




TU DARMSTADT, INSTITUTE OF PHILOSOPHY
MASTER THESIS, PHILOSOPHY OF TECHNOLOGY

HOW DOES RESPONSIBLE RESEARCH & INNOVATION APPLY TO THE CONCEPT OF THE DIGITAL SELF, IN CONSIDERATION OF PRIVACY, OWNERSHIP AND DEMOCRACY?

FIRST SUPERVISOR: PROF. DR. ALFRED NORDMANN (TU DARMSTADT)
SECOND SUPERVISOR: PROF. DR. JOCHEN KLUCKEN (FAU ERLANGEN-NÜRNBERG)

AUTHOR: SIJMEN VAN SCHAGEN MA

WS 2020/2021
STUDENT NUMBER: 2790787
SIJMENVANSCHAGEN@GMAIL.COM
28.01.2021



Contents

1 Executive Summary	2
2 Introduction	3
3 Scope and definitions	6
3.1 Digital Self	7
3.1.1 First layer of the DS: extension	8
3.1.2 Second layer of the DS: autonomy	12
3.1.3 Third layer of the DS: artificial intelligence	15
3.2 Responsible Research and Innovation	17
3.2.1 Underlying values and rights of RRI	18
3.2.2 My concept of RRI	18
3.2.2.1 Privacy	19
3.2.2.2 Ownership	20
3.2.2.3 Democracy	23
4 Relations between RRI and DS	25
4.1 In terms of Privacy	25
4.1.1 RRI and AI-related privacy issues	26
4.1.2 RRI and the privacy of the DS	28
4.1.3 Along the lines of informational self-determination	30
4.2 In terms of Ownership	31
4.2.1 RRI and the DS in relation to digital sovereignty	32
4.2.2 Along the lines of informational self-determination	33
4.3 In terms of Democracy	34
4.3.1 How AI affects democracy	35
4.3.2 Along the lines of informational self-determination	36
5 Case Study: how RRI meets the DS	36
5.1 Case of Digital Health: the Profile Basic App	37
5.2 Design approach of guidance ethics	40
5.2.1 Principles	40
5.2.2 Methodology	42
5.2.2.1 Step 1	43
5.2.2.2 Step 2	44
5.2.2.3 Step 3	46
6 Discussion & Conclusion	47
6.1 Discussion	48
6.2 Conclusion & Outlook	49
List of used literature	54
List of used online sources	59

1 Executive Summary

This master thesis studies to what degree Responsible Research & Innovation (RRI) can be applied to the concept of the Digital Self (DS). In order to examine this properly, it focuses on aspects of privacy, ownership and democracy. This work is inspired by the digital health domain, where a growing number of patients become enabled to benefit from AI-powered clinical decision support. Aim of this study is to provide insight into what cases can be considered for exploring new design requirements for digital health applications.

Increasingly, we use technology to manage our daily digital selves. Our smartphones, being examples of extensions of ourselves, are able to empower us as citizens within a democracy. The decision-making process belongs more and more to the realm of artificial intelligence (AI), powering our extended selves with algorithms. Although the era of the DS has already begun, it remains unnoticed by the paradigm of RRI. Instead of empowering informational self-determination, RRI tends to be overprotective of an outdated concept of what it is like to act as a human within a democracy which is obviously reshaped by digitization.

In order to maintain our general understanding of a democracy, there is a need for a concept of the DS in the near future. Not only will the DS-concept have the capacity to empower a democratic citizen, it will also be able to empower patients within a digital health environment. Due to the data hunger of large technology companies and governments, it has already become clear that the rise of digitization comes with a price for our society and our health. Therefore, it is the DS-concept that can make a difference in contributing to ethical and social standards. If the future of digital health offers us – additional to human medical stakeholders – AI-powered digital health apps to increase the quality of healthcare, we have the duty to take our values into account when designing such apps.

2 Introduction

With the growing impact of new and emerging science and technologies (NEST¹) on our society, for example AI, the ethical aspects of this impact come to public debate more and more often. During the recent Covid-19 debate, the Corona-Warn-App has been one of the most sensitive topics of discussion in Germany - similarly in other democracies around the globe². It highly touches upon privacy-, democracy- and ownership issues dealing with valuable personal data. Evidently, Covid-19 has significantly speeded up developments in digital health.

Nowadays, a majority of the world population is active online³. The smartphone, including all its (social networking) applications, might be the best example of a tangible object that uses algorithmic activity on a continuous basis to power our digital daily selves. It has truly become a mediator of our “thoughts and expressions and intentions and actions”⁴, increasingly making predictions and decisions on behalf of our physical selves. In other words, human agency is being reduced and “technological autonomy has developed into a system that takes over important decisions” (Demetis & Lee 2018). Including technologies such as 5G and the Internet of Things (IoT), smartphones and other wearable devices make key tools empowering citizens in data driven digital economies (EIT Digital 2020). That is why the smartphone is used in this thesis as a constant recurring concrete example of how the DS would be able to manifest in society.

Specifically, the domain of digital health⁵ has been influenced by new technologies like AI, as huge savings and major benefits are foreseen (Bjerring & Busch 2020). Supporting the decision-making process of both patients, doctors and other clinical stakeholders, AI has in the recent years strengthened its potential, its footprint and its reliability in healthcare. For example, self-learning

¹ This abbreviation is originally found within a Framework Programme description by the European Commission, symbolizing its great impact on society and ethical debates. Please see <https://cordis.europa.eu/programme/rcn/751/en>.

² Please see <https://www.nytimes.com/2020/07/08/technology/virus-tracing-apps-privacy.html>.

³ Please see <https://www.statista.com/statistics/617136/digital-population-worldwide/>.

⁴ Please see https://www.schiffhardin.com/Templates/media/files/publications/PDF/Prewitt_Circuit%20Rider_April2016.pdf.

⁵ Digital health can be described as the field where development and use of digital technologies (e.g. based on AI) with the aim of increasing the quality of healthcare, are brought into practice (WHO 2020).

algorithms designed to early and accurately detect patients with collapsed lungs by an AI-embedded X-Ray System⁶. Or the support of infected Covid-19 patients in shared decision-making throughout a clinical process (Debnath et al. 2020). Thus, a patient - or random democratic citizen – using AI-based (medical) applications on his smartphone increasingly leaves predictions and decisions to algorithms. The resulting view of this in the near future, is what the author refers to as the Digital Self (DS, please see the figure underneath): an autonomous agent (AA⁷) (Franklin & Graesser 1996) performing at an individual level of abstraction (LoA⁸) (Floridi & Sanders 2004), functioning as a digital extension of a (single) physical human being and connected to the identity of this human being (a democratic citizen). In a narrow sense, the DS consists of layers of extension, autonomy and artificial intelligence. These layers enable the working of respectively the dig-

ital identity of the self, its autonomous managing of personal data and the decision-making process.

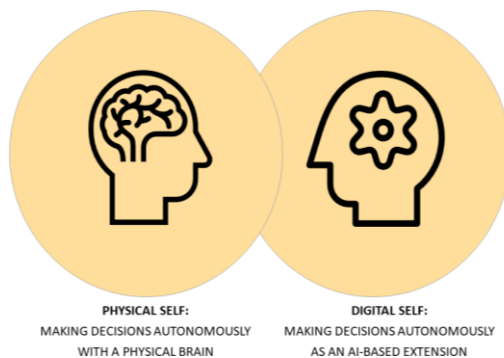


Figure 1: The Digital Self, an extension of the physical self

Using the above as a starting point, the author will first set the scene and build

the concept of the DS by elaborating on a philosophical analogy provided by the mediation theory (MT) by Verbeek (Verbeek 2005). In essence, the MT is an in-depth analysis of human-technology relations, showing that human actions and decisions are fundamentally technologically mediated. In addition, Verbeek uses a post-phenomenological approach to show how we use technology to extend ourselves in several ways.

⁶ Please see <https://www.gehealthcare.com/article/ai-embedded-x-ray-system-could-help-speed-up-detection-of-a-collapsed-lung>.

⁷ "An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future." (Franklin & Graesser 1996)

⁸ "A level of abstraction or LoA is a finite but non-empty set of observables, which are expected to be the building blocks in a theory characterized by their very choice." (Floridi 2004)

Assuming a DS in the near future, the need for protecting this digital extension of an individual would be evident. At first sight, the paradigm of Responsible Research and Innovation (RRI) aims to do this. One of the main motivations of this recently emerged governance approach is to manage ethical and social risks of technological innovation (Martinuzzi et al. 2018). However, RRI focuses on protecting and empowering the individual, in considerations of privacy and democracy, as it appears in our physical reality: a human being with a body including a brain. This human being has the right to be let alone (Warren and Brandeis 1890) and exactly this right has meanwhile become recognized a human right in Europe. In terms of democracy, being the key governance principle that informs the principles and practices of RRI (Gutmann in Stahl 2013), only physical citizens are considered.

Thus, RRI is based on European rights and values and most of these are protecting the (physical) democratic citizen and its personal data. When it comes to the DS, the notion of informational self-determination⁹ (Hummel et al. 2020) comes into play. Within this context, this notion can be explained as every citizen having the freedom what to do with their personal data, including core elements of privacy, ownership and democracy. How is RRI taking informational self-determination into account? And based on that, is RRI actually empowering the DS? As a democratic citizen would leave more and more decisions to his digital extension, the RRI picture is being questioned. Moreover, it may be so the DS is needed to maintain our general concept of democracy.

This thesis aims to examine until what degree RRI takes the DS into consideration, at the same time elaborating on a DS-concept and its potential benefits and concerns. The DS aims to empower the democratic citizen as well as the patient within a digital health environment. Simultaneously, as a DS depends on the working of sophisticated algorithms, one should at least understand their working - how they predict and decide. As transparency and explainability in algorithms have become a serious challenge, this goes for all DS-like tools as a completely new interface between DS and other DS's arises.

⁹ To be "understood as the ability of data subjects to shape how datafication and data-driven analytics affect their lives, to safeguard a personal sphere from others, and to weave informational ties with their environment." (Hummel et al. 2020)

In summary, this thesis would like to find out more about the potential implications of the DS in consideration of privacy, ownership and democracy within RRI. Therefore, the research question to be answered is the following:

How does RRI apply to the concept of the DS, in consideration of privacy, democracy and ownership?

In order to answer the above research question, not only a clear understanding of the question by its terms is essential but also an understanding of how considerations of privacy, democracy and ownership relate to the applicability of RRI on DS. The first part of this understanding will be covered in the next chapter. In addition, the author analyzes the relations between RRI and DS based on the scope and definitions and tries to identify how and until what degree RRI applies to the DS.

The outcome of this study aims to contribute to the concept of RRI, forming recommendations for new design requirements for digital health applications. Therefore, a digital health case is provided as well. The Profile Basic App (PBA) case is on the one hand based on the concept of the DS, on the other hand it can typically be seen as a RRI-example. In order to elaborate on these design requirements more specifically, the case will be subject to a recently developed concrete ethical design approach called the guidance ethics approach.

The final chapter includes an answer to the research question. Insights aim to contribute to future qualitative research in the domain of digital health.

3 Scope and definitions

Before starting an in-depth analysis of the DS, the following remarks are important to mention as the author will not specifically touch upon the related topics. First of all, as the DS is connected to the identity of a single physical human being, it is de facto its digital identity. This solves the problem of assigning responsibility and liability of an AA in case of failures or accidents – for example,

regarding self-driving cars (Chesterman 2020). Second, a short clarification for the 'Self' in DS must be given. The meaning and scope of the self used in this thesis refers to an informational interpretation: "The self is a complex informational system, made of consciousness activities, memories, and narratives. From such a perspective, you are your own information" (Floridi 2014, p.69).

3.1 Digital Self

Human beings increasingly use their smartphones to path their way through the digital world. In addition, most applications on their smartphone are well aware of the data of the user and are being continuously updated. Some of them even calculate, predict and make decisions on behalf of their user, based on their self-learning skills. For example, Uber using AI to determine what price your ride costs, AI-based credit decisions used by financial institutions, self-learning spam filters and Facebook using AI to decide what posts you see.¹⁰

The notion that humans use technological artefacts¹¹ (Verbeek & Vermaas 2009) to extend themselves in some way, is extensively described within the philosophy of technology. Therefore, in the next paragraph the author starts out and uses an analogy from the philosophy of technology to theoretically build the concept of the DS and sets the scene with how the DS is able to be an extension of the physical self. In the following paragraph, the author aims to give meaning to its autonomy. Finally, the third layer is about the main driver of the digital and autonomous dimension of the DS: artificial intelligence (AI). Further on in this thesis, the author aims to analyze how the DS could in fact be part of a democratic citizen who is not yet to be taken into account by the notion of RRI. All of the above-mentioned layers, including corresponding relations to the key considerations of privacy, ownership and democracy, are visualized in Figure 4 at the end of this capital.

¹⁰ Please see <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/>

¹¹ "Technological artifacts are in general characterized narrowly as material objects made by (human) agents as means to achieve practical ends. Moreover, following Aristotle, technological artifacts are as kinds not seen as natural objects: artifacts do not exist by nature but are the products of art." (Verbeek & Vermaas 2009) Throughout this thesis, 'technological artefact', 'tool', 'thing' and 'technology' all mean the same and are therefore randomly used.

3.1.1 First layer of the DS: extension

Explaining the concept of the DS starts with the analogy of the extended body, which has been firmly touched upon by the theories by Martin Heidegger (1962) and Maurice Merleau-Ponty (1962). Both are known to use a phenomenological approach¹². At a later stage, the Dutch philosopher of technology Verbeek (2005), inspired by the post-phenomenological approach of the American philosopher Don Ihde, builds further onto the theory of the extended body.

Starting out with Heidegger, it becomes clear that the early Heidegger – before ‘the turn’ in his thinking known as *die Kehre* - is the most relevant for analyzing his share in the analogy of the extended body. In short, in his 1927 published magnum opus *Being and Time*, Heidegger sets the scene by analyzing “the relation between human existence and its world” (Verbeek 2005, p.77). The key term he uses to describe that relation is “being-in-the-world”. Here the role of concrete technological artefacts, *Zeug* or equipment as Heidegger calls it, becomes significant – as a way of being-in-the-world, a way of being out there enabling this above-mentioned relation between humans and their world.

