prepared/presented to/accessed and/or downloaded by policy-makers, e.g., policy briefs, policy reports, position papers, blog posts.
- Number of interactions with policy makers e.g., evidence to Committees, interviews, emails.
- Attendance of policy-makers at project events
- Number of policy consultations and evidence sessions to which the project contributed
- Inquiries to project website/coordinator/team.
- Number of visits to the project policy page
- Trainings delivered to policy makers
- Referencing of project reports/quotes in international, regional or national policy reports and documentation.
- Consideration/inclusion of findings in new policies, regulations, funding programmes and guidance (direct referencing of project, results)
- Requests for project information and reports.

5. General approach on how RECs can support ethics in new emerging technology research

5.1 The role of RECs within new emerging technology research

Research projects involving human participants are in most countries reviewed by research ethics committees (RECs) which are multidisciplinary and independent groups of individuals whose role is to help researchers in carrying out research projects in an ethically sound way. RECs have to ensure that the autonomy, rights and dignity of the human participants involved in a research project are protected and safeguarded. Beside this RECs aim to support and facilitate the conduct of valuable research in general.[118] Most RECs are established to review health-related research projects. They are organized on a national level[119], but also on local and regional levels, e.g., at universities and hospitals.

A review by a REC is legally mandatory for all clinical trials in European countries. In many countries similar structures exist for medical research projects in general. Outside the medical field reviews by RECs are normally not legally binding. Nevertheless, RECs are established in other health-related research areas and also in the humanities, social and behavioural sciences. In these fields ethics assessment by RECs is often required by journals before publication and by financing bodies before funding.

Within the scope of this report, we want to examine what role RECs can play when it comes to ethically sound research on and with new emerging technologies. To answer this question, we need to define what new and emerging technology research projects are. The term "emerging technology" is widely known in the academic world and nevertheless difficult to define. If a technology is "emerging" depends on the subjective view of single persons or groups. Someone may see a technology as emerging, and others may take a different view. Therefore, it is difficult to come to a general definition. Rotol et al. identified five attributes that feature in the emergence of novel technologies: (i) radical novelty, (ii) relatively fast growth, (iii) coherence, (iv) prominent impact, and (v) uncertainty and ambiguity, and defined emerging technologies as:

> "a relatively fast growing and radically novel technology characterised by a certain degree of coherence persisting over time and with the potential to exert a considerable impact on the socio-economic domain(s) which is observed in terms of the composition of actors, institutions and the patterns of interactions among those, along with the associated knowledge production processes. Its most prominent impact, however, lies in the future and so in the emergence phase is still somewhat uncertain and ambiguous".[120]

In a nutshell this means that new emerging technologies can be understood as technologies that are still in development and this means that the outcome is still uncertain. Examples for emerging technologies are artificial intelligence (AI), Internet-of-Things, quantum computing, synthetic biology,

[118] https://eneri.mobali.com/research-ethics-committees-main-tasks-and-challenges
[119] For an overview on national RECs in Europe see the EUREC homepage: http://www.eurecnet.org/information/index.html
[120] Rotolo, Daniele, Diana Hicks and Ben Martin, "What is an Emerging Technology?", *Research Policy*, Vol. 44, Issue 10, 2015, pp.1827-1843. DOI: 10.1016/j.respol.2015.06.006, p. 1834.

3D printing, and smart materials.[121] Emerging technologies can also be new medical technologies or technologies under development in neuroscience, e.g., DNA vaccination, research on body implants, robotic surgery or brain-computer interface technologies. This means emerging technologies are developed in many different academic fields, like computer and information sciences, engineering sciences, natural sciences, medical sciences and others. A lot of emerging technology projects are already in focus of the work of RECs, especially health-related projects that involve human research participants. However, there are also a lot of new and emerging technology research projects that are not discussed by RECs yet. Especially for those research projects that do not involve human research participants it is often not obvious why a REC should be involved. Could RECs help ensure that such projects and their outcomes are ethical? What role could RECs have? Examples for such emerging technology projects are the *development of AI and smart information systems*, *research on the internet of things* and *developing smart cities.* Such projects are not involving human research participants. Nevertheless, they can have huge impact on humans and their autonomy, privacy and safety. If it is within the remit of RECs to support and facilitate the conduct of valuable research in general, as described earlier, they should play a role. In the following, we will show how RECs can be built in a way that they can support all types of research projects.

