Code of Digital Ethics

Version 1.0, July 2021

Merck KGaA, Darmstadt, Germany

Table of contents

1. Code of Digital Ethics

1.1 Preamble

1.1.1 Preamble – Introduction

People, machines, data and processes are increasingly interlinked in today's world. Technological advances based on the collection of data and its processing by algorithmic systems are confronting us with significant societal change. Merck KGaA, Darmstadt, Germany (hereinafter referred to as “Merck”, “we” and “us”), is committed to putting people at the center of its activities. Our goal is a clear orientation toward ethical principles in the digital age, just as bioethical principles have already guided us in our research and development to date.

People's trust is crucial if we are to seize the opportunities offered by digital progress and minimize the risks. Security, reliability and the acceptance of responsibility are just some of the aspects we want to take into account when designing digital innovations. We have therefore developed 20 digital-ethical principles that guide us in the design and development of digital technologies, products and services.

These principles have common basic requirements, which we have committed ourselves to strive to fulfil. Thus, the guidelines are linked to applicable law and an irrevocable commitment to human rights. We consider the human dignity contained in human rights to be a concept worth protecting. With the digital ethical principles, however, we even go beyond laws and regulations currently applicable to us. To the extent our principles go above and beyond applicable laws and regulations, they are a declaration of intent providing us with an ethical framework that supports us in responsible decision-making. It is crucial for us that these principles are incorporated into digital development from the very beginning. Furthermore, our businesses and developers of new technologies may approach our Digital Ethics Advisory Panel to receive advice and guidance on how best to adhere to these principles

1.1.2 Preamble – Terms of Use

The following principles, effective as of May 2021, apply to the development, commercialization, or utilization of all our digital products and processes. In practice, however, they will be of varying relevance to, for example, the various stages of development and not all principles will apply to all projects or processes. One reason for this is different reference levels of the underlying principles. Existing reference levels are data and algorithmic systems. The former focuses on digital data, which is used, for example, as a basis for machine learning. The latter looks at the architecture and dynamics of data-processing algorithmic systems.

The two perspectives apply to the design, application and use of data and algorithmic systems of any kind. Both reference levels also complement each other, so that a large part of the principles refers to both data and algorithmic systems. Exceptions are the principles traceability and explainability, which each refer to only one level.

Our five core principles are: autonomy, justice, beneficence, non-maleficence and transparency. Each of these principles is further defined by three subsidiary principles.

Based on our principles, we can reflect and classify our actions against the background of ethical considerations and align them accordingly.

1.2 Overview

1. Justice
„We stand up for justice in our digital offerings."

2. Autonomy
„We respect the autonomy of every single human in our digital offerings."

3. Beneficence
„We promote the needs and well-being of individuals and society through our digital offerings."

4. Non-Maleficence
„We avoid doing harm through our digital offerings."

5. Transparency
„We strive for the greatest possible transparency for our digital offerings."

1.3 Ethical Guidelines

1.3.1 Justice

Justice
„We stand up for justice in our digital offerings."
We recognize that justice is a fundamental part of human-centered digital offerings. We are committed to a fair distribution of digital resources and want to prevent discrimination in the long term. We therefore act in accordance with the three principles - *impartiality, equality* and *proportionality*.

Impartiality
„We actively combat prejudice in our digital offerings."
Every time data is collected and/or used, a bias may be present that leads to distortions in the results. Such a bias arises from cultural and personal prejudices, beliefs and life experiences. We are cognizant of prejudices in digital offers. To ensure that our processes minimize prejudice, we are deeply committed to involving diverse perspectives and data sources.

Equality
„We take measures to ensure equality in and through digital offerings."
We contribute to an equal opportunity to take social and economic benefit from the use of data and algorithms. We are committed to ensuring that the opportunity to benefit from our digital offerings is unimpeded by prejudices, preferences or other artificial and unjust barriers. We also demand the same from the digital services of third parties that we use in our own company.

Proportionality
„We are committed to proportionality in the collection and use of data."
We ensure that the collection and use of data is proportionate. Within data projects, data is therefore collected and used in a purposeful manner so that only necessary data is collected. Where data influences our decision-making, we recognize the limits of the meaningfulness of data and do not draw hasty and disproportionate conclusions.