For Heidegger, technological artefacts “disclose a world” (ibid. p.79). This becomes clear when Heidegger analyzes the behavior from the perspective of the use of tools. He looks at their usefulness, their purpose and directedness towards their potential users, and their role within the public environment. He uses two central terms when looking at the use of technology, namely present-at-hand (*Vorhandenheit*) and readiness-to-hand (*Zuhandenheit*). Concrete example: when actively working with your smartphone, it is ready-to-hand. You do not pay attention to the smartphone itself, but you focus on the activity you are performing on it: in fact, your smartphone is hiding within your relation with the surrounding world. At the moment your smartphone breaks down because the battery is too low, it changes from ready-to-hand into present-at-hand: the organic whole with the surrounding environment of you and your smartphone, suddenly breaks down. Thus, sticking with the smartphone example, it can be contended that as you are unconsciously using it and it is ready-to-hand, it becomes an extension of your body.

¹² “Phenomenology is the study of structures of consciousness as experienced from the first-person point of view. The central structure of an experience is its intentionality, its being directed toward something, as it is an experience of or about some object.” (Woodruff Smith 2013) Please see <https://plato.stanford.edu/entries/phenomenology/>.

The topic of bodily extension has always been close to the field of expertise of the French phenomenological philosopher Merleau-Ponty. His phenomenological theory of embodiment and perception focuses on the body and its consciousness as a way of perceiving and experiencing the world and as its point of reference. "For Merleau-Ponty the body cannot be viewed purely and as an object that is a collection of organs, fluid, muscle, bone, skin and so forth. Rather it is a schema that is an all-perceiving union of mind and body and is the bodily intentionality of being in the world. Embodiment for Merleau-Ponty therefore is a holistic sensing of the world not specific senses operating independently but in unison everything combines to interpret the life-world... For Merleau-Ponty, artefacts and technologies become part of the body schema." (Irving 2019, p.109)

Furthermore, Merleau-Ponty is able to give concrete examples to show how the extension of the body can take place (Verbeek 2005). He mentions a woman with a feather on her hat, who instinctively feels "where the feather is just as we feel where our hand is" (ibid. p.124). Although complementary in terms of considering relations between humans and their world, rather than Heidegger Merleau-Ponty assesses the way these relations can take place based on the presence of technological artefacts. Using our smartphone as an example again, it extends the spatiality of our body and at the same time our body perceives the world with it – e.g. when we are performing telemedicine with our therapist, we perceive the presence of our therapist. In other words, while making a video call our intentional relation with the world – in this case our therapist – is extended through our smartphone.

Yet, the above-mentioned examples and analyses of both Heidegger and Merleau-Ponty are pointing to the process what Verbeek calls technological mediation. The basis of his mediation theory (MT) is to focus on what technology 'does' with regard to human-technology relations, by "thinking forward": how it influences existence and society (Verbeek 2011). Technology should not be seen as objects (versus human subjects), rather as media – channels between humans and the world around them. When we use technology, Verbeek claims it shapes all kinds of relations between users and their environment. For example, medical robots who organize how we give

care to patients. In essence, the MT says we are all fundamentally mediated beings. At the heart of the MT lies the thought that technologies do not simply create connections between users and their environment, but actively help to constitute them.

To support this line of thinking, Verbeek uses phenomenology and post-phenomenology. According to Verbeek, human-technology relations with the world have a hermeneutic and an existential dimension: through technologies, human beings are present in the world and the world is there for human beings. Technologies, in other words, help to shape human experiences and practices. Verbeek uses a phenomenological perspective to analyze this human-world relation, overcoming the subject-object twofoldness. That way, it becomes clear that phenomenology focuses on the relations between humans and their world – as human beings are always directed at the world, experiencing it and acting in it.

Subsequently, Verbeek uses the post-phenomenological approach of the American philosopher Don Ihde (1993) to not only interpret the relations between humans and their world, but also include technologies as a mediator in those relations. The post-phenomenological approach maintains that human beings and the world constitute and co-shape each other. According to post-phenomenology, reality shows itself in relations and considers these relations as mediated. Technologies help to shape relations between humans and world, and in doing so they also help to shape how we are human beings and what the world means to us.

Thus, Verbeek claims technologies are part of our relations with the world. To show how technology is there between humans and the world, Ihde distinguishes the following four different intentional human-technology relations, each represented in their own scheme and with a typical example (Figure 2).

TYPE OF RELATION	SCHEME	EXAMPLE
Embodiment ¹³	(human - technology) → world	Pair of glasses, 'through' which you look to the world
Hermeneutic ¹⁴	human → (technology - world)	A thermometer, giving us an interpretation of the world
Alterity ¹⁵	human → technology (- world)	A cash machine, with which we interact, the technology being the other and the world behind not being important
Background ¹⁶	human (- technology - world)	A light, contextualizing our perceptions

Figure 2: Human-technology relations (Verbeek 2011, p.143)

According to Verbeek, contemporary technologies challenge Ihde's framework (1990). This is the case at both extremes of his framework, so on the side of the embodiment relation and on the side of the background relation. Let us focus on the side of the embodiment relation, as this relation can be used for conceptualizing the DS. On the side of the embodiment relation, Verbeek discovers a cyborg relation¹⁷. As "...this form of intentionality takes us into the realm of the "trans-human...it is located beyond the human being." This intentionality is "not entirely human" (Verbeek 2005, p.144) and basically that goes for both relations. In fact, both the cyborg relation and the embodiment relation contain characteristics that could be attributed to DS. However, the key difference between the embodiment relation and the cyborg relation lies in the transparency, whether the distinction between human and technology within the mediated experience can be made: in cyborg relations this distinction cannot be made anymore.

¹³ Embodied relations require transparency, based on conditions of technical serviceability, human skill and aiming for specific "mediated perception" (Verbeek 2005, p.126).

¹⁴ Hermeneutic relations require interpretation. In these relations "we are involved with the world via an artifact, but the artifact is not transparent" (Verbeek 2005, p.126)

¹⁵ "In alterity relations humans are not related, as in mediating relations, via a technology to the world; rather they are related to or with a technology." Technology here takes the role of the "quasi-other" (Verbeek 2005, p.126-127).

¹⁶ "...technological artefacts in background relations do not play a central role in our experience." (Verbeek 2005, p.128)

¹⁷ In this relation humans merge with technology, having a relation with the world. The difference with the embodiment relation of Ihde is the fact that contemporary technology, a brain implant for instance, merges with the human body instead of still being able to be distinguished from the human body (like the example of the pair of glasses) (Verbeek 2011).

A concrete example of the grey area between an embodiment relation and a cyborg relation within the digital health domain, is the MS Sherpa app¹⁸. In case of being a multiple sclerosis (MS) patient, you are able to use the MS Sherpa app on your smartphone, possibly connected to technology (e.g. wearables) attached to your body, to see how your daily lifestyle choices work out on the development of your disease. These behavioral kinds of choices can be made for you.

Using the above example, it becomes clearer that the DS might belong to the realm of the cyborg relation. At first sight, it seems the smartphone is used for personal experiencing, “broadening the area of sensitivity” of the patient’s body to the world. The smartphone “withdraws from my perceiving.” Using my smartphone as a telephone, it would still be an embodied relation as there is a distinction between on what my role is (talking through my phone) within the mediation experience. The smartphone with the MS Sherpa app, however, seems to extend the power of the patient’s cognition, predicting the course of disease via its artificial intelligence. In doing so, its algorithm uses the patient’s personal data (in both quantitative and qualitative ways) and blends cognitive outputs into its own calculations. Therefore, a distinction between the patient and the technology in the mediated experience cannot be clearly made anymore.

3.1.2 Second layer of the DS: autonomy

Evidently, the digitally extended self is autonomous. What is this autonomy like then? Evidently, it would consist of both human and nonhuman autonomy – the latter referring to the DS-layer of artificial intelligence. Presumably, the autonomy of the DS would always be limited, as it depends on its technical extension. How is this autonomy even possible?

To understand the concept of autonomy as simply the freedom of a random individual to take its own decisions would not dispense justice to the complexity, sensitivity and philosophical dimension of autonomy. In many ways it is a controversial and much debated concept. A more profound definition of individual autonomy would be the “idea that is generally understood to refer to the

¹⁸ The MS Sherpa app is a mobile health application for MS patients, which can be used for home monitoring in between clinical visits. It is developed by various institutes and third parties and is based on scientific research. Please see <https://www.mssherpa.nl/>.

capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces, to be in this way independent" (Christman 2020). In other words, autonomy can be understood as a form of self-government, where one acts by means of free will, based on one's own principles and values.

The American philosopher Joel Feinberg (1986) has an often cited and rather versatile definition of autonomy, including four different meanings:

1. Autonomy being a self-governing capacity
2. The actual state of autonomy
3. Autonomy being an individual ideal
4. Autonomy being a right standing for self-sovereignty

As mentioned before, the self-governing capacity of the DS is already limited by means of its technology. In other words, as my digital extension seems a mixture of my physical-me and the technology I use to 'be' and 'act' within the digital domain, the technology determines my autonomy at least for a (significant) part. This already makes the first point mentioned above a relevant one. Point two and three seem only indirectly relevant in building the DS-concept and its autonomy in particular, as they can be explained as rather personal features. In point four, autonomy in the sense of self-sovereignty, which can be defined as "the quality of living in accordance with one's inner nature or genius" (Trotter 2014, p.244), is a key point. Both the concept of ownership and informational self-determination are directly related to this: in the next chapter this will be elaborated on more deeply.

Starting out with the self-governmental aspect of autonomy, which seems to be the one mostly connected to technology, it is essential to take moral agency into account: when humans act ethically, they are self-governed. According to mainstream ethical theories, autonomy itself is a key feature of moral agency (Verbeek 2011). For example, the Kantian definition of autonomy¹⁹ is

¹⁹ Kant sees the concept of autonomy in relation to ethics: while acting in the everyday life, one has the moral right and capacity to make decisions in order to govern his or her own life (Sensen 2013).

basically the ability to adapt law to oneself and the capacity and space one has to abide by that law, making decisions: the more one is able to adapt this self-law, the better one is able to attribute moral value to one's actions. Just like Kant, Verbeek uses autonomy in the meaning of freedom.

As Verbeek proves with his MT that moral agency is a starting point of technological mediation and therefore highly influencing decision-making, it seems that it is not only for humans but also for technology possible to possess a certain amount of autonomy. In other words, as we rely on technology so much in our daily (digital) lives, it plays a significant role in our decision-making process. Evidently, autonomy is being "distributed among the human and nonhuman elements in human-technology associations" (ibid. p.60). In short, technology actively supports creating and shaping our autonomy. Self-sovereignty, in the sense of informational self-determination, is a good example of how technology does so. As large tech firms like Facebook have been scraping and exploiting the data we share online for nearly decades²⁰, it has become clear that the borders of our informational self-determination are being seriously challenged.

The above-mentioned Facebook example, in combination with the hybrid nature of autonomy, typically connects to the notion of autonomy by Foucault (Foucault in Verbeek 2011). He includes both nature and nurture as factors in forming autonomy, being a sort of self-mastery. Nurture, in this matter, includes interaction with technology. Therefore, main characteristics of the Foucauldian autonomy are developing "a new relation to power" and "a free relation to technology" (ibid. p.85). When these features are seen in the light of the above example of the dominant world power of large technology firms like Facebook, it becomes even more evident that the autonomy of the DS has a hybrid character. In other words, a random individual would also develop his autonomy in terms of self-sovereignty by experiencing interactions while using a highly influential social platform like Facebook; as the DS of this random individual potentially (partly) takes over these interactions, the hybrid character of autonomy becomes even more clear.

²⁰ Please see <https://medium.com/swlh/facebook-scraping-still-a-privacy-disaster-c70dd1896286>.

In the next paragraph, the layer of artificial intelligence of the DS will be elaborated on. AI is able to support our autonomy in the sense of “who we can become (autonomous self-realization)” (Floridi et al. 2018, p.691) within the frame of human flourishing, thus able to support the autonomy of the DS. By means of building a bridge, I would like to already introduce a workable definition of the autonomy of a morally acting artificial agent. For Floridi and Sanders (2004) this autonomy “means that the agent is able to change state without direct response to interaction: it can perform internal transitions to change its state. So an agent must have at least two states. This property imbues an agent with a certain degree of complexity and independence from its environment.” (ibid. p.357) Clearly, this definition implies self-governance being a key aspect of the autonomy of the DS.

In addition, a more general critique on AI e.g. used in Google’s services is that we gradually lose our autonomy by making use of those (Bar-Gil 2020). This critique can be countered by looking from the perspective of regaining autonomy by using the DS in the sense of governing our data in a more ethical way. In other words, personal data governance by the DS is able to increase our digital sovereignty²¹, being able to compensate the loss of autonomy we are suffering in our current onlife²² situation.

3.1.3 Third layer of the DS: artificial intelligence

“In information societies, operations, decisions and choices previously left to humans are increasingly delegated to algorithms, which may advise, if not decide, about how data should be interpreted and what actions should be taken as a result. More and more often, algorithms mediate social processes, business transactions, governmental decisions and how we perceive, understand, and interact among ourselves and with the environment” (Mittelstadt et al. 2016, p.1). Although the exact consequences of this algorithmic mediation are unclear, it is obvious that AI will have a “profound impact on human flourishing” (Umbrello & Van de Poel 2020, p.2).

²¹ “Digital Sovereignty is the ability of an entity to personally decide the future form of identified dependencies in digitalisation and to possess the necessary powers.” (Steiner & Grzymek 2020)

²² Term used by Floridi (2015) to articulate how our online profiles dominate our offline life, increasingly powered by smart and responsive technologies.

Starting off with a relatively basic but relevant definition of AI, it can best be described as “computers that mimic cognitive functions that humans associate with the human mind” (Russell & Norvig 2009). Over time, it has been hard to provide a more clear and narrow definition, as it heavily depends on the context what AI means. In addition, its meaning changes over time (Buiten 2019). Decision-making would be the most relevant example of a cognitive function as mentioned above, in this context. In essence, the decision-making process can be understood as an ongoing cognitive process, which is both conscious and complex on the one hand and rather automatic on the other (Cervantes et al. 2015).

