sciendo

# Designing technology in the age of digitalization: needs for technology assessment and ethics

**doi:10.2478/mape-2019-0060**
Date of submission to the Editor: 07/2018
Date of acceptance by the Editor: 08/2018
**MAPE 2019, volume 2, issue 1, pp. 612-623**

**Armin Grunwald***
Institute for Technology Assessment and Systems Analysis (ITAS), **Germany**
Karlsruhe Institute of Technology (KIT), **Germany**

## INTRODUCTION

The rapidly advancing digitalization increasingly affects almost all areas of human life. Digital economy and society, sustainable management and energy, as well as healthy living and civil security pose grand challenges and are strongly affected by the digital transformation. Technology research, design, and development is confronted with digitalization in two respects: (1) future technologies, e.g., in the fields of care robotics, autonomous driving, and Industry 4.0, have to be designed and developed, and (2) the respective R&D processes themselves will happen in a more digitized environment. Digitalization offers great potential for innovation and business. However, at the same time, new challenges arise, and changing boundary conditions must be taken into account when researching and developing new technology for and in the digital age.

Technology is not value neutral (e.g., van de Poel, 2009). Designing technology implies making decisions according to criteria which mostly not only include technical but also non-technical criteria such as economic, ecological, or ethical ones. In the age of digitalization, the significance of values in the design of future technology increases because more and more close and intimate interfaces between humans and technology have to be shaped. Designing human-machine interfaces is not only a functional issue but touches upon ethical questions such as the distribution of responsibility, but also upon anthropological issues related to human self-image. This observation applies to the human-machine interface at the individual level as well as the relations between digitized technology and society at large. In the respective research, design and development processes, value-laden issues such as control, privacy, empathy, responsibility, and accountability must be considered, far beyond technical issues of efficiency and reliability.

Against this background, the objective of this paper is to explore necessities and requirements for designing technology in the age of digitalization. First, as a more general issue, the task is to critically discuss the recent revival of technology determinism in the wake of digitalization (Sec. 2). Technology determinism denies the possibility to shape and design technology with respect to ethical values and societal goals. Instead, it supports strategies of adaptation to ongoing digitalization. In contrast, I will demonstrate the need for designing and shaping digital future technologies by three examples:

* armin.grunwald@kit.edu

- Many expect that future industrial production (in Germany called Industry 4.0) will be characterized by close cooperation between humans, robots, and other types of digital technology involving issues of responsibility, accountability, and liability.
- The field of self-driving cars raises many questions of control, of risk, responsibility, and liability, of how to deal with emergency situations, and how to take care of ethical dilemmas such as the so-called trolley problem.
- In the field of care, an intense debate is going on about the development and use of care robots. Obviously, ethical questions of autonomy and empathy need to be addressed in this context.

These cases demonstrate the need to (more closely) involve interdisciplinary approaches going for excellent and innovative technology, but simultaneously being aware of the social embodiment and integration of these technologies. As exemplary approaches, I will finally (Sec. 5) introduce technology assessment (TA), value sensitive design (VSD), and responsible research and innovation (RRI). The latter is among the more recent elements of research policy to appropriately tackle the challenge of co-shaping and co-designing new technology.

## PLEA AGAINST TECHNOLOGY DETERMINISM

In current public debate, the dominant impression is that digitalization develops its own dynamics. Especially business representatives and politicians like to talk about digitalization as an inevitable natural phenomenon, like a tsunami or an earthquake (Grunwald, 2019b). Taking this perspective, society and individuals can only adapt to digitalization. Active shaping is not an option then. This rhetoric argues with (supposed) practical constraints and an (also supposed) lack of alternatives (Grunwald 2019a, 159):

A widespread example of daily technology determinism today can be found in current narratives around digitization. Managers, scientists, and politicians use metaphors such as “tsunami” or “earthquake” in order to describe the coming digitization of almost all processes in everyday life, in the workplace, and in the economy. This metaphor conveys different messages, namely:

(1) That the technological development behind digitization cannot intentionally be influenced by society but comes into reality comparably to a natural event like an earthquake or tsunami governed by natural laws.
(2) That digitization will proceed at an incredibly high speed, suggesting that (a) any attempt to influence it will be doomed to fail, and (b) also any attempt to escape will fail.
(3) That digitization will be devastating for economies, enterprises, or the labor market if no pro-active measures are taken – which could be, e.g., pro-active education or establishing incentive systems to promote adaptation to further digitization.