1.3.2 Autonomy

Autonomy
„We respect the autonomy of every single human in our digital offerings."
We recognize that everyone has the right to live according to their own ideas and set their own goals. People should make their own decisions according to their own standards as long as this does not harm the interests of others. This also applies in the digital context when transmitting data or agreeing that data may be evaluated using an algorithm. We do not objectify human beings. The following three guidelines, based on the principles of *explainability*, *privacy* and *literacy*, guide the development of our digital services.

Explainability
„We explain our algorithmic systems."
We are convinced that algorithmic systems should be explainable. Everyone who uses or considers using our digital services should know whether they are directly or indirectly affected by an automated decision. If this is the case, we are committed to ensuring that they understand how the results of algorithm-based processes are achieved. Algorithmic systems that cannot be explained are carefully examined for their added value.

Privacy
"We protect the privacy of those who provide data."
We respect and protect the privacy of individuals, both during the collection of data and during interaction with our digital offerings. We therefore ensure that individuals can make sovereign decisions about the use of their data.

Literacy
„We promote digital literacy."
We actively promote the digital competence of those who develop, use, operate and decide on the deployment of algorithmic systems. Therefore, we provide these people with up-to-date knowledge regarding the handling of data and algorithmic systems. In addition, we support our employees in asking digital ethical questions and questioning the effects of our algorithmic systems.

1.3.3 Beneficence

Beneficence
„We promote the needs and well-being of individuals and society through our digital offerings."
We recognize that our digital offerings have the potential to promote the well-being of individuals and society. We want to achieve this positive impact short-, medium- and long-term. We want to ensure that users have clear benefits from using our data and algorithmic systems. Therefore, we act according to the following three guidelines based on the principles of *security*, *sustainability* and *responsibility*.

Sustainability
„We keep an eye on the impact of our digital offerings and design them to be sustainable."
We know that the decisions we make today about the use of data and algorithmic systems will have an impact on future generations. Therefore, we design our digital services in such a way that they can have a positive long-term effect on people. This includes ensuring the compatibility and readability of data for future software systems as well as a prudent and foresighted use of algorithmic systems.

Security
„We protect our algorithmic systems and the data entrusted to us."
We store and manage data in secure environments and protect it from unauthorized access. In addition, at the beginning of all our projects, we implement technical protection measures to ensure the security of data and algorithmic systems. In the case of algorithmic systems, we thoroughly check which data is included in the calculation procedures.

Responsibility
„We take responsibility for data and algorithmic systems."
We encourage and ensure that everyone in our company who develops, operates or decides on the use of digital services develops a sense of responsibility for working with data and algorithmic systems. We believe that responsibility can neither be delegated to an algorithmic system nor to those who provide data.

1.3.4 Non-Maleficence

Non-Maleficence
„We avoid doing harm through our digital offerings."
We recognize that in many decisions regarding digital services, a compromise must be found between feasibility and potential risk. For important decisions, we thoroughly assess the advantages and disadvantages of the individual options. Humans are always at the center of this assessment. The guidelines of *reliability*, *controllability* and *accountability* help us to avoid damage.

Reliability
„We ensure the reliability of our digital offerings."
We are aware that the reliability of our digital services is crucial for their use. We therefore continuously ensure that the data processes and algorithmic systems we establish or use work as they should.

Controllability
"We maintain control over our data and algorithmic systems."
We maintain control over our data and algorithmic systems and ensure that human intervention is always possible. We actively fight against paternalistic effects of data and algorithmic systems.

Accountability
„We assign clear responsibilities for our digital offerings."
We create clear organisational structures for our handling of data and algorithmic systems. We ensure clear responsibilities within our processes. We do not ignore negligent behaviour and prevent it through proactive measures. In addition, we discuss this with our business partners to determine responsibilities with them as well.

1.3.5 Transparency

Transparency
„We strive for the greatest possible transparency for our digital offerings."
We recognize that open communication about the use of data and algorithmic systems is important. We want to inform others about our intentions and our behavior to the appropriate extent. The principles *comprehensibility*, *interactivity* and *traceability* help us to achieve transparency.

Traceability
„We are able to trace the origin of our data."
We pay attention to the origin of our data, how it is linked together in the course of our use and how it is further processed. We want the history of the data to be recordable and communicable. We are committed to using data from qualified and verified sources.