5.2 General approach on how to build a REC

RECs are well established to review health-related research and also in the humanities, social and behavioural sciences RECs exist. However, in some research fields it is difficult for researchers to find a REC that fits well to review their research projects. Therefore, either new RECs need to be established or the old ones need to be extended. The experiences gained by medical RECs during the last decades can serve as basis for building new RECs or to adjust the already established ones. Suggestions for the composition of RECs can be found in different guidelines developed for the health-related work of RECs.[122] In general the following criteria for the composition of RECs can be drawn (based on work already done for SIENNA task 5.1):

- RECs need to be **multidisciplinary**. Members of RECs need to come from different disciplines.
- REC members should display an appropriate **balance of scientific expertise, philosophical, legal and ethical backgrounds.**
- The members of a REC should **collectively** have the qualifications and experience to review and evaluate a research proposal.
- RECs must invite **non-members with expertise** in special areas for assistance if needed.
- **Community members** or representatives and/or laypersons need to be members of a REC with equal standing.

[121] See also Brey, Philip, Brandt Dainow, Yasmin J. Erden, Owen King, Philip Jansen, Rowen Rodrigues, Anais Resseguier, and Amal Matar, D6.1 Generalised Methodology for Ethical Assessment of Emerging Technologies, SIENNA Project, page 13. Forthcoming at https://www.sienna-project.eu/publications/deliverable-reports/.

[122] For more information, see Tambornino, Lisa, Dirk Lanzerath, Philipp Hoevel, Tom Lindemann, D5.1: Report Documenting Elements to Open and Complement Operational Guidelines for Research Ethics Committees, SIENNA Project, 2021.

- When a proposed study involves **vulnerable** individuals or groups representatives need to be members.
- Woman and men should be represented equal (**gender balance**).
- No recommendation can be made about the exact number of members. However, a REC should not have too many members, otherwise organisational problems may arise.
- RECs need to be **independent**. Committee members need to be **resistant to bias and conflict of interest** or opt out in the event of conflict of interest. Procedures on how to declare a conflict of interest must be fixed.
- The discussions RECs have in their board meetings must generally be **confidential**.
- REC members should be **trained regularly** relevant to their role in the REC.
- RECs already established in the health-related fields can help to facilitate the development of RECs for other fields.
- RECs may be **established at universities, ministries, research institutions**, but also in **private companies**, like AI companies.
- Procedures for the **appointment of REC members** and the **duration of membership** should be clearly fixed.
- Clear procedures are also needed for making use of the committee, as well as for efficient processes by which the committee does its work (including structure of meetings and timeline for review).
- Researchers need to be aware that RECs issue statements, not research permits.
- RECs must provide information to researchers about the RECs policies and procedures.
- RECs are only one piece of the puzzle when it comes to ethical research on new technologies. RECs need to **work closely together with other ethics bodies and ethics advisory boards**. Different aspects of a research project could be reviewed by different bodies. Especially risks assessment of new emerging technology projects cannot be carried out without the help of other ethics bodies.

These criteria for a good composition of a REC can apply not only for the engagement of RECs in health-related research, but also for REC work in all other research fields. The **challenge**, though, is to define which disciplines need to be represented for specific research fields and specific research projects. Moreover, another challenge arises when it comes to community engagement and the involvement of laypersons. Defining the relevant community and/or stakeholders for a specific research project is not easy. It is often unclear what persons or groups will be affected by a specific research project. This is especially challenging when it comes to emerging technology research projects that do not involve human research participants directly but could have an immense impact for humans. Humans are involved as users of technology or otherwise affected by the development or deployment of a technology, and this means that their autonomy, privacy and safety and other ethical values may be violated.

