The work of physicians at the interface of big data, artificial intelligence and human experience

Opinion of the Bioethics Commission

The work of physicians at the interface of big data, artificial intelligence and human experience

Opinion of the Bioethics Commission

Vienna, 2020

Imprint

Media owner, publisher and editor:
Secretariat of the Bioethics Commission, Ballhausplatz 2, 1010 Vienna
Authors: Bioethics Commission
Translation: MMag. Felicitas Hueber
Vienna, 2020. Date: 18 May 2020

Content

1 Preamble

Artificial intelligence is one of the key concepts of our time. Artificial intelligence can enhance some aspects of our lives in previously unimaginable ways, but it can also evoke images of a dystopian future, linked with a feeling of helplessness. Some regard it as the ultimate cure-all, while others see it as a great danger. So what does the term really mean today? It was coined over half a century ago, and was originally used purely to describe "thinking machines." Today we think about artificial intelligence principally in connection with high-performance computers.

Although the idea of artificial intelligence is not fundamentally new, rapid developments in computer technology, and particularly the increasing availability of large volumes of data – big data – have made access to it much easier and led to heightened interest in this topic area. There is scarcely a nation, hardly any national or international advisory body on (bio-)ethics that has neglected to devote a considerable amount of attention to digitalization and artificial intelligence, weighing up the potential benefits and dangers.

The Bioethics Commission of the Austrian Federal Chancellery has already addressed the ethical aspects of this subject in the past, both internally and at public events. As far back as 2009 the Commission published an opinion paper on "Assistive Technologies – Ethical Aspects of the Development and Use of Assistive Technologies with Regard to Older People."

In 2016 the Commission held a public meeting in the Federal Chancellery on the use of robots in the care sector under the title "Of Humans and Machines: Robots in Care." In this context national and international experts, together with members of the Commission, discussed ethical implications and social consequences of automation in healthcare. This meeting of the Commission was the beginning of an intensive process of engagement with this topic, which resulted in the 2018 publication of a unanimous opinion on "Robots in the Care of Older People." Although this concluded the special focus on this topic, it was evident that all members shared a wish for further discussion on big data, artificial intelligence, and medicine. Under the Austrian Presidency of the Council of the European Union in the second half of 2018, the 23rd National Ethics Councils (NEC) Forum – which brings together the chairs and leaders of national bioethics councils and commissions – was held in Vienna on September 19 and 20, with representatives from the European Group on Ethics in Science and New Technologies (EGE), the Council of Europe, UNESCO and the WHO. At this meeting members of the EGE also presented an opinion statement on the topic of Artificial Intelligence.

In the first half of 2019 the topic of "Digitalization & ethics" was explored further in a joint workshop involving the Bioethics Commission and Think Austria (a strategy unit of the Federal Chancellery), and various international experts, as part of the "*Geist & Gegenwart*" Whitsun Dialogue at Seggau Castle in Styria.

Today it is rare to find a scientific conference that does not include expert presentations on artificial intelligence. Increasing amounts of public and private funding are being invested in this area, with great expectations of a better life in an ageing society. Medicine in particular is one of the fields with high hopes of new technologies. With growing frequency, artificial intelligence and complex data analyses, which physicians themselves do not understand in detail, are becoming part of the standard procedures used to help make medical decisions.

So what are the particular ethical problems that may arise in this kind of situation in healthcare? Many people fear that problems may result if medical staff trust computerized assistance systems based on artificial intelligence when treating patients, rather than making their own assessment. On the other hand, where there is divergence between a judgment made by artificial intelligence and one based on a physician's clinical experience, the question arises as to which are the most appropriate decision-making procedures. Finally there is also the question of how to program medical assistance systems in a way that is ethically justifiable, when medical and economic factors can be taken into consideration, hidden discrimination by algorithms can in theory not be excluded, and decisions are constantly being made which could potentially be life-and-death decisions.

Effective medical care is particularly dependent on scientific advances. These advances, however, come not only from the innermost sanctum of medicine, but from interdisciplinary collaboration between the life sciences and the humanities. Two hundred years ago, the work of physicians was still very limited. It was not until the late 18th and early 19th century that modern medicine began to emerge. The medical world began to reflect upon itself, with a new "physicians' perspective" – a different way of seeing things, based on a new structure between words and things, and facilitated by the medical work being done in the increasingly numerous hospitals, where patients were treated by physicians. Here – in contrast to the old-style medicine, which took place in isolation beside the sickbed – patients could be observed and their clinical histories compared. The new practice of observing symptoms, systematic methods for physical examination, and in some cases the subsequent postmortem resulted in new perspectives on disease, as well as on patients themselves. This was also where the professional skills of physicians began to be standardized. In Vienna, the achievements of the great physicians in the First and Second Vienna Schools of Medicine were a testimony to these developments.

These processes have continued. In the past just as today, physicians have always been required to constantly keep abreast of new understanding and methods, to continue to develop their skills, not to stand still, but instead to apply new scientific findings in their day-to-day work, while also maintaining a sensible respect for tried and tested methods, and continuing to use these. The difficulty, however, is that today our accumulated medical knowledge is so extensive that it is no longer possible for an individual to access the entirety of the existing and relevant knowledge at the moment of clinical decision-making.

Discussions within the Bioethics Commission are not focused purely on scientific research findings, but primarily on the work of physicians and the associated responsibilities. Ethical principles such as respect for patients' autonomy, the "do-no-harm" principle, care, and the priority of patient well-being, as well as the principles of equality are all important topics. The impact of artificial intelligence is no exception.

2 Introduction

Digital technologies have not only revolutionized medical research (e.g. by making it possible to analyze big data), enabled completely new treatment methods (e.g. personalized medicine) to be developed, and reformed the administration of the healthcare system (e.g. electronic patient records), but also have an effect on medical and nursing care in direct contact with patients.

"Digital technologies" include a broad spectrum of tools and services that use electronic information and communications technologies (ICT). In the healthcare context, relevant applications are "electronic health" (e-health) or "digital health," which include the following areas: telemedicine, teleconsultation, telecare (ambient assisted living, or AAL), telediagnosis, virtual laboratories, telemonitoring, electronic medication, digital documentation / clinical history, health portals, personal health management, social health networks, e-learning, electronic billing, e-payments. These ICT applications have to some extent become part of routine practice, and are being further developed with more innovative use of big data, machine learning and artificial intelligence (AI). Algorithms are a key tool of digital technologies, and form the basis of all computer programs; "algorithmic decision-making systems" describe the entire process from data collection to the eventual findings.

The most important manifestations include the following:

- Support for **diagnostic evaluation** with modern sensor technology, which, thanks to mobile devices, networked environments, and automated data processing, allows the diagnostic process to be extended into patients' homes; as a result procedures that were previously undertaken for a specific reason are evolving into continuous monitoring.
- Support for **diagnosis and treatment decisions** through the use of advanced algorithmic systems (rule-based programming and / or machine learning), which includes a range of functions – from patient-oriented triage systems to medical decision-making assistance software and algorithmically determined decisions for immediate application.

- Support for care-related or **medical procedures** using robotics (see the Bioethics Commission's opinion paper on "Robots in the Care of Older People").

Embedding digital technologies into situations involving direct contact between physician and patient **changes the relationships in both directions**. Certainly, digital technologies may turn patients into more pro-active and self-reliant partners in the process of preventive care, diagnosis and treatment. However, physicians increasingly assume the role of intermediary between the technology and the patient. Both these tendencies have direct consequences for medical training and continuing education, just as they do for the equipment needed in medical institutions and the structure of the healthcare system, as well as the relevant legal frameworks. The unique nature of the physician-patient relationship, however, also makes higher demands with respect to following **ethical principles in the design** of digital technologies ("ethics by design"). The observations on the following pages about the work of physicians apply equally to the nursing and care sectors.

It is also evident that modern medicine has increasingly become a form of data-intensive biomedicine, due to a series of technological tools that have been developed with the sequencing of the human genome and to the establishment of new branches of science such as bioinformatics and data science. These enable to more clearly capture and to a certain extent experimentally record the complexity of living systems, so that today we have a different understanding of complexity in the context of health and sickness from that of the late 20th century.

The possibility of finding "meaning" in large volumes of data through machine processes, which are increasingly also "learning" and evolutionary, i. e. "intelligent" processes, brings visions of individualized (personalized) medical care, and of a patient-centered, highly efficient healthcare system closer to reality than ever before, to the extent that scientific knowledge about their application can be made accessible in a way that is helpful to humans. Clearly this is an enormous challenge which is highly controversial and gives rise to a range of ethical issues.

3 Physicians and their work, big data and artificial intelligence

This opinion paper focuses on the implications of big data, artificial intelligence, and machine learning for the work of physicians, along with their use in diagnosis and treatment planning. For this reason it explores some fundamental reflections on the practices and responsibilities of physicians (including legal perspectives), on psychosocial aspects of the relationship between the physician and the patient, and on economically motivated efforts to make healthcare more efficient and effective.

The aim of this paper is to give physicians and decision makers in the healthcare sector

1. some insights into this topic area, and to describe the mathematical, technical, medical and legal aspects;
2. in particular to explain the ethical issues associated with the use of big data, artificial intelligence and machine learning; and
3. to provide recommendations for action on various levels.

Although other healthcare professions are also affected by this topic area, this paper concentrates specifically on the work of physicians.

3.1 The work of a physician

In comparison to the tried and tested forms of evidence, the use of big data, artificial intelligence and machine learning presents some new challenges for physicians. These concern the collection and recording of data, and the interpretation of computer-generated diagnoses and suggestions for treatment. In clinically-oriented research it also concerns the sharing and exchange of data.

3.2 Psychosocial aspects of the relationship between physician and patient

The physician's work begins in the waiting room, or perhaps even on the telephone when the patient calls to make an appointment; being kept waiting for a long time can already affect the patient's attitude towards the physician. The work of the physician

includes greeting and making eye contact, the tone of voice and handshake, and also involves listening in a particular way: with empathy, receptively, communicating warmth and security. So the physician's work begins at a much earlier stage than we might think.

Good communication between physicians and patients is a key element of clinical practice. It is a vital part of medicine and a central element of care provision (Epstein 2017; Street et al. 2009; Balint and Shelton 2002). It is a combination of good interpersonal relationships, effective information exchange and enables the patient to make a decision (Borck 2016; Epstein 2017; Müller-Mielitz 2017; Hehner et al. 2018).

3.2.1 The medical consultation

Physicians are generally highly trained but they may lack some communicative and psychosocial skills. A frequently quoted study by Beckmann and Frankel (1984) found that doctors interrupt patients after 18 seconds of conversation on average. There are also recent studies confirming that physicians have difficulty communicating with patients. For example Street and Haidet (2011) show that doctors frequently do not acquire sufficient understanding of the attitudes and values of their patients, if the latter are not pro-active in communication.

1. There are many obstacles to effective communication. From the physician's perspective this is not just a question of the heavy workload. Doctors frequently work in a context of anxiety: they worry about unrealistic expectations, and about patients' fears, and that they may not be able to help. In this atmosphere it is often difficult to be open about the unknown and to face the unexpected.
2. Having to convey bad news and withstand (strong) emotions is often depressing and energy-sapping. On the other hand, it is empowering and satisfying for physicians to be able to share decisions on the basis of informed consent or in the case of particularly complicated medical information. Patients can be passive, reluctant to give their own opinion or begin discussions about aspects of their treatment. Furthermore, doctors often find it difficult to create the right atmosphere for participative communication (Ubel et al. 2017). Reasons for this include "failing to communicate clearly to patients about decision-relevant information, overwhelming patients with irrelevant information, overlooking when patients' emotions made it hard to engage in choices, and making recommendations before discussing patients' goals" (Ubel et al. 2017, 31). Effective communication helps patients cope with their illness and the associated treatment. So it is important to empower them to make decisions.
3. Physicians tend to overestimate their own knowledge and the ability of patients to understand the facts (Tversky / Kahneman 1974). Consequently, doctors' general familiarity with medical terminology may mislead them into assuming that their patients understand their explanations (Gundling et al. 2019). This is compounded by a tendency for physicians to overestimate the benefit of what they are doing. This is described as the "therapeutic illusion." It is quite common for the work of

physicians to demonstrate what Epstein (2017) defines as a confirmation bias, namely the belief – which is difficult to correct – in the beneficial effect of their own orders and prescriptions (Epstein / Pro-Publica 2017). Often the diagnosis that comes into the physician's mind first is the one selected.

4. We know that the act of talking itself can have a therapeutic effect, by reducing anxiety as well as providing comfort and encouragement. However, the effect is often not directly observable, but is evident instead from patients' understanding, confidence, and agreement (Street et al. 2009). This leads to the realization that the doctor herself or himself is a "drug" although it is often not certain how frequently it should be administered, what the appropriate dosage is and what the side-effects may be (see below).

3.2.2 The dynamics of the physician-patient relationship

In his classic text "The Doctor, His Patient and the Illness," Balint (1964) identified six key factors that affect the development of the physician-patient relationship:

- The "basic fault", a concept originally from psychoanalysis, is a reaction to experiences in early childhood. It determines how an individual reacts to stress and affects the way people react to their environment. Doctors who are self-aware are better able to understand their patients. (The "basic fault" exists in every individual.)
- The "apostolic function" of doctors refers to their paternalism. Physicians have a tendency, as soon as patients have described their symptoms, to develop a fixed idea about what is right for them, what they should expect, what they will have to put up with, and how they should behave.
- The mutual "investment fund": the idea is that physicians and patients "train" each other. Balint calls this a "mutual investment fund". If this is successful, it results in a positive, trusting relationship that is beneficial for both parties.
- The role of physicians as a therapeutic agent: the "doctor as a drug" is in use constantly, but we know very little about its pharmacological effects and side-effects, or its dosage. However today we do know that the placebo effect plays a significant role.
- The "deeper diagnosis": the mutual "investment fund" provides the framework within which physicians are able to learn more about patients and so formulate a "deeper, more holistic diagnosis."
- The "collusion of anonymity": patients with multiple undefined symptoms are often sent to various different specialists for investigation. If none of these specialists is able to arrive at a diagnosis, the outcome is often an endless loop of "all clear" letters, and another round of visits to doctors when further symptoms emerge. Balint (1964) speaks of a "collusion of anonymity," a covert agreement that none of the physicians involved feels responsible for these patients in the appropriate sense.