Obviously, the AI-based calculations to make autonomous decisions are done by algorithms. Algorithms are formally described as the mathematical construct with “a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions” (Hill in Mittelstadt et al. 2016, p.2). The type of algorithms which would do the work for the DS are in such a way sophisticated, that they are able to make plausible decisions based on a complex calculation process, not necessarily easy to understand by human beings. The research field that focuses on training these algorithms e.g. to recognize patterns in data and subsequently generating a model doing predictions based on the same data, is called machine learning. It is machine learning that “is defined by the capacity to define or modify decision-making rules autonomously.” In other words, a machine learning algorithm obtains autonomy by learning.

Connecting the above to the former given definitions of the DS and of the self, it is a continuously learning AA acting as an ever more intelligent digital self. As it is connected to the physical self of a single human being, the DS is always able to be directly or indirectly supervised by its corresponding physical self. This is also the aspect that gives the DS a digital identity. Therefore, the DS is an AA best described as a digital personal trainer or a digital personal assistant: it is an extension of the physical self that needs autonomy in order to be able to make decisions. From this perspective, the aim of AI powering the DS is delivering augmentation rather than automation (Djeffal 2019). This means that the aim of AI used for the DS is not to replace the tasks of the physical self, but to strengthen the interaction between the physical self and DS by ways of adding on to human intelligence.

Particularly, the DS is able to make predictions and decisions on its own but can, for example, also function as a recommendation system (RS). The main goal of the RS is giving its user highly relevant suggestions as an output, based on an input consisting of metadata on the user's preferences (Milano et al. 2020). For example, when the DS communicates with or functions as a digital health app or assistant (as the DS of a doctor) it can help us to make constant decisions to stay healthy: this is a form of self-nudging that still requires human autonomy, but where AI e.g. supports in the implementing part of the decision-making process.

3.2 Responsible Research and Innovation

With the rise of nanotechnology in the US in the 90s of the last century and beginning of the current one, it became clear that societal, ethical and governance questions were asked related to the consequences of this emerging technology. Although the resulting upcoming public debate was a rather new phenomenon (Shelley-Egan et al. 2018), it has woken up the worlds of politics and science with regards to new, potentially disruptive technologies for good: the need for innovating responsibly was born.

Although forerunners of RRI e.g. ELSA or ELSI²³ have been widely endorsed within science, RRI has been most ambitiously taken up by the EU as a top-down instrument of policy strategy (Zwart et al. 2014). The EU uses RRI to tackle the so-called Grand Challenges²⁴ of our time. RRI builds onto technology assessment (TA²⁵) and emphasizes ethical reflections of science and technology within society. RRI, in general understood as an "adaptive and anticipatory governance" (Owen & Goldberg 2010) approach, is frequently defined in two different ways. On the one hand in a socio-empirically way and on the other hand in a substantive normative way (Ruggiu 2015). In a general

²³ Ethical, Legal and Social Aspects (or Implications) of Research. ELSA has been more commonly used in the EU, while ELSI has been initiated (through program funding) in the US (Zwart et al. 2014).

²⁴ Within the Horizon 2020 Program of the EU, its focus is on seven Grand Challenges: Health, demographic change and wellbeing; Food security, sustainable agriculture and forestry, marine and maritime and inland water research and the bioeconomy; Secure, clean and efficient energy; Smart, green and integrated transport; Climate action, environment, resource efficiency and raw materials; Europe in a changing world - inclusive, innovative and reflective societies; Secure societies - protecting freedom and security of Europe and its citizens.

²⁵ TA can be defined "as a form of policy research that examines short- and long-term consequences (for example, societal, economic, ethical, legal) of the application of technology." (Banta 2009)

sense, for this thesis the normative definition by Von Schomberg (2011) is most relevant. He defines RRI as “a transparent, interactive process in which societal actors and innovators become mutually responsive on each other, with a view to the ethical acceptability, sustainability and societal desirability of the innovation process and its marketable products”. This definition is commonly used within the EU and the academic field.

3.2.1 Underlying values and rights of RRI

In this thesis, the in-scope aspect of the definition by Von Schomberg is the ethical acceptability, to be understood “as requiring compliance to the fundamental values of the EU charter on fundamental rights” (Walhout & Kuhlmann 2013). In other words, these fundamental rights, which incorporate a set of key European values, should determine the amount of “ethical acceptability of both research and innovation” (Ruggiu 2015). In this thesis, the EU charter on fundamental rights form the scope of RRI in terms of the key concepts of privacy, ownership and democracy. The question remains, as mentioned before, whether these concepts as used by RRI are still up to date when looking at the development of our DS. In the next paragraph I will suggest a more specific and workable definition of RRI.

3.2.2 My concept of RRI

The central concept used to assess until what degree RRI takes the DS concept into account in consideration of privacy, ownership and democracy is informational self-determination. Therefore, my definition of RRI is inspired on the fields of Computer Science and the Ethics of IT and is for a greater part a literally used definition of RRI by Van den Hoven (2017): “RRI is an activity or process which may give rise to previously unknown designs...pertaining to the...conceptual world,...which – when implemented – expand the set of relevant feasible options regarding solving a set of moral problems.” Making a contribution to solve moral problems regarding privacy, ownership and democracy is exactly what the DS is aiming for. Enabling a DS, based on the principle of informational self-determination, is what RRI should aim for.

3.2.2.1 Privacy

Privacy, scoped within RRI rather as the protection of personal data, has been a fundamental right within the EU as of the Lisbon Treaty 2009²⁶. In the EU it is not only considered a fundamental right but also a social value. Article 8 of the European Charter of Fundamental Rights²⁷ is about the “protection of personal data”. It says literally:

1. “Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.”

Ever since large technology firms started to discover personal and behavioral data as a commodity and subsequently exploiting it by selling it to advertisers on a global scale, data has been higher valued than oil. Evidently, in the era of digital technology and its resulting surveillance capitalism privacy has gained significantly in importance (Stahl 2013). With regard to the future concept and impact of privacy, it seems the Collingridge dilemma²⁸ has a fundamental influence. This goes also for the future of other aspects of RRI being discussed in this thesis, namely ownership and democracy. Because it is until a certain degree unpredictable what kind of (unintended) consequences technological innovation will have, it remains a challenge how to ensure a satisfying societal quality of the aspects of privacy, ownership and democracy on the longer term – especially, as societally entrenched technology is hard to steer. Nobody thought of decent privacy regulation in the early days of the personal computer (Sutcliffe 2011). In other words, RRI is “linked to fundamental epistemological limitations.” (Stahl 2013, p.6)

²⁶ Please see https://edps.europa.eu/data-protection/data-protection_en.

²⁷ Please see <https://fra.europa.eu/en/eu-charter/article/8-protection-personal-data>.

²⁸ “When a technology is still at an early stage of development, it is still possible to influence the direction of its development, but we do not know yet how it will affect society. Yet, when the technology has become societally embedded, we do know its implications, but it is very difficult to influence its development. The dilemma is one of the biggest challenges for responsible design and innovation” (Kudina & Verbeek 2019, p.292)

In addressing key concerns of privacy, specific approaches and methodologies have been developed under the RRI umbrella. An example of a specific approach would be a privacy impact assessment or an ethical impact assessment. A concrete methodology and example of a RRI-strategy, incorporating privacy into technology, is called privacy by design²⁹. This methodology will come back in chapter five of this thesis, being part of an innovative way of designing digital health applications.

Privacy is in two opposite ways intrinsically interwoven within RRI: both as a driver for RRI and as a key aspect being part of the responsibilities making sure of the “desirability and acceptability of research and innovation.” (Stahl 2013, p.6) First, privacy is a key driver to set RRI in the spotlight: a lot of RRI issues have to do with privacy. This has accelerated the debate on RRI. At the same time, privacy has become an important part of this ongoing RRI debate. From the perspective of RRI, safeguarding privacy has become one of the key responsibilities which has led to serious legislation and regulating EU wide measures, ending up in the adoption of the GDPR in 2016 (effective in 2018).

In short, privacy is intensively connected to key aspects of RRI being “activities, actors and normative foundations.” (ibid.p.6) Activities contain research, assessments, foresights, stakeholder engagement and awareness and accountability structures. Actors include all kinds of researchers, civilians, educational and legislative organizations, policymakers and professional (non-)public organizations. Normative foundations comprise the concept of responsibility, which can be found in “existing norms and principles of governance” (ibid. p.4) and at a European level mainly to be found in human rights as processed in the European Charter of Fundamental Rights.

3.2.2.2 Ownership

In this thesis, the scope of ownership is data ownership. As this is an inherently complex topic, starting off with a basic definition might make it even more complicated. Taking the words ‘data’

²⁹ “Privacy by design (PbD) is an approach to systems engineering that seeks to ensure protection for the privacy of individuals by integrating considerations of privacy issues from the very beginning of the development of products, services, business practices, and physical infrastructures. It can be contrasted to an alternative process where privacy implications are not considered until just before launch” (De la Torre 2019)

and ‘ownership’ apart, might work better. Data, for example a set of personal data elements, can be owned in the sense of being controlled, accessed and shared. Ownership assumes a certain right to control, a right to access and sharing the rights to control and access. Therefore, data ownership must be about managing data rights. Clearly, there is a large legal dimension to the topic of data ownership, as well as an ethical dimension: these two dimensions are in scope here, as both are important dimensions of RRI too.

As RRI is a governance approach, and data ownership is an example of a data governance process, there must be a perspective on data ownership from the RRI paradigm. But clearly, from a legal point of view there are still no concrete strategy and policy for data ownership at an EU level³⁰. In the previous paragraph on privacy, it has become clear that the foundations of RRI mainly lie in European frameworks like the European Charter of Fundamental Rights and legislation like the GDPR. However, existing European legal frameworks do not or cannot take data ownership into account in a sufficient way.

The high complexity and corresponding uncertainty (Van Asbroeck et al. 2017) of finding a proper legal framework is one of the main reasons for the problem set out above. A short version of explaining this complexity: the problem first comes up as data heaps up in the beginning of the digital era, the early days of the data driven economy within the EU (Duch et al. 2017). Needless to say, the main characteristic of data is that it copies and spreads itself so easily. It is a huge challenge to assert ownership, especially when so many parties are involved in the data value chain, at the same time protecting the legal rights of the true owner of personal data – if there is one. What data ownership makes even more complicated, is the ambiguity in the definition of personal data and the question whether personal data is intrinsically owned by the individual (Van Asbroeck et al. 2017). In addition, the contrast between law in theory and practical implementation plays a large role in the complexity of data ownership.

In the previous chapter, privacy has been scoped as data protection. Data protection is, as the word itself implies, rather concerned with the protection of personal data than with the control

³⁰ Please see <https://medium.com/data-legally/eu-drops-data-ownership-807ca597fd62>.

over data. Data ownership is in that sense an extension piece of data protection: it concerns the possession and responsibility for information – being meaningful data. In essence, “the degree of ownership (and by corollary, the degree of responsibility) is driven by the value that each interested party derives from the use of that information” (Loshin 2001, p.29) In other words, the complexity of data ownership is mainly due to the significant number of stakeholders in the data. The more they claim to be the owner of data and wanting to profit from it, the more responsibility it involves.

From an ethical perspective, the challenge is in first instance to unravel and interpret the complexity of data ownership. Hummel et al. (2020) argue that within “a digitized and datafied life-world, claims to data are indispensable towards claiming fundamental rights and freedoms. These preliminary observations prompt us to clarify what data ownership exactly means, how it is justified, what it tries to achieve, and whether it succeeds in promoting its aims” (ibid. p.2).

Hummel et al distinguish “four dimensions of calls for data ownership”, each dimension including a contrast in itself:

1. “the institutionalization of property versus cognate notions of quasi-property”
2. “the marketability versus the inalienability of data”
3. “the protection of data subjects versus their participation and inclusion into societal endeavors”
4. “individual versus collective claims and interests in data and their processing”

Poles	Main perspective	Claims	Expectations
1 Property–quasi-property	Interplay between individual, rights, and resource	Incidents of (quasi-) property	Control data flows and outcomes of data processing
2 Marketability–inalienability	From the individual to the resource	Freedom whether or not to market what is mine	Benefit from resource, avoidance of harm from selling core aspects of my self
3 Protection–participation	From the resource to individual constitution, flourishing, and integrity	Protection, participation, inclusion	Maintaining a sphere of secrecy, weaving informational ties at one’s own discretion
4 Individual–collective	Interplay between individual, others, and resource	Consideration of interests, needs, and preferences	Harmonization between individual and common good

Figure 3: Poles of data ownership (Hummel et al. 2020, p.21)

For Hummel et al, the central term to assess these dimensions is informational self-determination. This is, amongst others, a starting point “for global discourses on human and foundational rights in digitization.” (ibid. p.8) Evidently, informational self-determination is a central term in this thesis. In the next chapter it will be discussed in a way which lays bare a RRI-deficit, in the sense of how RRI applies to the DS in consideration of ownership.

3.2.2.3 Democracy

As in general, democracy is a process and not a reachable goal, the core element of democracy in scope here is the decision-making process. This can be both as an individual, democratic citizen and as a group of individual stakeholders.

RRI is foremost connected to the theory of deliberative democracy³¹ (TDD), basically the kind of democracy where deliberation forms the key to decision-making. Essentially, RRI uses TDD as a solid basis and connects anticipation, reflection and inclusive deliberation to decision-making processes (Owen et al. 2012). In other words, TDD “may well provide a solution for governance”

³¹ “Although deliberative democracy encompasses a broad spectrum of ideas, the motivational aim of deliberative theory is to legitimize political decisions by creating procedures that allow democratic decisions to be a result of mutual understanding, publicly expressed reason, and broadened political inclusion” (Brown 2021).

(Reber 2017, p.39) on the one hand, while also covering ethical discussion and (individual) reflection regarding uncertainties raised by new technologies – specifically emerging neurotechnologies – on the other hand.