This everyday determinism is probably also widespread over large parts of society. Used by engineers and managers, there might be a strategic intention behind it; declaring something as determined excludes questions of possible alternatives.

However, this technology determinism seems to contradict simple facts. Every single line of a source code is written by humans. Software runs on hardware that is also produced by humans, or by machines that have been developed and programmed by humans for this purpose. Algorithms, robots, digital services, business models for

digital platforms, or applications for service robots are invented, designed, manufactured, and deployed by humans. Search engine software, algorithms for big data technologies, and *social media* have been developed and implemented by humans. These "makers" of digitalization usually work in companies, authorities, or secret services. They pursue certain values, have opinions and interests, follow a corporate strategy, political requirements, military considerations, etc. In this way, they influence the manner and direction in which digital development is driven and used. However, if *other* people with *different* values and interests could design, or at least co-design, digitalization would develop differently.

Seen in this light, the question of design arises. In the social sciences, it is therefore widely held as common sense that technology determinism has been proven false for both empirical and theoretical reasons (Grunwald, 2019a). There is no indication of laws or regularities governing technological advance behind the backs of the actors, such as computer scientists and engineers. However, when listening to public debate, to mass media articulations, and to voices from the economy and often also from engineering sciences, a completely different picture arises: a strong conviction that the course of development of new technology is ruled by technology determinism, in particular in the field of digitalization, but also beyond.

The problem is that if technology determinism were to dominate, it would unfold factual power. If, according to the tsunami example quoted above, many or perhaps most people believe that digitalization will arrive like a tsunami, and if, accordingly, many people prepare themselves and society for this imagined tsunami – then digitalization will simply *be like a tsunami*. In this story, possible alternatives of shaping digitalization as a social process with possibilities for intervention and options to choose among alternative pathways, e.g., according to ethical or social considerations, will simply be ignored, forgotten, or even suppressed. Technology determinism, therefore, can become self-fulfilling, similar to self-fulfilling prophecies (Merton, 1948).

This would be fatal, because there is not *the* digitalization or *the only* way of digitalization into the future. Instead, the future in general, but also the future of digitalization, is a space of possibilities full of alternatives. Which of them will become real has not yet been determined, but depends on many decisions at very different levels, in companies and data corporations, in politics and regulation, by computer scientists and engineers, in science policy, and in the agenda setting of research institutes. This idea is the first step towards taking a *shaping* perspective on the further development of digital technologies. And there is also a place for ethics, in order to draw attention to principles such as human and civil rights, justice and fairness, privacy and inclusion by comparing and weighing up different alternatives.

## DESIGNING TECHNOLOGY IN THE ERA OF DIGITALIZATION

Unlike traditional technologies, digitalization changes the relationship between (human) acting subjects and (technical) objects. As increasingly intelligent systems gain autonomy for decisions, far-reaching questions arise as to where responsibility lies when algorithms make decisions. This development, which has potential for numerous innovations, is likely to form a core element of many requirements for the design of digital technology. They no longer focus only on technical functionality and economic efficiency, but also on ethical and anthropological issues.

**The future of industrial production – Industry 4.0**

Industrial production will be digitalized in the context of Industry 4.0, in accordance with the principles of self-organization (Manzlei et al., 2016). The real-time networking of products, processes, and infrastructures is supposed to significantly change production processes, business models, products and services, as well as the world of work. The organizational concept of Industry 4.0 consists of four basic design principles (according to Hermann et al., 2016):

- *Networking*: Machines, devices, sensors, and human beings can network with each other in order to communicate and exchange data via the Internet of Things and the traditional internet.
- *Information transparency*: Sensor data expand the information systems of digital factory models to create a virtual image of the real world ("digital twins" of real objects) and enable, e.g., the smart factory.
- *Technical assistance*: AI assistance systems support human beings in making informed decisions and responding to problems more quickly.
- *Decentralized decisions*: Production systems make independent decisions and perform their tasks as autonomously as possible. Only in exceptional cases, e.g., in the event of disruptions or conflicting goals, they will delegate tasks to a higher level, e.g., a human supervisor.