So we can see that the relationship between physician and patient is easily hampered, and it is no surprise to find that these relationships are frequently tense, uncomfortable, and unsatisfactory.

3.2.3 The significance of transference, countertransference, and different attachment styles

Research studies have shown that a significant percentage of physically ill people suffer from severe anxiety and depression: here good mechanisms for transference and counter-transference – successful attachment – are particularly important. This presupposes, however, that patients have the right – which for good reasons is documented – to self-determination, which includes the right not to know, and the right to be irrational. There may be good reasons for refusing treatment or for stopping it, such that these kinds of decisions cannot necessarily be attributed to the patients' attachment style.

The relationship of the patient to the physician is seen as a **transference** of earlier relationships to important caregivers. If patients do not accept the treatment proposed, or want to be discharged early, this can in some circumstances be traced back to an unsatisfactory atmosphere in their dealings with care providers. The possibility must then be considered that early childhood experiences may have affected their relationship with key caregivers.

Physicians' feelings and reactions towards their patients are described as **countertransference**. This too is influenced by early relationships and can have both positive and negative effects on the relationship between physician and patient. Physicians therefore need a certain amount of introspection, to recognize and understand their own feelings and internal obstacles. From the medical perspective, they need to observe and analyze the way their patients' behave, but also to examine their own feelings (including challenging responses such as anger, helplessness or intense emotional impact).

At this point it is helpful to take a look at some research findings on attachment. Based on the attachment theory developed by John Bowlby and Mary Ainsworth (2003), Bartholomew and Horowitz (1991) defined four prototypes for adult attachment behavior:

- The **secure attachment style** is observed in people who had warm, emphatic caregivers who successfully communicated a fundamental feeling of security. This type of patient is receptive, open, and cooperative in contact situations.
- The **preoccupied attachment style** is characterized by dependent behavior in the patient, strongly expressed feelings, and an intense need for emotional intimacy. In the medical setting, a chronic sense of vulnerability or danger makes these patients pay very close attention to their physical symptoms. This can make it difficult for physicians to ascertain hard facts from the clinical history.
- The **dismissing attachment style**: in childhood these people experienced too little sense of being cared for and protected. In situations that demand a certain

amount of intimacy and dependency, these patients tend to react in a reserved, defensive and frequently distrustful way, because they feel helpless.

- The **fearful attachment style**: in childhood these individuals' needs were rebuffed by their caregivers, or they experienced unkindness. In case of illness this often results in an insecure, ambivalent response to the doctor. These patients make it clear that they are suffering, but refuse medical treatment. They need trust to be built up very carefully, and need a calm and gentle approach from the attending physician.

Sometimes doctors need to be organized, scientific, and decisive. Then sometimes they need to be more aware of their own reaction to the patient, for example to recognize something unexpected, a tension in the air, a lack of knowledge, conflicts related to the patient's current problems. Frequent, short contact times combined with a sense of long-term commitment help these patients to allow the physician into their confidence at their own rhythm, depth and pace, and this also allows the physician to define the appropriate boundaries for his / her involvement. The "ideal" consultation results in a special moment of mutual understanding.

To summarize, both at the emotional level and with regard to modern medical ethics, it is important to develop a trusting and productive human relationship / attachment.

3.3 Legal responsibility for the work of physicians

The aim of any physician's work is to contribute to the wellbeing of sick and healthy patients, in accordance with the current status of medical science and experience (see Austrian Physicians Act – ÄrzteG, Section 49).

If **harm** is caused to a patient by any action contrary to good professional practice, this may result in a **civil and possibly also criminal liability**, both for the **individual** and also in certain circumstances for the **legal entity** concerned. In any such case, a judgment is always made in retrospect (**ex post**) by the courts – with reference to expert opinion where necessary – as to whether the work of the physician(s) was **professionally correct at the time (ex ante)**, and whether all necessary precautions were undertaken to minimize risk. There is no legal assurance in advance to the effect that a specific course of action is also legally correct in all circumstances. This applies equally to technology-assisted surgical methods or computer-assisted diagnosis, as well as to traditional medical investigations and treatment. With this in mind, the following sections aim to explore first of all the general principles of liability in the work of physicians, followed by the liability situation in association with (the use or omission of) new scientific methods.

In the context of "big data," the clinical physician is also a **nodal point, gathering, receiving and sharing data**. These data are subsequently stored, with the result – due to the sensitivity of such data – that a significant legal responsibility effectively falls to

the physician, with regard to how these data are handled. Furthermore, the physician **requests data** of various kinds, which further intensifies the responsibility already outlined for protecting and processing such data. The regulations on the legal handling of health data are by no means new, and the responsibilities with regard to data protection have already intensified significantly in recent years, particularly in connection with the General Data Protection Regulation (GDPR). At the same time, a broad awareness must be generated about the enormous potential of professionally gathered, curated and networked data in healthcare.

3.3.1 General principles concerning responsibility for a physician's work

The **legal responsibility of an individual** is based – in both civil and criminal law – on social disapproval of a particular behavior and is generally expressed in terms of culpability. In addition there is also concept of liability without culpability, for example product liability or absolute liability. **Legal entities** are liable under civil law according to the same principles, whereby they are liable for the culpability of their so-called organs and authorized representatives (such as the culpability of selection or monitoring), in the context of a contract (e.g. a treatment contract) this also includes the culpability of every assistant. In criminal law, the responsibility of legal entities (corporate responsibility) is essentially dependent on the idea that the organization was a criminal accessory to culpable conduct by a specific person (individual liability). Furthermore, a kind of organizational culpability must be included here, when either the act must have been committed for the benefit of the organization, or certain obligations of the organization must have been contravened by the act (see Austrian Corporate Criminal Liability Act – VbVG, Section 3(1)). Whether or not the conduct was legally improper should initially be judged on whether it is forbidden by law or regulatory statute (**breach of legal provision**). If there is no such formal legal guidance, then reference is made to subsidiary professional rules on duty of care (**breach of good professional practice**). In this context guidelines by professional associations or other "standards agreed by qualified experts" can have particular significance (for general observations on the topic of guidelines and criminal liability, see Birklbauer 2019). If there are no such standards for a particular field of activity, a third possible criterion for conduct that breaches the duty of care is the lack of risk avoidance action which an average or more careful other individual working in the same field as the perpetrator would have taken (**conduct of an average or careful other individual**).

Section 49(1) of the Austrian Physicians Act (ÄrzteG) also mentions reference to professional standards and guidelines for the medical profession as a standard of care. The act specifies that a physician is required to respect "existing provisions and professional quality standards" and "to safeguard the wellbeing of the ill and to protect the healthy." A breach of the duty of care exists if the specific physician, for example in establishing the medical indication for a specific treatment, or during the course of a specific treatment, has deviated from the general standard as accepted by (interna-

tional) professional colleagues. The current valid standard for a professional group is thus effectively the criterion for legal responsibility. By the same token, professional **guidelines** also include a **presumption of due care**. The physician who treats patients in accordance with the guidelines is acting with due care providing he / she has no evidence that the principles recommended in the guidelines no longer correspond to the definition of good professional practice. To this extent professional associations also bear a certain amount of responsibility, namely to ensure their guidelines match the current status of specialist expertise. This is a considerable challenge, particularly in connection with research methods where due to the complex data structure it is difficult to ascertain the quality of the data.

In this context, medical computer programs, for example, which are categorized as **medical devices**, are accorded a certain presumption of trust, providing this categorization is based on a legally established conformity assessment procedure (see 3.3.2).

3.3.2 Intelligent software as a medical device

The use of "artificial intelligence" by physicians does not occur in a legal vacuum or on a purely experimental basis. On the contrary: software, including any form of artificial intelligence, is regarded as a **medical device** if it is specifically intended by the manufacturer to be used for one or more of the healthcare purposes included in the definition of medical devices. However, general-purpose software (e.g. office software) – even if it is being used in healthcare organizations – is in principle no more a medical device than lifestyle products (e.g. fitness apps) are.

Medical devices can only be introduced to the market and used by the physician if they have been through the prescribed **conformity assessment procedure** and they fulfill the "essential requirements." Until recently the essential requirements were defined in the relevant Annex I to Directives 90/385/EEC (active implantable medical devices), 98/79/EC (in vitro diagnostic medical devices) and 93/42/EEC (other medical devices). Effective May 26, 2020, these Directives (and the relevant national laws implementing them) have been replaced by Regulation (EU) 2017/745 (Medical Device Regulation, or MDR), and by Regulation (EU) 2017/746.

The appropriate conformity assessment procedure, and to what extent this needs to involve an independent testing and certification authority (the "appointed authority"), depends on the potential risk of the devices. Thus the MDR (in line with Directive 93/42/EEC) stipulates that **devices should be categorized into four classes** (I, IIa, IIb, III) according to the criteria specified in Annex VIII. Conformity assessment is carried out in according to Annex IX of the MDR, based on a quality management system and evaluation of the technical documentation.

The MDR also includes comprehensive provisions concerning software. Software designed to control a medical device or to influence its use is classified in the same way as the device itself. If the software is independent of other devices, it is classified separately. Software intended to **monitor physiological processes** belongs to Class IIa, unless it is intended for monitoring vital physiological parameters, where a change could

lead to an immediate risk for the patient (in which case it is Class IIb). Software designed to provide information to aid **decision-making for diagnostic or therapeutic purposes** belongs in principle to Class IIa. If such a decision could result in serious detriment to a person's health, or in the need for surgical intervention, the software is categorized as Class IIb. If the decision concerned could have effects that might result in the death or irreversible detriment to a person's health, it is categorized as Class III. All **other software** is categorized as Class I.

3.3.3 (Omission of) due consideration of new research findings

With regard to the connection between deviations from a defined standard and liability for negligence, reference can be made to the principles formulated for **off-label use** of treatment methods or the medications (see Kopetzki 2008; Mayrhofer 2014). The absence of market authorization for a specific treatment does not necessarily mean that its application would be a breach of good professional practice, and thus not permissible under Section 49(1) of the Austrian Physicians Act (ÄrzteG). On the contrary, the use of such a medication may even be advisable, as long as this application of the medication is indicated according to the relevant current status of medical and pharmaceutical understanding and there is a good prospect of success. However, the physician making "off-label" use of a medication has a heightened obligation to justify its use, because he / she cannot base this "simply" on the authorization, and must cite other sources instead. Here in addition to scientific publications the guidelines published by professional associations play an important role, which is vitally supported by the wide availability of real world data.

If a piece of software has not been evaluated by the appropriate conformity assessment procedure, but the physician still wishes to base a diagnostic and / or therapeutic decision on algorithmic calculations, caution is advisable. The use of new treatment methods or findings may in a given case be a basis for (civil or criminal) responsibility, if this can be seen as negligence. For example, if new, as yet **insufficiently researched methods are used** and the patient is harmed as a result, a legal responsibility arises if such treatments are considered experimental, without an adequate evidence base. Liability might not apply if the patient has consented to the possible outcomes with adequate knowledge of the full risk; where a criminal issue is concerned the justifying consent is limited by the corrective of the violation of moral principles (see Austrian Penal Code – StGB, Section 90).

On the other hand, legal liability may also be established if a **new, promising treatment method is withheld** from the patient because the physician has not yet familiarized himself with it and does not wish to do so. Liability would then only be inapplicable if the relevant treatment method is not accessible, or cannot be arranged, since no one can be obliged to do something that is not actually available to him / her.

In order to provide patients with comprehensive and thorough information to form the basis on which they can accept or decline possible treatments, it is essential that **physicians** themselves undertake **adequate continuing education**, so that they

can suggest and explain these sorts of options. When using algorithmic support for diagnoses and treatments it is particularly important for the physician to have sufficient understanding of how the relevant data and the underlying algorithm etc. are derived, and that he / she can trust the quality of published studies and the methods used. In this light, the review process in advance of a publication has particular significance. Here too the standard of care is generally based on the **physician with average knowledge and skills** who has also fulfilled his / her obligation for regular continuing education. For all legal considerations of liability however, it must be kept in mind that the legal scope of liability should not be stretched too far. Negligence means the failure to uphold "due" care, rather than the neglect of "possible" care.

3.3.4 Responsibility vacuum with the use of artificial intelligence?

In connection with civil liability when artificial intelligence is used, concerns are often expressed about a responsibility vacuum. In the context of fault-based liability, the physician (or operator of the healthcare institution where the physician is employed) is liable only for his / her own culpability and the culpability of people deployed in fulfillment of the duties of their employment (Austrian Civil Code – ABGB, Section 1313a). If instead of people it is machines that are deployed, however, liability only applies to the **individual culpability of the physician**, which may consist for example of negligence in the selection, use, updating or monitoring of the relevant software. If the physician has used recognized and certified software, applied it in accordance with the guidelines and might not necessarily know about any faults in the software, then the physician (or operator of the healthcare institution) is **not liable for faults in the software as such**. Consequently a growing body of opinion claims that Section 1313a of the Austrian Civil Code (ABGB) should apply equally to faults in intelligent software (see Expert Group on Liability and New Technologies, Report on Liability for Artificial Intelligence and other Emerging Digital Technologies – New Technologies Formation 2019, Recommendations [18] and [19]; Opinion of the German Federal Government's Data Ethics Commission 2019, Recommendation no. 74).