The following additional criteria for the composition of RECs need to be added to meet these challenges (also presented in **figure 4**):

- RECs need to be constituted by a **core committee** and an **extended committee.**
- Members of the **core committee** must ideally be **philosophers trained in ethics or ethicists from other disciplines** and **individuals who understand how scientists and innovators think**. These individuals may be scientist themselves or they may be actors that have a first-hand experience of the kind of work they are supposed to discuss.
- The core committee must discuss and decide together what **expertise** is needed to advise on the specific research area. The committee needs to be extended with the appropriate **experts**.
- The core REC together with the invited experts need to discuss which **community/stakeholders/stakeholder groups** could be affected by the research project. Citizen representatives that could bring in the **public view** need to be invited as REC members.
- The REC needs to work closely together with the researcher, but also with other **ethics bodies, ethics advisory boards** and **data protection officers**.

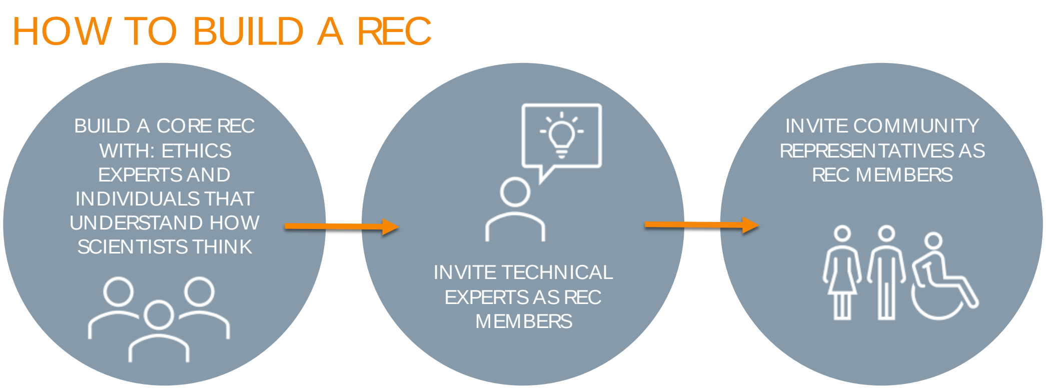

Figure 4: *Building a REC with a core committee and an extended committee*

5.3 General approach on how guidance for RECs should be structured

As already outlined above two types of emerging technology research projects have to be distinguished for our purpose: research with human participants and research without human participants. Whilst for the first type already RECs exist, for the second type RECs are relatively new and rare.

Research on new emerging technologies which include human participants are within the scope or similar to health-related research projects that are approved by many RECs in many different countries. Standards and procedures for RECs on how to give researchers advice in such projects exist.

Some examples are the Declaration of Helsinki[123], the guideline for good clinical practice E6(R2)[124], the CIOMS International Ethical Guidelines for Health-related Research Involving Humans[125] and the Guide for Research Ethics Committee members from the Steering Committee on Bioethics[126]. These and other relevant documents provide guidance for RECs on:

- Informed consent
- Fair selection of participants
- Inclusion in research of vulnerable participants
- Research with children
- Risks and harms for participants
- How to minimize risks
- Risk assessment
- Data protection and privacy
- Storage of biological material
- Social and economic value
- Risks and benefits for the society

This guidance is particularly relevant for research involving human participants. For the ethical assessment of research projects on emerging technologies without human participants, the guidance offered in health-related guidelines alone is not sufficient.[127] The existing guidance needs to be supplemented with additional information that addresses the specific ethical challenges associated with some research projects on emerging technologies. But what should this general guidance for RECs ideally look like? The challenges are field-specific, and therefore it is not useful to create one big guide for RECs on how to review research on emerging technologies - this would be an incredibly long and impractical document. Developing a field-specific guide for RECs is also not practical as this would be a huge task and again the documents would be too long and impractical. In order for RECs to quickly and easily find the information, it makes much more sense to develop a decision tree that should be constantly updated with the latest findings. The European Network of Research Ethics Committees (ENERI) started with the development of such a decision tree.[128] The ENERI decision tree aims to help

[123] World Medical Association (WMA), "WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects", 2013. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

[124] European Medicines Agency (EMA), "Guideline for good clinical practice E6(R2)", 2016, p. 10. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf

[125] Council for International Organizations of Medical Sciences (CIOMS) and World Health Organization (WHO), "International Ethical Guidelines for Health-related Research Involving Humans", 2016. https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf

[126] Steering Committee on Bioethics (CDBI), "Guide for Research Ethics Committee Members", 2012. https://www.coe.int/t/dg3/healthbioethic/activities/02_biomedical_research_en/Guide/Guide_EN.pdf

[127] We analysed this in Tambornino, Lisa, Dirk Lanzerath, Philipp Hoevel, Tom Lindemann, D5.1: Report Documenting Elements to Open and Complement Operational Guidelines for Research Ethics Committees, SIENNA Project, 2021.

