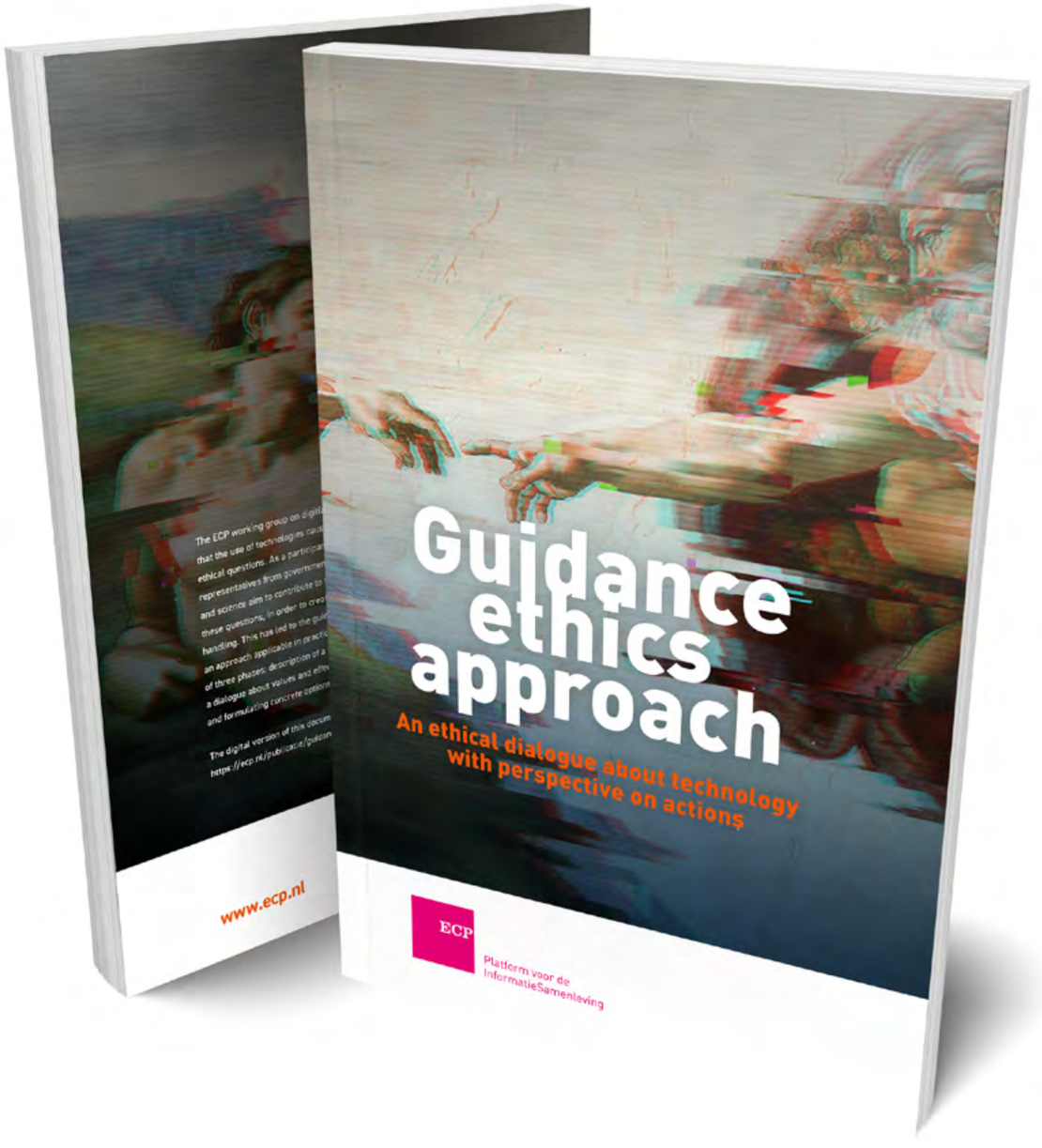


Guidance ethics approach

By Peter-Paul Verbeek and Daniël Tijink

An ethical dialogue about technology with perspective on actions

ECP
Platform voor de InformatieSamenleving

Colophon

Authors
Peter-Paul Verbeek
Daniël Tijink

Design
Brand new something

Photography
Unsplash, Shutterstock

Printing:
Veldhuizen grafisch effect

With special thanks to
Turnaround Communicatie

Contents

Introduction

Guidance ethics approach

We live in a society where digitization causes major changes. The smart phone in everyone's hand, algorithms suggesting or sometimes making smarter decisions, robots supporting people or replacing part of their work. Many of these developments raise ethical questions. Are they good developments? Do they fit in with our values? Can we give them a responsible direction?

These are urgent questions. The development of digital technology involves not only major social, but also economic interests. Companies want to keep up with developments, Europe cannot stay further behind China and America. People ask for new solutions, but there are also concerns about the impact of rapid digitization. How can we connect ethics and innovation in a fruitful way?

A joint venture has been established around this question between ECP | Platform for the Information Society and philosopher of technology Peter-Paul Verbeek. Professor Verbeek is a leading thinker in the (inter) national discussion on ethics and technology. He has introduced insights from the philosophy of technology into the ethical discussion. The basis of his ideas, supported by a great deal of research, is that it is better to consider people and technology as intrinsically connected rather than opposed to each other. Technologies, after all, have always helped to shape how we are humans: from pencil to the printing press, and from the steam engine to artificial intelligence. They are not alien to human existence, but rather mediate the way in which we live our lives and organize our societies. This approach to technology, often indicated as 'mediation theory', has far-reaching consequences for ethics. Ethical questions are often framed as dilemmas: should we or should we not accept this technology? But from the perspective of technological mediation, the main question becomes: how can we deal in a responsible way with our connections to technology? Verbeek calls this alternative to classical ethics: guidance ethics. Technology is guided in its roles in society and, conversely, society is guided in its dealing with technology. Chapter 1 further explains this approach to technology.

The aim of ECP's initiative to collaborate with professor Verbeek was to make the 'guidance ethics' approach applicable in practice. To this end, a working group was formed with parties from the (IT) business community (KPN, IBM, Facebook, Microsoft), from governmental bodies (the Ministries of Foreign Affairs, Economic Affairs, Justice &Safety and the VNG (the Association of Dutch Municipalities) and from various societal sectors: Alliander (energy), the National Police Force (security), Siza/Academy Het Dorp and the Rijnstate hospital (healthcare). There was also a participant of the Royal Family Service. From the ethics of technology field, researchers from Leiden University, TU Delft, Tilburg University and Erasmus University participated.

This publication explains and elaborates the approach that was developed by this working group. After an introductory chapter on the basic ideas of the approach (Chapter 1), we present four cases (in Chapter 2), which serve as examples for the rest of the document. Subsequently, we explain the starting points for the concrete implementation of guidance ethics (Chapter 3) and then present the 'Guidance ethics approach' itself (Chapter 4). After this, we will elaborate how the approach can be used in practice (Chapter 5). The final chapter contains a brief retrospective and a look at the future (Chapter 6).