The liability of the software manufacturer is also far from certain. There is some dispute as to whether strict liability as defined by the **Product Liability Act** (PHG – Produkthaftungsgesetz) applies at all to software alone, or if it is only valid for physical objects in which software in embedded. Furthermore, liability only applies to faults that were inherent in the product at the time it was released on the market, while faults that arise from later updates or from the lack of security updates are not covered. A manufacturer can also be exempted from liability if he / she is able to show that the fault was not detectable in terms of the current status of knowledge and technology at the time when the product was introduced to the market. **Fault-based liability of the manufacturer** as defined by the Austrian Civil Code (ABGB) is similarly fraught with difficulties, especially as proof of negligence and / or causality is often not possible.

For this reason there is growing demand for the introduction of **absolute liability** for the use of this kind of system. An additional idea came into play briefly – suggested

by the European Parliament (resolution of February 16, 2017, on civil legislation in the area of robotics, Recommendation no. 59 et seq.) – which was to attribute the status of legal personality to robots and "autonomous" software; this idea has by now quite rightly been rejected, however, by majority view (see for example the 2018 Open Letter to the European Commission, Artificial Intelligence And Robotics; Opinion of the German government's Data Ethics Commission 2019, Recommendation no. 73). It seems now that the European Parliament is more inclined to support a two-strand concept of liability, encompassing (amended) product liability on the one hand and a new (yet to be introduced) AI liability on the other (draft dated April 27, 2020, for a Report by the Committee on Legal Affairs, 2020/2014(INL)).

3.3.5 Data protection responsibility

With regard to the handling of patients' data, the physician is subject to the provisions of the **General Data Protection Regulation** (GDPR), supplemented by the Austrian Data Protection Act (DSG). Also of significance for physicians is the **Austrian Health Telematics Act** (GTelG) 2012. In the field of medical research the legal position is also defined by a series of **additional regulations**, particularly the Austrian Research Organization Act (FOG).

The **Electronic Health Record** system ELGA enables electronic networking of patients' ELGA health data recorded in various areas of the healthcare sector. ELGA creates comprehensive links between inpatient institutions such as hospitals, local physicians, pharmacists and care institutions. The Austrian Health Telematics Act (GTelG) 2012 provides clear regulations about who is permitted to access ELGA health data: apart from the patient himself / herself, it is exclusively those physicians and ELGA health service providers who are actually treating or advising the patient concerned. The data cannot be accessed for example by public authorities, company physicians, or any physicians who have been specifically excluded from access by the patients themselves. The first data available through ELGA are medical and care notes on discharge from public hospitals, laboratory results, radiology results, and medication data, i. e. prescribed medications must be recorded in the e-database of medications. Pharmacists are also obliged, from a specified point in time, to record the dispensation of prescription-only medications and drugs that have the potential to interact with other medications. As a result of the Austrian Health Telematics Act (GTelG) 2012, there are numerous other requirements concerning the collection, storage and transmission of data that physicians need to consider.

Most of the data collected and processed by physicians is **health data** and therefore falls into the "special categories of data" as defined by Article 9 of the GDPR, which are subject to stricter requirements for processing. Any processing of such data – even just viewing on a screen – requires **legal justification**. This may be the explicit consent of the person affected, necessity for fulfillment of the treatment contract, or necessity for fulfillment of a legal obligation (such as for example according to the Austrian Health Telematics Act (GTelG) 2012 or the Austrian Physicians Act (ÄrzteG)).

In practice, these data are largely processed for fulfillment of the treatment contract or the statutory duty of documentation, such that explicit consent from the patients is generally not required. However, detailed **information** must be provided to patients, including about how their personal data have been processed, and for what purposes, and about patients' **rights** with regard to their personal data. This may for instance also include the right to deletion.

In any medical practice the responsibility for legally correct data processing lies with the relevant (independent) physician, or in a healthcare institution with the operator and its agents. Any breach of data protection regulations may result in an administrative criminal liability and civil liability for the responsible person. As is well known, substantial fines are possible under the GDPR.

There is often a **potential conflict** between the pursuit of the best possible quality of medical care with the help of modern digital technologies, and the pursuit of data privacy. They are particularly the technologies that are widely used on the domestic level (e. g. data collection for health purposes by means of smart watches, smartphones or smart home appliances) which are in many cases inherently associated with enormous potential for intrusion. While there are many instances where the protection of life and health is assessed as having priority over digital self-determination, this cannot be regarded as absolute priority. Adequate data protection is not only required by law, but is also an important factor in creating trust and acceptance amongst patients in connection with data processing in healthcare, and so also adds important leverage for further digitalization in and of healthcare.

3.4 Economic factors and motivations

Digital technologies in the healthcare sector are also seen as system-changing from an economic perspective. They influence the configuration of the care process, risk profiling, preventive care, diagnostics, therapeutic approaches, public health, nursing and care, as well as research (Jannes et al. 2018, 15 et seqq.).

The economic motivator is to make the healthcare sector more effective and efficient through the use of digital technologies, increasing the overall cost-effectiveness of healthcare provision (Müller-Mielitz 2017), and delivering an appropriate return on investment:

- A potential increase in effectiveness is anticipated from the way digital technologies can help to enhance our understanding of the causes of disease, associated factors and effective treatment approaches. The aim is to reduce inappropriate and costly treatments.
- It is hoped that increased efficiency through the use of digital technologies will result in a more positive input-output ratio (particularly by reducing expenditure by tailoring services better to the needs of individual patients).

- Returns on investment resulting from potential increases in effectiveness and efficiency benefit different stakeholders to a different extent.

These prospects are generating a high level of investment confidence (Forbes Insights Team 2019; Taylor 2015; Hipp et al. 2018) in digital infrastructure and other digital technologies. Pharmaceutical companies for instance are increasingly investing in apps for information and training on dealing with medications (Taylor 2015, 10). New stakeholders are also emerging in the healthcare sector. Technology companies such as Google, Apple and Microsoft are now becoming substantial innovators in the healthcare sector. Linked to these are numerous start-ups, app developers, and small technology firms that are appearing in the healthcare market with what are sometimes disruptive innovations (Kaltenbach et al. 2016; Hipp et al. 2017).

In the German healthcare sector the potential for increased cost-effectiveness is estimated to be worth 12 % of total expenditure overall, if existing digital technologies were fully exploited (Hehner / Biesdorf et al. 2018), of which 47% would be attributable to inpatient care (Hehner / Liese et al. 2018). Corresponding data on the situation in Austria is not available at the present time. The following technologies are regarded as relevant for this potential improvement in cost-effectiveness (in order of impact; Hehner / Biesdorf et al. 2018; Hehner / Liese et al. 2018):

- electronic data transmission instead of on paper: standardized electronic patient records, electronic prescriptions, communication of health care professionals within hospitals, electronic payments;
- online interaction: teleconsultation, remote monitoring of chronically ill patients, electronic (algorithm-based) triage at the interface between inpatient and outpatient care;
- work processes: digital networking for nursing staff (access to patient details across intra- / extramural boundaries, e. g. in-home care), barcode-based administration of medications, RFID tracking, monitoring of vital parameters, robots for hospital logistics, automation of simple processes, electronic payments;
- decision-making support: performance dashboards (i. e. digital information systems that provide internal information about care procedures by clinical teams, and their outcome, and so help to identify the potential for improvement), control of patient flows, electronic (algorithm-based) triage, support for clinical decision-making (i. e. for treatment decisions), decision-making support in connection with genetic tests;
- self-care for patients: tools for the management of chronic illnesses, medical chatbots, tools for disease prevention, patient-supported networks, digital diagnosis tools, virtual reality applications (e. g. for pain management);
- patient self-service: making appointments electronically.

Most of the potential for increased cost-effectiveness is currently regarded as being in the process management of clinical, administrative and logistics areas (Forbes Insights Team 2019). Digital technologies are used in these areas with the aim of reducing transaction costs. Direct provision of healthcare (preventive care, diagnostics, therapy, rehabilitation) plays a comparatively minor role at present. However, this is expected to change in the future (Price 2019; Hipp et al. 2018, 8 et seqq.; Hipp / Schlude et al. 2017, 17 et seqq.; Prainsack 2019, 11–16).

The economic motivation behind the use of big data and machine learning for diagnosis and treatment planning also brings some of the ethical considerations into focus, particularly with respect to some emerging tensions:

- concentration on the core business of clinical medicine: reducing transaction costs through the use of digital technologies frees up resources for the core business of medicine;
- equal opportunity for access: increased or reduced inequality in healthcare through the use of digital technologies ("digital divide"; Fischer 2017, 147);
- increased efficiency of structural healthcare provision: support for healthcare research through the use of big data analyses (Jannes et al. 2018, 26);
- efficiency at the expense of effectiveness: failure to differentiate between correlation and causality; encouraging automatism (Jannes et al. 2018, 25);
- responsibility for and during use: obligation for service providers to engage with digital technologies as part of their professional standard of care (Gründinger et al. 2019,12); increased transaction costs and resulting delays in development and use of digital technologies due to inadequate allocation and distribution of responsibility (Jannes et al. 2018, 29);
- risk stratification in the health insurance sector: implications of data processing with regard to pre-existing health conditions and patients' adherence to treatment (Gründinger et al. 2019, 17);
- competing algorithms: tendency towards monopolization in data-driven services; formulation of data-sharing obligations (Gründinger et al. 2019, 41).

3.5 Big data: Sources and quality

3.5.1 What does big data mean?

Critics argue that the current focus on "big data" gives the misleading impression that this term refers to completely new approaches. The use of the term "big data" may obscure the fact that the work of physicians has, for several decades – at least since the introduction of evidence-based medicine – been based on the systematic evaluation of data, either directly (by physicians consulting resources such as systematic reviews) or indirectly (via clinical guidelines). So this criticism should be taken seriously. At the same time, there is no question that the rapid development of diagnostic and therapeutic

technologies and tools, particularly since the turn of the millennium, and the "digital revolution" in general, have resulted in significant changes to established practices, standards, and options.

There is no universally accepted definition of "big data." The following definitions, from various disciplines and contexts, are some of the most established:

- Big data as an **incremental phenomenon**: "big data" signifies datasets that are larger than those for which earlier databases and software systems were designed. They are larger in terms of volume, velocity (i. e. the speed with which they are generated and move through systems) and their variety (e. g. Laney 2011).
- Big data as a **computational problem**: closely linked to the previous definition, this one refers to datasets that are so large that they exceed existing capacities for collection, storage and analysis (McKinsey 2011). By this definition, "big data" means all datasets for which storage and processing cannot be handled by the means currently available.
- Big data as a **techno-social phenomenon**: according to this definition, the term "big data" describes not just datasets and the associated technologies, but also the recording of data about more and more aspects of our bodies and our personal and social lives ("datafication").
- Big data as a **methodological approach**: by this definition "big data" approaches are those that look for correlations in large datasets, without a specific hypothesis (e. g. "what correlations can be found between genetic markers and a phenotype in the data of 300,000 patients?" (Antes 2016)). This is in contrast to traditional methods that use a working hypothesis or a specific research question when analyzing data. If correlations are found, further research can be undertaken to explore whether these are spurious correlations or linked by a causal connection; i. e. the working hypothesis is generated after the data is analyzed, not before.
- Big data denotes **all data that constitute a phenomenon.** This is the most "radical" definition of big data, and the one that is least established, relatively speaking. According to this definition, the term "big data" would only be used if for example a person's entire body was captured by imaging technology, and not just the organ where a tumor is located that needs to be treated. Similarly, proponents of this definition would only talk about big data if a person's entire genome was sequenced, and not just the gene in which there is a suspected mutation.
- Big data **as an asset**: the multinational IT company Gartner defines big data as "high volume, high velocity, and / or high variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation" (Gartner 2019). By characterizing big data as an asset, the company treats data as something that can be seen as property and that can be sold (see Birch 2017, for a critical view).

- A particularly relevant definition in the **Austrian context** is that of the Research Organization Act (FOG): here the concept of big data is described as "the processing of large volumes of data that are largely or completely unstructured" (Austrian Research Organization Act – FOG, Section 2b(3)). The notes to the Act explain that this definition was based on that of the National Institute of Standards and Technology, U.S. Department of Commerce (NIST) from September 2015 (incorrectly dated in the notes as September 2017). There is now a revised version of the NIST definition, dated June 2018.

Building upon the work of researchers such as Gernot Rieder and Judith Simon (2016), the Austrian Bioethics Commission considers any interpretation of "big data" which refers only to the technical aspects of data-intensive practices in science and medicine, to be unhelpful. In the words of Rieder and Simon:

> "it seems more productive to think of it as the terminologically contingent manifestation of a complex socio-technical phenomenon that rests on an interplay of technological, scientific, and cultural factors. While the technological dimension alludes to advances not only in hardware, software, but also infrastructure and the scientific dimension comprises both mining techniques and analytical skills, the cultural dimension refers to (a) the pervasive use of ICTs in contemporary society and (b) the growing significance and authority of quantified information in many areas of everyday life." (Rieder / Simon 2016, 2).

In this spirit we understand big data as a socio-technical practice that should be seen in connection with the political, economic and social factors that enable it.

3.5.2 Where do the data that are used for big data analyses in medicine come from?

Due to the increasing "datafication" of our society, today's researchers have access to data on diverse aspects of our lives that were not recorded in this form in the past. Previously the only option for researchers who wanted to know about a person's nutrition, how much they exercised every day, or what effect a new medication was having on their gut bacteria, was to question people directly. Today such information can be recorded on a semi- or fully automated basis using smartphone apps or other wearable sensors. This means that information is no longer obtained in dialogue with the patient, but instead by direct interrogation of the body. At the same time, researchers today have access to datasets about the healthcare of large segments of the population (Hoeyer 2016; Kautzky-Willer et al. 2017). This means that the data analyzed for medical purposes no longer – as used to be the case – come exclusively from hospitals or from medical research. Today these data come from many different sources, as well as from smartphones (apps, exercise tracking, etc.), from public archives, from weather and environmental sensors, from private companies that make their data available for medical research either for money or as a philanthropic "donation," etc. (Vayena et al. 2018).