The aim is to enable individualized production according to customer requirements, as a radical counter model to Fordist mass production. The required automation technology is to be based essentially on AI through the introduction of methods of self-optimization, self-configuration, self-diagnosis, and independent cognition. In addition to individualized products, also reductions in production times and efficiency gains are to be realized through accumulated learning during production. Geographically distributed production capacities and the respective planning and control systems should cooperate autonomously and thus make better use of existing resources. In realizing these expectations, AI plays a crucial role in data processing, learning processes, and self-organization.

Here, technology design is confronted with considerable complexity. Technical standards and norms must be developed to enable human-machine or machine-machine communication. Coordination and cooperation between humans and machines require clearly defined interfaces; among other things, it must be clarified who has to adapt to whom. Data security and ownership are given high priority and must be legally protected, just as liability issues for the complex allocation of responsibilities between human and machine must be legally clarified. The resulting comprehensive transformation of the world of work (Börner et al., 2018) does not only involve the necessity of early education and training initiatives but also affects safety at work without boundaries (crowd working), the future role of trade unions or new employee representation organizations, the international division of labor in a global labor market, and the development of social security systems. In light of the expected far-reaching impacts on the economy and world of work, a foresighted and responsible design perspective that includes society (Mainzer 2016, 222) is urgently needed in order to identify promising developments and anticipate undesirable developments at an early stage and then avoid them, if possible.

**Autonomous driving**

The future of mobility is closely related to the ongoing digital revolution. In recent years, human drivers have been provided with more and more assistance based on advanced sensors, real-time evaluation of the collected data, and actuators implementing conclusions made by algorithms. Processors and sensors are increasingly able to observe the traffic situation in the surroundings of a car in real-time and determine the next steps to be taken in order to adapt the car to the respective traffic conditions. This development has already led to a partial automation of driving in new vehicles. Highly automated systems can autonomously change lanes and exert other functions without human intervention. In some countries, test fields have been set up where highly or fully automated vehicles can operate.

It is important to recognize that autonomous driving opens up a wide range of new mobility options beyond traditional individual mobility with private cars: new mobility conceptions and patterns, new business models for mobility providers, and new combinations of private and public transport, or even a blurring of the traditional borders between them, could become possible. This property alone makes self-driving cars a possibly disruptive innovation. Despite the many and far-ranging positive expected consequences of autonomous driving concerning safety and comfort, it must not be forgotten that this implies certain risks, some of which are well-known from traditional driving, while others are new and often related to the digitalization of driving necessarily involved (Maurer et al., 2016).

Research is ongoing in particular on the co-evolutionary dynamics between automation/digitalization and daily mobility patterns and routines. The diffusion of the various incarnations of automated vehicles (AVs) will, depending on how they are expected and designed to act and interact with each other and with humans, increase complexity in road traffic, in particular in mixed traffic with AVs and human drivers co-existing. Since road traffic is a result of a network of permanently negotiated and reordered social relationships that goes far beyond simple rule-obeying behavior, AVs at a certain level of automation will become social actors within this network (this refers to the issue of technology taking the role of subjects mentioned at the beginning of this section). This creates numerous research challenges, such as whether humans will attribute agency to AVs in traffic, whether they have to – and should – be able to negotiate in certain traffic situations, whether automation-compliant behavior of humans in traffic should become a future regulatory principle, and whether values could be implemented in AVs, and if so, which and how (Grunwald, 2018).

In particular, the changing responsibilities require special consideration, both in ethical and legal terms: “In the case of non-driverless systems, the human-machine interface must be designed such that at any time it is clearly regulated and apparent on which side the individual responsibilities lie, especially the responsibility for control. The distribution of responsibilities (and thus of accountability), for instance with regard to the time and access arrangements, should be documented and stored. This applies especially to the human-to-technology handover procedures” (Ethics Commission, 2017). This applies to all forms of highly and partly automated driving, whereas fully automated or autonomous driving is not a problem here, since control is completely exerted by technology.