[128] https://eneri.eu/decision-tree/

RECs and researchers to think about ethical questions and challenges that might arise during a planned research project. Guidance on ethical principles and points to consider, but also on relevant guidelines, codes and other helpful references for many different research areas are given. For instance, researchers doing research with human participants can click on the appropriate button and will come to the following aspects to be considered:

- How will the informed consent be articulated (language, clearness, completeness, age appropriate)?
- What is the state of the research participants (able to give consent, not able to give consent, minors, vulnerable groups etc.)?
- Are the participants patients or healthy volunteers?
- What are the possible benefits for the participants?
- What are the possible risks and burdens for the participants?
- How is a clarification about voluntariness of participation and opportunity of withdrawing at any time communicated?
- Will a compensation be paid and is this an incentive to participate in the study?
- What is the type of information asked or investigated about and how sensitive is this information (personal, life style, health etc.)?
- How is personal data protected (data protection: anonymisation, pseudoanonymisation etc. / data used in interviews, videos, tapes etc.) and who is responsible for them?
- How will the process of recruitment of volunteers be organised?
- What are the criteria of inclusion and exclusion of human participants?
- How is the balance of power between the researcher and the human participant addressed?

For research in AI and engineering the following aspects could be relevant for RECs to consider:

- May the newly created technology harm organisms, the environment, or human beings?
- To what extent are the technologies attuned to the needs, aspirations and views of society? How are risks and benefits for society balanced?
- What kind of impacts do the developed technologies have on society, also on a global scale?
- May the developed technologies be subject to secondary or dual use?
- What kind of societal impact will a potential industrial implementation of the results have?
- Are there any culturally or socially controversial or sensitive impacts (privacy, property rights, justice)?
- Are aspects of access and benefit sharing carefully considered?

- ➢ Does the research have an international impact and may conflict with international regulations?
- ➢ Are all data protection and privacy aspects considered?
- ➢ Are all necessary approvals stored in the records?
- ➢ Does this research involve the use of methods that may cause harm to humans including research staff? Are alternative safer mechanisms considered?
- ➢ Are possible unforeseen long-term-effects or side-effects on the environment, public health, public safety, and/or future generations considered?
- ➢ Are there any economic interests in particular when there are co-operations with industry planed?

The ENERI decision tree offers a lot of more helpful guidance for RECs. Further work on structure, content and also on the technical realization are necessary in order to make the tool even more practical. We suggest continuing the work started within the ENERI decision tree in other EU funded projects in order to develop a general tool giving guidance for RECs.

5.4 What should guide the RECs decision-making?

The work of RECs is often guided by principles. The principles used for ethical reviews of health-related research projects with humans are respect for persons namely autonomy, beneficence, non-maleficence and justice. These principles are very general and in biomedical research using these principles can be a good basis. The question is whether these principles are sufficient to review new emerging technology research projects, regardless of the scientific field. This question was already analysed by the SATORI project and published in the CEN workshop agreement.[129] The SATORI project concluded that in "different scientific fields, different special conditions may arise, and with differing frequency. In addition, fields may include field-specific methods, approaches, practices and conventions that also necessitate field-specific principles."[130] There is a general discussion on what principles are most important when it comes to concrete emerging technology research and some suggestions have been made. For the field AI for instance the EU funded SHERPA project identified nine key ethical principles that include privacy, autonomy, freedom, dignity, safety and security, justice/fairness, responsibility/accountability, well-being (individual, societal and environmental) and transparency.[131]

We suggest adopting the four basic principles for the review of all research projects, regardless of the scientific field. However, beside autonomy, beneficence, non-maleficence and justice, also social and environmental well-being need to guide the RECs decision-making. This is especially important with view to new emerging technology projects with high-socioeconomic impact.