Thanks to the digital and computational tools available today, data from various different sources can be combined and analyzed such that general trends can be extrapolated for the whole population (e. g. that taking lipid-reducing medications reduces the risk of cancer in diabetes patients (Kautzky-Willer et al. 2017). Furthermore, probabilistic statements can be made about individuals (e. g. whether a patient has a higher or lower risk of suffering from a chronic illness than the average of the population).

Another factor worth mentioning in this context is the blurring of the dividing line between medical research and medical practice in the use of data. Data that were collected for clinical purposes, such as for example details of diagnoses, medications prescribed, laboratory tests, and other information kept in patients' records, can today be interrogated systematically to find hidden patterns – such as previously unrecognized links between patient characteristics and prognostic parameters (e. g. patients treated with a certain medication recover particularly quickly after an operation; or patients with a particular comorbidity more frequently have complications after a specific intervention). Insights such as these that can be derived from hypothesis-free queries in large datasets, can be used to inform the treatment of individual patients and at the same time also enhance medical research (Wooden et al. 2017). Respecting the right to self-determination with regard to personal information of the individual patients from whom these data was sourced, and other aspects of data protection, are particularly important in the context of new possibilities like these.

3.5.3 Quantity and quality

Another significant consideration in this context is the quality of the data collected. In an era when all data used for medical research and clinical practice were recorded by clinicians or professional researchers, in most cases it was clear how good or poor the quality of the data was that were being used. In a situation where data come from many different sources, it is increasingly difficult to judge the quality of the datasets. This kind of evaluation is often impossible for both practical and technical reasons, because the company that provides the sensor, the app, or the device used for data collection treats details of the data gathering process as a trade secret. A further challenge in the context of data quality is the question of representativity and thus also comparability. In the past, large proportions of the population would be included in the datasets of a regional health insurance provider, for example – and thus also in the research projects using these datasets. Today, if telecommunications companies or fitness app operators make data about the movements of their customers available to researchers (e. g. to reconstruct the distribution of a virus), then these datasets miss out all the people who do not have a cellphone account or do not use fitness apps. Since these are often marginalized groups of people, big data approaches in medical research run the risk of focusing on the comparatively "healthy and wealthy" group, and overlooking the people who would be most likely to need support.

An additional problem for the interpretation of digital data is the availability of good metadata, i. e. information telling the data analyst who or what these data repre-

sent, what the strengths and weaknesses of the datasets are, etc. The notion that data speak for themselves is not valid in the context of digital big data (Leonelli 2016). To be able to make meaningful use of the data it is essential to have good and transparent information about the context of data collection, e. g. exactly how were the individuals selected whose data are recorded in the database? How were key categories such as for instance diagnostic, therapeutic, demographic and other parameters defined? Were these categories applied in a verifiably standardized way? etc. If this kind of contextual information is not available, there is a risk that the data cannot be interpreted in any meaningful way (or facilitate wrong interpretations).

In general, on the question of data quality, it is evident that criteria for measuring the quality of data cannot be formulated on an abstract basis, but must instead always be determined on a basis that is practice-, context-, and purpose-specific. If data from cellphone network operators is used for research in digital epidemiology, the key criteria are different from those in a context where the aim is to use analysis of electronic health records to find out which people would benefit most from a particular medical intervention.

3.5.4 Data quality and artificial intelligence (AI)

In medical research it has been well known for a long time that a medical application for pattern / rule recognition can reach false conclusions due to problems with the underlying data. In 1997, for example, a study was carried out with the intention of making methods of machine learning useful for medical purposes. The aim of the study was to calculate a survival prognosis for patients with pneumonia, so that patients with an increased risk would be treated as hospital inpatients while those with a lower risk would be treated on an outpatient basis (Caruna et al. 2015; Cooper et al. 1997). On the basis of the training dataset, the rule-based system developed the (incorrect!) rule that patients with a history of asthma had only a very small risk and therefore could be treated on an outpatient basis. This could be explained by the fact that patients with pneumonia who also suffered from asthma were admitted immediately to the outpatient emergency room, where they received intensive care; consequently their survival prognosis improved compared to patients (without asthma) who did not receive intensive care. While the artificial intelligence recognized correctly that asthma patients, with the intensive care they were receiving, had a lower risk, it was not able to recognize that this outcome was dependent on the intensive care received. So a key problem in the area of data quality lies in the fact that data, just like the algorithms used, can be “biased,” which can be attributed to various different causes (e. g. collection and subsequent selection of data, initial design and autonomous further evolution of the algorithm, incorrect use etc., see also 3.7.3).

The problem of and the requirement for data quality in the field of medical research is therefore not limited just to big data, but also affects all collections of data (patient records, data from medical studies etc.; also those that are not interlinked). This again shows the close relationship between the technological opportunities on

the one hand, and the transparency of the decision-making process on the other (black box problem).

With regard to the collection of data from the patient by means of sensors, existing regulations seek to ensure that the sensors are of an appropriate quality (e. g. Austrian Medical Devices Act – MPG, Austrian Product Liability Act – PHG, Austrian Product Safety Act – PSG). The Austrian Medical Devices Act (MPG), and the Austrian Medical Device Regulation (MPVO) which succeeds it, imposes strict requirements on manufacturers, designed to safeguard the security and reliability of medical devices. Before a medical device can be approved, its effectiveness and security must be established in a clinical testing process. The law also treats software as a medical device in many contexts. This government safety standard is accompanied by civil liability law, such as the provisions of the Austrian Product Liability Act (PHG), to protect patients from faulty products that could pose a risk to their health. In the area of product liability there are serious demarcation problems in this context, as to when a product is subject to the jurisdiction of the Austrian Product Liability Act (PHG). The Austrian Product Safety Act (PSG) adds subsidiary protection on the quality of sensors.

We agree that further analyses of these data by private companies (e. g. in the case of fitness trackers) should only be carried out for the purposes of medical research if the decision making process is traceable and the AI system used by a private company is therefore "explainable," which is currently only possible to a limited extent with data-based systems using neural networks and is the subject of intensive research. Nonetheless, impressive results can sometimes be obtained, even with isolated raw data from the sensors – for instance the DeepHeart application depends on neural networks and the sensor data from smart watches (e. g. Apple Watch), which enable it to highlight certain cardiovascular risks (high blood pressure, diabetes, sleep apnea, high cholesterol) with increased sensitivity [98 %] and specificity [90 %].

3.6 Key concepts: association, correlation and causality

Association, correlation and causality are concepts that can have several meanings, depending on the context. In everyday use, the term "association" refers to a mental connection, while "correlation" is the mathematically described relationship between two or more occurrences, and "causality" refers to the reason for an occurrence, identifying a cause and its effect. Broadly speaking, causality – however difficult it is to ascertain – plays a particularly significant role in medicine and healthcare. Even the simple question of whether it was the tablet which resolved the headache, or if the headache would have stopped even without the tablet, illustrates that causality is more complicated than is commonly assumed, since it is not easy to identify whether it was the medication, the passing of time, or both, or something completely different, which made the headache go away.

3.6.1 Changing definitions

With the development of data-intensive biomedicine, which is ultimately the basis of genetic / genomic medicine – important in many clinical contexts today (Horton / Lucassen 2019) – the concepts of association, correlation and causality also gained in significance; this partly reflects growing scientific understanding with regard to causality in complex living systems, and the development of methodology and technology in statistics, epidemiology, and computer science, which is also the basis for machine learning (see 3.7). This also results in evolving definitions of related scientific and philosophical concepts, such that causality in particular is a very multifaceted concept. If, for example, the efficacy of a treatment or undesired side effects have to be identified, a randomized clinical study can provide information, but the results often do not allow to prove direct causal relationships. Therefore, proving causality is a major challenge for biomedical informatics (Kleinberg / Hripcsak 2011).

Causal relationships have three important fields of application in medicine, namely prediction, explanation (diagnostics, and the justification for treatment), and policy (in the sense not only of treatment decisions, but also healthcare policy). Explanations concern the relationship between phenomena and why they are associated with each other, as well as specific occurrences (why they happened at all or why they happened in the way they did; Kleinberg / Hripcsak 2011). Since learning algorithms need to use causality models in order to make decisions (see 3.7), these underlying models are immensely important.

The concept of complexity, which in common parlance is often a synonym for "difficult," requires a deeper understanding of biological systems, which – in contrast to mechanical systems – continue to develop through learning processes: complexity in the context of medicine and healthcare is much more than just "difficult" (Sturmberg 2018). An interesting aspect in this context is the convergent development of different fields of science, which are ultimately leading to a new way of looking at life processes – one which in the ideal case is free of value judgments and stereotypes, and to that extent is more "objective" than previously. However, this results in numerous uncertainties and ethical questions concerning basic research in biomedicine in particular, which relies increasingly on the methods of bioinformatics and data science (Überall / Werner-Felmayer 2019).

3.6.2 The example of genetic association analysis

The following example illustrates the importance of association, correlation, and causality in the interpretation of genetic / genomic data for the expression of complex traits (phenotypes), such as height, or for example neurodegenerative diseases (Alzheimer's, Parkinson's disease), diabetes (type 2), cardiovascular disease, and cancer. Here genetic association analyses provide important insights. With complex traits, models based on "one gene, one mutation, one outcome" (Gallagher / Chen-Plotkin 2018, 717) – as they are used for "traditional" genetic disorders such as Huntington's Disease – are not adequate. This is because numerous genes and non-genetic factors are involved in the manifestation of complex diseases, and their inheritance pattern is not Mendelian (Gallagher /

Chen-Plotkin 2018). Association analysis was initially one approach to data mining in large sets of genomic data, which is of relatively high quality because it is structured. This method looks for patterns that suggest a statistical relationship between genetic variables that appear together. In the medical context, association analyses are carried out for example with gene expression data or with data on genetic variation such as single-nucleotide polymorphisms (SNPs), but can also be used to predict biologically relevant interaction networks, e. g. of proteins (Atluri et al. 2009). Over the last few years, genome-wide association studies (GWAS) have revealed thousands of genetic variants that are associated, with a certain degree of probability, with the manifestation of complex traits or with disease risks, i. e. are found with a certain statistically relevant frequency in people with certain diseases.

However, as there have so far only been a few cases where a causal relationship between the occurrence of a variant and the manifestation of the trait could be found, the biological significance of this kind of variant remains unclear in many instances (Gallagher / Chen-Plotkin 2018). Generally the SNPs that are statistically associated with a certain disease or complex trait are located outside protein-coded genes, so that it is often unclear which genes are affected, which molecular mechanisms are involved, and what impact the changes in the affected gene function or gene regulation have on the risk of disease (Gallagher / Chen-Plotkin 2018). Another factor is that every variant found through GWAS is associated with hundreds or even thousands of other variants, which are in turn associated in a statistically significant way with the occurrence of a trait, without shedding any light on their biological function, i. e. their causality.

This phenomenon was described as the "missing heritability problem," and occurs for instance in a supposedly simple relationship, i. e. expressed as a person's height. Height is a trait – as we have discovered in the era of GWAS – associated with a very large number of genes, which each have a relatively small effect on the way it is expressed, and where environmental influences are also significant; despite analyses of the genome data of a very large number of people (on the scale of 1 million), most of the genetic variants found show no statistically significant correlation, i. e. are "unmappable" (Barton et al. 2019). To enable these genetic variants to be considered in spite of this, polygenic risk scores are now used. These scores are the product of risk analyses in which all variants are added together and weighted according to the strength of their effect, so that complex relationships, which are not understood in detail or in the sense of causality, can be evaluated to produce a measure of probability, or the risk of expression of a particular trait, such as height, or a specific disease. GWAS studies are also always population studies, which means that in addition to inheritance and various environmental factors, a confounding factor, i. e. a so-called third variable or confounding variable, plays a role in the prediction of risk scores, e. g. the structure of the population. Polygenic risk scores, which are based on the analysis of many small effects, can be susceptible to error, as shown in the context of GWAS studies on height (Barton et al. 2019). Because of the additional conjecture of polygenic adaptation, i. e. the concept that adaptation occurs in the course of evolution by natural selection on

the basis of many genetically influenced traits, polygenic risk scores can also lead to the overestimation of population differences with regard to their disease risks and other complex characteristics (Rosenberg et al. 2019; see 3.7.3 for further detail).

Polygenic risk scores have caused a shift from the original aim of genomic analyses, which was to identify genes that are causally involved in the expression of a trait, towards the prediction of a phenotype on the basis of predispositions associated with certain traits. This is a fundamental difference to previous conceptualizations of genetic inheritance, which were oriented towards clinical diagnoses. It is therefore extremely important to be aware of this shift from a diagnostic to a predictive understanding of a variety of genetic and other causes of certain traits in order not to fall into the trap of reductionism, which ultimately leads to the violation of ethical standards with regard to the dignity of the individual in his or her own particular form. Another aspect is that the concept of prediction implies that occurrences can actually be predicted. In the context of genetic variants however, it is often unclear to what extent these affect the expression of a trait. Polygenic risk scores may provide a figure that links statistical relationships with a possible risk, but they do not simplify the complex functional relationships between genes, the environment and evolution. Moreover, as shown by the latest academic literature on the subject, they can foster the misunderstanding of a causal relationship between genetic variations and complex traits (Rosenberg et al. 2019) (see also 3.7.3). In the clinical context however, polygenic risk scores can still be helpful, particularly in the early identification and prevention of common diseases in adults, such as coronary heart disease for example (Torkamani et al. 2018). It is evident that association studies have evolved in just under 20 years from pattern recognition in genomic data to prediction studies with a high potential for relevance, which need a deeper mathematical understanding of risk and causality, and need to be carried out in precisely defined contexts.