Many technological developments raise ethical questions

Chapter 1
Guidance ethics

Ethical discussions about technology often have the character of a dilemma: is it acceptable or not to apply this technology? In the public debate, ethical concern about technology therefore primarily leads to radical criticism: 'The power of Big Tech,' 'the spying government,' 'the end of labor by AI,' et caetera. As a result, it is hard to constructively deploy ethical questions and concerns to give technological developments a desirable direction, instead of merely fully embracing or rejecting them.

An interpretation of technology ethics as accepting or rejecting technology places technology and society in opposition. In that approach, technology poses a potential threat to society and it is the responsibility of ethics to determine which technology may be allowed and which may not. However, this picture is not correct. Technology and society are fundamentally intertwined. Technology is developed by people with a view on a certain role of that technology in society. And society has always taken shape through interactions with technology, often in ways that were not explicitly intended by designers. For example, the printing press did not only bring the possibility to reproduce texts more easily, but also contributed to the reformation, the emergence of modern science and universities, the importance of knowledge, et cetera. We are just as connected with technology as with language or gravity: technology helps to make us the people we are. Technology and society shape each other: that is the lesson we can learn from the past 50 years of research in Science and Technology Studies.

This interconnectedness of technology and society entails a different role for ethics. The standard model of 'ethical assessment', in which normative theories help to decide whether a technology is acceptable or not, does not do justice to this interconnectedness. In most cases, the question is not whether a technology should be allowed or not, but how we can deal with it in a responsible manner. Moreover, a consequence of the interconnectedness of people and technology is that also the ethical frameworks with which we assess technology develop in interaction with that technology. What we understand by 'privacy', for example, is developing hand in hand with the technologies that are shifting the boundary between private and public.

Instead of seeing ethics as some form of 'assessment', then, it should also be seen as the normative 'guidance' of technology in society. And at the same time, ethics can also guide society in dealing with technology. Such an approach does not place ethics outside of technology, as an external 'assessor', but right in the middle of it. It is 'ethics from within', not from the outside. This type of ethics is not primarily focused on the question whether a technology is acceptable or not, but rather asks whether and under what conditions a technology can be given a responsible place in society. The central question in guidance ethics is not 'yes or no?', but 'how?' It does not focus on rejecting or accepting, but on the valuable design, implementation and use of new technology.

Instead of seeing ethics as some form of 'assessment', it could also be seen as the normative 'guidance' of technology in society.

Central to this guidance ethics is the inventory of the possible social implications of a technology, and the central values that are at stake. This is done in a deliberative process. In the guidance ethics approach, it is important to always start from concrete technologies and their specific effects and consequences. It is not about making generic analyses of 'digitization' or 'artificial intelligence' as such, but about the concrete applications thereof in a societal domain. After all, the concrete interaction between humans, technology and society has a central place in this ethical approach.

When making an inventory of effects, it is important to not only look at the consequences for individual users, but also at the social implications for, among other things, education, healthcare, the judiciary, legislation, policing and law enforcement. Moreover, technologies also influence frameworks of interpretation: they help to shape the meaning of central values such as privacy and autonomy.

Chapter 2
Four cases

The guidance ethics approach that we present in this document takes as its starting point the connectedness between technological developments on the one hand and human beings and society on the other. The best way to illustrate that approach is to discuss a number of cases for which the approach has been used. The core idea of guidance ethics, after all, is not to focus on abstract questions about 'Technology', but on concrete technologies that function within a concrete context.

Case 1 **AI in psychiatric care** [1]

UMC Utrecht is investigating an AI application that can estimate the risk of aggression in patients in the psychiatric department. The application is not being used yet, but has been successfully tested. The department is ready to put it into practice. The advantage of a good risk estimation is that professionals, together with patients (and possibly family) can choose from a number of measures to use, for example medication, more sports or family that can stay over.

Aggression, both verbal and physical, is common; there are hundreds of reports per year in this institution. It has an impact on the patient, but also on the relatives and care professionals. That is why professionals, patients and relatives have chosen to see whether data analysis can help.

The AI application receives anonymous clinical texts as input (intake report, incident reports and other reports). These are converted into mathematical representation, after which machine learning can be applied; this gives the outcome that there is a chance that (for example) aggression will occur within thirty days.

Currently, doctors and nurses estimate the risk of aggression. It appears that the AI application predicts aggression slightly better than the questionnaires that are currently used for this and significantly better than the subjective assessment by healthcare professionals. After good results at the UMC Utrecht, the tool was validated in a large mental health care institution with comparable good results.

1 The first case we present here was elaborated in a workshop organised by ECP and NICTIZ, the three other cases were brought forward in the ethics and working group. We thank Karin Haagoort (UMCU), Edo Haveman (Facebook), Brigitte Boon (Siza, Academy Het Dorp) and Marc Noordhoek (BZK) for the input of their case and the explanation thereof. The responsibility for the text about the cases lies with the writers of this document.

Case 2 **Facebook: algorithmic timeline/newsfeed**

Everyone knows Facebook. Originally created as an online facebook for students, it is now the largest social network in the world. The technological application discussed in this case is the newsfeed of this medium. The activities that users see on their newsfeed (when they have made friends and follow pages) are processed by an algorithm that puts them in a specific order. Engagement is an important factor in this process. The algorithm puts items on top that people really don't want to miss, such as wedding photos and baby photos.

This 'newsfeed' technology makes many things possible: user-generated content, freedom of information (everyone can post almost anything), democratisation and approachability of politicians, quick questions and answers, mobilization (women's march), and denouncing new issues (#metoo).

As a result of the development of the internet and platforms, part of the gatekeeper role of newspapers or TV news programs is shifting to algorithms and the user. They help to determine the relevance and reliability of information. Giving people what they want increases the chance that intellectual content will be ranked lower or sensational messages higher in the newsfeed. When it comes to misinformation and polarising content, this is undesirable.

Case 3 **Feeding robot**

Academy Het Dorp supports institutions in long-term care with (research into) the use of new technologies. This includes research into the ethical values that play a role in the use of technology. One of the care institutions, and founder of Academy Het Dorp, is Siza. Siza is a care institution that provides care to people with different types of disabilities, including people with physical disabilities, non-congenital brain injury, intellectual disabilities and severe multiple disabilities. Siza wants to offer its clients as much autonomy as possible and uses new technologies for this.

Clients with physical disabilities and non-congenital brain injury live in Het Dorp, one of the locations of Siza, in their own house that is adapted as much as possible to their specific needs. Some clients cannot eat by themselves and need help with that. In addition to the support provided by healthcare workers, they use different types of technology. Think of a feeding robot or a robotic arm (mounted on a wheelchair). With the robot (arm), the food or drink can be brought to the client's mouth. The health care worker now only needs to prepare the food and clean up afterwards, and clients can have their meal as they please: at their own pace, with privacy - or, if they so choose, with the help of a health care worker.