---

[129] European Committee for Standardization (CEN), "Ethics assessment for research and innovation - Part 1: Ethics committee", Workshop Agreement CWA 17145-1, May 2017. ftp://ftp.cencenelec.eu/EN/ResearchInnovation/CWA/CWA1714501.pdf, p.19.

[130] Ibid, p. 19.

[131] Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, Guidelines for the development and use of SIS, 2019, Deliverable D3.2 of the SHERPA project. https://doi.org/10.21253/DMU.11316833

We would like to conclude our remarks with a comment on the role of RECs. Researchers conducting research on new emerging technologies outside the medical field should see the work of RECs as an offer to help them conduct ethically valuable research. They should not necessarily see RECs as a review body (as they are for medical research with human participants). The collaboration between RECs and researchers should ideally help researchers to become "good" researchers in a moral sense. This means researchers with moral sensitivity for ethical issues, important ethical values and good thinking. Together, RECs and researchers need to find out what ethical issues of new and emerging technology research projects with high socio-economic impact are and how to solve or avoid them before starting a research project.

References

AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI", 2019. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines

Alix, *Working Ethically At Speed*, Medium, 2018.

Anderson, Philip, and Michael L Tushman, "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change", *Administrative Science Quarterly*, 1990, pp. 604–633.

Beck, Kent, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, … David Thomas, *Manifesto for Agile Software Development,* AgileManifesto, 2001.

Brey, Philip and Brandt Dainow, "Ethics by Design and Ethics of Use in AI and Robotics", Annex 2 of Resseguier, Anaïs, Philip Brey, Brandt Dainow, Anna Drozdzewska, Nicole Santiago, David Wright, SIENNA D5.4, *Multi-stakeholder Strategy and Practical Tools for Ethical AI and Robotics*, 2021. https://www.sienna-project.eu/publications/deliverable-reports/

Brey, Philip, "Anticipatory Ethics for Emerging Technologies", Nanoethics 6, 1–13 (2012).

Brey, Philip, "Engineering Design: Its Nature and Its Social, Political and Moral Dimensions", undated.

Brey, Philip, "Ethics of Emerging Technologies", in Sven Ove Hansson (ed.), *The Ethics of Technology: Methods and Approaches, Philosophy, Technology and Society*, Rowman & Littlefield International, 2017, pp. 175–192.

Brey, Philip, "Research Ethics Guidelines for Artificial Intelligence", Annex 4 to Resseguier, Anaïs, Philip Brey, Brandt Dainow, Anna Drozdzewska, Nicole Santiago and David Wright, SIENNA D5.4, *Multi-Stakeholder Strategy and Tools for Ethical AI and Robotics*, forthcoming.

Brey, Philip, "Research Ethics Guidelines for the Engineering Sciences and Computer and Information Sciences", in Kelly Laas, Elisabeth Hildt, and Michael Davis (eds.), *Codes of Ethics and Ethical Guidelines: Emerging Technologies, Changing Fields*, Springer, Dordrecht, Netherlands, 2021.

Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, *Guidelines for the development and use of SIS,* 2019*,* Deliverable D3.2 of the SHERPA project. https://doi.org/10.21253/DMU.11316833

Brey, Philip, Björn Lundgren, Kevin Macnish, and Mark Ryan, *Guidelines for the Ethical Use of AI and Big Data Systems*, SHERPA project, July 2019.

Brey, Philip, Brandt Dainow, Yasmin J. Erden, Owen King, Philip Jansen, Rowena Rodrigues, Anais Resseguier, and Amal Matar, *D6.1 Generalised Methodology for Ethical Assessment of Emerging Technologies*, SIENNA Project. Forthcoming at https://www.sienna-project.eu/publications/deliverable-reports/.

Brey, Philip, Philip Jansen, Jonne Maas, Björn Lundgren, and Anais Resseguier, *D4.7 An Ethical Framework for the Development and use of AI and Robotics Technologies*, 2020, SIENNA Project, 2020.