3.7 Machine learning / algorithms / technical foundations

Generating "meaning" from big data requires mathematical models and statistical methods for pattern recognition, classification, prediction, and — most importantly — causal relationships (causal inference), taking a large number of variables and their interrelationships into account. Developments in this area have recently fueled expectations that autonomous systems with human-like intelligence could soon become a reality. Judea Pearl, a pioneer in causality research, nevertheless still sees a number of fundamental obstacles, such as the inability of these systems to respond to new situations that have not been preprogrammed, the inability to explain the predictions and recommendations they make (i. e. "black boxes", see 3.7.4), and a lack of understanding of cause-effect relationships. However, these issues could be solved with computational tools for causal modeling (Pearl 2019).

Methods of artificial intelligence have been used in medicine since the development of computer-aided clinical decision-making systems in the 1970s. Since then, researchers have been increasingly successful in using large quantities of unstructured data, such as natural language texts and images, within the framework of various types of machine learning. Machine learning has developed in recent years into one of the most successful forms of artificial intelligence. This refers to giving computers the ability to learn by means of their own activity, such as by analyzing data from surveillance cameras. This implies that computers "reprogram" themselves and that therefore not all of their actions were programmed in advance. When such a computer is fed with data, it generates and refines complex analytical models, optimizing them based on a learning model to improve the accuracy of a desired solution to a problem.

While relatively simple, linear models with only a few variables were used in clinical practice prior to these new possibilities of machine learning, this required extensive, laborious effort to extract data from various sources and standardize it. Today, some learning models utilize millions of data points and tens of thousands of predictors to arrive at more accurate prognoses (Rajkomar et al. 2018). The research literature describes numerous examples of machine learning that promise better diagnoses (such as for rare diseases), timely identification of high-risk patients, and prevention of undesired side effects of medications as well as a more efficient allocation of resources.

3.7.1 What is machine learning?

The "learning" part of machine learning is based on algorithms capable of optimization. Algorithms are nothing new and are certainly not an invention of the computer age. They are rule-based procedures that allow humans or computers to solve a problem systematically. Algorithms consist of a sequence of clear instructions, such as an "if-then" decision tree with multiple branching points. The algorithms used in the context of machine learning generally serve to minimize errors or maximize the accuracy of an output, such as a prognosis. Typically, algorithms take the form of a formula, a diagram (such as a decision tree) or a scale. Simple examples from medicine are the body mass index or the Framingham risk score for cardiovascular disease, which are calculated based on algorithms.

The results of a machine learning process depend on the quality of the input (the data) and the algorithm. The first predictions are typically not very accurate or even false, but it is possible to measure their deviation from the correct result, and then use the error to improve the algorithm. However, this is only possible if suitable data sets exist for validation of the results. **Neural networks** have this capability. They measure the deviation (or the error) and modify the parameters until a higher accuracy is achieved. They represent a form of **optimization algorithm**.

"Deep learning", a subcategory of machine learning based on artificial neural networks, refers to algorithms that have achieved a special accuracy in solving numerous problems. These are primarily mathematical models that imitate the method by which the human brain interprets information to draw conclusions. A neural network consists

of many individual neurons, which are usually arranged in multiple connected layers. The number of layers determines the level of complexity that an artificial neural network can depict, with many layers making a network “deep”.

It is also important to understand the various types of machine learning (Jannes et al. 2018; Deo 2015):

- Human-supervised machine learning algorithms apply what they have learned to new data in order to predict future events. In this case, the learning algorithm is trained with existing examples for which the intended result is already known, and with sufficient training the algorithm is capable of independently generating results based on new data. The algorithm also compares the achieved results with the intended output to identify errors and modify the model.
- In contrast, unsupervised machine learning algorithms are used for problems for which the data has not yet been classified or described. The learning algorithm is able to independently identify new patterns and correlations without relying on existing prototypes. Features are used for spontaneous classification.
- Reinforced learning lies between the two learning methods described above. A goal is specified as with supervised learning, but the algorithm must find the realization independently as with unsupervised learning. The algorithm carries out a kind of trial and error process to optimize the result.

A simple example of supervised learning is the automated interpretation of ECG or X-ray images based on a pattern recognition method. The algorithm is provided with a limited selection of diagnoses and given the task of correctly classifying the data. In this case, where the correct diagnosis is known, it is also possible to evaluate the quality of the algorithm, which is often impossible in other contexts. The advantage over diagnosis by an experienced physician is the potential for a more accurate diagnosis based on the analysis of a large volume of data.

With unsupervised learning, the goal is to identify patterns in previously unstructured and uncategorized data. In view of the heterogeneity and multifactorial nature of many diseases, there is great interest in identifying and defining their variants (with the goal of precision medicine; see National Research Council 2011; Prainsack 2015), thereby enabling the development of more targeted therapies.

Despite the enormous progress that has been made in analytical methods, machine learning has so far been only partially successful in solving the challenges it faces. These challenges concern:

- the high complexity of the data to be processed;
- the risk of faults and bias in the algorithms used;
- the tendency to dispense with (causal) substantiation in favor of sometimes uninterpretable correlations.

3.7.2 Data complexity

Machine learning in connection with "big data" is confronted not only with large volumes of data but also with highly heterogeneous data originating from diverse sources. In the field of medical research, sources include research and development data (such as laboratory or pharmaceutical data), clinical administrative data, electronic health data, and patient-generated health data. This latter type of data originates from social media or other online resources, smartphone apps, and "wearables" (devices with sensors that are integrated into clothing or implanted). The modality of the data is correspondingly heterogeneous: previously coded information, free text, and images mixed together with other types of signals or graphs. This data complexity is one of the motors behind the development from classical multivariate data analysis to the new "data science", a discipline concerned with "cleaning up", preparing, and analyzing data.

Closely related to the large volume and heterogeneity of the data is a phenomenon called noise accumulation. Predictive analytics with the aid of statistical methods frequently involves simultaneous estimations of multiple parameters. For example, cardiovascular disease is associated with numerous factors, such as smoking, overweight, diabetes, elevated cholesterol level, high blood pressure, and other risk factors. If the accumulated estimation error (or noise, in other words the signals that have no significance for the question to be answered) of these parameters is high, the learning algorithm may overlook variables with significant explanatory value. This is a mathematical problem that must be solved in order to strengthen trust in automated medical decision systems (Gandomi / Haider 2015).

Another associated problem is what has been termed spurious correlation. This refers to the phenomenon that a statistical analysis can lead to false results if the massive data volume leads to the identification of correlations between the data, although they actually have no meaningful relationship to each other (e. g. the number of storks in a region relative to the number of births; both might increase in the same period without being meaningfully related). The more variables there are, the more correlations may be statistically significant. This is related to the fact that in large data sets, large deviations can be attributed more to variance (or noise) than to information (or signal) (Calude / Longo 2017). In a paper entitled "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," Anderson (2008) argues that "… with enough data, the numbers speak for themselves."

3.7.3 Bias

One problem under much discussion is the potential systemic distortion of algorithms known as bias. Citron and Pasquale (2014) describe this as a general problem of increasing digitalization.

> "By scored society, they mean the current state in which unregulated, opaque, and sometimes hidden algorithms produce authoritative scores of individual reputations that mediate access to opportunity. These scores include credit, criminal, and employability scores. Citron

> and Pasquale particularly focused on how such systems violate reasonable expectations of due process, especially expectations of fairness, accuracy, and the existence of avenues for redress" (Osoba / Welser 2017, 11).

Machine learning algorithms are said to be slaves to the data from which they learn. The principle of "garbage in, garbage out" is well known. It is possible to differentiate between various sources of potential systemic distortion in the use of machine learning algorithms (Gianfrancesco et al. 2019):

- **Bias in the algorithm**: This bias has nothing to do with the data set; it is a mathematical property of the used algorithm. Its opposite is variance: algorithms with high variance (sensitivity to small fluctuations in the training data) are in fact compatible with higher data complexity but are also more sensitive to noise, making them less able to handle data outside of the training set.
- **Bias in the sample**: Bias exists in the sample if the data set used for training the algorithm does not adequately represent the problem space for the model. Various techniques are employed to avoid such bias, such as validation of samples with respect to their representative character or identifying characteristics of the population that are intended to represent the sample. With respect to the USA, for example, Cohen and Grave (2017) argue that ethnic minorities, women, people with low socio-economic status, and immigrants are only insufficiently represented in data sets (electronic patient files, genome databases, presence on the internet and in social media), meaning that their specific health problems are also underrepresented. According to Need and Goldstein (2009), for instance, 96 % of the participants in genome mapping studies were of European origin.
- **Prejudicial bias**: This refers to the influence of cultural or other stereotypes on patient data. For example, Hammerlund (2018) shows an ethnic bias in the predictors for invasive surgical interventions in the case of acute myocardial infarction. This bias, which influences the health data of patients of African-American origin, is related to the fact that such persons are more likely to be treated by surgeons with higher (risk-adjusted) mortality rates as well as that this patient group finds it more difficult to identify the symptoms of an infarction and subsequently seek medical attention. It is important to emphasize that prejudicial bias need not be intended or even conscious; implicit bias (Greenwald / Krieger 2006) arises, for example, when people who are not aware of any prejudice translate the dominant expectations of hierarchies or social orders into their actions without being conscious of doing so (FitzGerald / Hurst 2017). One example here is the practice of taking complaints by women that indicate cardiovascular disease less seriously because it is considered a "male" disease. Conversely, osteoporosis was long considered a problem affecting women after menopause, although one-third of broken hips associated with this condition occur in men, who also experience twice the mortality rate of women (Looker et al. 1997). Neither humans nor

machines (which are supplied with already distorted data) are immune to such systematic distortions.

- **Measurement bias**: Misclassification of illnesses and measurement errors are frequent sources of bias in observational studies and analyses of patient files (Pot et al. 2019). Also Gianfrancesco et al. (2019) refer to the example of the unequal diagnosis and treatment of women and men suffering from coronary heart disease. A learning algorithm based on such distorted data will reflect the bias existing in actual medical practice and therefore misclassify patients.

The topic of gender bias in medicine is currently receiving significant attention. It offers a good illustration of the various forms of bias that are due to the insufficient representation of women in clinical studies as well as prejudice and measurement errors. The topic of gender has been addressed in all discussions of the Bioethics Commission since 2007. A study demonstrates considerable differences of the ways in which women and men describe somatic symptoms (Barsky et al. 2011). In "Pain and Prejudice", Jackson argues: "Centuries of female exclusion has meant women's diseases are often missed, misdiagnosed or remain a total mystery" (Jackson 2019). Despite extensive efforts to counteract this neglect of women in research and the resulting gender bias, these efforts are still limited to a restricted set of illnesses, with considerable national differences. One example is Alzheimers disease, where women's advantage of better memory performance on average often results in their being diagnosed (too) late (Sundermann et al. 2017). Gender bias is particularly pronounced in the area of mental health. Examples include how pain is expressed by women and men and how this is interpreted by the treating physicians (Samulowitz et al. 2018) as well as the fact that men and boys more frequently receive a diagnosis of ADS than women and girls (in a 4:1 ratio) because the diagnostic instruments were primarily developed and validated on male subjects (e. g. Beggiato et al. 2017).

These and many other studies show that sample bias (as well as prejudicial bias) potentially impacts a wide range of illnesses and their representation in "big data" data sets. Chin Yee and Upshur (2018) argue that these various forms of bias correspond to the inherent sources of uncertainty in diagnosis and therapy planning in medical practice. There are few good reasons to assume that these uncertainties and errors could be compensated for by the sheer volume of data.

3.7.4 Interpretability

Machine learning does not fundamentally diverge from conventional statistics. But while the latter is focused on testing hypotheses with regard to causal relationships, machine learning is less concerned with the interpretability of models. Its focus lies much more on the predictive performance and generalizability of the models. However, the core of a physician's work, according to Chin-Yee and Upshur (2018), lies precisely in providing reasons – "clinical judgment refers to the range of complex reasoning tasks and actions

performed by clinicians in the context of offering diagnosis, therapeutic options, and prognosis to patients regarding their health and illness" (638).

Machine learning algorithms are frequently referred to as "black boxes" because the methods of generating data models are difficult (or even impossible) to interpret since the functions that connect the data (input) to a specific result (output) are too complex. In this way, a fully automated, algorithm-based process can undermine the desire to find a causal explanation. This means that physicians are confronted with the problem of having to trust the system without understanding the system's conclusions and without being able to explain them to their patients (Vellido 2018).

On the other hand, it is argued that rigidly insisting on interpretable causal relationships could compromise the advantages of algorithm-generated processes, since the necessary simplifications would undermine a model's efficiency. Models could prove accurate and efficient in practice despite not being transparent. In addition, reservations are expressed concerning the exclusion of correlations that cannot be explained based on current medical understanding but still have a high predictive power (Zarsky 2018). However, these correlations can only be significant within their specific contexts and not permit any generalization. Refraining from identifying a causal relationship in a correlation, or an underlying mechanism, can lead to overlooking possible side effects. It also may mean relinquishing the opportunity to gain fundamental insights into, for example, the mechanisms of an illness.

It is also argued that focusing on otherwise unexplainable correlations could reveal patterns that may be made use of in potentially interesting ways; nevertheless, additional sources of error are also introduced in this way. For example, such correlations could allow to group individuals in various ways with the consequence that they are treated unfairly or stigmatized (Zarsky 2018). One classic and widespread example is the use of "race" (generally self-identified) as an imprecisely defined proxy for belonging to a population that is primarily defined socio-culturally and to make predictions concerning the health of individuals on this basis. This has, for example resulted in the misdiagnosis of hemoglobinopathies (sickle cell anemia, beta thalassemia) or underdiagnosis of cystic fibrosis among African-Americans (Yudell et al. 2016). This practice, which, in recent years (in the so-called post-genomic era) has again and increasingly set foot in medical journals, replaces an understanding of the genetic diversity of humans by the entirely outdated concept of continental "races" instead of finally abandoning this concept and to identify valid, causally explainable, correlations of genetic markers (Bonham et al. 2018; Cooper et al. 2018). Because this is an exceptionally complex domain, no solution can be expected here anytime soon, although the call for new solutions and attention to the topic is gaining voice (Yudell et al. 2016; Bonham et al. 2018; Cooper et al. 2018). Simplified, genetically deterministic explanations of intelligence and human behavior have become fashionable in the age of "big data" and are given preferential treatment in media reports (Barton et al. 2019; Rosenberg et al. 2019; Torkamani et al. 2018; Comfort 2018).