Algorithms and robots are increasingly being used in 'normal' work processes

Case 4 **Undermining**

Undermining criminality is a collective term for offences that damage public and social structures as well as trust in them. An important element of such crime is the abuse of legal/upper world persons, organisations and institutions.

In the 'undermining' project, the aim is to gain more insight into where undermining is concentrated, by using (big) data from different sources. A growing number of larger municipalities are cooperating with the Public Prosecution Service and the Ministries of Internal Affairs, Justice & Security and Finance. Sources that are used are, for example, the Land Registry and the Statistics Netherlands. The intentions and preconditions are laid down in the Government Gazette under the name: CITY DEAL insight into undermining. This guarantees a clear definition and a solid legal basis. For example, the analyses are scientifically verifiable, the process complies with the GDPR and the algorithms are tested for validity. There is no black box.

The project focuses on three themes: drugs, fraudulent foundations and real estate fraud. An example of the latter are the so-called wind catchers. People without income and without own funds who own real estate. Their analysis maps their concentration up to neighbourhood level.

Chapter 3
Principles for the implementation of guidance ethics

The purpose of this publication is to come to a method based on which the ideas of guidance ethics can be applied in practice. To this end, we first take an intermediate step. We state the principles that go with this way of looking at the development of technology and of society. In the discussion of these principles, we use elements from the previous cases as an illustration.

The how-question is central

Humans and technologies are connected and continuously influence each other, as two partners in a dance. The core of guidance ethics is therefore, unlike many other ethical approaches, not the human assessment of technological development. The core is the interaction, the 'how' question, instead of the 'yes or no' question: how can people and technology develop in a valueable way?

When we look at the case of the feeding robot, the primary question is not if we should use a feeding robot or not, but how we can use it the best way. Can we have a debate on how we can deal with the feeding robot instead of merely discussing what is right and wrong? Can we find options for action for good use? Instead of the question 'can we delegate care for vulnerable people to machines?' the question is: 'is there a way to take the core values in care as a starting point when developing and using care robots?'.

The focus on the how-question does not mean that it is never possible to say 'no' or that in principle every technology can be used or introduced everywhere, however, the focus is not primarily on the 'yes or no' judgment. That question is quickly limiting, which means that options for action are not discussed. Also, the values or interests of certain groups are often not sufficiently taken into account. Focusing on the how-question makes it possible to actually connect ethics with technology. This question makes room to look for conditions under which a technology can function in a responsible manner. And those conditions are in the design of the technology itself, in its social embedding and in the way people use it. But in the end, 'not' also remains an answer to the 'how' question. If it appears that the technology is not compatible with our values, the model's outcome is that the technology should not be used.

Small, continuous steps

Closely connected to asking the how question is the awareness that improvement comes in (small) steps. If one is looking for a 'yes or no', 'is this allowed or not', one is looking for the ultimate answer. The guidance ethics approach allows us to see that the interconnection between human beings and technology develop in small, continuous steps. Technologies always adapt to how human beings interpret them, but users also adapt to the new possibilities.

There are all kinds of small steps in the development and implementation of the feeding robot. The robot is becoming more and more precise, partly based on the input of the users. In addition, the role of the feeding robot in the care process and what is expected of the care professionals is changing.

Some ethical approaches believe that the development of technology and its impact on human beings can be stopped and that ethics should determine when it has to be stopped. From the guidance ethics perspective, we think that this situation rarely occurs. What is possible, however, is a continuous consideration of what can be improved and how to take concrete steps to achieve that.

Technology and society are in a process of continuous mutual adjustment

Technology in context

Because humans and technology are so closely connected, it makes little sense to speak about technology without involving human beings and their context. Indeed, it matters where and by whom the dialogue is conducted.

With guidance ethics, we want to start a discussion about concrete technologies that function within a specific context, with real people. That is why we rather not talk about 'Technology' or 'Human Beings' as if these were single, well defined concepts. Discussions about AI or block chain can be interesting, but only become relevant when they touch on practice.

This puts demands on the level of abstraction that we choose when talking about a technology. The case of feeding robots is a good example. We are not talking about 'robotics' (too general) or about a brand of robot of a certain serial number (too specific), but about 'feeding robots'. More specifically: feeding robots within a certain context, in this case an institution for care for the disabled. Lessons learned will (in large part) also apply to the use of other feeding robots in other institutions. But they may also apply to a wider e-health context, or to wider or more limited applications of robotics; yet very specific for this particular institution and this particular feeding robot.

In fact, we cannot assess without context. The patterns observed in undermining are only meaningful if agents and policy-makers can do something with it, if analysts give an interpretation of it and if conversations with residents and other stakeholders explain how they see them. Ethical tension exists both in the network around the technology and in the data analysis itself.

Human values

The purpose of guidance ethics is to give human values a guiding role in the development, implementation and use of technology, ranging from justice, autonomy and speed to sustainability, safety, effectiveness, et cetera.

Emotions, both positive and negative, can play an important role in the search for the values that are central to a certain technology. Fear, enthusiasm, astonishment, concern: these are all indications that the technology is putting something valuable at risk or enables it. As a result, emotions are an indicator of the normative frameworks that should be given a place in the design, implementation and use of this technology.

Which values are relevant, depends on the specific technology and context. In Chinese culture, for instance, different values prevail than in the culture of the United States of America, and other values prevail in healthcare than in construction.

Moreover, the advent of technology can change values. Before the introduction of mobile telephony, values such as 'reachability' and 'attention' had a different interpretation than nowadays.

There is almost always tension between the different values that play a role in certain technology in a specific context. In Facebook's newsfeed, there is tension between the freedom to publish anything and the negative consequences of reinforcing statements that are less relevant or even untrue. In the undermining project, there is tension between the desire to obtain the most accurate information and the protection of the privacy of individuals.

The pursuit of a *value-able* co-development of technology and human beings is, in short, complex. It is important to recognise this complexity and to find a way to deal with it. Reasoning based on one specific value, without being aware of this complexity, does not bring much.

What will the future bring: in a positive and in a negative way

We do not know what new technology will bring us or where we will bring technology. When we look back at future predictions, they often turn out to be wrong. The rise of the internet was hardly predicted in the 1950s, but flying cars were. Nevertheless, images of the future are needed, they are an important part of the interaction between technology development and social change.

Many ethical discussions focus on the possible disadvantages of a technology. In the guidance ethics approach, we assume that technology has both positive and negative consequences. It is important to give both enough room in a dialogue. If we only talk about how a feeding robot can never offer patients human contact the way a healthcare professional can, we ignore the opportunity the robot offers to patients to regain something of their human dignity by being able to eat by themselves again.