Another focus of the debate is on the so-called trolley problem. It refers to seemingly hopeless situations where there are only more or less bad solutions rather than good

ones. Again and again, the question is asked whether an AV should, in such a situation, run down, e.g., two children or three elderly people, if there is no third option. So far, we have considered these situations as tragic occurrences where someone was at the wrong place at the wrong time. However, when it comes to autonomous driving, the different descriptions of the trolley problem gain a completely new and, above all, *practical* relevance. When using autonomously operating machines, tragic constellations in road traffic suddenly seem to turn into dilemma situations, which bear a striking resemblance to the theoretical thought experiment of the trolley problems. Long before a potential dilemma situation occurs, it must be clarified how or according to what criteria an autonomous vehicle should decide in such a situation.
The basic criteria on which algorithms should control the behavior of autonomous machines must still be determined by humans (Wallach and Allen, 2009). Moral decisions (about people's lives) of autonomous machines in road traffic must ultimately be made by humans. Accordingly, responsibility for life and death (in the case of dilemma situations) or responsibility for accidents caused by autonomous vehicles cannot be delegated to technology. However, this leaves open the question of who is or should be responsible for the behavior of autonomous vehicles in road traffic and its consequences.

**Care robots**

Health and care are important fields of application for AI-based technologies. In light of an increasingly aging society and a growing share of people in need of care in the total population, the future of care is a major societal challenge. Autonomously operating service or care robots and assistance technologies in combination with neurotechnologies (e.g., exoskeletons) are considered to have great potential to support care. AI plays an important role here in enabling technologically autonomous systems to act adequately in complex environments, e.g., through real-time detection of relevant and sometimes rapidly changing framework conditions during active operation or of the condition of the affected persons, e.g., in the case of variable dementia (Decker et al., 2017). Since humans and technology come into close contact in these fields of application and the affected people are often in need of care and thus might be helpless against potential malfunctions of technology, an AI-based recognition of the persons' condition is extremely important. Because individuals and their relatives are directly affected in all fields of health and care and questions of dignity, autonomy, and humanity arise, it is obvious that ethical and legal questions must be considered when developing such technologies. However, also the practical needs of the patients, their relatives, and the nursing staff must be taken into account early in the development in order to avoid purely technocratic solutions and prevent the occurrence of acceptance problems later on. Not only is the problem- and addressee-related approach to technology design crucial here, but also the question of the right timing (Decker and Fleischer, 2010). Therefore the questions of how the respective AI-based assistance systems and, above all, a governance of technology development and the practical use of technology that appropriately responds to the care challenges could look like, are issues which cannot be dealt with by developers, ethicists, and lawyers alone, but need a complex transdisciplinary network of stakeholders and those affected. Technology design in this field must therefore include the groups of persons involved to a particularly high degree.

Overall, these examples show that technology design – here and in many other fields of digitalization – should not be developed exclusively by computer scientists and engineers. In addition to technical expertise and excellence, there is also a need for impact awareness, ethical sensitivity, and legal competence, including the knowledge of the respective disciplines.

## INTER-DISCIPLINARY CO-DESIGN FOR THE ERA OF DIGITALIZATION

Technology and society mutually influence each other (Rip et al., 1995). Society contributes to shaping science and technology by providing incentives and developing regulation, by market forces under global competition, and by consumer demands. At the same time, science and technology have consequences for society. For shaping technology for the digital age, new developments must be assessed by exploring technical potentials and non-technical issues in a systemic interplay of increasingly digitized technology and the digital transformation of society, taking into account ecological and economic factors, ethical aspects and societal acceptance, and political framework conditions.

### Technology assessment

Technology assessment emerged in the 1970s in the United States as a science-based and policy-advising activity (Bimber, 1996) with the Office of Technology Assessment at the US Congress as the first TA institution. In its first period, technology was considered to follow its own dynamics (technology determinism, Sec. 2 above) with the consequence that shaping technology was not an issue. The main task of TA at that time was seen in its early-warning function in order to enable political actors to take measures, for example, to compensate or prevent anticipated negative impacts of technology. This changed completely in the 1980s, following the social constructivist paradigm, which emphasized opportunities for shaping technology according to social needs and values (Bijker and Law, 1994). In this framework, the approach of constructive technology assessment (CTA) was developed (Rip et al., 1995). CTA started considering technology development and innovation processes (Smits and ten Hertog, 2007). TA as a guide to designing new technology and possibly resulting innovations has since then been part of the overall TA portfolio.