There are also efforts to counteract the "black box" effect that arises from the difficulty of explaining or interpreting results. Examples of this include rule-based

representations that are compatible with medical justifications or nomograms, which are used by doctors for visualizing the relative weights of the symptoms underlying a diagnosis. Wachter et al. (2018) discuss the possibility of increasing transparency by providing counterfactual explanations. According to them, it is not absolutely necessary to make the algorithm itself understandable, which would often be entirely impossible due to its complexity. Counterfactual explanations arise from simple "if-then" statements, such as: "If your 2-Hour serum insulin level was 154.3, you would have a score of 0.51; If your Plasma glucose concentration was 158.3 and your 2-Hour serum insulin level was 160.5, you would have a score of 0.51" (Wachter et al. 2018, 21). Wachter et al. consider counterfactual explanations to be a reasonable "lightweight form of explanation"; they are also easy to compute automatically:

> "Unlike existing approaches that try to provide insight into the internal logic of black box algorithms, counterfactual explanations do not attempt to clarify how decisions are made internally. Instead, they provide insight into which external facts could be different in order to arrive at a desired outcome" (43).

Counterfactual inference can also be used to make predictions by machines "approximately fair" by introducing fairness classifiers that function independently of the causality model used by the algorithm (Russell et al. 2017).

4 Case studies

As already described in the preceding sections, great expectations have been pinned on big data analysis and artificial intelligence with regard to improved healthcare.

Examples show that the use of machine learning and artificial intelligence yields advantages particularly with regard to more standardized and optimized evaluation of imaging data in diagnosis. Additional clinical aspects are still relevant for the final diagnosis, however, meaning that an expert is still always essential for the final decisions. One major advantage of machine learning and artificial intelligence can be seen in the ability to make optimal diagnostics available to more people around the world. However, such developments absolutely depend on the existence of a highly standardized data basis, as was shown in a study by the Department of Dermatology at the Medical University of Vienna (see below). A number of individual examples of the use of AI in clinical work are presented below, although these are primarily taking place within the framework of research projects and have only been incorporated into routine medical practice to a very limited extent.

4.1 Public health / health planning

Data analysis can make it possible to respond more quickly to changes in the health of the population, to develop corresponding prevention models, and to establish the necessary healthcare resources when they are needed. One example is a project under the auspices of the Austrian Social Insurance System together with Stefan Thurner at the Institute of the Science of Complex Systems at the Medical University of Vienna aimed at making predictions of health developments on the basis of anonymized data sets. Every disease and every medical treatment of eight million Austrians in the years 2006–2007 was analyzed, regardless of whether this took place in a hospital or private clinic. This makes possible a prediction of the risk for every single disease (a total of 1642) in various segments of the population (Sauter et al. 2014; Klimek et al. 2016), depending on age and gender. The analysis was anonymized and enables a mapping of the "disease demographics" of Austria.

Furthermore, Stefan Thurner and Peter Klimek worked with Alexandra Kautzky-Willer of the Department of Internal Medicine III of the Medical University of Vienna to investigate the personalized disease risks of diabetes patients. More than 100 "disease pairs" were identified (Klimek et al. 2015). This gives physicians the ability to plan preventive therapies on the basis of risk scores and to specifically ask patients concerning possible secondary ailments in order to respond early to potential developments. Such analyses also enable the estimation of future costs in the healthcare system.

4.2 Imaging procedures

Methods of machine learning are of particular interest in this area, with algorithms providing diagnostic assistance in oncology (e.g. detection of breast or lung tumors). New, sensitive techniques (PET / MRI) and multiparametric evaluations permit better differentiation and classification of malignant and benign breast lesions (Vogl et al 2019). The first step here is improved "recognition" of a disease, while the second step involves prognosis and predicting the course of the disease in order to ensure an optimally individualized therapy (European Society of Radiology [ESR] 2019).

Another application is concerned with recognizing "phenotypes" in large clinical routine populations. Imaging data is used here in order to identify patients with similar characteristics and thereby to investigate diagnostic categories and establish them as tools for decision-making in machine learning models.

Future applications are focused primarily on the use of new medications and therapies since methods that enable the prediction of treatment effects (such as within the framework of clinical studies) can contribute to better characterization of patient groups.

4.3 Neurology

In neurology, imaging procedures are a key tool for clinical diagnosis. One important area of application is structural magnetic resonance imaging (MRI) for differential diagnosis of neurodegenerative diseases, such as differentiating atypical parkinsonism from Parkinson's disease (Scherfler et al. 2016). Because the diagnostic is currently aimed primarily at the early stages of these diseases, recently developed procedures for automated identification of brain areas related to the disease and the 3D volumetry of these areas are of great importance. As recently shown in a study at the Medical University of Innsbruck, AI-aided image analysis can yield a significant improvement in the detection and classification of atypical parkinsonism, which is of particular relevance for treatment as well as the performance of clinical studies (Krismer et al 2019).

Despite the increased efficiency and sensitivity made possible by a form of artificial intelligence, validation by experts who are well versed in neural anatomy remains essential (Scherfler et al. 2016).

Such systems help free up resources while also offering a more precise diagnostic basis. However, they also generate additional work. Current estimates expect a roughly 20–30 % improvement in diagnosis in this area thanks to computer-aided imaging procedures.

4.4 Dermatology

A study by the Department of Dermatology at the Medical University of Vienna has shown that artificial intelligence is superior to humans in the diagnosis of pigmented skin lesions such as moles and melanomas. The objective was to differentiate between benign and malignant pigmented skin lesions. The training data came from the image database HAT10,000, which contains over 10,000 digitized reflectance confocal microscopy images of seven different types of pigmented skin lesions (Tschandl et al. 2018). The performance of 511 dermatologists from 63 countries and of varying skill levels, from beginners to experts with years of experience, was compared with the diagnostic algorithms of 77 laboratories working with automated analysis methods.

The best human experts correctly identified 18.8 of 30 images, while the best machines achieved 25.4 correct classifications. The study impressively demonstrated that the most important basis for such results is a comprehensive, well standardized, and high-quality database (Tschandl et al. 2018).

However, the authors note that while computers clearly achieved superior performance in this experiment, they cannot replace humans since the computer only analyzes an optical image from a specific point in time, while the diagnosis of a patient depends on observation of the progression, an estimation of whether the individual is a high risk patient based on various factors, how the skin lesion feels to the touch, and a comparison with other moles on the patient's body. The interpretation of the results is therefore still left up to the human physician.

4.5 Ophthalmology

One project that has gained significant international attention is an imaging procedure developed by the Department of Ophthalmology and Optometry in collaboration with the Center for Medical Physics and Biomedical Engineering of the Medical University of Vienna. High-resolution optical coherence tomography (OCT) enables the early diagnosis of retinal conditions to allow for specifically directed treatment.

This imaging procedure produces precise, layered images of the retina without direct contact for analysis by automated algorithms. In just a few seconds, this method can diagnose retinal conditions at high resolution in order to initiate the necessary treatment within the framework of personalized medicine. For the development of the OCT method, Christoph Hitzenberger and Adolf Fercher of the Center for Medical Physics and Biomedical Engineering were recognized in 2017 with the Dolores H. Russ Prize, the “Nobel Prize for engineering”.

It is hoped that every ophthalmologist will soon be able to make use of this technology, improving diagnosis and therapy for the roughly 170 million people around the world who suffer from macular degeneration.

The projects described in sections 4.4 and 4.5 are examples of how artificial intelligence and new technologies open up the possibility of lowering the threshold for improved, standardized care of many patients (Gerendas et al. 2018).

4.6 Oncology

The potential of machine learning in the field of oncology can be illustrated by a basic research project of the Christian Doppler Laboratory for Applied Metabolomics (CDL-AM) at the Medical University of Vienna. This project is focused on developing a non-invasive diagnostic method for optimizing therapy and follow-up in oncology. The goal is to obtain an algorithm that can predict the further progression of the disease with high probability based on individual patient data.

Within the project, algorithms are being developed that combine functional imaging data from positron emission tomography (PET) with histopathological imaging and genetic data from tissue samples of the same patients. This data is used to optimize the PET-generated cancer diagnosis with machine learning programs such that invasive diagnostics (biopsy) is only still required in exceptional cases. In addition to eliminating the difficult extraction procedure, this bypasses the limited informative value of a biopsy of an ever-enlarging, heterogeneous tumor. Mutations in the genetic material as well as epigenetic changes of the chromatin influence the metabolism and thereby also the architecture and appearance of the tumor. The texture and development phases as well as the metabolism of a tumor can be determined via PET, and metastases can be detected. This enables the structural and functional measurement of the entire tumor mass in real time.

The image data therefore reflects the status of the tumor metabolism, from which predictive statements can be derived concerning the further progression of the disease. The procedure is being developed together with the corporate partner Siemens Medical Solutions. Frequently occurring cancer types such as colorectal carcinoma, prostate carcinoma, and aggressive tumors in the otorhinolaryngology area are being analyzed first. In the initial step, learning algorithms are being fed with PET data from patients treated at the Vienna General Hospital for whom the histopathological and molecular pathological diagnoses based on tissue samples are available and a complete medical history is known. The result is a retrospective correlation between imaging parameters, the tissue-based morphological diagnosis, and (epi)genetic parameters as well as the corresponding disease progressions.

In parallel to this, animal models are being generated in accordance with the investigated cancer types to enable a proof of principle check of the correlation between specific genetic defects and PET data based on a µPET machine built specially for mice.

4.7 Intelligent aggregation and visualization of health data

The use of intelligent applications is not limited to direct diagnosis. It is also suitable for improving the presentation of information on which medical, therapeutic, and patient care activities in healthcare facilities are based.

For example, every diagnostic, medical, and therapeutic action is documented in healthcare facilities today. This documentation is stored and managed in a hospital information system (HIS). This extensive, detailed documentation serves as the basis for subsequent diagnosis and therapy decisions as well as for therapeutic and care activities. Manual, non-automated, and unfiltered queries of numerous and diverse documents in the original format (e. g. X-ray images, laboratory results, other texts, and scans) as well as repeated queries of the same patient information are nevertheless viewed by healthcare personnel as an additional burden, not least due to the increasing volume of available information. The information problem is particularly large at the interfaces. This applies to inter-professional (e. g. between physicians of various in-patient departments) as well as multidisciplinary (e. g. between physicians and therapists) information exchange between the various types of healthcare personnel. Although modern hospital information systems increasingly collect, store, output, and manage data in a structured fashion, the querying of relevant information still requires individual, manual, and non-automated actions that lead to unfiltered results. The time required for obtaining the information also depends on the access permissions, the searcher's knowledge of the system itself, and the speed of the system. This cuts into the actual treatment time.

Various efforts have therefore been initiated to improve this interface through the use of intelligent (technology-based) solutions. A research project of the Austrian Research Promotion Agency (FFG) entitled "Smart Aggregation and Visualisation of Health

Data" (SMARAGD) aims to develop technical components for the intelligent aggregation and visualization of information from electronic health data that fully takes into account the needs of specific professional groups (the project is focused on information relevant to occupational and physical therapists by way of example). In accordance with its objectives and subject matter, this research project is supported by a wide range of disciplines (IT, medical and health sciences, social sciences, legal), which explains the large number of research partners. The project is being directed by the IMC University of Applied Sciences Krems, and other participating research institutions include SYNYO GmbH, the University of Graz, the Linz Institute of Technology of the Johannes Kepler University Linz, the Medical University of Graz, Know-Center GmbH, and the University of Vienna. It is partially funded by the Austrian Research Promotion Agency.

5 Ethical aspects of artificial intelligence

The relationship between ethics and medicine is ancient and has undergone various transformations over the course of history. If one distinguishes "ethics" from "morality", understanding the former to be the theory of morality and morals, "morality" can be understood as the totality of ethically related principles, rules, and norms that are of importance to the interactions between different societies and individuals. The objectives of "morality" and "ethics" are closely related to the objectives of medicine. Put simply, it can be stated that the objectives of "ethics" and "morality" lie in establishing the foundation for a good and successful (communal) life, while the objective of medicine lies in preserving and restoring health. However, all this is not solely about knowledge but rather the application of this knowledge in connection with the respective objectives.

"Medical ethics" is currently understood to refer primarily to the area of applied ethics, which is focused on providing guidance on "good and correct medical behavior" in the sense of the main principles of medical ethics (respect for autonomy, doing no harm, doing good, and justice).

The shift in medical ethics since ancient times (Hippocrates) is due to a number of causes. First, one must consider the infinitely expanded range of activities resulting from technological development. At the same time, the shift has been influenced by the plurality of lifestyles and moral perspectives that have driven this expanded range of activities throughout the course of historical and cultural changes.

As the flip side of the new diagnostic and therapeutic possibilities, made possible in part through big data and artificial intelligence, a growing mistrust of the health system can be observed, based on an objection to the anonymity and loss of humanity

as experienced both by patients and practitioners. In no small part, this has its roots in the growing power of bureaucracies in our healthcare systems as well as the lack of transparency concerning their economically based decisions.

Special challenges for medical ethics are found in exceptional and extreme situations such as in vitro fertilization, technology-based artificial prolonging of life, organ transplants, and intensive medical interventions directed not only towards survival but also to improve the patient's quality of life. Such problems cannot be resolved simply through the application of algorithms because the associated decisions cannot be made solely based on quantitative criteria.