In fact, it is the TDD that gives most body and potential to RRI than the other way around – and this is not only because RRI is the inheritor of TDD. Deliberation forms “a condition for RRI”, as the TDD “believes that citizens have the capacity to search for and collectively formulate the common good within public deliberations that link common good, justification and legitimacy, and respect the autonomy of citizens.” In terms of the RRI process, this means all relevant stakeholders have “the capacity to justify and perhaps present arguments for their decisions. They expect citizens (or participants) to be able to justify their choices, rather than stick to their frequently vague preferences.” (ibid. p.51) Obviously, TDD feeds the type of empowerment RRI aims to provide democratic citizens by concrete means.

TDD can, in turn, benefit from the precautionary principle³² in the sense of responsibility to give its meaning more body related to technological uncertainty in the future. Thus, a continuous and reflective broad societal debate on the uncertainties of new (e.g. cognitive enhancement) technologies has a high potential influence on the maintenance of the understanding of democracy.

³² The precautionary principle is based on the German *Vorsorgeprinzip*, meaning “a distinction is to be made between human actions that cause ‘dangers’ and those that merely cause ‘risks’: in the case of danger, the government is to prevent these by all means; in the case of risk, the government is to carry out a risk analysis and may order preventative action if deemed appropriate” (Morris 2000, p.1).

Layer	Function	Relation to key considerations
Extension	On the one hand, technology we use becomes unconsciously an extension of our body. At the other hand, it becomes an extension of our relation with the world. The DS is a digital extension of our physical self.	The DS as an extension corresponds well to an extended concept of our <i>privacy</i> , in the sense of managing and protecting personal data.
Autonomy	The self-governing and self-sovereign capacities of autonomy are the most important characteristics here. The first with regard to the ability of the DS to make autonomous decisions, the latter with regard to a true control of its digital identity.	Data <i>ownership</i> matters evolve around personal data governance and data process results. Therefore, data ownership from the DS-perspective is a form of autonomous data processing.
Artificial Intelligence	It is the capacity to learn and think as a human that makes the DS a complete Digital Self. The cognitive competencies to predict, to recommend and to decide are inevitably powered by AI algorithms.	Making autonomous and independent decisions is key within a <i>democracy</i> . In order to maintain our understanding of a democracy radically shaped by digitization, there is a need for the DS and its capacity to make such decisions.

Figure 4: The DS-layers and corresponding relations

4 Relations between RRI and DS

In order to lay bare the RRI-deficits with regard to the DS in this chapter, the relations between RRI and the DS are examined in terms of privacy, ownership and democracy. This is done by using the notion of informational self-determination as the guiding principle, working as a barometer at the background of these relations.

4.1 In terms of Privacy

Although privacy and RRI have been intrinsically interwoven, it has become obvious that the concept of privacy within the scope of RRI is very limited as only a physical person and not their digital extension is considered in the European Charter of Fundamental Rights. During time, privacy has become a rather sensitive topic within the EU, on which the epistemological limitations of RRI might have an inhibitory effect. With the GDPR, the EU has anticipated on this and therefore it can be claimed that the GDPR plays an important role in this renewed conception of privacy related to RRI.

In order to answer the question if RRI truly has a deficit when it comes to privacy as seen from the DS perspective, it is in first instance important to realize the DS is closely connected to a physical person. The DS is, however, autonomous – just as our brain in a sense. It is de facto our digital identity, powered by AI. Thus, the first important question is until what degree privacy related to autonomous agents is considered by RRI. The stance on AI by RRI can be considered as the EU stance on AI, articulated within the “Ethics guidelines for trustworthy AI” put together by the High-Level Expert Group on AI (AI HLEG 2019). Second, the question until what degree the privacy of the DS is taken into account by RRI – considering the imbalance between data protection and data governance within the concept of informational self-determination – is key.

4.1.1 RRI and AI-related privacy issues

Recommendation systems, as used by e.g. Big Tech³³ to give you unsolicited customized suggestions, is the first concrete AI application that pops up in most people’s minds when talking about privacy and is therefore a useful example in this context. Such a recommendation system invades the privacy and the more algorithms based on personal data are used, the creepier people experience these practices (Watson & Nations 2019). Violations of privacy within this perspective have to do with “unfair or otherwise malicious uses of personal data” (Milano et al. 2020). Thus, privacy in relation to AI in any case contains elements of non-intrusion, also known as “accessibility privacy” (Watsons & Nations 2019) which “refers to being left alone and free from government intrusion.”

The next logical question would then be: what kind of AI would take the above into account? The three key characteristics of trustworthy AI, according to the EU (AI HLEG 2019), are lawfulness,

³³ Big Tech is a generic term for five of the largest technology companies in the world, also known as GAFAM: Google, Apple, Facebook, Amazon and Microsoft. “A Big Tech company wants to be the One Stop Shop for your digital (and increasingly analog) life. An advantage in one area gets leveraged for adjacent services from the same company, doing stuff aside their core competency. A known brand and large user base make almost every endeavor reasonable” (Stegmann 2020)

ethicalness and robustness. Ethical AI is built on five principles, namely beneficence³⁴, nonmalificence³⁵, justice³⁶, autonomy³⁷ and explicability³⁸ (Floridi et al. 2018). Guidelines for trustworthy AI entail seven core requirements for AI systems, of which privacy and data governance³⁹, transparency⁴⁰ and accountability⁴¹ are most relevant to mention here.

Why are these so important to take into account? Because there are high risks regarding human dignity at stake (Floridi et al. 2018). Concerning privacy and data governance, not taking consent into account has the risk of unlimited data hunger used for gaining state or corporate control over individuals – whose safety can be in danger or whose behavior can be manipulated. The biggest risk of a lack in transparency is black boxing, meaning that the functioning of algorithms is untransparent and not to be understood by its designers themselves – let alone the public, regulators and researchers. This can possibly do damage to our democracy and our health, for example when introducing a revolutionary AI breast cancer screening diagnostic tool but not sharing the secret behind its methodology⁴².

Related to the transparency risk, a well-known risk of a lack of accountability is bias within AI. In order to deal with this properly, decisions by AI systems need to be placed “in a broader context and by classifying them along moral values.” (Dignum 2018) Bias is, however, a persistent and inherently human characteristic and therefore practically impossible to get rid of. From that perspective critics are questioning the support of AI development as it covers the real field of interest, namely the fact that the underlying data is largely in private hands⁴³ and comes at a societal cost.

³⁴ Doing only good, here meant as “Promoting Well-Being, Preserving Dignity, and Sustaining the Planet” (Floridi et al. 2018)

³⁵ Doing no harm to “Privacy, Security and “Capability Caution” (Floridi et al. 2018)

³⁶ Here meant as “Promoting Prosperity and Preserving Solidarity” (Floridi et al. 2018)

³⁷ Here meant as “The Power to Decide (Whether to Decide)” (Floridi et al. 2018)

³⁸ Here meant as “Enabling the Other Principles Through Intelligibility and Accountability” (Floridi et al. 2018)

³⁹ Includes “respect for privacy, quality and integrity of data, and access to data” (AI HLEG 2019)

⁴⁰ Includes “traceability, explainability and communication” (AI HLEG 2019)

⁴¹ Includes “auditability, minimisation and reporting of negative impact, trade-offs and redress” (AI HLEG 2019)

⁴² Please see [Researchers take issue with study evaluating an AI system for breast cancer screening \(medicalx-press.com\)](https://www.medrxiv.org/content/10.1101/2020.03.10.20048888v1)

⁴³ Please see [The Seductive Diversion of ‘Solving’ Bias in Artificial Intelligence | by Julia Powles | OneZero \(medium.com\)](https://www.onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-1e1e1e1e1e1e)

In summary, it is the combination of always (humanly) understanding exactly what data is used and how a decision is supported or made within the AI decision-making process. For Morley et al (2019) these components form the guiding principle of explicability within ethical AI to prevent recommendations playing a large privacy invading role. This principle of explicability does work out for privacy matters relating to AI, seeing RRI from the perspective of the GDPR. However, it becomes obvious that the latter has prioritized “privacy and explicability over the promotion of autonomy in design choices”: it provides information for individuals what is being done with their data, but does not help “developers provide meaningful explanations that give individuals greater control over what is being inferred about them from their data.” In other words, what privacy means in theory is being considered towards the individual, but in practice the individual has been left to his or her lot in terms of how privacy within the AI-powered digital world actually works out.

4.1.2 RRI and the privacy of the DS

As in the previous paragraph it has become clear that AI-related privacy concerns are for an important part tackled by explicability (in the sense of transparency and accountability), this in fact touches upon the DS in the way that it is being protected against algorithms but not empowered by algorithms as it is. For example, the EU report on Trustworthy AI (AI HLEG 2019) does mention the limitations and duties of AI systems e.g. that they “must be identifiable” but does not consider their benefits for citizen empowerment. In short, RRI does shape a safe landscape for the DS communicating with other DS’s and other AI systems but it clearly fails to take its empowerment (in fact its existence) into account.

In addition, the question rises what the risks of a lack in privacy, transparency and accountability imply for the DS itself – both internally (e.g. in DS-configuring processes towards the physical identity it belongs to) and externally (e.g. in dealing with the outside world or other DS’s). Here the author focuses mainly on external implications, as these are most relevant in the context of the thesis.

The risk of a lack in privacy and data governance would evidently mean a lack of autonomy of the DS, enabling manipulation by whatever external powers – with possible consequences also for the physical self. Privacy and data governance of the DS are in this regard and until a high degree safeguarded by means of consent, as mentioned in the previous paragraph. Consent is a method to preserve autonomy. As there are various forms of consent and the concept is getting more sophisticated, dynamic and adapted to digital reality every day, a positive effect of consent on the autonomy of the DS can be identified. Specifically, dynamic consent⁴⁴ would be suitable for being dealt with by the DS as “the goal for AI consent should be one of partnership development between parties, built on responsive design and continual consent” (Jones et al. 2018, p.64). Thus, key characteristics of dynamic consent include “fully informing” the participant as well as engaging the participant on a continuous basis (Dankar et al. 2020).

In terms of transparency, black boxing by the DS would seem a contradiction as you have to know how your DS comes to a decision. Clearly, black boxing would be undesirable, alienating and out of the question. Thus, it seems obvious that algorithms powering the DS need to be completely known, checked and reproduceable as both the connected physical self and external entities need to understand the DS` decisions. At first sight, this makes a fair and hard requirement for the design of the DS. On second thought, the physical self includes a brain which has a line of reasoning we do not completely understand either: how do humans make decisions? It does not necessarily take Kahneman (2011) to realize human beings basically act irrational, switching continuously between on the one hand impulsive and instinctive thinking and on the other hand slow and logical thinking. The first kind, accordingly to Kahneman known as “System 1 Thinking”, would be most relevant for typing the DS in its early days in the sense that it acts like a rather fast automatic intuitive pilot, subconscious and unable to really motivate its decisions. System 2 is more associated with “complex computations” and related to making “deliberate choices between options”.

⁴⁴ “Dynamic consent is a strategy that is oriented to involve participants, support the principle of informed consent, and solve the stationary aspect of consent. It is designed to support personalized consent via a technological construct at its base in the form of a communication platform. And it is upon this construct that it aims to facilitate the consent process; specifically by establishing a continuous two-way communication between investigators and participants” (Dankar et al. 2020, p.914).

Therefore, System 2 seems to be related to the more advanced DS, being consciously able to explain its reasoning. The latter precisely supports the argument that expecting full transparency of the DS' algorithms is only realistic on the longer term.

Accountability relates mostly to the responsibility of taking moral values into account within the process of deliberation. In other words, the ethical kind of decision-making with a focus on explaining and justifying actions and decisions. Clearly, this would require an advanced, strong AI-powered⁴⁵ kind of DS, as it needs “both the function of guiding action (by forming beliefs and making decisions), and the function of explanation (by placing decisions in a broader context and by classifying them along moral values)” (Dignum 2018). The fact that the DS is connected to the physical self, does not have to mean that it represents the same moral values. As it is acting autonomously, moral values by the DS can be derived from external societal experiences too – just like the physical self seeks moral growth based on nature and nurture. This would imply a certain extended form of consciousness, which sceptics (still) highly doubt if ever possible (Clark & Chalmers 1998).

4.1.3 Along the lines of informational self-determination

Evaluating the DS and RRI relation with regards to privacy along the lines of informational self-determination, it becomes obvious as mentioned in the beginning of this paragraph that RRI covers the passive component (protection of personal data) but is less concerned to take the (pro)active component (data control as part of personal data governance) into account. Clearly, the DS relies on both components as it is built for being empowered by an autonomy ruled by algorithms. Consent seems, however, a key to mold these two components into one. It might be a very basic form of data governance, but it is (and can be in future, based on personal data-store concepts like Tim Berners-Lee's SOLID⁴⁶) indeed more: primarily, meaningful (dynamic) consent

⁴⁵ Strong AI is also called artificial general intelligence. “Realization of strong AI is not yet within reach: The objective underlying the concept of strong AI is to allow natural and artificial intelligence media (e.g. humans and robots) to establish a level of mutual understanding and trust when working in a joint field of activity. Thus, efficient human-machine collaboration could be learned and facilitated, for example. Strong AI is able to independently recognize and define tasks and independently develop and expand upon knowledge in the corresponding application domain. Strong AI studies and analyses problems in order to find an adequate solution, and the problems can also be new or creative” (FHWS 2021)

⁴⁶ Please see [Home · Solid \(solidproject.org\)](https://solidproject.org/)

on access and use forms the individual data sovereignty. The latter is an important part of personal data governance, which is in scope of a main approach on personal data control here (EIT Digital 2020). Key elements of this approach include “support for the deployment of platforms or application ecosystems based on personal data-stores and introduction of expiration dates for exclusive access to some data assets” and encouraging “cloud infrastructures for personal data-stores, foster agreements/standards on the structure of personal data stored on online social networks and other online platforms and support personal data portability across online platforms.” (ibid. p.28) In short, new forms of consent like dynamic consent can become a key part of a decentral approach to tame the dominate data slurping mechanism by large technology companies and governments, enabling citizens to take control of their own data or have their data controlled – adequately and upon their consent – by their DS’s.

With regard to transparency, informational self-determination makes clear that it is essential to know upfront what is going on within the DS when it is predicting and deciding. Obviously, one wants to know exactly what personal data is concerned and subsequently is being transferred in the digital world. However, as previously mentioned within the physical self, a lot of decisions are – despite being switched off with analytical thinking – taken based on impulsive thinking, gut feeling or intuition: not only the easy decisions either. From the latter perspective, informational self-determination is not a valid principle at all.