Interactivity
„We enable interactive communication."
We are committed to an interactive approach to the users of our digital services. We make sure that users of our digital services can contact us directly. We give users the opportunity to voice problems or concerns and we take these seriously.

Comprehensibility
„We ensure that our digital offerings are understandable."
We make sure that our digital offerings are understandable, and we want to explain them adequately. When interacting with the those who use our digital services or decide on their use, we pay attention to the different social and cultural backgrounds and tailor our communication accordingly.

2. Appendix

2.1 Principle Definitions

2.1.1 Justice

Justice is a core principle to which the sub-principles of *impartiality*, *equality* and *proportionality* are assigned in the context of digital solutions.

Justice can be understood as the moral obligation to make a fair judgement between competing claims. As such, it is about distributive fairness and respect for human rights to prevent the systematic exclusion of users or groups of users by unjustified barriers.

An injustice can occur when a person is denied his or her entitlement to a benefit without good reason or when a disproportionate burden is imposed on him or her. In the handling of data and/or algorithmic systems, for example, there is a risk that certain people will have unequal opportunity to benefit from digital solutions or will be structurally discriminated against in the collection/usage of data.

Organisations that commit themselves to the principle of *justice* have established measures to uphold the principles of *impartiality*, *equality* and *proportionality*. Therefore, they reduce prejudice in digital solutions, expose, reduce and prevent structural inequalities and build trust through the appropriate usage of data.

Impartiality
Level of reference: Data and algorithmic systems

In the context of digital solutions, *impartiality* means that data is collected and/or used without prejudice or discrimination. Prejudice and discrimination can result from unrecognized or ignored distortions of perception (also called bias). A bias can arise from cultural and personal prejudices, beliefs and life experiences and is also reflected in digital solutions.

The design of the data acquisition may include one or more biases. Also, the way data is processed by an algorithmic system may be distorted by the algorithmic design. For example, historical data may be based on outdated world views that are no longer valid today. In addition, the selection of criteria for data analysis by algorithmic systems can be based on unintended prejudices against certain people.

Organisations that commit themselves to the principle of *impartiality* recognize the existence of prejudices and are aware that dealing with data is not a neutral process. They therefore take specific measures to promote the inclusion of diverse perspectives and data sources.

Since promoting the sub-principle of *impartiality* actively combats prejudice in digital solutions, it contributes to preserving the core principle of *justice*.

Equality

Level of reference: Data and algorithmic systems

In the context of digital solutions, *equality* means that no one ought to be excluded from the social and economic benefits of using data and algorithmic systems based on artificial and unjust barriers.

Unequal access to digital solutions can reinforce structural inequalities at social and economic level or create new inequalities.

Organisations that commit themselves to the principle of *equality* strive to ensure that unjustified structural barriers that determine who can benefit from digital solutions are identified, reduced, or prevented. Therefore, attention is paid to the issue of accessibility – for example, by providing high-contrast presentations or read aloud elements.

Since the promotion of the principle ensures *equality* in and through digital solutions, it contributes to preserving the core principle *justice*.

Proportionality

Level of reference: Data and algorithmic systems

In terms of digital solutions, *proportionality* refers firstly to the extent to which data is collected and secondly to the limitations of data regarding their interpretability by algorithmic systems.

A risk in the use of data is that data collected today may be used for unforeseen purposes in the future. Unnecessary or excessive collection of data increases that risk of misuse.

Organisations that commit themselves to the principle of *proportionality* firstly only collect the data necessary to achieve their objectives. Secondly, they are aware of what data is processed in algorithmic systems to avoid disproportionate conclusions. To ensure that data is used proportionately, organisations document their entire process within the data chain.

Since promoting the sub-principle of *proportionality* ensures a fair use of digital resources, it contributes to preserving the core principle of *justice*.

2.1.2 Autonomy

Autonomy is a core principle to which the sub-principles *explainability, privacy* and *literacy* are assigned in the context of digital solutions.

Autonomy has its origin in the ability of people to develop their own ideas of a good life and to pursue goals in a self-determined way. Accordingly, every human being has the right to make his own decisions and to act according to his own standards as long as this does not harm the interests of others.