The role of medical ethics must therefore be to set limits to the dominance of technological, economic, and administrative factors and to require inclusion of the human aspect – the "dignity" of the individual – with equal weight in all medical considerations. The danger of viewing doctors as "health engineers" and patients as "objects to be repaired" has long been a topic in medical ethics discussions.

5.1 Changes in the physician-patient relationship

The increased use of decision-making systems based on machine learning is changing the physician-patient relationship. This is not necessarily an entirely new phenomenon since this relationship has changed continuously over time due to a wide range of developments, such as the proliferation and partial technologization of medical diagnosis and therapeutic procedures, the increasing specialization of the medical profession and the organizational complexity of the healthcare system. The electronic health record was originally designed to simplify billing. It was not intended for use by physicians during their conversations with patients. Very often, physicians experience it as compromising their well-being, causing mental fatigue or even burnout symptoms. They feel that there is too much to document, especially in the first consultation, which is so important on an interpersonal level (e. g. Ventres et al. 2006). For example, Lown and Rodriguez (2012) concluded from an empirical study: "Screen-driven communication inhibits patients' narratives and diminishes clinicians' responses to patients' cues about psychosocial issues and emotional concerns" (392). The fundamental need of the (ill) person, on the other hand, is to receive attention from the physician.

The relationship between physician and patient is based on a process of deliberation in which findings, observations and decisions are communicated, justified, and discussed not merely with regard to "facts" but also values, preferences, and attitudes (e. g. Wirtz et al. 2005). There is a fear that the use of machine learning in medicine could displace this relationship based on trust and dialogue in favor of statistical arguments; especially if the results of predictive analytics cannot be causally interpreted and made transparent (see 2.3.4). But even if one succeeds in establishing alternative explanatory models, such as counterfactual explanations, in medical practice reservations remain with

regard to quantitative-mathematical approaches to medical justification, as illustrated by Cohen et al. (2017):

> "Imagine a woman undergoing treatment for breast cancer and trying to decide whether to opt for partial or the much more invasive radical mastectomy. The doctor recommends the radical mastectomy. When asked why, he says 'for patients like you we know from the data that it tends to be the best option.' When she asks 'what is it about my case that makes you think that', shall he respond 'the algorithm has examined 10,000 variables from your EHR and, based on its validated model, determines this is what is appropriate in your case?' To be sure that may be much better than answering as the current physician might that 'given the limited number of patients I have seen in my practice, what I learned in medical school, and what I have read in the literature, I think this will be better for you' – but will the patient accept the former answer as better?" (452).

Another ethical problem is related to the fact that methods of predictive analytics tend, as already mentioned, to direct the perspective from the individual patient to a patient population. Rules that health organizations are already attempting to implement today may gain additional weight in this way. Chin-Yee and Upshur (2017) argue that this would push the ethical, patient-centered focus (which distinguishes a physician's work) on the questions "What is the significance for this patient?" and "What should be done for this individual?" to the background:

> "Such questions can be addressed, but require us to move beyond the quantitative evidence emphasized by data-driven approaches, contextualizing these approaches with the qualitative, personal evidence which emerges through a process of dialogue between physician and patient, and then applying the totality of the evidence through the exercise of phronesis" (644).

In the end, humans are beings that require engagement, attention, and care. Illnesses, especially severe illnesses, often produce regressive states in people since they frequently induce great fear and prompt child-like behavior, dependence, and a need for security. At the same time, the advanced technical capabilities – such as scientific precision and optimal therapy – should be made available, including with the help of artificial intelligence. Physicians should know the patient and apply their knowledge in a caring and competent fashion. The greatest hope of artificial intelligence lies in the possibility of using the valuable time that can be saved to deepen the trust between the physician and the patient. There would then be sufficient time during consultations to build deeper and more compassionate connections. It would also be worth considering how physicians could be trained to gain a better understanding of their patients. In his book "Deep Medicine", Eric Topol shows how the proliferation of machines increases the diagnostic power, thereby expanding the fund of medical knowledge for all physicians.

5.2 "Whose responsibility" in the case of machine-based decisions?

In the debate surrounding the implementation of machine learning methods, references abound to the frequent occurrence of medical errors by doctors, associated with the assumption that such errors could be reduced through the use of artificial intelligence. In a now classic study, Fagerhaugh et al. (1987) described the sources of errors in clinical settings (persons, organization of the workflow, procedures, equipment) as well as the difficulties that healthcare organizations have in addressing and handling these problems. The literature also refers to numerous sources of error in diverse medical technologies, including those associated with big data analytics: systemic errors (such as bias in data), false or unverifiable data, etc. Physicians, who continue to be responsible for the basis of their decisions and actions, can only account for this by critically evaluating their information sources, insofar as this is possible in their everyday work.

The use of big data, artificial intelligence, and machine learning opens up multiple areas of concern that must be resolved to determine whether and how physicians can live up to this responsibility.

5.2.1 Freedom of action on the part of physicians

Exercising responsibility is getting more difficult, as automated decision-making processes are narrowing physicians' spaces for action. One possibility is to "suggest" to physicians specific selected options in the event of a diagnosis by incorporating default rules or framing effects that reflect the preferences of the healthcare organization into the underlying model. This possibility goes back to the idea of "nudging" – to present the decision options in a specific situation in a way that the person chooses the most "reasonable" option – proposed by Thaler and Sunstein (2008) an idea that has met some criticism. Applied to the situation of physicians, this means that – for instance given a risk of sepsis – they are "steered" by the system to the variant viewed as most economical by the healthcare organization, with other variants requiring special justification, in other words additional effort (Cohen et al. 2014).

The chances of physicians to exercise the responsibility associated with their role therefore depends on the decision-making freedom that is offered by the implementation of the predictive analytics system and on how much physicians are supported or hindered in the exercise of this freedom. Cohen et al. (2014) discuss the variants of opting out and opting in:

> "The model could trigger a patient care intervention that would occur unless the physician or nurse overrode it (opting out) or could suggest an option that the physician or nurse would have to order or carry out (opting in). The choice of default options becomes even more challenging if the recommended action carries substantial risks as well as benefits" (1145).

5.2.2 The criterion of system criticality

The concept of system criticality is used in this context. It combines the risk associated with a decision with the structural decision criticality. Here, "risk" means the combination of the likelihood of harm occurring (e. g. due to a poor decision) and the severity of the potential harm. On the other hand, structural decision criticality refers to the complexity of the decision (from the pure representation of reality to a value-based assessment of the reality to the multifactorial prediction of a future reality), the effects of the decision (from a purely abstract conceivable action context to a concrete action context to the direct implementation), and the reversibility of the effects (from full reversibility to irreversibility). The higher the system criticality, the higher the system requirements with regard to the role of the physician in the decision process, the transparency and explainability of the results, and the monitoring of the system.

Another important decision criterion is the level of automation. A taxonomy introduced by Pasumaran et al. (2000) differentiates four levels of automation: (1) the system filters the information classified as relevant for an analysis; (2) it uses this information for a diagnosis; (3) it proposes a sequence of actions; (4) it executes these actions. Only at this last level is the decision concerning diagnosis and therapy actually fully automated. The critical question with regard to the introduction of a decision system based on machine learning is therefore about the "tradeoff in which more automation yields better human-system performance when all is well, but induces increased dependence so that it will produce more problematic performance when things fail" (Wickens et al. 2010, 389). This means that a system, as the described case studies suggest, may in certain cases achieve better results than the physician, but only "when all is well" – 2when all prerequisites for an optimal decision on how to proceed are met.

5.2.3 Handling bias and lack of transparency

One central aspect for exercising responsibility lies in the options available to physicians for handling bias and the lack of transparency. Machine learning algorithms are frequently referred to as opaque "black boxes". This means that physicians are confronted in some circumstances with the problem of having to trust a predictive analytics system without any insight into the system's conclusions and without being able to explain them to their patients (Vellido 2018). Another problem lies in the various forms of bias – bias in the algorithm, bias in the sample, prejudicial bias, measurement bias (see 3.7.3) – which limit the informative power of the results of predictive analytics and can even lead to incorrect results.

In consideration of these limitations to machine learning and the data sets and algorithms on which they are based, it is apparent

> "that it is important to have 'humans in the loop' in AI systems, even as they become more and more advanced and accurate. Humans are needed to check the quality of the data as well as interpret and contextualize the results" (Ferryman 2019, 1).

However, this requirement cannot be directed only at physicians. It also applies to data specialists and system developers. Among the proposed approaches for realizing the requirement "to have humans in the loop" are: to insist on transparency (providing insight into how a model makes decisions) or – where this is not possible due to the complexity of the models – to provide other forms of justifiability (e. g. counterfactual explanations), to strengthen the methods that ensure "algorithmic fairness", to require researchers to carry out a gender analysis, etc.

All in all, it is necessary to find a balance between the potential of artificial intelligence of finding new diagnostic methods and the fundamental dialogical component in the work of physicians that places great value on providing reasons and making comprehensible, i. e. "telling a story". Zarsky (2018) speaks of an "obsession with causality" and contrasts these two approaches as follows:

> "Insisting on the development of mechanisms brings the risk of rejecting correlations that current science cannot prove or even explain and yet nonetheless have predictive value. Such a policy choice thus will tilt the balance of expected outcomes in favor of existing knowledge
>
> versus
>
> […] data must be explained through the telling of a story […] to prevent the analyst from mistaking noise for a signal, an error that would delude others as well" (Zarsky 2018, 50).

From an ethical perspective, it can be seen that the use of big data, artificial intelligence, and machine learning opens up a new dimension to the problem of responsibility. Responsibility can be interpreted from an ethical perspective in terms of just three aspects (WHO, WHAT, and TO WHOM / WHAT?) or of six aspects (WHO, WHAT, TO WHOM / WHAT, WHY, FOR WHAT and WHEN?) (Ropohl 1987, 155).

5.3 Social justice – the unavoidability of distributive effects?

Big data runs the risk of failing to systematically collect data on specific marginal groups within society – migrants, persons with low economic status, etc. The insufficient representation of these groups within the data sets is related to the fact that members of these groups visit a doctor less often, search less often (if at all) for health information online, and do not own credit cards or wearable devices that continuously collect data about life habits and health status. Very often these are precisely the people whose situation would require more research and intervention (e. g. Cohen et al. 2017). Furthermore, a gender bias that is historically deeply anchored within biomedicine – despite

all efforts to the contrary – has resulted in insufficient representation in the data sets of the gender-specific aspects of many illnesses.

This bias in the data sets is exacerbated by the fact that a machine learning-based model is not neutral. The development of such a model includes decisions about which problems should be prioritized, which algorithms should be used, and how the model should be used.

> "A machine learning model is architected from the programmers that create it, the algorithm and metrics used, and the data it takes as input. When a development team programs a machine learning model they must choose carefully: what type of algorithm is used, how the algorithm is set up, what metrics and parameters are used, and on what data the algorithm is trained and tested. Creators' influence can show up in unexpected ways" (Shadowen 2017, 9).

Given the scarcity of resources in the healthcare system, such decisions have unavoidable distributive effects. These effects are particularly felt in cases in which the result of an analysis could generate disadvantages for individual patients while advantaging others. Such a decision would be justified by the assessment that it has an overall positive impact on the health of a specific patient population. One example cited in the literature concerns a physician who is faced with deciding whether to treat a patient with a moderate organic dysfunction using resource-intensive medical procedures while the predictive analytics identifies other patients as more highly prioritized for such a treatment. In view of the scarcity of intensive care beds, this is a situation in which one patient who could profit from such treatment might be refused a medical service because an algorithm has identified other people as having conditions that put them at greater risk:

> "A different set of patients are admitted in a world where the analytics help guide the decision-making and they too may benefit, neither benefit nor be harmed, or be harmed. If the predictive analytics approach is to be used and be justified it must be because overall more patients benefit than under the status quo" (Cohen et al. 2014, 463).

Of course, physicians already regularly make decisions concerning scarce resources that could have distributive effects. However, this situation could be systematized or intensified by the possibilities of predictive analytics as well as the "choice architecture" built into the therapy proposals.

Because models are not neutral, lack of diversity among those who develop artificial intelligence applications can also pose a problem. One example is that Apple developed a tool for comprehensive recording of patient health status that – until 2015 – did not include menstruation cycles, "which may not be surprising considering the engineering team is predominately male" (Ferryman 2019, 1).

5.4 New requirements for physicians

The adoption of information infrastructures in healthcare (IIH), strengthened by the availability of "big data", leads to a situation in which healthcare personnel (in some cases also patients) increasingly takes on tasks that could be classified as data work. An increasing number of people use mobile apps that collect health-related data for personal use as well as for professional purposes. Physicians are expected to feed various systems with data that do not only document interactions with individual patients but are also used for various secondary purposes, including big data applications (Pine et al. 2018).

In addition, physicians must be capable of interpreting the results of predictive analytics for their decisions, balancing them against other sources of information, and interpreting them in cooperation with colleagues and members of other healthcare professions and in consultation with patients.

The positive aspects of the availability of predictive analytics can only be capitalized on if healthcare organizations offer physicians the support and the time necessary for reflection and critical analysis. It has also been proposed that new professional groups be created whose primary task it is to support physicians in the analysis and interpretation of health data. Fiske et al. (2019) define the job profile of such "health information counselors" as follows:

> "HICs would have broad knowledge of various kinds of health data and data quality evaluation techniques, as well as analytic skills in statistics and data interpretation. Trained also in interpersonal communication, health management, insurance systems, and medico-legal aspects of data privacy, HICs would know enough about clinical medicine to advise on the relevance of any kind of data for prevention, diagnosis, and treatment." (37)

In addition, the new requirements arising from the use of predictive analytics systems should also be included in medical education and training. For example, Cohen et al. (2017) ask:

> "Is the current state of medical education adequate to make physicians (as well as nurses, hospital administrators, etc.) wise users of predictive analytics? Medical education is already densely packed with a myriad of kinds of learning, but data science has traditionally not been a focus. Would widespread adoption of predictive analytics be met with widespread improvements in data science education in medical school, or would (and should?) it become a specialized set of learning for a subset of physicians with others just told to 'trust the algorithm'?" (451).