CEN, *Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework*. CEN workshop agreement, CWA 17145-2, 2017.

Council for International Organizations of Medical Sciences (CIOMS) and World Health Organization (WHO), "International Ethical Guidelines for Health-related Research Involving Humans", 2016. https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf

Dainow, Brandt, “Threats to Autonomy from Emerging ICTs”, *Australasian Journal of Information Systems*, Vol. 21, No. 0, 2017.

Daly et al., “Comparing Ideation Techniques for Beginning Designers”. Element Method, 2016.

Erden, Yasemin J. and Philip Brey, *D5.3: Methods for promoting ethics for human enhancement*, forthcoming.

Eriksson, S, ‘Rationality of rules’, 2021. (unpublished)

European Commission, *Horizon 2020 – Work Programme 2016-2017 General Annexes*, Publications Office of the European Union, 2016.

European Committee for Standardization (CEN), “Ethics assessment for research and innovation - Part 1: Ethics committee”, Workshop Agreement CWA 17145-1, May 2017. ftp://ftp.cencenelec.eu/EN/ResearchInnovation/CWA/CWA1714501.pdf

European Medicines Agency (EMA), “Guideline for good clinical practice E6(R2)”, 2016, p. 10. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf

Fitzgerald, Brian, Gerard Hartnett, and Kieran Conboy, “Customising Agile Methods to Software Practices at Intel Shannon”. *European Journal of Information Systems*, 15(2), 200-213, 2006.

Friedman, Batya, and David G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*, Cambridge, MA: The MIT Press, 2019

Friedman, Batya, David G. Hendry, and Alan Borning, *A Survey of Value Sensitive Design Methods*, Boston Delft: Now Publishers Inc, 2018

Friedman, Batya, Peter Kahn and Alan Borning, “Value Sensitive Design and Information Systems”, in P. Zhang and D. Galletta (eds.), *Human-Computer Interaction in Management Information Systems: Foundations*, Armonk, NY: M.E. Sharpe, 2006, pp. 348-372.

Gauttier, Stéphanie, Robert-Jan Geerts, Sinan Senel, Michael Nagenborg, "Building capacities for policy and industry outreach", Report, 4TU, November 2017.

Gebru, T., J. Morgenstern, B. Vecchione, J.W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford, “Datasheets for Datasets”, ArXiv:1803.09010 [Cs], March 19, 2020.

Gilman, SC, "Ethics codes and codes of conduct as tools for promoting an ethical and professional public service: Comparative successes and lessons", Prepared for the PREM, the World Bank, 2005.

Gonçalves, Luis, *What Is Agile Methodology*, Evolution4all, 2019.

Hamlyn, Rebecca, *D4.5 Public Views on Artificial Intelligence and Robots across 11 EU and Non-EU Countries*, SIENNA Project, 2020.

Haupt, R., M. Kloyer and M. Lange, “Patent indicators for the technology life cycle development,” *Research Policy* 36(3):387-398, 2007.

High-Level Expert Group on Artificial Intelligence, “The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment”, 2020. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Ho, Jonathan C., and Chung-Shing Lee, “A Typology of Technological Change: Technological Paradigm Theory with Validation and Generalization from Case Studies”, *Technological Forecasting and Social Change*, Vol. 97, August 2015, pp. 128–139.

Hoven, Jeroen van den, Pieter E. Vermaas, and Ibo van de Poel, "Design for Values: An Introduction", in Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel (eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, Dordrecht: Springer Netherlands, 2015b, pp. 1–7.

Hoven, Jeroen van den, Pieter E. Vermaas, and Ibo van de Poel, eds., *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, Springer Netherlands, 2015.

Huldtgren, Alina, "Design for Values in ICT", in Jeroen van den Hoven, Pieter Vermaas, and Ibo van de Poel (eds.), *Handbook of Ethics and Values in Technological Design: Sources, Theory, Values and Application Domains,* Dordrecht: Springer Netherlands, 2015.

Hutchby, Ian, "Technologies, Texts and Affordances", *Sociology*, Vol. 35, No. 2, 2001. Pp.441-456.

IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition, IEEE, 2019.