The common denominator between accountability and informational self-determination lies in the responsibility and the capacity to decide for yourself. Presumably, based on powerful AI and memory, the DS is able to have a positive effect on both the responsibility and (mainly) the capacity to decide.

4.2 In terms of Ownership

As discussed in the previous chapter on data ownership, it has become clear that there is more to this concept than only the legal matters e.g. the right to full data ownership. Exactly this becomes obvious when studying the four mentioned calls for data ownership, particularly highlighted within pairs of poles as already mentioned in the previous chapter, by Hummel et al (2020). Both

the poles of property versus quasi-property and protection versus participation, touch upon key points in relations between RRI and the DS from the perspective of data ownership. The next paragraph will cover this topic.

4.2.1 RRI and the DS in relation to digital sovereignty

Essentially, the pole of property versus quasi-property is most relevant within the context of the DS and deals with controllability of data. Typically, controllability is about sovereign personal data access and data processing and therefore falls under the domain of digital sovereignty. The latter forms an important condition for stimulating informational self-determination. But why controllability? Primarily, because ownership implies in most cases multiple rights and covers multiple actors. Therefore, in most cases can be spoken about quasi-ownership. As this type of ownership puts “individuals in a position to distribute, retract, shield, but also share their data for a variety of purposes, including but not limited to personalized medicine, algorithmic applications, biomedical research, and their own clinical care within a patient-centered health system” (Hummels et al. 2020), the issue here is about data sovereignty and specifically controllability – being “the availability of effective means for data subjects to exercise control over her data.”

In short, debating and subsequently empowering controllability falls in the light of data sovereignty within the scope of “ethically sound data governance” and can therefore typically be associated with RRI. The latter, however, does not have that in scope – yet. As mentioned in the previous paragraph, personal data governance is a proactive element whereas RRI tends to focus too much on the passive component of protection. Given that calls for data ownership need to be discussed in public to be taken seriously, it is exactly the political dimension of RRI being able to organize this and make this happen on an EU-level. As the DS forms the digital identity of a single person, it is able to autonomously operate (managing personal data autonomously) but based on that person’s personal preferences on data governance. Ethical data governance belongs par excellence to the digital self.

Yet, the question rises how to make this controllability possible. This is where the pole of protection versus participation comes in. Data ownership, thus forms of quasi-ownership, makes controllability, personal data governance, possible. In other words, (quasi) data ownership makes an important condition for privacy and data governance. This is where protection and participation become relevant: as personal data is not owned by anybody before a person is born and starts generating data, personal data deserves to be protected.

Subsequently, as humans need personal space to shape their lives in a socially embedded context, they choose to participate and share their data. It is the DS that seems to be able to grasp a role here. It makes a proper concept in shaping and capitalizing this kind of data ownership: because data flows never stop, its algorithms can bring structure and guidance in what is (partly) owned and in which embedded context. This is an important prerequisite for proper data governance on a daily basis.

4.2.2 Along the lines of informational self-determination

With regard to data ownership, informational self-determination of data subjects can be simplified by supporting a public discussion on “claims on the redistribution of resources” and “recognition of proclaimed data owners”. Redistribution of (personal) data resources relates to, simply said, the spreading mechanism of personal data copies. Recognition relates to valuing this data.

It is RRI that is able to shape the discussion on redistribution and recognition of personal data. As the dimension of data ownership goes beyond legal frameworks, it is in the interest of RRI to extend the scope of the concept of data ownership and its diverse forms. As suggested in the previous paragraph, the DS makes a good enabler for operationalizing the outcome of this discussion. As personal data piles heap up, physical individuals cannot value their data properly anymore – let alone supervise redistribution on a daily basis. In addition, the larger the amount of infor-

mation to process for the physical brain, the lower the quality of the decision based on that information: in other words, information- or cognitive overload⁴⁷. Thus, the need for the DS here is evident in the sense of structuring, processing and participating in the decision-making process.

4.3 In terms of Democracy

Considering democracy, central to RRI is the aspect of joint and inclusive deliberative decision-making. In the case of the DS, being a decision-making entity, it seems on first sight the kind of decision-making is really framed individually and autonomously. However, as the interface between the one DS and other DS's (e.g. the DS of a patient interacting with the DS of a doctor) evolve, it becomes a process of shared decision-making. True, inclusive decision-making would seem too far-fetched for the AI-powered DS in the near future: analyzing and coming deliberately to a decision with more entities or stakeholders would be more typical for strong AI. Decisions based on a rather simple form of problem-solving weak AI⁴⁸, which is nowadays close to being mainstream in many societies across the globe, would be the first step for the DS in taking decisions in the digital domain. In essence, the DS is out there to make life less complicated, as being a digitally sovereign citizen in a democracy has become increasingly complicated for the physical self (Keymolen 2016).

Still, there is much to say for the fact that the DS is able to decide freely and independently – usually brought up as the key factors of what makes us human. Based on a digitally sovereign model of personal data governance as discussed in the previous paragraph, the DS deciding free

⁴⁷ “In ordinary language, the term ‘information overload’ is often used to convey the simple notion of receiving too much information. Within the research community this every day use of the term has led to various constructs, synonyms and related terms as for example cognitive overload, sensory overload, communication overload, knowledge overload, or information fatigue syndrome” (Eppler & Mengis 2003, p.7).

⁴⁸ “Weak AI (also known as narrow AI) does not exhibit any creativity, nor does it have the explicit ability to independently learn in the universal sense. Its learning abilities are mostly limited to training of detection patterns (machine learning) or comparison and search operations with large quantities of data. Using weak AI, clearly defined tasks can be handled based on a defined methodology in order to solve more complex problems which, however, are recurrent and precisely specified. The benefits of weak AI are especially relevant in automation and controlling of processes as well as in speech recognition and processing. For example: Text and image recognition, speech recognition, translation of text, navigation systems, etc. Digital assistant systems like Alexa, Siri and Google Assistant also belong to the category of weak AI” (FHWS 2021)

from outside influences and in an autonomous way should actually be possible. The type of decisions will, at first, only involve relatively basic decisions requiring weak AI. Also, basic forms of forward-looking responsibility can be involved here. In the next chapter, a case of digital health will touch upon the topic of the DS empowering patients.

Then again, complex moral decision-making also needed for deliberating within a democracy are too far-fetched for the DS in an early stage. For example, backward-looking responsibility or reflection. The latter is an inherently complex aspect: reflecting on a decision requires consciousness and for AI and thus the DS this is (still) only theoretically possible. Take for instance the kind of reflective decisions that evaluates the (individual) quality of life, in the sense of what life makes meaningful including selecting fundamental criteria of human experience where this evaluation is based on. Such decisions can be life-changing e.g. when offered a life-sustaining cancer treatment, as being devastating for the human body.

In summary, looking at an aspect that in fact can be relevantly discussed here, is the fundamental democratic element of autonomous decision-making. Before identifying the RRI deficit – which is in this regard directly connected with informational self-determination, the way autonomous decision-making by AI affects our democracy will be elaborated on in the following paragraph.

4.3.1 How AI affects democracy

In the past decade, democracies all across the world have been suffering from AI-powered digital surveillance capitalism (Zuboff 2015). More specifically, digitization both weakening social cohesion and voting processes clearly undermines democracy (Budd et al 2019). Seen from a societal angle, it is therefore important to look at the democratization of AI, starting by understanding AI “as a set of general-purpose technologies that can be used in very different circumstances and very different ways to achieve multiple tasks.” (Djeffal 2019) In other words, seeing AI in terms of benefits and concerns would help in the democratization process. As AI-based technologies have great potential to support and empower democratic processes, they can also have a negative or disastrous influence on democracy. The first when it comes to e.g. empower patients suffering

from disabilities within democracy, examples regarding the latter e.g. the Brexit and the US elections in 2016 are widely known. In creating a future of democracy with positive impacts by AI, the usual “democratic toolbox” is able to safeguard and steer democratization of AI. Specifically, an ethical concept like explainable and meaningful AI (meaning respecting human autonomy) would become a valuable and vital option for society. Such a concept puts people in the center of the fast developments within AI. An important normative component for meaningful AI to aim at, is the special standing humans have by holding human rights. As the connection between RRI and human rights is set out before, this part is clearly covered by RRI. What remains, is the consideration that the DS may well bear human rights too and exactly this consideration supports the argument why we need the DS to maintain our understanding of democracy.

4.3.2 Along the lines of informational self-determination

In addition to the final point made above, other key components of the concept of meaningful AI can be found in informational self-determination. Namely, at the heart of the concept of informational self-determination lie the notions of autonomy and decision-making. Evidently, independent and free decision-making are vital for humans to function within a democracy. Therefore, both of these key components of the decision-making process within the DS are essential and needed to function in a democracy highly shaped by digitization. The RRI-deficit is, consequently, the empowerment of decision-making by the DS. Clearly, enabling and supporting of empowerment of the DS should be typically addressed by the governing mechanism of RRI.

5 Case Study: how RRI meets the DS

As a result of the previous chapters, a case is provided. The case is based on the concept of the DS, integrated into a future patient-centered digital health ecosystem (Figure 5, p.37). The case is about patient data governance within healthcare, being a highly sensitive and highly actual topic. It is a clear RRI-case, as it fulfills a great need of responsibly innovating within digital health in dealing with personal data. It involves the societal impact of how emerging health technology is able to take patients suffering from a chronic disease into account. Specifically, this case is related to Parkinson’s disease (PD). It aims to identify new design rules for digital health applications. In order to do this properly, the case will be subject to a recently developed concrete ethical design

approach. This design approach, called the guidance ethics approach, can form an interesting strategy for RRI within digital health.

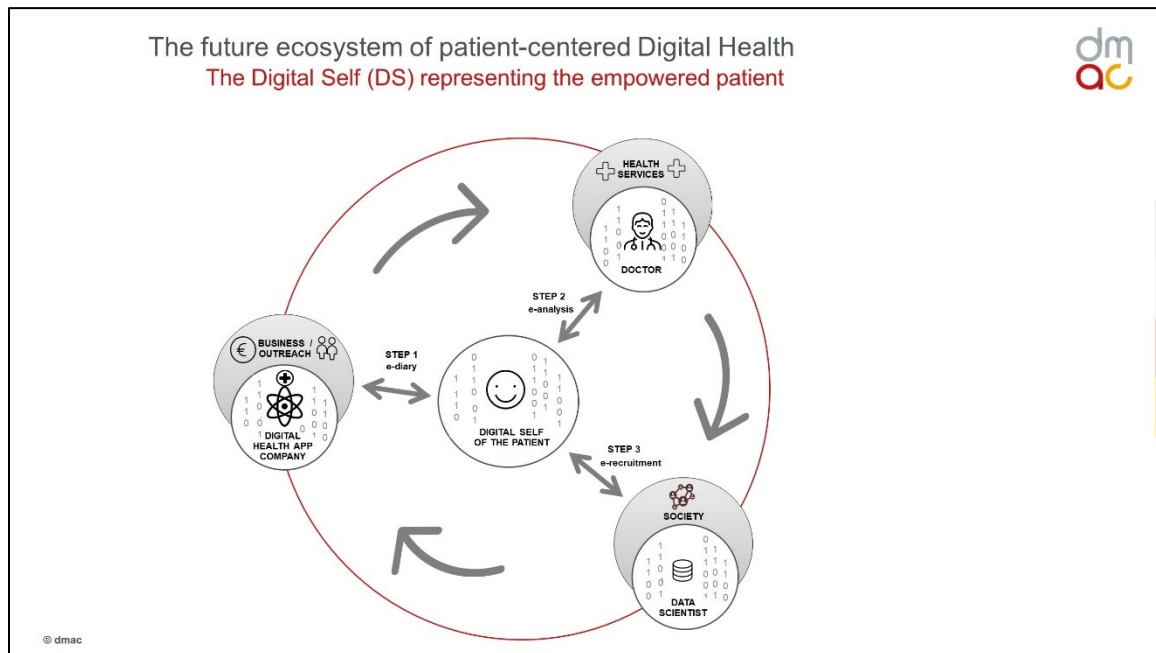


Figure 5: An impression of “The future ecosystem of patient-centered Digital Health” (dmac⁴⁹)

5.1 Case of Digital Health: the Profile Basic App

Together with other organizations in a consortium, dmac investigates in putting a prototype patient-centered health application into practice which takes the digital sovereignty of a patient into full account by enabling the concept of the DS as presented in this thesis. As a patient already spends a significant part of his time in the digital world as part of patient engagement, it is more than likely he would leave more and more medical decisions to his DS.

Evidently, the topic of personal data governance is already highly sensitive in offline health and it becomes even more sensitive when discussing how to deal with it in digital health: the benefits e.g. in the sense of clinical decision-making and for patient empowerment purposes might be

⁴⁹ Dmac stands for Digital Medizin Application Center and aims to accelerate digital health innovation by e.g., supporting so-called DiGAs to develop and certify health apps in Germany. Please see [Digital Health Application Center GmbH | Medical Valley Bamberg \(mv-dmac.com\)](https://www.digital-health-application-center.com/)

evident, however the digital sovereignty of the patient remains key. That is why dmac aims high with developing this prototype app, targeting at creating a whole new patient-centered digital health ecosystem (Figure 5) for all officially regulated health apps (so-called DiGAs⁵⁰) in Germany. The app is called the Profiling Basis App (PBA). It is meant for patients to onboard and build their profile (step 1, e-diary), subsequently have it standardized by their doctors (step 2, e-analysis) and then see if it can be matched with new studies or findings by data scientists within research institutions and the pharma industry (step 3, e-recruitment). The PBA is in full development and has not been tested yet.

As healthcare starts with generating trust between doctor or therapist and patient, it becomes clear that personal health data should be treated with great caution within the whole patient data chain. So far, all patient health data have been commonly stored in different places e.g. hospitals in a central way, posing several societal risks within the rise of digital health. This is the immediate cause for dmac to address these risks in the design of new health apps, by considering both a blockchain-based decentral design⁵¹ and dynamic consent by the patient within the PBA. These considerations aim to increase digital sovereignty and open the door to a future enabling of the DS-environment of the patient. At the same time, as a result of this process – being a safe and secure way of sharing personal health data – optimal trust between patient and doctor can be generated.