There may be a parallel here with movements in other scientific areas: positive design (designing for new possibilities), positive psychology (focusing on how someone could thrive instead of focusing on someone's problems), positive health (not just focusing on what's wrong and not possible, but also on what is possible). Guidance ethics aims to move beyond 'negative ethics', centred around negative aspects that should be avoided or prevented, towards 'positive ethics', centred around the values that should be fostered in the design, implementation and use of technology.

Action options

The guidance ethics approach looks for concrete options for action in order to achieve a more valuable interaction between people, society and technology. We distinguish three types of options for action to achieve that valuable technological-social development.

Designing technology: *ethics by design*

Every technology has built-in values, so to speak, technology invites certain behaviour. A technology can therefore be designed in such a way that it better matches certain values. For example, the value of privacy can be guaranteed by allowing the user to control the cookies that are stored, or by giving cameras on crowd control drones a low resolution that makes it possible to count numbers of people but not to recognise individual people.

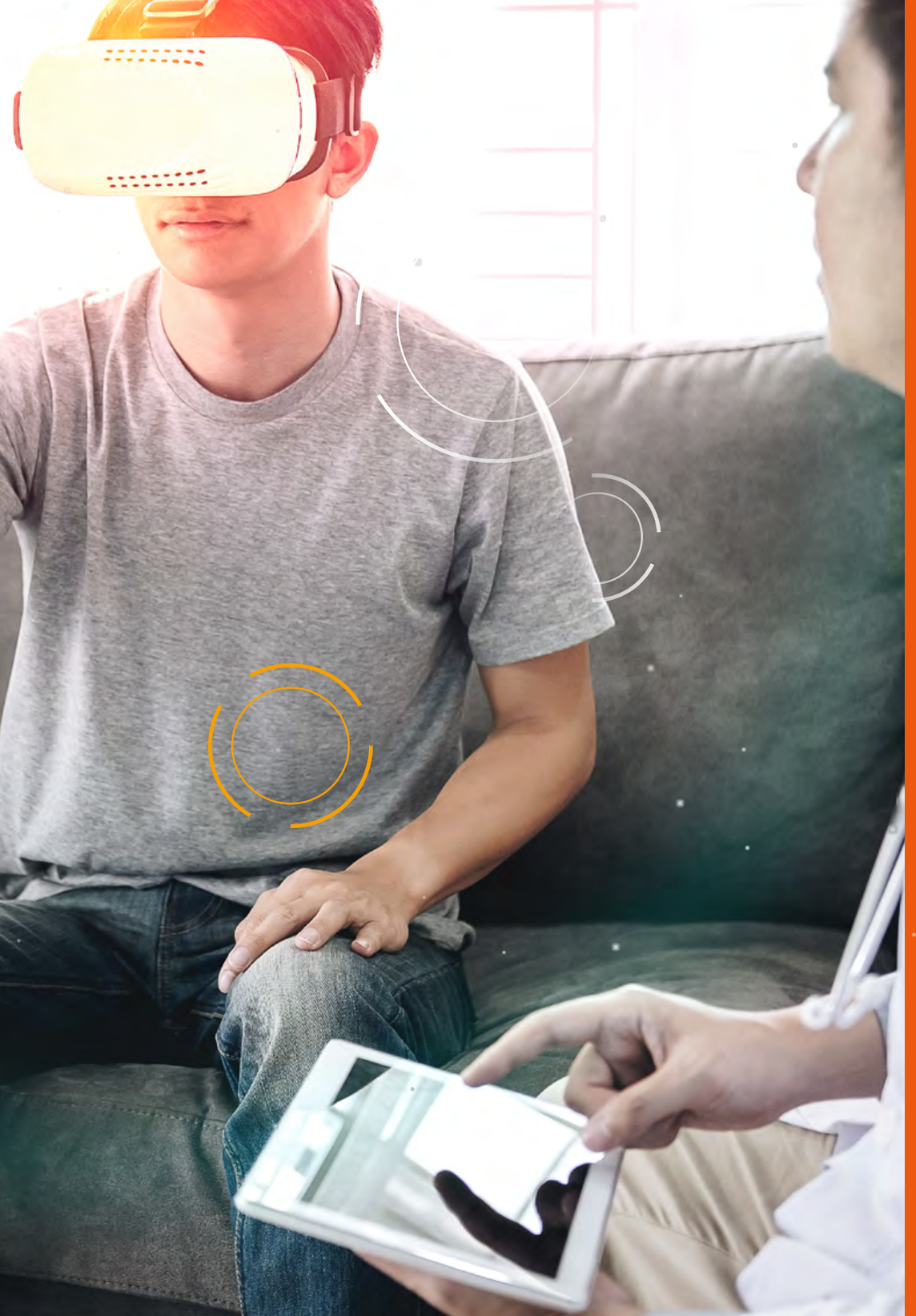
Environment: *setting up the environment (physical aspect) and making agreements (social aspect); ethics in context*

Every technology is used in a context: physical, social, organizational/ legal. With new technologies, that context/environment is also adjusted. The increasing use of the car entailed the construction of (physical) sidewalks and traffic lights. Also socially, new agreements were made: pedestrians on the sidewalk, cars on the road, pedestrian crossings and refuge hills as safe places for pedestrians. These agreements have also been legalized through traffic laws.

User: *awareness and behaviour adaptation: ethics by user*

When it comes to the use of technology, people can display more and less valuable behaviour. In traffic, for example, awareness and training take place through traffic lessons at school, driving lessons can lead to a driving license and awareness is created by designated driver campaigns and traffic signs with children playing on it.