Technology assessment is an interdisciplinary research field aiming at, generally speaking, providing knowledge for better informed and well-reflected decisions concerning new technologies (Grunwald, 2019a). Its initial and still valid motivation is to provide answers to the emergence of unintended and often undesirable side effects of science and technology. TA is intended to add rationality and reflexivity to technology governance by integrating any available knowledge on possible side effects, by supporting the evaluation of technologies according to societal values and ethical principles, by elaborating strategies to deal with inevitable uncertainties, and by contributing to constructive solutions to societal conflicts around science and technology. There are three partially overlapping branches of TA addressing different targets in the overall technology governance (following Grunwald, 2019a):

(1) TA has initially been conceptualized as *policy advice* (Bimber, 1996), and still many TA activities are located in this field (Grunwald, 2019a). The objective is to support policy makers in addressing the above-mentioned challenges by implementing political measures such as adequate regulation (e.g., the

precautionary principle), sensible research funding, and strategies toward sustainable development involving appropriate technologies. In this mode of operation, TA does not *directly* address technology development, but considers the *boundary conditions* of technology development and use.

(2) It became clear during the past decades that citizens, consumers and users, actors of civil society, stakeholders, the media, and the public are also engaged in technology governance in different roles. Participatory TA developed approaches to involve these groups in different roles at different stages in technology governance (Abels and Bora, 2016; Joss and Belucci, 2002). *Citizen science* is a current field of participation.

(3) A third branch of TA is more directly related to concrete technology development and engineering. Departing from analyses of the genesis of technology conducted in the framework of social constructivism (Bijker et al., 1987), the idea of *shaping technology* according to social expectations and values came up. Different approaches, including system-analytical TA (Grunwald and Achternbosch, 2013), aim at adding prospective knowledge about possible consequences and impacts of technology to the design, development, and engineering processes. The overall aim is to strive for "better technology in a better society" (Rip et al., 1995).

In order to make TA work in specific projects, a set of methods has been developed in the form of a "method toolbox" (see Decker and Ladikas, 2004). The methods applied in TA are research methods, interactive methods, and communication methods. *Research methods* are applied to TA problems in order to collect data, to facilitate predictions, to do quantitative risk assessment, to allow for the identification of economic consequences, to investigate social values or acceptance problems, and to do eco-balancing. *Interactive, participatory, or dialogue methods* are needed to organize social interaction in such a way as to facilitate conflict management, allow for conflict resolution, bring scientific expertise and citizens together, involve stakeholders in decision-making processes, and mobilize citizens to shape society's future. Values play a crucial role in all of these fields. Therefore, TA cannot be value neutral but has to be careful and transparent while conducting its assessment and evaluation processes (Grunwald, 2019a). The field of digitalization is, according to the examples mentioned but going far beyond, a major challenge to TA.

**Value Sensitive Design**

*Value Sensitive Design* (VSD, Friedman et al., 2006; van den Houven et al., 2015) can goes back to the revelation of value structures implemented in technology (e.g., van de Poel, 2009) which raised the question of whether and how this would allow for explicit technology design according to values (Brey, 2009). Accordingly, VSD aims to translate relevant values into design requirements and specific technology (van de Poel, 2013). To this end, it includes the iterative and interrelated steps of identifying relevant groups of users and affected persons including their moral concepts, the interpretation of values as well as philosophical and ethical reflections in the respective technical context, the transfer or operationalization in design requirements, the resolution of potential incompatibilities or conflicts between value fulfillments and, eventually, the review of the implementation.