Jansen, Philip, Philip Brey, Alice Fox, Jonne Maas, Bradley Hillas, Nils Wagner, Patrick Smith, Isaac Ouoch, Laura Lamers, Hero van Gein, Anais Resseguier, Rowena Rodrigues, David Wright, and David Douglas, *D4.4: Ethical Analysis of AI and Robotics Technologies*, SIENNA Project, 2019.

Jansen, Philip, Stearns Broadhead, Rowena Rodrigues, David Wright, Philip Brey, Alice Fox, Ning Wang, *D4.1: State-of-the-art Review: Artificial Intelligence and Robotics*, SIENNA Project, 2019.

Kantar (Public Division), *D4.6: Qualitative Research Exploring Public Attitudes to AI and Robotics*, SIENNA Project, 2019.

Kneuper, Ralf, *Software Processes and Life Cycle Models: An Introduction to Modelling.* Cham, Switzerland: Springer Nature Switzerland, 2018.

Kohlberg, L, "Stage and sequence: The cognitive-developmental approach to socialization", Handbook of Socialization Theory and Research, Vol. 347, 1969, pp. 347-480.

Lawton, Alan. "Developing and implementing codes of ethics.", *Viešoji Politika ir Administratyvimas*, Vol. 3, No. 7, 2004, pp. 94–101.

Lighthouse, 'Developing a code of conduct: A step by step guide', 2013. https://www.lighthouse-services.com/documents/Developing%20a%20Code%20of%20Conduct%20[A%20Step-by-Step%20Guide].pdf

Liversidge, Ed, The Death Of The V-Model, *Harmonic Software Systems*, 2015.

Luis Gonçalves, *13th Annual State of Agile Survey*, Evolution4all, 2020.

Martin, Clancy, Vaught, Wayne and Solomon, Robert C. (eds.), *Ethics across the Professions: A Reader for Professional Ethics*, New York: Oxford University Press, 2010.

Messikomer, CM, Cirka, CC, "Constructing a code of ethics: An experiential case of a national professional organization", Journal of Business Ethics, Vol. 95, No. 1, 2010, pp. 55-71.

Miller, Catherine, and Rachel Coldicutt, *People, Power and Technology: The Tech Workers' View*, London, Doteveryone, 2019.

Ministry of Science and Technology of China. *Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence*, State Council, 2019.

Newton, LH, Hoffman, W, Frederick, R, 'The many faces of the corporate code', Taking Sides: Clashing Views on Controversial Issues in Business Ethics and Society, 3rd Edition (Dushkin Publishing Group, Guildford, CT), 1994, pp. 81-88.

Nissenbaum, Helen, "Values in the Design of Computer Systems", *Computers and Society* 28 (1), 1998, 38–39.

Norman, Donald A., *The Design of Everyday Things*, Basic Books, New York, 2002.

OECD, *Recommendation of the Council on Artificial Intelligence.* OECD, 2019.

Olechowski, Alison, Steven D Eppinger, and Nitin Joglekar, "Technology Readiness Levels at 40: A Study of State-of-the-Art Use, Challenges, and Opportunities", *2015 Portland International Conference on Management of Engineering and Technology (PICMET)*, IEEE, 2015, pp. 2084–2094.

Oliver, K., P. Cairney, "The dos and don'ts of influencing policy: a systematic review of advice to academics", Palgrave Commun 5, 21, 2019. https://doi.org/10.1057/s41599-019-0232-y

Pardi, Tommaso, "Fourth Industrial Revolution Concepts in the Automotive Sector: Performativity, Work and Employment", *Journal of Industrial and Business Economics*, Vol. 46, No. 3, September 2019. pp. 379-389.

Phoenix, J.H., L.G. Atkinson, and H. Baker, "Creating and communicating social research for policymakers in government", Palgrave Commun 5, 98, 2019. https://doi.org/10.1057/s41599-019-0310-1

Reddy, J. N., "Introduction to the Finite Element Method", McGraw-Hill, Hightstown, 2005.

Rodrigues, Rowena, Konrad Siemaszko, and Zuzanna Warso, *D4.2: Analysis of the Legal and Human Rights Requirements for AI and Robotics in and outside the EU*, SIENNA Project, 2019.