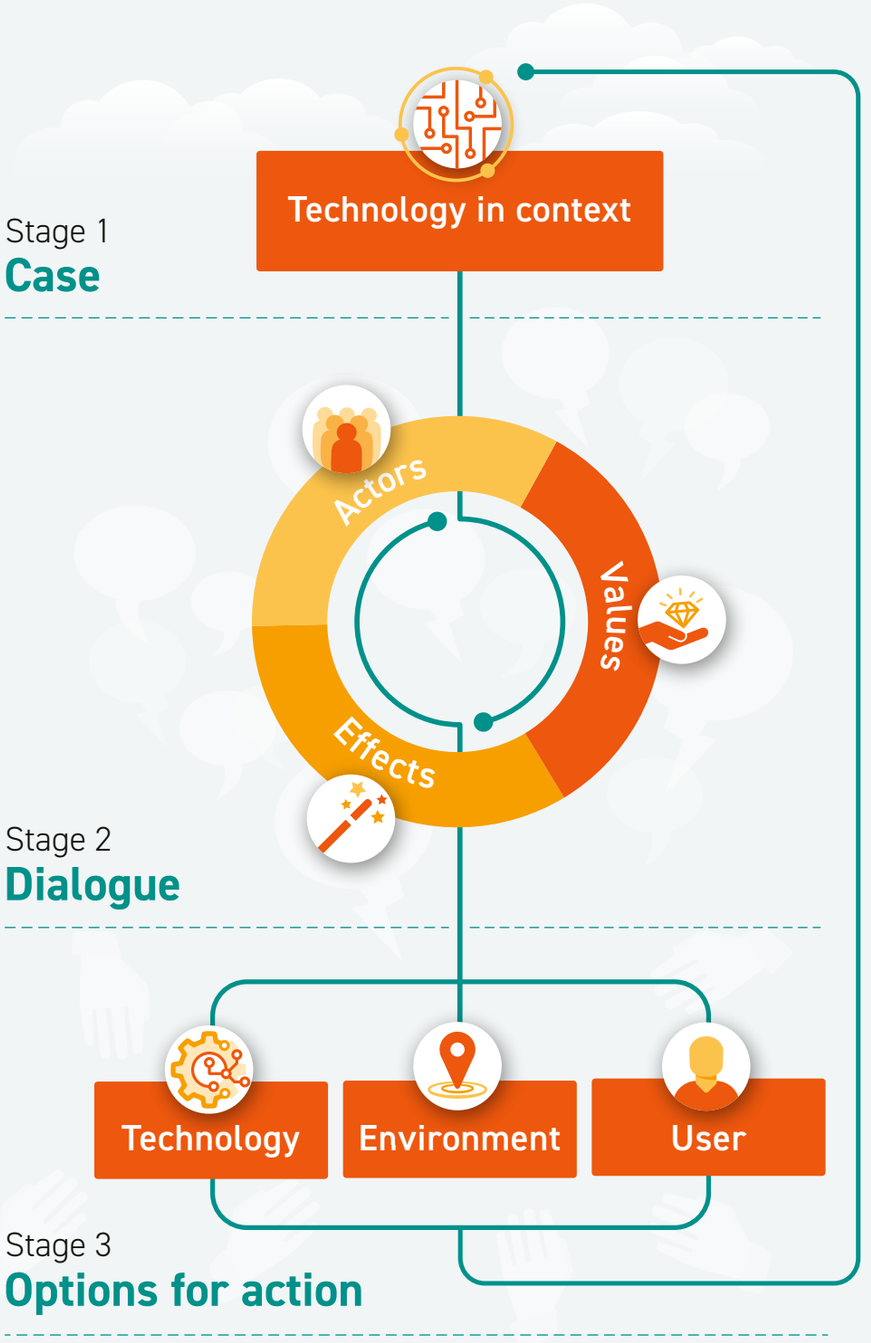
In this chapter, various elements were discussed that are important for the practical translation of guidance ethics. These elements are the building blocks for the approach presented in the next chapter.



Chapter 4

Explanation guidance ethics approach

In this chapter, the principles are translated into a method: the guidance ethics approach. The figure on page 32 shows the main elements of the model in three steps.



Stage 1 **Case: technology in context**

Describe the technology and the context in which that technology functions. The point is to get a close understanding of what we are dealing with. Some discussions about technology are more about (positive or negative) images of that technology than about its actual functioning and its meaning for people. Because we opt for a focus on a concrete technology in a concrete context, we are able to come to a fairly precise description, both of the technology and of what the use of technology means in its context. The point is to make a clear, understandable description, without too much jargon or technical details. The description must be understandable for interested outsiders.

Stage 2 **Dialogue: actors, effects, values**

This step focuses on a further elaboration of the case. After having developed a closer understanding of the technology and its context, we need to investigate the possible effects of using a technology in that context. We want to know who is involved and which values play a role in the practices around the technology and its potential impact and implications.

Actors

In a specific context, it is usually quickly clear who the relevant actors are. The parties involved may also be asked who else could be relevant actors. At a generic level, relevant actors often include clients/citizens, professional users, policy-makers, designers. For example, healthcare cases typically involve patients, care professionals and caregivers as relevant actors.

Ideally, the people actually involved have input in the dialogue. They don't always have to be people who are actively involved in the use of technology. The use of a technology can also have a huge impact on non-users; think of the influence of cars on pedestrians. If it is not possible for everyone involved to participate, there may be people who represent them

or who are willing and able to think from their perspective. Academics or other experts having experience with or expertise in the subject can also be involved, to develop a broad social perspective.

Effects

The use of a technology has all kinds of effects. Some effects can be immediately clear, where others might occur, for example in the future or under specific circumstances. The first step is to collect the potential effects of the technology as openly as possible, without taking desirability or likelihood into account. Next, it is good and pragmatic to identify which effects are most relevant.

The dialogue continues to be based on concrete technology and context, and from there various potential effects are discussed. Distinguishing different effects can help in obtaining a rich and realistic image:

- positive and negative effects,
- known and foreseeable effects,
- direct and indirect effects,
- effects for different actors,
- effects on different levels: individual (micro), social (meso) and social (macro).

Values

Technology is always surrounded by different values. Think of justice, applicability, reliability, solidarity, respect, autonomy. They often remain implicit in discussions, because criticism of technology is typically formulated more concretely. If the AI application in the GGZ institution makes someone feel that they are being monitored, the underlying value may be autonomy, or privacy.

In most cases, several values play a role simultaneously. As with the effects, the first thing to do is to make an open inventory, followed by the identification of the values that are considered to be the most relevant, whereby it remains important to keep an eye on the 'less relevant' values.

At present, various ethical codes or guidelines are being written in many sectors, in companies and by the government. A small grasp of AI alone leads to the following examples: Google AI Principles, Microsoft AI Principles, UK initial code of conduct for data driven health and care technology, European Commission High Level Expert Group on AI, Smart Dubai AI Principles, Nesta principles for public sector use of AI, Code of conduct AI the Netherlands ICT, AI Impact Assessment (ECP), Responsible Innovation: 7 principles for the public sector (Ministry of the Interior and Kingdom Relations). As an example, the text box shows the list established by the EU High Level Expert Group on AI. For the guidance ethics approach, these are sources of inspiration to identify the values playing a role in technology in context.

European Commission High Level Expert Group on AI

1. Accountability
2. Data Governance
3. Design for all (by all - include diversity)
4. Governance of AI Autonomy (Human oversight)
5. NonDiscrimination
6. Respect for Human Autonomy
7. Respect for Privacy
8. Robustness
9. Safety
10. Transparency

https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

Dialogue

This stage is called 'dialogue'. In an open exchange between the actors involved, it becomes clear what the possible effects and important values are regarding the use of a technology. A dialogue is an important part of the approach and often takes place in a workshop setting, certainly if the parties involved feel a joint responsibility for a technology and see that they need each other to take the next steps. Parts of the dialogue naturally also take place outside the workshop. The conversation continues.

In addition to and in preparation of the actual dialogue, other means may also be used. Interviews provide a different kind of insight into how the discussion partners perceive the effects and values. Literature research can often be a valuable addition as well, certainly if followed by a good analysis.

A good dialogue stage has several types of outcomes. First, it often takes away uncertainties. The input of different types of knowledge gives everyone a better picture and therefore a better idea of the possible effects. By thinking through the effects, it becomes clear where the expectations and fears lie and, possibly, how realistic they are.