When *autonomy* is restricted, the human can become objectified. In the context of digital solutions, this danger exists both in the collection and processing of personal data and in automated decision making by algorithms.

Organisations committed to the principle of *autonomy* have established measures to uphold the principles of *explainability, privacy* and *literacy.* To this end, they explain their algorithmic systems, respect the status of personal information and support users in acquiring user skills.

Explainability
Level of reference: Algorithmic systems

In the context of digital solutions, *explainability* means that users can understand how the results of algorithm-based processes are achieved. If this principle is not respected, there is a risk of restricting the right of self-determination and, as a consequence, of a growing mistrust towards algorithmic systems and organisations using them.

Organisations committed to the principle of *explainability* ensure that users know the reason for their decision wherever they are directly or indirectly affected by automated decisions. For example, in the application of an algorithmic system, it is understood which data are linked together to achieve the corresponding result. In addition, users receive sufficient information about the models used and the architecture of the algorithmic system.

Since promoting the sub-principle of *explainability* enhances the decision-making capacity of individuals, it contributes to preserving the core principle of *autonomy*.

Privacy
Level of reference: Data

In the context of digital solutions, *privacy* refers to information about, for example, an individual's family, friends, health and financial circumstances.

Respect for *privacy* means that each user is able to determine what information is shared with others. This concerns information originally provided by the user as well as information generated about the user in the course of his/her interaction with digital solutions. Otherwise, there is a risk, for example, that users may feel that their *privacy* is being invaded.

Organisations committed to the principle of *privacy* respect and guarantee privacy at every stage of the data chain: whether at the beginning when determining the type of processing or at the time of processing within the data chain. This is where measures such as anonymization of personal data, the use of synthetic data or software dashboards to control privacy settings by users come into play.

Since the promotion of the sub-principle of *privacy* recognises the self-determination of the user, it contributes to preserving the core value of *autonomy*.

Literacy

Level of reference: Data and algorithmic systems

In the context of digital solutions, *literacy* means sufficiently educating users about digital solutions. The goal of the sub-principle differs in terms of whether the users are external to the organisation or internal to the organisation.

For the former, *literacy* means imparting knowledge of data and algorithms to help them build user competence. This user competence enables the risks and opportunities of digital technologies to be generally understood. If *literacy* is not ensured, there is a risk that organisations will exploit their users. Organisations committed to the principle of *literacy* share knowledge and publicize the social and ethical implications of digital solutions.

Within the organisation, *literacy* means ensuring and promoting the competence and qualification of employees in dealing with digital solutions. Those who develop, use, operate or decide on the deployment of algorithmic systems need a sufficient understanding of how these systems work and their possible (ethical) implications. Otherwise, there is a risk of improper use and misinterpretation of data and algorithmic systems. Organisations committed to the principle of *literacy*, for example, provide competence-building measures such as training, further education, and certification of employees.

Since promoting the sub-principle of *literacy* enables users to make informed choices and encourages critical reflection by staff and decision-makers, it contributes to maintaining the core principle of *autonomy*.

2.1.3 Beneficence

Beneficence is a core principle to which the sub-principles of *security, sustainability* and *responsibility* are assigned in the context of digital solutions.

Beneficence has its origin in the demand to promote the needs and well-being of individuals and society through action.

If this is not the main motive for developing and using digital solutions, there is a risk that organisations will lose the opportunity to make a positive impact on society. In addition, trust in digital solutions can be permanently lost.

Organisations committed to the principle of beneficence have established measures to uphold the principles of *safety*, *sustainability*, and *responsibility*. To this end, they protect data and algorithmic systems from unauthorized interference, set long-term goals and prevent negative effects by promoting individual attention to ethically questionable situations.

Sustainability
Level of reference: Data and algorithmic systems

In the context of digital solutions, *sustainability* means keeping in mind their transformative and long-term effects on the individual and society.

Certain developments in the field of data and algorithmic systems that were started with good intentions may have a negative impact on individuals and society in the future. Another aspect of *sustainability* concerns the long-term availability of data. Data is supposed to be available as digital resources for future generations.

Organisations committed to the principle of *sustainability* regularly review the potential impact of their data and algorithmic systems. They also recognize data as a digital resource. Data is stored in formats that are conducive to collaboration and sharing.