The Profiling Basis App, as the name already implies, involves the profiling of a patient in order to offer him the best possible healthcare. Specifically, this profiling is done in the following three stages:

1. The e-diary stage: questionnaires to be filled out (anonymously) by the patient, in order to build a health profile. Only a limited amount of patient data is needed.

⁵⁰ Please see [BfArM - Digital Health Applications \(DiGA\)](#)

⁵¹ On the one hand, “Blockchain or Distributed Ledger Technology is a disruptive technology that provides the infrastructure for developing decentralized applications enabling the implementation of novel business models even in traditionally centralized domains” (Pop et al. 2020, p.1). On the other hand is the “rise of blockchain technology as a responsible and transparent mechanism to store and distribute data...paving the way for new potentials of solving serious data privacy, security, and integrity issues in healthcare” (Khezzar et al. 2019, p.1). Therefore, blockchain-based decentralized design offers a potential win-win for digital health applications.

2. The e-analysis stage: the doctor of the patient receives an expert structured analytical overview based on the patient's profile and the profile can be completed.
3. The e-recruitment stage: based on the profile and its clinical picture, the matchmaking process with studies and findings of research institutions and the pharma industry take place. Here, new research can find its way helping the patient in his current situation in order to get the best possible treatment.

How does the PBA work for the DS of a future Parkinson's disease (PD) patient?

Imagine John, a 51-year-old PD patient, heavily dependent on healthcare from various stakeholders. John is aware of his situation and knows about his own disease well. He uses medical devices and sensors connected to his smartphone by means of a specific PD app, which is based on the PBA. In doing so, John feels in control of his own health, being safeguarded and medically monitored 24/7. His doctors and therapists know exactly about the status of the situation of the digital John – when he wishes so.

John trusts this way of working and is happy with how it works, as he knows his valuable medical data is not lying on the streets in the hands of private and public parties he is not aware of and did not share his data with. Also, by means of the PBA he is able to share his data with researchers, who make use of his data in order to improve his health and that of all PD patients. Besides monitoring, based on the PBA his specific PD app also provides various solutions and self-evaluation tests for assessing cognition, mood and nutrition to motivate him in becoming more active in managing his disease. All data from the mobile application and the sensors sticks with him and is saved only on his devices.

For data requests by his doctor or other medical stakeholders, confirmation is continuously asked by an algorithm which makes clear in an easy and transparent way why his data is to be shared. This type of dynamic consent is being executed by John's DS, as he configured his DS in a way that always respects his personal preferences regarding health data sharing management. This forms a new dimension of patient empowerment, assuming an autonomously operating digital identity connected with an individual, physical identity.

This means his privacy is protected in a compliant way and every form of data sharing is a legal contract – the blockchain aspect. Involved clinicians can access this information as well using the same PBA-based PD app, specifically designed for their respective needs to provide faster and more accurate care for John. In future, machine learning techniques can be used to estimate symptoms and disease progression trends to further enhance the provided information. The PD app includes a decision support system notifying clinicians for the detection of new symptoms or the worsening of existing ones.

5.2 Design approach of guidance ethics

As stated in the beginning of this thesis, new technology and digitization deeply shapes society and raises ethical questions. This goes for the case as set above as well. For addressing these questions into the design of new health technology and subsequently provide recommendations, a proper design approach that brings about responsibility is needed. One that uses a practical starting point in the sense of a specific technology and its outcome within a societal context. The author proposes, therefore, the design approach of guidance ethics by Verbeek & Tijink (2020). The basis for this approach lies in the mediation theory by Verbeek, as elaborated on earlier in this thesis. Central for this approach is answering the “how to deal with new technology responsibly”- kind of question, instead of the traditional ethical “should we or should we not accept this new technology”- kind of question. This central idea makes the outcome of this approach not a typical assessment, but rather a “normative ‘guidance’ of technology in society.” Therefore, values are given “a guiding role in the development, implementation and use of technology, ranging from justice, autonomy and speed to sustainability, safety, effectiveness, et cetera.” In the next paragraphs it will become clear how the principles and methodology of the guidance ethics approach can be adapted to the above case, in order to elaborate on the case more deeply and show what added value an ethical design approach can have.

5.2.1 Principles

Before the methodology is to be elaborated on, the principles or so-called building blocks of the guidance ethics approach are important to consider when using the guidance ethics approach.

The author aims to connect these principles directly to the case. Most relevant principles to be considered here are the following (cf. Verbeek & Tijink 2020 p.21-28):

- The how-principle: focusing on the how-question as stated above, implies considering continuous gradual improvement. As technology develops gradually, and relationships between people and technology also develop gradually, it is important that this approach is implemented ongoing, taking one small step at a time. This means the design process does not end when the product (e.g. a health app) is delivered.
 - In relation to the case: not only the PBA-based PD app will adapt gradually to the patterns of a specific PD patient, but also the role of the app in the care-process and expectations of medical stakeholders e.g. caregivers changes throughout the process.
- The contextual-principle: technology always comes into being within a certain context, including humans. Doing ethical guidance of technology means considering and analyzing ethical questions with regard to the context of technology, being the outputs from both the outside environment of the technology (stakeholder network) and the inside core of the technology (data).
 - In relation to the case: not only considering the consequences of the PBA-based PD app for the individual patient, but also the societal implications for (digital) healthcare and legislation.
- The value-principle: directly related to the purpose of this approach, as stated above. What kind of values to be used within the design, implementation and use of the technology, often depend on the cultural context of the technology considered. In addition, values can change over time and are often conflict with each other. The key is to be found in dealing with this on a continuous basis.
 - In relation to the case: there is a clear tension between optimal privacy for the patient and improving the quality of health by doing research on extensive patient data and have patients profit from the outcome.
- The action-principle: there are three “concrete options for action” that can be used here, shortly to be mentioned as “ethics by design”, “ethics in context” and “ethics by user”. The first is about processing values into technological design, the second about taking the

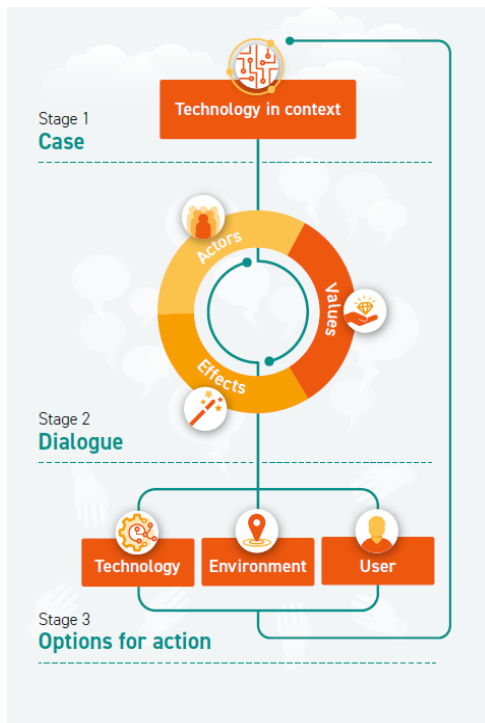
physical context of the technology into account and the third about “awareness and behavior adaptation”.

- In relation to the case:
 - Ethics by design: values (in the sense of non-functional design criteria) like privacy, autonomy (empowerment) and trust seem central in order to design a PBA-based PD app in a responsible manner.
 - Ethics in context: as more and more PD patients are able to profit from the freedom that comes with successfully managing their disease actively by the PBA-based app, the environment has to be adapted to this as well. For example, creating wide hallways in senior residences without device disturbing materials or too much stimulus. Potential social agreement here is e.g. to prioritize device depending patients for device charging points in these residences.
 - Ethics by user: as a PD patient using a PBA-based app, this will train the patient to use devices and monitoring services responsibly. Informing patients about when the emergency function should be used, or connecting to patient interest networks like ParkinsonNet⁵², will create patient awareness.

5.2.2 Methodology

The methodology of the guidance ethics approach can be defined in three stages, as visualized on the next page.

⁵² ParkinsonNet is a patient-centered healthcare and networking concept, which was founded in the Netherlands. It aims to “train and educate healthcare professionals, make them Parkinson’s disease experts and create networks of those professionals around and with those people living with Parkinson’s.” The concept of ParkinsonNet has also been implemented internationally, as a best practice. Please see [Home - ParkinsonNet International](#)



The first stage is about making clear what kind of technology is used and what it exactly does. This concerns the main design requirements for the PBA-case, which are discussed underneath.

The second stage is about stakeholder and value mapping: this will also be elaborated on.

The third stage is about defining actions in terms of ethics by design, ethics by context and ethics by user as described above. With regard to the case, please also see underneath.

Figure 6: Methodology of the guidance ethics approach (Verbeek & Tijink 2020, p.32)

5.2.2.1 Step 1

As mentioned within the case, the technology powering the PBA has the aim of creating a whole new patient-centered digital health ecosystem. It has two main design features, that are of most relevance in consideration of privacy, ownership and democracy. In addition, the following features clearly complete each other:

1. Decentral design
2. Dynamic consent

Ad 1

As opposed to a centralized design, where users trust their data with a third-party using cloud storage, within a decentralized design the technology e.g. the database is split between and ruled by users themselves (Liu et al. 2020). This is based on blockchain technology, which is in fact “a

decentralized database.” In addition, blockchain-based decentral design centers the user or patient and offers by its mechanism a transparent way of saving data transactions. Thus, decentral design generates the value of trust by design. Trust is one of the most important values within healthcare, specifically in a patient-doctor relationship. In concrete, the PBA offers a blockchain-based decentral design allowing the patient to keep all his data stored on his smartphone (or other personal devices).

Ad 2

The above makes a bridge to dynamic consent, as every time a stakeholder (doctor, therapist, expert etc.) needs a piece of patient data, an algorithm sends a request to the patients’ smartphone. When the patient agrees with the request, only that piece of data will be sent matching the exact demand of the stakeholder. This is where the blockchain-idea comes in: each request is in fact a legal contract. Every request for patient is done continuously and therefore typically a dynamic form of consent. This process is digitally enabled by the PBA and offers patients real-time control over sharing their data with relevant stakeholders within the patient journey according to their preferences (Kaye et al. 2015) – in an ethical way. The ‘dynamic’-part relates mostly to a flexible way of making decisions on your own data over time including a clear overview.

5.2.2.2 Step 2

In this second step, the potential consequences of the PBA are to be examined. This will be done in terms of:

1. Stakeholders
2. Consequences
3. Values

Ad 1

In this case a multi-stakeholder engagement is needed to address and debate ethical dilemmas from several perspectives, which can be seen as typically a strategy for RRI. PBA-stakeholders can

be roughly divided into internal and external stakeholders and the goal here is to come to a balance. Most important requirement would be to make sure all stakeholders involved are decision-makers (preferably at C-level).

It becomes clear that for the design and development of the PBA health app, the following internal stakeholders can be identified:

1. Technology company that produces the general app functionality
2. Technology company that produces the core e.g. the decentral design and dynamic consent feature
3. dmac, that pushes digital health innovation by consulting and medical expertise
4. A doctor or similar medical stakeholder, as a representative of the regional hospital

In addition, the following external stakeholders can be identified:

5. Patient (and in future his digital self), as a user of the app
6. Random other doctor or external medical expert
7. Data scientist
8. Pharma industry expert and representative
9. Patient interest group representative
10. Societal IT-interest group representative

Ad 2

Inventorizing potential consequences of the PBA in a transparent and exhaustive way, is first of all important here. Thereafter, listing the most applicable. Subsequently, identifying and elaborating on the following sorts of effects of the PBA are able to “help in obtaining a rich and realistic image” (Verbeek & Tijink p.34).

- “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)”

Due to reasons of confidentiality and too much irrelevant detail, the PBA-case will not be discussed with regard to potential consequences.

Ad 3

Technology is unavoidably connected to and influenced by human values. In the first part of this thesis, it has become clear that incorporating values into the design of AI-systems can be of societal significance (AI HLEG 2019). In fact, to a great extent the same values can be identified being of relevance within the context of the PBA-based technology. Specifically, the following values:

- Trust
- Privacy
- Autonomy
- Safety
- Security

5.2.2.3 Step 3

As the guidance ethics approach helps with an ethical embedding of PBA within a digital health environment and at the same time is able to help dmacc and healthcare in a broader sense with the implementation of PBA, this final step provides a first concrete incentive for thought.⁵³

The following options for actions, related to the earlier mentioned action-principle, can be formulated:

1. PBA “ethics by design” action option: **transparency guarantee needed in blockchain-based algorithmic-powered dynamic consent by the patient.**
 - a. Explanation: as patients using the PBA receive data requests by their doctors or by other medical stakeholders, confirmation is continuously asked by an algorithm which should make clear in a transparent way what part of his data and why his data is to be shared. In case a PD-patient is not understanding this for

⁵³ Note from the author: as the guidance ethics approach cannot be fully adapted to the case within the scope of this thesis, it mostly forms an exploration showing how an ethical approach can be used for designing health apps and in a broader sense, co-designing a digital health ecosystem.

whatever reason e.g. a reason that directly refers to his disease itself, a guaranteed level of transparency is needed in order to be sure that the patient is well aware of the grounds on which the algorithm bases its data request.

2. PBA “ethics in context” action option: **additional profile assessment by medical professional within the PBA-process required.**
 - a. Explanation: as an arbitrary patient uses PBA to acquire insight on a potential (chronic) disease, he starts by filling out the e-Anamnesis-based e-diary questionnaire most likely on his own. Possible misunderstanding of the questions or context can possibly lead to false diagnosis of a chronic disease and false hope of help after the matchmaking process based on his profile. An additional professional assessment, for example within a face-to-face conversation with a doctor, is therefore still a requirement within the PBA-process.
3. PBA “ethics by user” action option: **gaining consciousness on what the PBA can do for you as a (e.g. PD) patient.**
 - a. Explanation: as patients give shape to the societal impact of the PBA, their awareness with regard to the use of the PBA is of high importance. If patients are conscious of what both decentral design and dynamic consent can mean to them and how this is able to work out in usage, chances of acceptation, disciplined- and responsible use of the PBA will be higher. Awareness and acceptance of the PBA can be endorsed by networks of healthcare professionals like ParkinsonNet as well, showing patients (based on training) clearly what the PBA can do for a specific group of patients.