A field of special interest in the design of digital systems is related to their often action-

regulating and institution-like character. Institutions are established social rule systems, which enable, organize, and limit social interactions and are equipped with means of enforcement, such as incentives or sanctions, and thus contribute decisively to the concrete realization of social values (Brey, 2009). Software and AI-based networks and services are increasingly taking over the role of such regulatory elements. Because the progressive shift of interpersonal interactions into digital systems also means that the associated social rules and underlying values are "programmed into the systems". This enables the automation of rule enforcement, the technical avoidance of rule deviations, and the establishment of rules in fine granularity.

Against this background, complex algorithms are attributed an independent institutional effect analogous to social rule systems. Examples can be found in social rules of internet filters, digital rights management systems, internet architectures and protocols, search engines, e-commerce systems, social networks and online communities, AI systems, decision rules in big data analytics, or scoring systems implemented by information technology (Orwat et al., 2010).The literature on software as institution points out that regulations using software tend to "overwrite" or jeopardize conventional rules, values, and expectations, that they can gradually infiltrate public rule systems, even though they have no democratic legitimacy if generated by globally operating corporations. The design of AI-based systems as services can build upon experiences of VSD with information and communication technologies (e.g., van den Hoven, 2007; Brey, 2009). In the field of digitalization, VSD can be easily combined with TA.

**Responsible research and innovation (RRI)**

The ideas of "responsible research" in scientific and technological advance and "responsible innovation" in the field of new products, services, and systems, which have been discussed intensively for years (Siune et al., 2009, Owen et al., 2013) and gave rise to the phrase "responsible research and innovation" (RRI). The concept of responsible innovation adds explicit ethical reflection to technology assessment (TA) and science, technology and society (STS) studies and incorporates them all into integrative approaches to shaping technology and innovation. Responsible innovation thus brings together TA with its experience in assessment procedures, actor involvement, foresight, and evaluation with ethics, in particular in the context of responsibility, and also builds on the body of knowledge about R&D and innovation processes provided by STS and STIS studies (science, technology, innovation and society). Ethical reflection and technology assessment are increasingly taken up as integrative part of R&D programs (Siune et al., 2009).

The emergence of RRI (Siune et al., 2009) reflects the diagnosis that available approaches to shaping science and technology still do not meet all of the far-ranging expectations toward technology governance and achieving a "better technology in a better society" (Rip et al., 1995). The hope behind the responsible innovation movement is that new – or further developed – approaches could add considerably to existing approaches such as TA and engineering ethics. Indeed, compared to earlier approaches there are shifts of accentuation and new focuses of emphasis (Grunwald, 2011):

- "Shaping innovation" complements or even replaces the former slogan "shaping technology" which characterized the social constructivist approach to technology. This shift reflects the insight that it is not technology *as such* which influences society and therefore should be shaped according to society's needs, expectations, and values, but it is *innovation* by which technology and society interact as has been pointed out by many STIS studies.
- Now, a closer look is taken at societal contexts of new technology and science. Responsible innovation can be regarded as a further step toward taking the demand-pull perspective and social values more seriously in shaping technology and innovation.
- Instead of distant *observation* following classical paradigms of science, there is now a clear indication for *intervention* into the development and innovation process: Responsible innovation projects shall "make a difference" not only in terms of research but also as interventions into the "real world".
- The spectrum of stakeholders to be involved in participatory processes and dialogue must be broadened further due to new forms of science and technology governance (Siune et al., 2009).

Following the above-mentioned issues, responsible innovation can be regarded as a radicalization of the well-known post-normal science (Funtowicz and Ravetz, 1993), being even closer to social practice. Responsible research and innovation unavoidably requires more intense inter- and transdisciplinary cooperation between engineering, social sciences, and applied ethics, similarly to, but often exceeding, TA and VSD. The major novelty in this interdisciplinary cooperation might be the integration of ethics (normative reflection on responsibilities) and social sciences such as STS and governance research (empirically dealing with social processes around the attribution of responsibility and their consequences for governance).

## CONCLUDING REMARKS

New digital technologies open up numerous new possibilities in almost every aspect of modern society. However, they also create new challenges. Firstly, the rapid growth and vital importance of information and data in all branches of society generate enormous requirements for data handling, processing, and computing. Secondly, digitalization processes are an enabler for significant societal transformation processes, which must be designed responsibly. Thirdly, the high dependence on digital information and communication technologies in infrastructures such as the energy supply system, the mobility system, or information processing and storage capacities in many fields may create new threats to security and privacy. Fourth, new human-machine interfaces will require ethical and anthropological consideration. Furthermore, all these transformation processes may have huge implications for employment, inclusion, and distributive justice, which could trigger future social conflicts.