Because by promoting the sub-principle of *sustainability*, digital solutions are designed to improve people's lives in the long term, it contributes to preserving the core principle of *beneficence*.

Security
Level of reference: Data and algorithmic systems

In the context of digital solutions, *security* means that organisations take measures to protect data and algorithmic systems against attacks, unauthorised access, and manipulation.

The special nature of algorithmic systems, especially in the form of self-learning systems, raises further security aspects that organisations take into account. For example, the robustness against hostile input data must be strengthened and the manipulation of self-learning systems has to be prevented.

Organisations committed to the principle of *security* use active protection measures for data and algorithmic systems. Such protective measures are, for example, the strict control of the allocation of access rights or the renunciation of an insecure Internet connection when using algorithmic systems.

Since promoting the sub-principle of *security* minimizes risks and dangers of authorized access to data and thus creates added value for individuals and society, it contributes to maintaining the core principle of *beneficence*.

Responsibility
Level of reference: Data and algorithmic systems

In the context of digital solutions, *responsibility* means that members of an organisation are aware of the potentially negative impact of their technology on society. They are able to take *responsibility* for them and efforts to limit them.

Responsibility can neither be transferred to the algorithmic system nor to those who provide data. Otherwise, there is a risk of potentially negative impacts of technology being overlooked and thereby creating a general uncertainty regarding digital offerings.

Organisations committed to the principle of responsibility empower those who responsibly develop, operate, or decide on the deployment of digital solutions. This can be done, for example, through qualification measures.

Since promoting the sub-principle of responsibility actively encourages members of the organisation to point out potential risks, it contributes to preserving the core principle of *beneficence*.

2.1.4 Non-Maleficence

Non-maleficence is a core principle to which the sub-principles *reliability, controllability* and *accountability* are assigned in the context of digital solutions.

Non-maleficence means avoiding harm to people or minimising risks. In the development of digital solutions, compromises are made with regard to the advantages and disadvantages for the people concerned. Many factors can influence a compromise, but the protection of people always takes priority over all considerations of usefulness.

If compromises are not based on the principle of *non-maleficence*, there is a risk, for example, that some users will be disadvantaged or even harmed by new digital solutions.

Organisations committed to the principle of *non-maleficence* have established measures to uphold the principles of *reliability*, *controllability*, and *accountability*. To do this, they deliver on the promises made within digital solutions, they ensure that a human being has ultimate control over data and algorithmic systems, and that an internal authority structure is provided.

Reliability
Level of reference: Data and algorithmic systems

In the context of digital solutions, the sub-principle of *reliability* means that warranted characteristics and performance are actually fulfilled by such solutions.

If the principle of *reliability* is not actively promoted, there is a risk that potentially harmful impacts of digital solutions cannot be adequately anticipated or avoided, which also leads to a long-term loss of trust in digital solutions in general or in the organisation providing them.

Organisations committed to the principle of *reliability* regularly check the constancy within their data chain and algorithmic systems. One measure to ensure *reliability* is, for example, the regular execution of planned queries to the system. The queries are structured according to the guaranteed characteristics and performance.

Since promoting the sub-principle of reliability supports the safe functioning of digital solutions, it contributes to maintaining the core principle of *non-maleficence*.

Controllability
Level of reference: Data and algorithmic systems

In the context of digital solutions, the *controllability* sub-principle refers to the fact that humans always have ultimate control over data and algorithmic systems.

If *controllability* is not credibly guaranteed, the risk increases that unintended harms cannot be prevented by human agency. It may also become more difficult for organisations to obtain data in the future.

Organisations committed to the principle of *controllability* ensure that they know exactly what data is stored where and who has access to it. Organisations that use algorithmic systems do not use systems whose functionality is so complex that they cannot be changed by human intervention. For example, it should always be possible to stop algorithmic systems.

Since with the promotion of the sub-principle *controllability* paternalistic effects of data and algorithmic systems are reduced, it contributes to preserving the core principle of *non-maleficence*.

Accountability
Level of reference: Data and algorithmic systems

In the context of digital solutions, *accountability* means always ensuring that the actions of a person, a group of persons or an organisation can be clearly attributed to data and algorithmic systems.