Secondly, the actors bring in different perspectives, which brings up the most important values that play a role in this technology within this specific context. This often leads to mutual understanding, because it enables people to put themselves in the perspective of the other. They do not necessarily have to agree with the importance of the values of the other, but can better understand the importance that the other attaches to them.

There are also other methods available to support such a dialogue stage, such as the methods of a 'moral deliberation', Socratic debate et cetera. In fact, this stage is valuable in itself, besides its role in the 'Guidance Ethics' approach. Within the guidance ethics approach, though, this stage is an essential bridge between the first stage - technology in concext - and the third stage: the identification of options for action.

Stage 3 **Options for action**

The core of the guidance ethics approach is to guide technology in society in an ethically valuable way and to guide society in the ethically valuable embedding and use of new technology. This requires action. That is why the emphasis in the guidance ethics approach is not only on having a good conversation or establishing an ethical code, but also on arriving at options for action. Three types of options for action are available in the guidance ethics approach: connected to the technology, to the context and to the user. After a brief explanation, a box will give examples of options for actions in each of the the four cases of chapter 2.

Technology, ethics by design

Ethics by design has been in the spotlight for some time. It has now become a best practice to include moments of ethical reflection in the design process of a technology, so that ethical values are actually included in the design of technology. Ethics are not only a matter of human beings, but also of technologies. Every technology influences the choices and behaviour of human beings. An 'ethics by design' approach deliberately shapes that influence, based on explicit ethical reflection.

Ethics by design sometimes happens without being labelled as such. In a well-functioning market, customers indicate what they expect from a product and producers will try to adapt to it, including the values that go with it, such as safety, sustainability, aesthetics, applicability, et cetera.

However, not all markets are perfect and not all social practices are markets. In healthcare, for example, customers have little purchasing power, because this power has been handed over to insurance companies; and with digital media platforms, monopolistic situations often arise. Obviously, governments can play a role here: they can impose conditions and requirements through legislation and regulations. This is a fairly slow and not very accurate instrument, though. Moreover, not all values can be optimally translated into market mechanisms. This makes it interesting to also try to enable designers to incorporate ethical values into their work at an early stage.

Options for action: design

AI in psychiatric care:

- Professional judgment required: there is a fear that psychiatrists will blindly trust the AI application. A technological solution is that the AI judgment only becomes accessible after they have completed their own analysis.
- More transparency: the algorithm makes predictions that are good, but not transparent. Transparency could be added to the system, so that it becomes more understandable and easier to explain on what basis the algorithm makes its recommendations.

Newsfeed Facebook:

- Removing fake accounts: a lot of 'pollution' on social media comes from fake accounts. Facebook's 'real name policy' is a method to prevent that.
- Recognising Fake News and giving it a lower ranking: clickbait and sensationalism lead to similar patterns (omissions in title, disappointment with user) and are therefore recognizable by AI algorithms. These algorithms can then rank this type of content lower on timelines.

Feeding robot:

- Better washable: it appears that the robotic arm is not only used to eat, but also to scratch your head (just like human hands).
- Talking means not eating: can AI be built into the feeding robot, so that food is only brought to the mouth when the user is not talking?

Insight into undermining

- Because of privacy, the choice was made for representation at neighbourhood level, not at the individual's level.
- The choice has been made not to create a blackbox. This limits the number of possible AI applications.

Environment, ethics in context

The ethical dialogue about technology often focuses on that technology, while environment or context in which that technology functions is just as important. That context is often also different. It makes a difference when Facebook is used by someone aged 13, 35 or 80, and whether it is used in a work context or privately. It makes a difference whether an AI application is used in healthcare or in the construction industry.

It is therefore important to pay more attention to that environment, also because it contains part of the solutions. When developing technology, a designer and a user will have to think about what that means for the system in which that technology is applied. An organization or a system will change after the introduction of a technology. How is this change shaped? Can it be shaped in a way that takes into account the values that were identified?

The adaptation of the environment can be physical, social or legal. Think of the introduction of the car, a new technology at the time: it entailed physical adjustments such as sidewalks, roundabouts or traffic lights, social adjustments such as mutual agreements in traffic, and legal adjustments such as traffic law.

In 'ethics by design', designers and (tech) companies often hold the key, but when it comes to adapting the environment, decisions are mainly made by organisations (meso) and the government (macro). A company that imports robots, will make adjustments for the safety of employees, for training, new processes, et cetera. Examples from the government side are the construction of roads, the introduction of the general data protection regulation (GDPR) and the setting of preconditions for a personal health environment, by setting (MedMij) standards.

Options for action: environment

AI in psychiatric care

- Setting up a 'red button'/reporting point where a patient or person concerned can have the judgment of the AI application ignored if they think it is really wrong. In that case, the 'old' way of working starts again.
- Agreement that there will always be a double assessment, one from the professional and one from the AI application.
- Agree with family, client and colleagues when to use AI and when not. Possibly with the option not to use it.

Newsfeed Facebook

- Collaboration with external fact checkers who can offer disclaimers with messages.
- Preparation of legislation on transparency of political advertisements.

Feeding robot

- Agreements about safety are needed. Who is responsible if something goes wrong, if someone is injured by using the feeding robot.
- Reimbursement of technology (feeding robot, robotic arm) that encourages autonomous eating and drinking, is desirable. This requires clear reimbursement rules.
- Appointments and working methods of care professionals must be adjusted. For example about hygiene.

Insight into undermining

- All analyses on the data take place on the basis of a four-eye principle to prevent bias.
- CBS uses an output check where automatic and manual checks are made to ensure that no information that can be traced back to a person is taken from the analysis of the environment.

- Logging of all analyses by the analysts in which 'unlawful' investigations (which are outside the research question) are identified.
- Analysts may not have worked in investigations in the past two years.
- Analysts must sign a comprehensive confidentiality agreement.

Proper use, ethics in use

Also the user plays a central role in shaping the social impact of a technology. People can handle technology with care or recklessly, can be well trained or poorly trained. The first step is awareness. What does a technology do, what can it do and what can I do as a user? Action options can take the form of information or awareness campaigns, but one can also become aware of a new technology and its implications through school, word of mouth, news carriers.

The second step is actual behavioural change. Often that means training and exercise. This can vary from reading a manual to taking a course. Sometimes, behaviour is not difficult to carry out, but may be difficult to change. It requires a different habit or discipline, for instance not to drink and drive. Everyone can do it, but it's about the acceptance that this is indeed dangerous behaviour and that it is not uncool not to drink. On the other hand, getting a driver's license is an example of something that requires a lot of training for most people before the use of the technology is safely mastered.

Options for action: user

AI in psychiatric care

- Being able as a healthcare professional to explain what you do and why, in the same way as with an MRI scan. They need training for this.
- Knowing as a patient that AI is being used and knowing that it is possible to ask for a second opinion.