Rodrigues, Rowena, Stearns Broadhead, Philip Brey, Zuzanna Warso, Tim Hanson, Lisa Tambornino, and Dirk Lanzerath. SIENNA D1.1: The Consortium's Methodological Handbook (Version V0.6), 2018. https://doi.org/10.5281/zenodo.4247383

Rotolo, Daniele, Diana Hicks and Ben Martin, "What is an Emerging Technology?", *Research Policy*, Vol. 44, Issue 10, 2015, pp.1827-1843. DOI: 10.1016/j.respol.2015.06.006

Ryan, Mark, Philip Brey, Kevin Macnish, Tally Hatzakis, Owen King, Jonne Maas, Ruben Haasjes, Ana Fernandez, Sebastiano Martorana, Isaac Oluoch, Selen Eren, and Roxanne Van Der Puil, *D1.4 Ethical Tensions and Social Impacts.* SHERPA project. 2019.

Ryu, Jiyeon, and Soon Cheon Byeon, "Technology Level Evaluation Methodology Based on the Technology Growth Curve", *Technological Forecasting and Social Change*, Vol. 78, No. 6, 2011, pp. 1049–1059.

Schwab, Klaus, *The Fourth Industrial Revolution*, World Economic Forum, Geneva, Switzerland, 2016.

Shelley-Egan, Clare, David Wright, Rok Benčin, Jelica Šumič Riha, Gregor Strle, Daniela Ovadia, Adelina Pastor Cañedo, Christine Angeli, Menelaos Sotiriou, "SATORI Deliverable D2.1 Report (handbook) of participatory processes", July 2014, p.32. https://satoriproject.eu/media/D2.1_Report-handbook-of-participatory-processes_FINAL1.pdf

Sinclair, C, Poizner, S, Gilmour-Barrett, K, Randall, D, "The development of a code of ethics for canadian psychologists", Canadian psychology = Psychologie canadienne, Vol. 28, No. 1, 1987, pp. 1-8.

Stahl, Bernd Carsten, Richard Heersmink, Philippe Goujon, Catherine Flick, Jeroen van den Hoven, Kutoma Wakunuma, Veikko Ikonen, and Michael Rader, "Identifying the Ethics of Emerging Information and Communication Technologies: An Essay on Issues, Concepts and Method", *International Journal of Technoethics (IJT)*, Vol. 1, No. 4, 2010, pp. 20–38.

Steering Committee on Bioethics (CDBI), "Guide for Research Ethics Committee Members", 2012. https://www.coe.int/t/dg3/healthbioethic/activities/02_biomedical_research_en/Guide/Guide_EN.pdf

Stilgoe, Jack, "Self-Driving Cars Will Take a While to Get Right", *Nature Machine Intelligence*, Vol. 1, No. 5, May 2019, pp. 202–203.

Tambornino, Lisa, Dirk Lanzerath, Philipp Hoevel, Tom Lindemann, *D5.1: Report Documenting Elements to Open and Complement Operational Guidelines for Research Ethics Committees*, SIENNA Project, 2021.

Tambornino, Lisa, Dirk Lanzerath, Rowena Rodrigues, and David Wright, *D4.3: Survey of REC Approaches and Codes for Artificial Intelligence & Robotics*, SIENNA Project, 2019.

Taylor, Margaret, and Andrew Taylor, "The Technology Life Cycle: Conceptualization and Managerial Implications", *International Journal of Production Economics*, Vol. 140, No. 1, 2012, pp. 541–553.

UKRI, Institute for Government, "How to engage with policy makers: A guide for academics in the arts and humanities", 2020. https://www.instituteforgovernment.org.uk/publications/how-engage-policy-makers

US Government Accountability Office, "Technology Readiness Assessment Guide", 2020.

Wellens, Koen, The Seductive and Dangerous V-Model, *Testing Experience*, December 2008.

West, Sarah, Myers, Meredith Whittaker, and Kate Crawford, *Discriminating Systems. Gender, Race,* and Power in AI, AI Now Institute, 2019.

World Medical Association (WMA), "WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects", 2013. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/