Therefore, interdisciplinary approaches will have to focus on the social, political, and economic preconditions as well as on the impacts of socio-technical change by applying technology assessment, systems analysis, innovation research, and ethics. To address the grand challenges that society, science, and the economy are facing, new technology must be embedded successfully into its social environment. This general statement applies particularly to ongoing and accelerating digitalization as

well as to its promise to profoundly transform a variety of social, economic, and scientific activities and competencies. The successful integration of new technologies into their societal target systems requires approaches such as applied ethics, technology assessment, or RRI.

## REFEENCES

Abels, G. and Bora, A. (2016). Ethics and Public Participation in Technology Assessment. [online] DOI 10.13140/RG.2.2.35586.89282.

Bijker, W. and Law, J. (eds) (1994). Shaping Technology Building Society. Cambridge: MIT Press.

Bijker, W., Hughes, T. and Pinch, T. (eds) (1987). The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology. Cambridge/London: MIT Press.

Bimber, B.A. (1996). The Politics of Expertise in Congress: the Rise and Fall of the Office of Technology Assessment. New York: State University of New York Press.

Börner, F., Kehl, C. and Nierling, L. (2018). Chancen und Risiken mobiler und digitaler Kommunikation in der Arbeitswelt. TAB-Arbeitsbericht Nr. 174. Berlin: Büro für Technikfolgenabschätzung beim Deutschen Bundestag.

Brey, P. (2009). Values in Technology and Disclosive Computer Ethics. In: L. Floridi, ed., The Cambridge Handbook of Information and Computer Ethics. Cambridge: Cambridge University Press, pp. 41–58.

Decker, M. and Fleischer, T. (2010). When Should There Be Which Kind of Technology Assessment? A Plea for a Strictly Problem-Oriented Approach from the Very Outset. Poiesis & Praxis, 7, pp. 117–133, DOI:10.1007/s10202-010-0074-6.

Decker, M. and Ladikas, M. (2004). Technology Assessment – Method and Impact. Berlin: Springer

Decker, M., Weinberger, N., Krings, B. and Hirsch, J. (2017). Imagined Technology Futures in Dmand-oriented Technology Assessment. Journal of Responsible Innovation, 4(2), pp. 177–196.

Ethics Commission (2017). Automated and Connected Driving. Final Report. Berlin: Federal Ministry of Transportation and Digital Infrastructures. Available at: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.html [Accessed 19 Aug. 2018].

Friedman, B., Kahn, P. and Borning, A. (2006). Value Sensitive Design and Information Systems. In: P. Zhang and D. Galletta, eds, Human-Computer Interaction in Management Information Systems: Foundations. New York/London: M.E. Sharpe.

Funtowicz, S.O., Ravetz, J. (1993). Science in the Post-Normal Age. Futures, 25(7), pp- 739–756.

Grunwald, A. (2011). Responsible Innovation: Bringing Together Technology Assessment, Applied Ethics, and STS Research. Enterprise and Work Innovation Studies, 7, pp. 9–31.

Grunwald, A. (2018). Self-Driving Cars: Risk Constellation and Acceptance Issues. DELPHI – Interdisciplinary Review of Emerging Technologies, 1, pp. 8–13.

Grunwald, A. (2019a). Technology Assessment in Practice and Theory. London: Routledge.

Grunwald, A. (2019b). Der unterlegene Mensch. Die Zukunft der Menschheit angesichts von Algorithmen, künstlicher Intelligenz und Robotern. Munich: RIVA Verlag.

Grunwald, A. and Achternbosch, M. (2013). Technology Assessment and Approaches to Early Engagement. In: N. Doorn, D. Schuurbiers, I. van de Poel, M. Gorman, eds, Early Engagement and New Technologies: Opening Up the Laboratory. Dordrecht et al.: Springer, pp. 15–36.