Newsfeed Facebook

- Media literacy training: knowing that fake news exists, ways to recognise and test it.
- Popularize the editorial role of newspapers and publishers so that more people are able to play the gatekeeper role themselves.
- Setting up a place where people can find correct and reliable information about vaccinations, nutrition or health.

Feeding robot

- Healthcare workers also need help and training. They need to know what can be done with the technology and what is safe.
- Healthcare workers have a different role when a feeding robot is used. They provide care less often or in a different way. This may require training.
- The robotic arm can be used in many different ways, more than just for feeding. For users, it can be helpful to find out together how the robotic arm can be used/operated. A suggestion is to 'freestyle' an afternoon with some super users, so you can learn what is possible and what fits you.

Insight into undermining

- Training policy officers and other users of the results of the analyses about the value and limitations of data analyses.
- A very limited number of people have access to the micro-data from CBS (Statistics Netherlands) and there are high demands on the level of education. Statistical scientific analyses are always well documented, checked and tested on repeatability.

Significantly, for each of the cases it is possible to achieve options for action in all three categories.

The users in a variety of roles

We see that the user shows up in different roles for all three options for action. In discussions, those roles are often confused, so we want to name them here again.

- The key in ethics by design, is that the user has a say in the design of technology.
- The key in context and environment, is that the user can participate in the discussion about how that environment is adapted. There are plenty of examples of computer systems that are introduced without consultation with the employee, with great frustrations as a result.
- Proper use is about how users deal with technology and context. The user may show more or less desirable behaviour.

We see those roles in internet banking, for instance. Technologically, payment via the internet becomes increasingly easy. This ranges from a better interface via the website, to the introduction of an app on the phone, the possibility of not only paying, but also sending payment requests (tikkies).

An example of setting up the environment was when a bank wanted to resell customer data anonymously. That resulted in a lot of opposition. The technology made something possible, but clients set preconditions. They did not want their data to be shared, even if the data were anonymized.

The use of internet banking has become increasingly easier, but many (elderly) people had to learn it, with the support from, for example, senior citizens' unions. The banks also try to encourage 'good behaviour', for example with the campaign where people are encouraged to look at the 'lock' symbol on the website, to check whether they are on a secure site.

Chapter 5

Use of the guidance ethics approach

In the previous chapter, we described the guidance ethics approach. In this chapter we will discuss the question of when and how one can use that approach.

It is clear that the guidance ethics approach can play a role when ethical questions arise about a technology. This may be the case with the introduction of a new technology, but it can also be when that technology manifests itself in a broader or different way in society. This chapter examines the potential of the use of the method within organisations or in the public sphere and we discuss how a workshop can be prepared.

Within an organization or in a social domain

The guidance ethics approach can be used both within one organization and within a social domain with multiple parties.

Organizations

Within an organization, the first question is who the problem owner is. Ethics is usually not part of the primary process, so it will probably require some effort to bring it to the attention. Ethics is often seen as an extra, as a show stopper or something that comes from a few enthusiasts. By its nature, the guidance ethics approach addresses a number of objections and opens up perspectives for engaging in ethics in a constructive manner. The strength of the approach within an organization is that options for action can be converted into real actions relatively quickly.

The question remains in which way, at which moment and from which processes the use of the approach can be used best. Some organisations have ethics committees, with the task of analyzing ethical dilemmas, while in other organisations, ethics more likely fits in with HR departments, training, client councils or client panels or is a matter for the boardroom.

When the decision has been made to use the guidance ethics approach, it is important to engage the right people. This will typically involve people from outside the organization (users, policy makers).

Large companies usually have a business code of conduct, design principles and sometimes also development patterns. Codes of conduct are rules of conduct for the entire company, guiding people's own behaviour. Design principles are about the actual design of products, and development patterns apply to implementation, the way solutions are applied at customers. The guidance ethics approach can be an important addition to this. It connects inside and outside, not only by giving the user a better role in the design process, but also by involving the company in the context in which the solution will function and by thinking about what this means for the user's skills.

Social playing field

Many ethical debates transcend the level of one organization. They concern topics that appeal to many in society and have to be discussed in the public sphere. Parties that can be at the forefront of this, are ministries, municipalities, advisory councils, social groups, action groups, debate centres, et cetera.

The guidance ethics approach is particularly suitable to give shape to this. The various actors are invited to participate in the debate from their perspective and are invited to think in terms of options for action. The guidance ethics approach requires serious interest in the problems of both technology and context and challenges to come up with suggestions for options for action. If all goes well, the various parties that can do something are also at the table. In this way, steps can be taken and the world becomes a little more Valuable.

Workshop: preparation, follow-up

Large or small, public or private, almost all dialogues involve the organization of a workshop. In this section, some suggestions are made for setting up a workshop involving the guidance ethics approach. We follow the three stages.

Technology in context:

- **Sharp definition of technology**
  - Is everyone sufficiently aware of the technology?
  - Do documents have to be submitted in advance, based on interviews and literature?
  - Does an expert have to be present to give an explanation and answer questions on site? Is the expert an outsider or one of the participants?

- **Sharp definition of environment**
  - Does everyone know the environment in which the technology is used?
  - Do documents have to be submitted in advance, based on interviews and literature?
  - Does an expert have to be present to give an explanation and answer questions on site? Is the expert an outsider or one of the participants?

Dialogue

- **Participants**
  - As many stakeholders as possible at the table, preferably with the power to act.
  - Balanced composition of the group in which different perspectives are sufficiently represented.
  - In addition, possibly a number of experts (academics) for generic input and interpretation or very specialist knowledge.
  - The group size must be such that it is possible to have a dialogue. Depending on the design of the workshop between 10-30 people, in one group 15 is the maximum.
  - Sense of relationships between the participants: are there tensions or strong connections between people?
  - Option: interviewing participants in advance often provides a sharper focus of the meeting and has the additional advantage that participants have already prepared themselves.

- **Possible information to be brought before or during the workshop:**
  - None: during the meeting, everyone is sufficiently informed and the choice is made to appeal to the creativity of the group.
  - The technology and the context.
  - List with relevant values.
  - List with possible effects.

- **Points of interest workshop set-up:**
  - Good moderator.
  - Clear, shared goal.
  - Distinguish between the creative process and the selection process.
  - Possibly subgroups to deepen and accelerate.
  - Ensure a common conclusion with clarity about the follow-up.
  - Good reporting.

Action options

- This stage can be discussed per type of options for action: technology, environment, behaviour.
- This stage can also be discussed from different actor perspectives; with the feeding robot: which options for action have been discussed from the patient's point of view, which from the caregiver's point of view?
- The golden rule for options for action is: the more concrete, the better. If suggestions are vague, keep asking, but don't get lost in too much details, that will be for later.
- It is easy to put another person in charge of options for action. Try to see if options for action can be addressed with various people at the table.
- Try to create an atmosphere in which people help each other take their responsibilities and use action potential.