If the principle of *accountability* is not taken into account, there is a risk that unstructured processes may arise. These unstructured processes invite the occurrence of misconduct or the misconduct remains undetected.

Organisations committed to the principle of *accountability* define short, medium and long-term strategies. The organisation itself and other parties involved take combined actions to ensure *accountability*. To ensure this, for example, an ethics board can be established to discuss critical cases and assess *accountability*.

Since promoting the sub-principle of *accountability* exposes potential organisational shortcomings, it contributes to maintaining the core principle of *non-maleficence*.

2.1.5 Transparency

Transparency is a core principle, to which the sub-principles *comprehensibility*, *interactivity* and *traceability* are assigned in the context of digital solutions.

As such, *transparency* refers to the extent to which access is granted to information about the underlying intentions or behaviour of an organisation. This is done either through the digital solutions themselves or through communication channels of the organisation.

A certain minimum level of *transparency* is necessary so that users are adequately informed about the consequences of using digital solutions and can make full use of them. If this minimum level is not maintained, a person cannot exercise the right to self-determination. As a consequence, there is an imminent risk of loss of trust in the product or service and/or the organisation and/or ultimately in digital solutions in general.

Organisations that commit themselves to the principle of *transparency* have established measures to maintain the sub-principles of *comprehensibility, interactivity* and *traceability*. In this way, users are informed about the results and possible consequences of digital solutions in an appropriate manner. They also give individuals the opportunity to interact with the organisation in various ways. Finally, organisations pay attention to where certain data comes from, how it is brought together and how it is processed.

Traceability
Level of reference: Data

In the context of digital solutions, the principle of *traceability* refers to the fact that attention is paid to where the data comes from, how it is brought together and how it is further processed. The aim of *traceability* is therefore to identify and trace the origin of data in a data chain.

If the *traceability* of data is not ensured, there is a risk that data of inferior quality will be processed and/or that linked data sets can no longer be clearly separated, possibly leading to outcomes of inferior quality with the potential to harm others.

Organisations that commit themselves to the principle of *traceability* undertake to use only data from qualified and verified sources and to trace any changes to the data. The latter can be ensured, for example, by means of comprehensive documentation.

Since the results of data processing can be defended or challenged on a sound basis by promoting the sub-principle of *traceability,* it contributes to preserving the core principle of *transparency*.

Interactivity
Level of reference: Data and algorithmic systems

In the context of digital solutions, the sub-principle of *interactivity* refers to interpersonal and human-machine-interaction. 'Interpersonal' means that there is direct communication between at least two people about data and algorithmic systems. Human-machine-interaction refers to the extent to which a person can influence or adapt a digital solution.

A lack of interaction increases the risk of critical issues being overlooked. There is also a risk that users are forced into a passive role in which their perspectives are not considered.

Organisations that commit themselves to the sub-principle of *interactivity* ensure sufficient communication in different ways. For example, they implement various simple and low-threshold communication channels (customer hotline, customer chat, e-mail, etc.). Another example are certain setting options within the digital solution.

Since users of digital solutions are given the opportunity to ask questions and also raise concerns by promoting the sub-principle of *interactivity*, it contributes to preserving the core principle of *transparency*.

Comprehensibility
Level of reference: Data and algorithmic systems

In the context of digital solutions, *comprehensibility* means that processes and effects are communicated in a clear and easy-to-understand way. Otherwise, there is an imminent risk that they will be not understood or misunderstood.

Organisations that commit themselves to the principle of *comprehensibility* actively promote credible communication. They take into account the different backgrounds of those who use, develop, operate or decide on the deployment of digital solutions. For example, comprehensive explanatory texts or videos can provide the user with sufficient information to handle the digital solutions responsibly. In this way, organisations additionally support an informed public debate.

Since promoting the sub-principle of *comprehensibility* makes users aware of the functionalities of digital solutions, it contributes to preserving the core principle of *transparency*.

2.2 Bibliography

Underlying Guidelines (Excerpt)

Accenture (2016): *Universal Principles of Data Ethics*. URL: https://www.accenture.com/_acnmedia/pdf-24/accenture-universal-principles-data-ethics.pdf Last access: 25.03.2020.

ADEL (2017): *Algorithm Data Ethics Label*. URL: http://www.adel-label.com/adel/la-demarche/ Last access: 25.03.2020.