Hermann, M., Pentek, T., Otto, B. (2016). Design Principles for Industrie 4.0 Scenarios. In: IEEE, 2016 49th Hawaii International Conference on System Sciences (HICSS), pp. 3928–3937, doi:10.1109/HICSS.2016.488.

Joss, S. and Bellucci., S. (eds) (2002). Participatory Technology Assessment – European Perspectives. London: Westminster University Press.

Mainzer, K. (2016). Künstliche Intelligenz – Wann übernehmen die Maschinen? Berlin/Heidelberg: Springer.

Manzlei, C., Schleupner, L. and Heinz, R. (eds) (2016). Industrie 4.0 im internationalen Kontext. Berlin: VDE Verlag.

Maurer, M., Gerdes, J., Lenz, B. and Winner, H. (eds) (2016). Autonomous driving. Technical, Legal and Social aspects. Heidelberg: Springer Open.
Merton, R. (1948). The Self-Fulfilling Prophecy. The Antioch Review, 8(2), pp. 193–210.
Orwat, C., Raabe, O., Buchmann, E., Anandasivam, A., Freytag, J.-C., Helberger, N., Ishii, K., Lutterbeck, B., Neumann, D., Otter, T., Pallas, F., Reussner, R., Sester, P., Weber, K. and Werle, R. (2010). Software als Institution und ihre Gestaltbarkeit. Informatik-Spektrum, 33(6), pp. 626–633.
Owen, R., Bessant, J. and Heintz, M. (eds) (2013). Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society. London: Wiley.
Rip, A., Misa, T. and Schot, J. (eds) (1995). Managing Technology in Society. London: Pinter.
Siune, K., Markus, E., Calloni, M., Felt, U., Gorski, A., Grunwald, A., Rip, A., de Semir, V. and Wyatt, S. (2009). Challenging Futures of Science in Society. Report of the MASIS Expert Group. Brussels: European Commission.
Smits, R. and den Hertog, P. (2007). TA and the Management of Innovation in Economy and Society. International Journal on Foresight and Innovation Policy, 3, pp. 28–52.
van de Poel, I. (2009). Values in Engineering Design. In: A. Meijers, ed., Philosophy of Technology and Engineering Sciences, 9, Amsterdam: Elsevier, pp. 973–1006.
van de Poel, I. (2013). Translating Values into Design Requirements. In: D. Mitchfelder, N. McCarty and D. Goldberg (eds): Philosophy and Engineering: Reflections on Practice, Principles and Process. Dordrecht: Springer, pp. 253–266.
van den Hoven, J. (2007). ICT and Value Sensitive Design. In: P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa and V. Laurent, eds, The Information Society: Innovations, Legitimacy, Ethics and Democracy. Boston: Springer, pp. 67–72.
van den Hoven, J., Vermaas, P. and van de Poel, I. (eds) (2015). Handbook of Ethics, Values, and Technological Design. Sources, Theory, Values and Application Domains. Dordrecht: Springer.
Wallach, W. and Allen, C. (2009). Moral Machines: Teaching Robots Right from Wrong. New York: Oxford University Press.

**Abstract.** Technology research, design, and development is confronted with rapidly advancing digitalization in two respects: (1) digitally supported or enabled technologies need to be designed and developed, and (2) the respective R&D processes themselves will happen in a much more digitalized environment. Technology design generally must take into account the values involved and possible consequences of the development and use of the resulting products, services, and systems. In a digitalizing environment, the issue of values gains even more significance because more and more close and intimate interfaces between humans and technology have to be shaped. Designing human-machine interfaces is not only a functional issue but touches upon ethical questions such as the distribution of responsibility, but also upon anthropological issues related to the human self-image and ideas about future society as well. In the respective research, design, and development processes, value-laden issues such as control, privacy, empathy, responsibility, and accountability must be taken into account beyond technical issues of efficiency and reliability. The need for designing and shaping digital future technologies involving ethics and technology assessment will be demonstrated by three examples: future industrial production and the fields of self-driving cars and care robots. Value sensitive design and responsible research and innovation will be introduced as approaches to deal with these challenges.

**Keywords:** digitalization, technology design, technology assessment, ethics, responsible research and innovation