Of course, it is important what will happen after the workshop. Very concretely, a number of options for action could be converted into actions. Some of the options will require a process of years, for example a new design or the introduction of a law. In that case it is of course possible to take the first steps. Another part will be 'low-hanging fruit'; options that can be converted directly into actions. For example, a freestyle afternoon can be organised for the users of a feeding robot or a consultation can be arranged to decide on its safety.

It may also be that the dialogue component requires a follow-up. That the conversation that took place in the workshop requires continuation. This is possible through multiple workshops, a publication, lectures, et cetera.

After all, the guidance ethics approach is interactive. We do not believe in final solutions, but in improvement steps.

Chapter 6
Follow-up

The guidance ethics approach as described in the previous chapter is a product of the ECP working group consisting of people from the business community, government, social sectors and ethicists/technology philosophers. With great thanks to the contributors of the cases, we have succeeded in making the philosophy practically usable. For us, however, this is only the first step, even an invitation, also to you as a reader, to take the next steps.

Distribution and improvement

The guidance ethics concept gives the opportunity to think about technology and human values in a different way. In discussions, in newspaper articles and conversations about innovations, we notice that even today, a fundamental separation between humans and technology is almost always assumed, a 'top-down' thinking, with 'technology' as the starting point, instead of 'bottom-up' thinking, by putting many different technologies in a specific context. The methods used to deal with ethical dilemmas are also often based on the traditional scheme.

We notice that the concept of guidance ethics catches on and is recognised, but also that it requires a strong dialogue to master this way of thinking. An intense and extensive debate about this is needed.

This debate about guidance ethics is not just an academic discussion, it is extremely relevant to practice as well. This is the time where digitization developments find their place in society, in all areas. For example, artificial intelligence (AI) has long been a promise and is now beginning to deliver on it, with all the ethical questions that come with it. And if we want to continue to innovate, we need a method that guides us, that innovates along in terms of ethics. And that is the guidance ethics approach.

Wide use of the guidance ethics approach

We developed the guidance ethics approach, based on theory and by using cases. We believe that the approach has a solid base and we know that it must continue to develop. That can only be achieved by implementing it in practice. There are still many aspects about which much can be learned. For example the use of the workshop within organisations, the impact of workshops, the coaching of the workshops, the delineation of the case, the way in which the results can be used in a comparable situation, the power of the approach for a social discussion.

We think the approach is mature enough to benefit organisations or processes that want to use it. At the same time, we want to keep improving the approach. We hope to be able to assist in this and to build a supporting network.

Ethics factory with options for action

If the guidance ethics approach is implemented regularly, many options for action will arise: options for action for technical design, for adapting the environment and for better use. These options for action are probably more useful than just for that one particular case. That is why we have the ambition to collect the options for action, so that others can also use them. That brought us to the term ethics factory.

The point is the understanding that options for action can be used in a wider context, become available to other parties and that there are many options. Distribution can of course also be targeted. For example: let's say that a healthcare institution identifies different options for action for the feeding robot. Then it is interesting to use those options and find further options in other institutions, and also together with other developers of the feeding robot and other users, so that the solutions found are widely implemented and shared.

Use the guidance ethics approach, take part in the ethics factory

About the authors

Daniël Tijink

Project manager ethics & digitalisation and member Management Team, ECP | Platform for the information Society

Peter-Paul Verbeek

Professor of philosophy of technology and co-director DesignLab, University of Twente

Appendix 1

Participants working group digitization and ethics

Appendix 1 – **Participants working group digitization and ethics**

Pallas Agterberg	Alliander
Brigitte Boon	Siza, Academy Het Dorp
Hans Bos	Microsoft Nederland B.V.
Roxane Daniels	Vereniging Nederlandse Gemeenten (VNG/ Association of Netherlands Municipalities)
Valerie Frissen	SIDN foundation
Edo Haveman	Facebook
Jos Huigen	KPN
Esther Keymolen,	Leiden University
Lana Klok	ECP | Platform voor de InformatieSamenleving
Natalja Krijgsman	ECP | Platform voor de InformatieSamenleving
Anja Lelieveld	Ministry of the Interior and Kingdom Relations (2019)
Roelof Meijer	SIDN
Rob Nijman	IBM Nederland BV
Marc Noordhoek	Ministry of the Interior and Kingdom Relations
Bettine Pluut	Erasmus University Rotterdam
John Post	Topsector Energie
Sabine Roeser	Delft Technical University
Daniel Tijink	ECP | Platform voor de InformatieSamenleving
Arie van Bellen	ECP | Platform voor de InformatieSamenleving
Erik van de Poel	Royal Family Service
Jeroen van der Ham	National Cyber Security Centre (NCSC)
Theo van der Plas	National Police
Sandra van der Weide	Ministry of Economic Affairs and Climate Policy
Angeline van Doveren	Rijnstate Hospital
Peter-Paul Verbeek	Twente University
Focco Vijselaar	ministry of Economic Affairs and Climate Policy

Appendix 2
Read more

Appendix 2 – **Read more**

ECP | Platform voor de InformatieSamenleving, *Het verhaal van digitaal, samen vormgeven aan onze digitale samenleving* (The story of digital, shaping our digital society together), ECP, 2018

ECP | Platform voor de InformatieSamenleving, *Artificial intelligence impact assessment*, ECP 2018

Esther Keymolen and Simone van 't Hof, *Can I still trust you, my dear doll? A philosophical and legal exploration of smart toys and trust*, Journal of Cybersecurity, issue2, vol 4, 2019, p. 143-159

Sabine Roeser, *Risk, Technology and moral emotions*, Routledge 2018

F.E. Scheepers, V. Menger, K. Hagoort, *Datascience in de psychiatrie* (Datascience in psychiatrics), Tijdschrift voor psychiatrie 60(2018)3, 205-209

Peter-Paul Verbeek, *Op de Vleugels van Icarus, hoe techniek en moraal met elkaar meebewegen* (On the Wings of Icarus, how technology and morality move together), Lemniscaat, 2014

Peter-Paul Verbeek, *De grens van de mens, over techniek, ethiek en de menselijke natuur*, Lemniscaat, 2011

Jan Van Zanen e.a. *Verlenging CITY DEAL Zicht op Ondermijning (nr 48699)*, Staatscourant (Government Gazette), 14 February 2019

The ECP working group on digitization and ethics notices that the use of technologies causes more and more ethical questions. As a participant of the working group, representatives from governmentental parties, businesses and science aim to contribute to the issues addressed in these questions, in order to create a mutual and fruitful handling. This has led to the guidance ethics approach. an approach applicable in practice. The approach consists of three phases: description of a technology in context, a dialogue about values and effects with stakeholders and formulating concrete options for action.

The digital version of this document can be found at:
https://ecp.nl/publicatie/guidance-ethics-approach/

www.ecp.nl