6 Discussion & Conclusion

This final chapter starts off with a discussion, followed by the conclusion answering the research question of this thesis as promptly as possible. Within the discussion, the findings of the case and suggested design approach are addressed in light of the thesis as a whole. An outlook is integrated into the conclusion.

6.1 Discussion

In this thesis, an effort is made to bring about the concept of the DS. The PBA-case has been used to showcase the DS concept, based on emerging health technology that aims to take essential human values into account. The guidance ethics approach allowed for understanding the PBA-case in relation to enabling the DS of a patient.

The applicability of the guidance ethics approach on the PBA-case can be cautiously confirmed, suggesting that PD patients and caregivers rely on health technology that takes values into account and at the same time gives them benefits in managing their disease. Specifically, in terms of practicability and using a holistic perspective, the guidance ethics approach shows its added value. Practicability in two ways: both in translating ethical principles into concrete guidelines and in step by step guiding of stakeholders throughout the process.

Furthermore, the approach utilizes a pragmatic order of typical steps in how to come to a result, aligning well with commonly used project- and change management methods within digital health innovation practices. That way, with regard to an implementation of the approach within the PBA-case the question whether there are ethical skills among the relevant designers to adequately make use of this approach can be overcome. Also, its multi-stakeholder approach is able to neutralize technological incompetence or potential resistance within the implementation.

In order to not just see a part of the added value of the guidance ethics approach but the whole picture, follow-up research covering the full methodology of the approach including its implementation within the PBA-case is needed. This kind of research in the field of applied ethics is quite recent and scarce. Moreover, further research from the RRI-perspective in the sense of retrospection is needed. Widening the RRI-concept to take informational self-determination-based concepts like the DS into account is not only essential for maintaining an understanding of democracy, but also for maintaining an understanding of healthcare.

In terms of general design requirements for health applications, decentral design and dynamic consent are able to form important solutions to deal with the essential principles of informational

self-determination and digital sovereignty. The PBA-case shows that embedding key values as trust, privacy, autonomy, safety and security into the design of health technology is actually possible. How these values can be processed as non-functional design requirements within a decentral design on a more detailed level, further research is needed.

Decentralized storage avoids cloud-based solutions in healthcare, making sure tech giants, other third parties and governments are not able to own and control sensitive medical patient data. At the same time, democracy assumes an individuum, an individual entity, who can speak for himself deciding about his own data. Dynamic consent is a patient-centered way to deal with continuous data requests to patient on a daily basis. In terms of privacy this creates an environment, where patients are able to tick these request boxes with trust and confidence.

Blockchain-based technology is for both decentral design and dynamic consent of beneficial use. In essence, both design requirements open the door for treating patient data in a new way, taking the concept of the DS of the patient serious within the digital health ecosystem of the (near) future and taking key ethical values into account. As a consequence, the enabling of shared-decision-making processes between doctor and patient by the DS of the patient has a positive effect on patient empowerment and -participation.

6.2 Conclusion & Outlook

Although the radical shaping of society by digitization intensifies every day, the concept of a DS has for a great deal (still) not been considered within the scope of RRI. Overall, RRI focuses too much on protecting the physical citizen, while “responsible innovation should explore ideas about what the good life is like, what the world should be like, and what people are or should be” (Ben- nink 2020). This directly relates to the RRI-deficit in consideration of democracy, which is all about what it means to act as a democratic citizen making independent and autonomous decisions, in a world dominated by information. As participating in shared decision-making processes by the DS seem only years away for e.g. the future patient, the consequences for society on the long term are incalculable and therefore worthwhile to explore. Fundamental human rights should include

the DS as well. From this perspective, maintaining a clear understanding of democracy lies within the grasp of RRI.

For RRI, leaning on the building block of anticipatory governance of emerging technology, this should be a clear sign to embrace the concept of informational self-determination when applied to the DS. Instead, RRI rather awaits than addresses. RRI works as an abstract concept in a rather passive way. The DS shows the urgency for RRI to work in a more proactive, addressing and dynamic way at the heart of emerging technology practices e.g. digital health. Only in that way informational self-determination can be respected, at the same time enabling and empowering the DS of democratic citizens and patients.

Regarding the deficit of RRI considering both privacy and (quasi-) ownership of the DS, the imbalance between data protection (relating to privacy) and data governance (relating to ownership) forms the central issue. The first is mainly about security and based on fundamental rights, therefore RRI concentrates too much on this one. The latter is mainly about controllability and thus digital sovereignty, here to understand as an ethical sound form of personal data governance. Digital sovereignty requires proactive addressing by RRI, taking informational self-determination into full account. As digital sovereignty is a key condition for autonomy, it seems a conflict between the values⁵⁴ of privacy and autonomy is the main issue to be addressed. For the time being, it is e.g. the AI-related aspect of explicability that makes this value conflict turn in favor of privacy.

Dynamic consent, however, is a means to restore the balance between privacy and autonomy. It forms a key condition for digital sovereignty, as shown by the PBA-case where dynamic consent functions as a key enabler for the DS of the future patient. The decentral design of the DS of the patient, can also be seen as key condition for autonomy. In addition, the key ethical AI aspect of accountability can be of importance here. The DS is able to bridge the so-called gap between principles and practice of ethical AI (Floridi 2018) in the sense of accountability, freeing the way

⁵⁴ According to Van de Poel (2009), "two or more values conflict in a specific situation if, when considered in isolation, they evaluate different options as best".

to independent and free decision-making by itself. This will provide the kind of autonomy needed to restore the above-mentioned imbalance.

Still, there are two ways RRI potentially applies quite well to the DS. First, viewing data ownership from a technical perspective along the lines of informational self-determination. Here, it is both the public discussion redistribution as well as the recognition of personal data where RRI and the DS become complementary: RRI as the organizer and stimulator of this societal discussion, the DS as a powerful way of bringing the outcome into being.

Another, more general way of how RRI enables the DS concept, is in light of the RRI-definition used in this thesis by Van den Hoven (2017): here, RRI works as a stimulus for concrete design approaches addressing in how to deal with fundamental values in technology practices. The guidance ethics approach (Verbeek & Tijink 2020) has shown to work like this, adapted to a digital health case about how the DS is able to empower PD patients. This methodology makes clear that considering incorporating values into the design of health technology from an early stage and within a multi-stakeholder environment, is essential. Moreover, the guidance ethics approach shows it is a RRI-strategy that can actually be applied to the DS.

In digital healthcare, more and more innovation is about delivering the right type of care by delivering the right information for the right patient, on the right moment through the right channel (Waag 2020). Evidently, with the parental aim of increase the quality of patient-centered (digital) healthcare by doing data-driven research – making the innovation acceptable to society. For PD patients, such easily accessible and made-to-measure type of care innovations would be a of great benefit. As data amounts in digital health are heaping up every day and patient data is in the current situation spread across a number of hospitals and other medical institutions (without the patient having overview and control), this can only be realized by a sustainable, robust, scalable and futureproof form of patient data governance – responsibly designed from the perspective of societal and ethical values. To realize this sustainability and robustness, science is in the lead; for scalability, commercial parties have most expertise. This kind of RRI-strategy involves multiple

stakeholders, as science, industry and other relevant (societal) stakeholders need to find each other and collaborate.

Based on the picture outlined above, the PBA-case offers a key instrument for generating trust. Both decentral design and dynamic consent form essential design requirements for the future digital health ecosystem (Figure 5, p.37), stressing the need for and enabling the concept of the DS of the patient, subsequently supporting patient empowerment and participation. The guidance ethics approach shows to be a practical way of addressing the influence of the above-mentioned emerging health technology on patients and other medical stakeholders. For example, it would be beneficial to a model of ownership of patient data if the first line of healthcare e.g. the general practitioner acts as a steward that proclaims dealing with data in a responsible way. At the same time, through this mechanism the general practitioner is able to also share this role with and assign to patients – including a feedback loop. This model could be an attractive image of what a future patient truly wants and expects. It is, however, an image that could take years and years before it can come true. The fictive case of Inga and Otto by Clark & Chalmers (1998) makes clear that Otto, who suffers of Alzheimer-disease, is empowered by and able to benefit from a piece of technology (in this case a laptop) being a very primitive form of his own DS.

Obstacles in digital health innovation have for an important part to do with compliance and new laws. Thus, it is a matter of thinking big and making small steps. The PBA-case can be an interesting, experimental use case to learn about and deal with patient data in a totally different way – at the same time shake up the bigger discussion on this topic. Important is to consider informational self-determination of the patient as a starting point.

In doing so, shaping of the DS of the patient (and potentially the DS's of other medical stakeholders) is done in a both organic and responsible way. At the same time, this process enables maintaining of and contributing to our understanding of privacy, ownership and democratic decision-making within digital health. Still, the awareness of the patient – amongst other stakeholders – in relation to his data is not to be forgotten in this picture: it requires focus and attention of the patient himself to see the value of this.

In future, an AI-powered PBA enabling the DS of a patient is able to play a significant role in prevention, diagnosis and treatment of diseases. In that scenario, it has to deal with the back side of AI – a lack of transparency and accountability foremost – in an adequate manner. For example, a faster diagnosis based on AI-powered clinical decision-making is within the future grasp of the PBA. When it comes to clinical decision making, data bias can become a hurdle as well: patient data of high reliability to train AI-systems is still scarce. Therefore, PBA-based patient data of high quality can be of great importance. In addition, the more transparency and explicability of AI, the more trust of patients in AI-powered decision-making (Bjerring & Busch 2020) – and in their DS.

List of used literature

1. Banta, D. (2009). "What is technology assessment?". *International Journal of Technology Assessment in Health Care*. 25 Suppl 1: 7–9. Cambridge University Press. <https://doi.org/10.1017/S0266462309090333>
2. Bennink, H. (2020). Understanding and Managing Responsible Innovation. *Philosophy of Management*, 1-32. <https://doi.org/10.1007/s40926-020-00130-4>
3. Bjerring, J. C. & Busch, J. (2020). Artificial Intelligence and Patient-Centered Decision-Making. *Philosophy & Technology*. Springer Nature B.V. <https://doi.org/10.1007/s13347-019-00391-6>
4. Budd, B., Midzain-Gobin, L, Goodman, N., Ouelette, D., Shoker, S., Tabassum, N., ... & Joseph, A. (2019). Digitization & Challenges to Democracy. [working-paper-oct-2019.pdf \(mcmaster.ca\)](#)
5. Buiten, M. C. (2019). Towards intelligent regulation of Artificial Intelligence. *European Journal of Risk Regulation* 10 (1), 41-59. Cambridge University Press. <https://doi.org/10.1017/err.2019.8>
6. Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19. Retrieved December 15, 2020, from <http://www.jstor.org/stable/3328150>
7. Chesterman, S. (2020). Artificial Intelligence and the Problem of Autonomy. *Notre Dame Journal on Emerging Technologies*, 1, 210–250. NUS Law Working Paper No. 2019/016. <https://ssrn.com/abstract=3450540> or <http://dx.doi.org/10.2139/ssrn.3450540>
8. Dankar, F.K., Gergely, M., Malin, B., Badji, R., Dankar, S.K. & Shuaib, K. (2020). Dynamic-informed consent: A potential solution for ethical dilemmas in population sequencing initiatives. *Computational and Structural Biotechnology Journal*, Volume 18, 913-921. <https://doi.org/10.1016/j.csbi.2020.03.027>
9. Demetis, Dionysios and Lee, Allen S. (2018). "When Humans Using the IT Artifact Becomes IT Using the Human Artifact," *Journal of the Association for Information Systems*: Vol. 19: Iss. 10, Article 5. <https://aisel.aisnet.org/jais/vol19/iss10/5>
10. Djefal, C. (2019). AI, Democracy and the Law. *Andreas Sudmann (Hg.): The Democratization of Artificial Intelligence. Net Politics in the Era of Learning Algorithms. Bielefeld*, 255-284. [Djef-fal 2019 - AI Democracy20200210-86051-vg90fk.pdf \(d1wqtxts1xzle7.cloudfront.net\)](#) , checked on 15.12.2020.
11. Duch-Brown, N., Martens, B., Mueller-Langer, F. (2017). The Economics of Ownership, Access and Trade in Digital Data. JRC Digital Economy Working Paper 2017-01. <https://ssrn.com/abstract=2914144> or <http://dx.doi.org/10.2139/ssrn.2914144>
12. EIT Digital (2020). *European Digital Infrastructure and Data Sovereignty – A Policy Perspective*. EIT Digital. [Full report: European Digital Infrastructure and Data Sovereignty // EIT Digital](#)
13. Eppler, M.J., Mengis, J. (2003). A Framework for Information Overload Research in Organizations. *Insights from Organization Science, Accounting, Marketing, MIS, and Related Disciplines*. [A Framework for Information Overload Research in Organizations \(rero.ch\)](#)
14. Feinberg, J. (1986). *Harm to self*. New York: Oxford University Press.