Agency for Digital Italy (2018): *AI White Paper*. URL: https://ai-white-paper.readthedocs.io/en/latest/doc/raccomandazioni.html Last access: 25.03.2020.

AI4People (2018): *Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. URL: https://www.eismd.eu/wp-content/uploads/2019/11/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society_compressed.pdf Last access: 25.03.2020.

BertelsmannStiftung & iRights.Lab (2019): *Algo.Rules – Regeln für die Gestaltung algorithmischer Systeme*. URL: https://algorules.org/typo3conf/ext/rsmbstalgorules/Resources/Public/assets/pdf/Algo.Rules_DE.pdf Last access: 25.03.2020.

Bitkom e.V. (2015): *Leitlinien für den Big-Data-Einsatz*. URL: https://www.bitkom.org/sites/default/files/file/import/150901-Bitkom-Positionspapier-Big-Data-Leitlinien.pdf Last access: 25.03.2020.

Bitkom e.V. (2018) *Empfehlungen für den verantwortlichen Einsatz von KI und automatisierten Entscheidungen*. URL: https://www.bitkom.org/sites/default/files/file/import/180202-Empfehlungskatalog-online-2.pdf Last access: 25.03.2020.

BMVI Ethik-Kommission Automatisiertes und Vernetztes Fahren (2017): *Bericht*. URL: https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile Last access: 25.03.2020.

CIGREF & Syntec Numérique (2018): *Digital Ethics - A Guide for Professionals of the Digital Age*. URL: https://www.cigref.fr/wp/wp-content/uploads/2019/02/Cigref-Syntec-Digital-Ethics-Guide-for-Professionals-of-Digital-Age-2018-October-EN.pdf Last access: 25.03.2020.

DATAETHICS (2018): *Principles and Guidelines for Companies, Authorities & Organisations*. URL: https://dataethics.eu/wp-content/uploads/Dataethics-uk.pdf Last access: 25.03.2020.

Data for Good (2018): *Serment d'Hippocrate pour Data Scientist*. URL: https://hippocrate.tech/ Last access: 25.03.2020.

Deutsche Telekom AG (2018): *KI-Leitlinien*. URL: https://www.telekom.com/resource/blob/532444/87e1e54df08cce6f4483985bd25250b6/dl-180710-ki-leitlinien-data.pdf Last access: 25.03.2020.

Die Bundesregierung (2018): *Strategie Künstliche Intelligenz der Bundesregierung*. URL: https://www.bmbf.de/files/Nationale_KI-Strategie.pdf Last access: 25.03.2020.

DSK – Datenschutzkonferenz (2019): *Hambacher Erklärung zur Künstlichen Intelligenz*. URL: https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf Last access: 25.03.2020.

See "Development of Merck's Code of Digital Ethics – Process Documentation" for complete list.

Scientific Papers

AlgorithmWatch (2019): *AI Ethics Guidelines Global Inventory*. URL: https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/ Last access: 25.03.2020.

Datenethikkommission (DEK) (2019): *Gutachten der Datenethikkommission*. URL: https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=4 Last access: 25.03.2020.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena E. (2018): *AI4People's Ethical Framework for a good AI Society: Opportunities, Risks, Principles and Recommendations.* URL: https://doi.org/10.1007/s11023-018-9482-5 Last access: 25.03.2020.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020): *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*. URL: http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420 Last access: 25.03.2020.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2019): *Introducing the Principled Artificial Intelligence Project.* URL: https://clinic.cyber.harvard.edu/2019/06/07/introducing-the-principled-artificial-intelligence-project/ Last access: 25.03.2020.

Institute for Digital Transformation in Healthcare (idigiT) (2019): *Analysearbeit des Projektes: Erstellung eines Code of Digital Ethics für die Merck KGaA.*

Mittelstadt, B. (2019): *Principles alone cannot guarantee ethical AI.* URL: https://doi.org/10.1038/s42256-019-0114-4 Last access: 25.03.2020.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016): *The ethics of algorithms: Mapping the debate.* URL: https://doi.org/10.1177/2053951716679679 Last access: 25.03.2020.

Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2019): *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*. URL: https://doi.org/10.1007/s11948-019-00165-5 Last access: 25.03.2020.