15. Floridi, L. (2014). *The fourth revolution: how the infosphere is reshaping human reality*. Oxford: Oxford University Press.
16. Floridi, L. (2015). *The Onlife Manifesto. Being human in a Hyperconnected Era*. Springer International Publishing. [1001971.pdf \(oopen.org\)](#)
17. Floridi, L., Cows, J., Beltrametti, M. *et al.* (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
18. Floridi, L., Cows, J., King, T.C. *et al.* (2020). How to Design AI for Social Good: Seven Essential Factors. *Sci Eng Ethics* 26, 1771–1796. <https://doi.org/10.1007/s11948-020-00213-5>
19. Friedman, B. *et al.* (2002). *Value sensitive design: Theory and methods*. UW CSE Technical Report, University of Washington. [Microsoft Word - vsd-theory-methods-tr.doc \(psu.edu\)](#)
20. Heidegger, M. (1962). *Being and Time* (J. Macquarie & E. Robinson, Trans.). Oxford: Blackwell.
21. Ihde, D. (1990). *Technology and the Lifeworld*. The Indiana Series in the Philosophy of Technology. Bloomington: Indiana University Press.
22. Ihde, D. (1993b). *Postphenomenology*. Evanston: Northwestern University Press.
23. Ihde, D. (2009). *Postphenomenology and Technoscience: The Peking University Lectures*. Suny Press.
24. Irving, L. (2019). Virtual worlds in Higher Education: Embodied Experiences of Academics. *The Translational Design of Universities. An Evidence-Based Approach*, 107-130. Koninklijke Brill NV, Leiden, The Netherlands. https://doi.org/10.1163/9789004391598_008
25. Floridi, L. & Sanders, J.W. (2004). On the Morality of Artificial Agents. *Minds and Machine* 14, 349–379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>
26. Floridi, L. (2014a). *The fourth revolution: how the infosphere is reshaping human reality*. Oxford: Oxford University Press.
27. Franklin, S. & Graesser, A. (1996). Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*. Springer-Verlag.
28. Hummel, P., Braun, M., Dabrock, P. (2020). Own Data? Ethical Reflections on Data Ownership. *Philos. Technol.* <https://doi.org/10.1007/s13347-020-00404-9>
29. Jones, M. L., Kaufman, E., Edenberg, E. (2018). AI and the Ethics of Automating Consent. In *IEEE Security & Privacy*, vol. 16, no. 03, pp. 64-72, 2018. <https://doi.ieeecomputersociety.org/10.1109/MSP.2018.2701155>
30. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
31. Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European journal of human genetics*, 23(2), 141-146. <https://doi.org/10.1038/ejhg.2014.71>

32. Keymolen, E. (2016). A Utopian Belief in Big Data. In: Lisa Janssens (Hg.): *The Art of Ethics in the Information Society*. Amsterdam: Amsterdam University Press 2016, S. 67–71. <https://doi.org/10.25969/mediarep/13397>
33. Khezr, S., Moniruzzaman, M., Yassine, A., Benlamri, R. (2019). Blockchain Technology in Healthcare: A Comprehensive Review and Directions for Future Research. *Applied Sciences*. 2019; 9(9):1736. <https://doi.org/10.3390/app9091736>
34. Kudina, O., & Verbeek, P.-P. (2019). Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy. *Science, Technology, & Human Values*, 44(2), 291–314. <https://doi.org/10.1177/0162243918793711>
35. Liu, L., Zhou, S., Huang, H., & Zheng, Z. (2020). From Technology to Society: An Overview of Blockchain-based DAO. [2011.14940] *From Technology to Society: An Overview of Blockchain-based DAO (arxiv.org)*, checked on 11.12.2020.
36. Loshin, D. (2001). *Enterprise Knowledge Management: The Data Quality Approach*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.
37. Martinuzzi, A., Blok, V., Brem, A., Stahl, B., Schönherr, N. (2018). Responsible Research and Innovation in Industry—Challenges, Insights and Perspectives. *Sustainability* 10(3), 702. MDPI, Basel, Switzerland. <https://doi.org/10.3390/su10030702>
38. Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge & K. Paul [Original 1945].
39. Milano, S., Taddeo, M., Floridi, L. (2020). Ethical aspects of multi-stakeholder recommendation systems. *The Information Society*. <https://doi.org/10.1080/01972243.2020.1832636>
40. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. (2016). The ethics of algorithms: Mapping the debate. In *Big Data & Society July-December 2016: 1-21*. Sage publishers. <https://doi.org/10.1177/2053951716679679>
41. Morley, J., Floridi, L., Kinsey, L., Elhalal, A. (2019). From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. ArXiv:1905.06876 [Cs]. Retrieved from <http://arxiv.org/abs/1905.06876>
42. Morris, J. (2000). *Rethinking risk and the precautionary principle*. Oxford | Boston, Butterworth-Heinemann.
43. Owen, R. & Goldberg, N. (2010). Responsible Innovation: A Pilot Study with the U.K. Engineering and Physical Sciences Research Council. *Risk Anal.* 2010;30(11):1699-1707. <https://doi.org/10.1111/j.1539-6924.2010.01517.x>
44. Owen, R., Macnagten, P., Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society, *Science and Public Policy*, Volume 39, Issue 6, December 2012, Pages 751–760. <https://doi.org/10.1093/scipol/scs093>
45. Pop, C. et al. (2020). Blockchain based Decentralized Applications: Technology Review and Development Guidelines. <https://arxiv.org/abs/2003.07131>

46. Reber, B. (2017). RRI as the inheritor of deliberative democracy and the precautionary principle. *Journal of Responsible Innovation*, 5:1, 38-64. <https://doi.org/10.1080/23299460.2017.1331097>
47. Ruggiu, D. (2015). "Anchoring European Governance: Two Versions of Responsible Research and Innovation and EU Fundamental Rights as 'Normative Anchor Points'." *NanoEthics* 9 (3): 217–235. <https://doi.org/10.1007/s11569-015-0240-3>
48. Russell, S. & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, Third Edition. Pearson.
49. Sensen, O. (2013). *Kant on Moral Autonomy*. Cambridge University Press.
50. Shelley-Egan, C., Bowman, D.M., Robinson, D.K.R. (2018). Devices of Responsibility: Over a Decade of Responsible Research and Innovation Initiatives for Nanotechnologies. In *Sci Eng Ethics* 24, 1719–1746. <https://doi.org/10.1007/s11948-017-9978-z>
51. Stahl, B. (2013). Responsible research and innovation: The role of privacy in an emerging framework. *Science and Public Policy*. 40. 708-716. <https://doi.org/10.1093/scipol/sct067>
52. Sutcliffe, H. (2011). A Report on Responsible Research & Innovation. [RRI Report Hilary Sutcliffe Final copy \(unimi.it\)](#) , checked on 15.12.2020.
53. Trotter, G. (2014). *Autonomy as Self-Sovereignty*. Springer Science+Business Media Dordrecht. <https://10.1007/s10730-014-9248-2>
54. Umbrello, S. & Van De Poel, I. (2020). Mapping Value Sensitive Design onto AI for Social Good Principles. [UMBMVS.pdf \(philpapers.org\)](#)
55. Van Asbroeck, B., Debussche, J., & César, J. (2017). Building the European Data Economy: Data ownership. *White Paper, Bird and Bird*. [https://sites-twobirds.vuture.net/1/773/uploads/white-paper-ownership-of-data-\(final\)](https://sites-twobirds.vuture.net/1/773/uploads/white-paper-ownership-of-data-(final)) , checked on 15.12.2020.
56. Van Den Hoven, J. (2008). Moral methodology and information technology. In K. E. Himma & H. T. Tavani (Eds.), *The handbook of information and computer ethics* (pp. 49–68). New Jersey: Wiley. <https://10.1002/9780470281819>
57. Van Den Hoven, J. (2017). Ethics for the Digital Age: Where Are the Moral Specs?. In Werthner H., van Harmelen F. (eds) *Informatics in the Future*. Springer, Cham. [1001939.pdf \(oapen.org\)](#)
58. Van de Poel, I. (2009). Values in engineering design. Meijers A, editor. *Philosophy of technology and engineering sciences*, vol. 9. Amsterdam: Elsevier B.V., p. 973–1006. <https://doi.org/10.1016/B978-0-444-51667-1.50040-9>
59. Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Pennsylvania State University Press.
60. Verbeek, P.-P. (2008). *Morality in design: Design ethics and the morality of technological artifacts*. Philosophy and design. Springer Netherlands.
61. Verbeek, P.-P. & Vermaas, P.E. (2009). *Technological Artefacts. A companion to the Philosophy of Technology*. Blackwell Publishing Ltd.

62. Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*: University of Chicago Press.
63. Verbeek, P.-P. & Tijink, D. (2020). *Guidance Ethics Approach*. The Hague: ECP. [guidance-ethics-approach-1.pdf \(wordpress.com\)](#)
64. Von Schomberg, R. (2011). Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields. <https://ssrn.com/abstract=2436399>
65. Von Schomberg, R. (2013). A Vision of Responsible Research and Innovation. In ... *Managing the Responsible Emergence of Science ...* , edited by Richard Owen, John Bessant, and Maggy Heintz, 51–74. John Wiley & Sons. <https://10.1002/9781118551424>
66. Walhout, B., Walhout, A.M., Kuhlmann, S. (2013). In search of a governance framework for responsible research and innovation. Paper presented at 2013 IEEE International Technology Management Conference & 19th ICE Conference 2013, The Hague, Netherlands. [In search of a governance framework for responsible research and innovation - CORE Reader](#)
67. Watson, H.J., Nations, C. (2019). Addressing the Growing Need for Algorithmic Transparency. *Communications of the Association for Information Systems*, 45. <https://doi.org/10.17705/1CAIS.04526>
68. Zuboff, S. (2015). Big Other: Surveillance Capitalism and the prospects of an Information Civilization. *Journal of Information Technology* 30, no. 1 (March 2015): 75–89. [Big Other: Surveillance Capitalism and the Prospects of an Information Civilization - Article - Harvard Business School \(hbs.edu\)](#)
69. Zwart, H., Landeweerd, L. & van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the European research funding arena from 'ELSA' to 'RRI'. *Life Sci Soc Policy* 10, 11. <https://doi.org/10.1186/s40504-014-0011-x>

List of used online sources

1. High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). *Ethics Guidelines for Trustworthy AI*. [Ethics Guidelines for Trustworthy AI | FUTURIUM | European Commission \(europa.eu\)](#) , checked on 08.12.2020.
2. Bar-Gil, O. (2020). *Defining our google self: How information technology mediates self-perception*. [מציגת של PowerPoint \(researchgate.net\)](#) , checked on 08.12.2020.
3. Berners-Lee, T. (2019). *Using Web standards to let people control their data, and choose the applications and services to use with it*. [Home · Solid \(solidproject.org\)](#) , checked on 19.01.2021.
4. Brown, G.W. (2021). *Oxford Reference. Deliberative democracy. Quick Reference*. [Deliberative democracy - Oxford Reference](#) , checked on 12.12.2020.
5. Bundesinstitut für Arzneimittel und Medizinprodukte - BfArM (2020). Welcome to the BfArM's website on digital health applications (DiGA or "apps on prescription"). [BfArM - Digital Health Applications \(DiGA\)](#) , checked on 10.12.2020.
6. Christman, J. (2020). *Autonomy in Moral and Political Philosophy*. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/autonomy-moral/> , checked on 10.09.2020.
7. Cuny Graduate School of Public Health and Health Policy (2020). *Researchers take issue with study evaluating an AI system for breast cancer screening*. [Researchers take issue with study evaluating an AI system for breast cancer screening \(medicalxpress.com\)](#) , checked on 09.12.2020.
8. De la Torre, L.F. (2019). *What is "Privacy By Design" (PbD)?* [What is "Privacy By Design" \(PbD\)? | by Lydia F de la Torre | Golden Data | Medium](#) , checked on 24.01.2021.
9. Dignum, V. (2020). *The ART of AI — Accountability, Responsibility, Transparency*. [The ART of AI — Accountability, Responsibility, Transparency | by Virginia Dignum | Medium](#) , checked on 09.12.2020.
10. Digital-Medizinisches Anwendungs-Centrum (dmac) (2020). [Digital Health Application Center GmbH | Medical Valley Bamberg \(mv-dmac.com\)](#) , checked on 10.12.2020.
11. European Commission (2014). *New and emerging science and technologies (NEST): Specific activities covering wider field of research under the Integrating and Strengthening the European Research area (2002-2006) (Last update 2014)*, <https://cordis.europa.eu/programme/rcn/751/en> , checked on 26.08.2020.
12. GE Healthcare (2018). *AI-embedded X-Ray system could help speed up detection of a collapsed lung*, <https://www.gehealthcare.com/article/ai-embedded-x-ray-system-could-help-speed-up-detection-of-a-collapsed-lung> , checked on 26.08.2020.
13. Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt (FHWS) (2021). *Weak vs. Strong AI*. [Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt \(fhws.de\)](#) , checked on 25.01.2021.
14. ParkinsonNet (2020). [Home - ParkinsonNet International](#) , checked on 10.12.2020.

15. Powles, J. & Nissenbaum, H. (2018). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. *Trying to "fix" A.I. distracts from the more urgent questions about the technology*. [The Seductive Diversion of 'Solving' Bias in Artificial Intelligence | by Julia Powles | OneZero \(medium.com\)](#), checked on 09.12.2020.
16. Prewitt, M.F. (2016). *Consumer Data encryption and the autonomous digital self. The Circuit Rider (April 2016)*, https://www.schiffhardin.com/Templates/media/files/publications/PDF/Prewitt_Circuit%20Rider_April2016.pdf , checked on 27.08.2020.
17. Singer, N. (2020). *Virus-Tracing Apps Are Rife With Problems. Governments Are Rushing to Fix Them*. <https://www.nytimes.com/2020/07/08/technology/virus-tracing-apps-privacy.html> , checked on 27.08.2020.
18. Statista (2020). *Global digital population as of July 2020*. <https://www.statista.com/statistics/617136/digital-population-worldwide/> , checked on 26.08.2020.
19. Stegmann, A. (2020). *What is (Big) Tech? A Taxonomy. Why the Definition matters*. [What is \(Big\) Tech? A Taxonomy.. Why the definition matters | by Andreas Stegmann | hyperlinked | Medium](#) , checked on 12.01.2021.
20. Steiner, F. & Grzymek, V. (2020). *Digital Sovereignty in the EU*. Bertelsmann Stiftung (publisher). [BST-Vorlage 2013 \(bertelsmann-stiftung.de\)](#) , checked on 08.12.2020.
21. Waag (2020). *Waag at Dutch Design Week 2020: A donor codicil for your data?* [Waag at Dutch Design Week 2020: A donor codicil for your data? | Waag](#) , checked on 02.12.2020.
22. Wei, J. (2020). *Definition - What does C-Level Executive mean?* [What is a C-Level Executive? - Definition from Techopedia](#) , checked on 11.12.2020.
23. Woodruff Smith, D. (2013). *Phenomenology*. Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/entries/phenomenology/> , checked on 29.08.2020.
24. World Health Organization (2020). *Global Strategy on Digital Health 2020-2025*. [gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf \(who.int\)](#) , checked on 24.01.2021.