6 General principles

Embedding digital technologies into situations that involve direct contact between physician and patient changes the relationships in both directions. On the one hand, digital technologies turn patients into more pro-active and self-reliant partners in the process of preventive care, diagnosis, and treatment. On the other hand, physicians are increasingly assuming the role of intermediary between the technology and the patient. Both these tendencies have direct consequences for medical training and continuing education, just as they do for the equipment needed in medical institutions and the structure of the healthcare system, as well as the relevant legal frameworks. The unique nature of the physician-patient relationship also poses greater demands on the ethical configuration of digital technologies ("ethics by design").

Based on the considerations presented in sections 1 to 5, the Bioethics Commission at the Federal Chancellery is submitting the following general assessment and recommendations to the Austrian Federal Government. The observations on the following pages about the work of physicians apply equally to the nursing and care sectors.

6.1 Improvement of medical care through the use of digital technologies

Modern digital technologies offer the prospect of improving the quality of medical care in accordance with the principle of beneficence. However, it must be considered that the work of physicians within the healthcare system must not just take data into account; it must also be based on knowledge and experience. Moreover, digitalization is not an end in itself; rather, it should be guided by the ambition to improve the physician's possibilities for action in significant ways. Ethically appropriate use of these technologies fundamentally presupposes such an improvement. In other words, the correctness of the diagnostic evaluation, the confidence in the accuracy of the diagnosis, the probability of the success of the recommended therapy, or the success rate of a medical intervention making use of such technologies must fundamentally be at least as good and ideally better than when using conventional technologies and only human actors. It follows from this that, given an achievable improvement in therapeutic success, use of the new technologies is not only permissible but ethically advisable. This is subject to clear proof of such an improvement in therapeutic success on the basis of appropriate medical evidence. Another ethically relevant argument for the adoption of digital technologies and for establishing the necessary conditions for their use is the prospect of achievable medical progress to the benefit of future generations of patients, even in the case that no improvement in therapeutic success can be achieved for the particular individual patient. The use of modern digital technologies is on principle ethically impermissible

if it results in a decrease in the medical quality standard, even if there were gains in efficiency. Digitalization must never compete with or in any way offset the two maxims of non-maleficence "do no harm" and beneficence "do good."

The quality of modern digital technologies is determined by a wide range of factors. In the case of tools based on artificial intelligence and machine learning, these factors include the quality of the analysis and training data, adequate training of the medical personnel, and consideration of the roles of the physician's experience and his / her dialogue with the patient. The emotional wellbeing of the patient and of the physician must always be considered in evaluating the quality of medical care. This wellbeing depends in turn on a number of factors, such as establishing a relationship of trust and successful communication between the physician and patient as well as the time available for consultations and inquiries. These prerequisites for adopting digital technologies must be met by developing an appropriate organizational structure. Potential time savings should be used to strengthen trust and emotional wellbeing and never be simply "rationalized away".

6.2 Distributive effects and bias

Artificial intelligence and machine learning do not automatically lead to increased social justice in the healthcare system. It is increasingly apparent that algorithmic systems can contribute to intensification of bias and discrimination, on the basis of gender, age, ethnicity, or socioeconomic status. Identifying and reflecting on the sources of such bias and counteracting their effects through a bias-sensitive design and use of the system pose challenges as well as opportunities for social change.

6.3 Changes to the physician-patient relationship

The increasing use of technology has changed the way physicians work with machines. While many decision-making processes have shifted in the direction of technology, physicians have to master entirely new types of tasks. Their influence is strong as long as the machine offers support in decision-making while the actual decision about the diagnosis and / or therapy is still left to the physician. Their influence becomes low when the machine generally "acts" autonomously and physicians are at best present in the background, ready to intervene in the event of complications. Between these extremes lies a broad spectrum of different possibilities. When using algorithmic systems, a diagnosis or therapy recommendation calculated by a machine should not lead directly to a de facto decision. Physicians should continue to be obligated to justify why they follow the recommendation of an algorithmic system or deviate from it. Particularly in the case of a high system criticality, i. e. decisions with high risk, organizational safeguards must be put in place.

Outsourcing tasks to machines provides physicians with additional spaces of action. It would be ethically questionable to deploy these spaces exclusively for increasing efficiency in the sense of reducing medical personnel. Rather, these freed up capacities should be used to offer individual patients more attention and care in the personal physician-patient relationship. New digital technologies will therefore ideally lead not to a reduction of the human factor in the physician-patient relationship but actually strengthen this factor and a focus on what matters most.

One prerequisite for actually realizing this potential for improved personal attention and care, strengthening the human factor, is an appropriate design of the healthcare system, including the financial compensation structures in place. The time that is gained in this process can only then be truly invested in the relationship between the physician and the patient, if what is termed "conversational medicine" gets more valorized in conjunction with corresponding financial incentives.

Their university education and the ensuing practical work should provide physicians with the knowledge and skills that are necessary for making decisions concerning the integration of algorithmic systems into their daily patient-focused activities, (see also 15 et seqq.). The use of modern digital technologies increasingly places the physician in an intermediary role between the patient and the technology, where it is important to disclose the uses of technology and explain to the patient the criteria underlying a recommendation made by the machine. It is important here that physicians be knowledgeable about the data quality, the capabilities of algorithmic decision-making, and the informative power of probabilities and other results in order to, on one hand, make their own appropriate assessments of the automated result and, on the other, to provide corresponding explanations to the patient.

The availability of high-quality data is a fundamental prerequisite for medical research and for patient care in line with the capabilities of modern medical science. It is critical to establish the legal, organizational, and technical prerequisites for the legally secure collection, curation, and use of data within the intramural and extramural healthcare systems. It is also critical for data to be shared at the national and international level and be made available for purposes of research as well as clinical practice. In pursuing these goals, it is necessary to ensure compliance with the fundamental, legally mandated level of data protection – especially by means of the most extensive possible synthetization or at least anonymization of data – as well as preservation of ethical standards (see also 17). Sufficient data protection is required in particular at the interfaces between hospitals, research institutions, private practices, and patients. These measures should strengthen the patients' acceptance and trust in data-supported applications in the short and long term.

The use of modern digital technology gives rise to a number of other ethical questions which are not new per se but which take on new quantitative or qualitative dimensions. This concerns, for instance, the handling of incidental findings, which can occur with greater quantitative frequency since machines are capable of processing and analyzing a larger information bandwidth. Physicians are obligated to consider carefully

how to handle incidental findings. Patients must be informed of the impacts of possible incidental findings. The "right not to know" must always also be preserved.

6.4 Consequences regarding responsibility and system design

The ethically justifiable use of modern digital technologies requires that the final medical decisions be explainable and transparent; the associated requirements are increasing with the level of system criticality. The transparency requirement applies to data sources and data quality. This includes the key parameters involved in a decision, such as age, gender, health history, image findings, and the extent to which certain assumptions are based on correlation or causality, possibly combined with counterfactual explanations. Ensuring explainability and transparency is primarily the responsibility of the system designer and must be taken into account in the approval processes for algorithmic systems as well as in the user information provided.

One central measure for ensuring data quality consists of furnishing the data sets used by algorithmic systems with metadata concerning the origin of the data. The ability to trace the context in which data are collected for a specific purpose is recommended in particular when the data may be used for analysis purposes in other contexts. This should make it possible to identify and correct any possible bias in the data set.

The use of artificial intelligence in medicine must give due consideration to the legal rights of all involved – in particular patients and physicians but also the administrators of healthcare institutions. Legal regulations should establish – for instance in the form of a dynamic system – control mechanisms that are adequate to a particular technology and its field of use. This could take various forms, from pure ex post verifications to audits in parallel with the development process to approval and certification requirements.

Hospital administrators and department heads should take full responsibility for the use of autonomous systems that do not merely provide support for the decision of a human being; they must be equally liable under civil law as in the case of decisions taken by human medical personnel. From an ethical perspective, the use of modern digital technologies in direct contact with patients requires the avoidance of a responsibility vacuum. The applicable civil and criminal legislation must be evaluated in this regard to ensure that accountability and liability are appropriately distributed even in the context of autonomous and networked systems. This includes accepting responsibility for technical aids that supplant human deliberation and decision-making processes according to principles similar to those that apply to human assistants (see Section 1313a Austrian Civil Code – ABGB). The introduction of an "electronic person", i. e. the recognition of robots and artificial intelligence as legal persons, must be rejected.

6.5 Consequences for medical education

Enabling physicians to integrate algorithmic systems into daily patient-related activities should be given due consideration in the medical curriculum. Medical education consider the fact that modern digital technologies question the preservation of the competence of the medical profession, insofar as often decades of clinical practice are required in order to become an experienced physician. It is primarily a matter of medical education to ensure that future generations of physicians continue to possess the requisite competence.

Decisions concerning the use and design of modern digital technologies must be sensitive to the preservation of medical competence (in the specific and general sense) in the way of "competence sensitive design"; insofar as these technologies directly impact on which human control functions will be maintained (e. g. the human in the loop principle), on the quality of medical care in exceptional situations (e. g. catastrophes or cyber-attacks), as well as on the capacity of future generations to innovate (e. g. the development of new medical technologies). Preservation of competence in the specific sense must be ensured through appropriate system design. Role swapping must be incorporated into the protocols; in other words, the physician must be regularly called upon to make an initial decision without knowledge of the recommendation of the algorithmic system. The preservation of competence must be safeguarded by intentionally establishing mandatory training modules and capitalizing on the opportunities for continuing medical education afforded by global networking and the availability of case studies and visual materials.

Fundamental knowledge of data law (e. g. data protection law) and data ethics must become a fixed component in the education of all medical professions. It will also be necessary to teach entirely new skills that have previously been absent or only barely included in the education of physicians (and nursing staff). These skills involve a fundamental understanding of the functioning of modern digital technologies, especially the importance of data quality, the risks of bias and discrimination, the informative power of algorithmic results, and the limits of algorithmic systems in general.

6.6 Recommendations

1. Modern digital technologies offer the prospect of improving the quality of medical care in accordance with the principle of beneficence. Whenever these technologies result in improved therapy success, their use is not only permissible but ethically advisable.
2. The emotional wellbeing of the patient and the physician must always be taken into account in evaluating the quality of medical care.
3. General medical progress to the benefit of future patient generations also presents an ethically relevant argument for the use of digital technologies.

4. Artificial intelligence and machine learning do not automatically lead to increased social justice in the healthcare system. Identifying and reflecting on possible sources of bias and discrimination and counteracting their effects through bias-sensitive design and use of the system pose challenges as well as opportunities for social change.
5. The increasing use of technology has changed the way physicians work with machines. When algorithmic systems are used, a diagnosis or therapy recommendation calculated by a machine should not lead on its own to a de facto decision.
6. Outsourcing tasks to machines frees up capacities of the physician. These freed up capacities should be used to enable more human interaction in the personal relationship between physicians and individual patients.
7. Saved time can only truly be invested in the relationship between the physician and the patient if there is increased valorization of "conversational medicine" in conjunction with corresponding financial incentives.
8. Within their university education and ensuing practical work, physicians must learn the necessary knowledge and skills to make decisions concerning the integration of algorithmic systems into their daily patient-focused activities.
9. All legal, organizational, and technical prerequisites must be established for the legally compliant collection, administration, and use of data in the intramural and extramural healthcare systems.
10. The ethically justifiable use of modern digital technologies requires that the final medical decisions be explainable and transparent.
11. The use of artificial intelligence in medicine must give due consideration to the legal rights of all involved – in particular patients and physicians but also the administrators of healthcare institutions.
12. Hospital administrators and department heads should take responsibility for the use of autonomous systems that go beyond merely providing support for the decision of a human being and must be liable under civil law just the same as if human medical personnel had been relied upon. This should guarantee that no responsibility vacuum is created.
13. Enabling physicians to integrate algorithmic systems into daily patient-related activities should be given due consideration in the medical curriculum.
14. Decisions concerning the use and design of modern digital technologies must be sensitive to the preservation of medical competence in the specific and general sense.
15. Fundamental knowledge of data law (e. g. data protection law) and data ethics must become a fixed component in the education of all medical professions.

7 Literature

Monographs and articles

Aichberger S, Thurner S (2016). Disentangling genetic and environmental risk factors for individual diseases from multiplex comorbidity networks. Scientific reports 6, 39658.

Anderson C (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired Magazine 16.07, 2.

Antes G (2016). "Is the age of causality over?". The Journal of Evidence and Quality in Health Care (ZEFQ) 112, 16–22.

Atluri G et al. (2009). Association Analysis Techniques for Bioinformatics Problems. In: Rajasekaran Sanguthevar (Ed.): Bioinformatics and Computational Biology, lecture Notes in Bioinformatics 5462, Springer Berlin Heidelberg, 1–13.

Balint M (1964). The Doctor, His Patient and the Illness; Stuttgart, Klett.

Barsky AJ, Peekna HM, Borus JF (2001). Somatic symptom reporting in women and men. Journal of general internal medicine 16, 266–275.

Bartholomew K, Horowitz LM (1991). Attachment styles among young adults: Test of a four category model. Journal of Personality and Social Psychology 61, 226–244.

Barton N, Hermisson J, Nordberg M (2019). Why structure matters. eLIFE 8, e45380.

Beckman HB, Frankel RM (1984). The Effect of Physician Behavior on the Collection of Data. Annals of Internal Medicine 101 (05), 692–696, https://doi.org/10.7326/0003-4819-101-5-692.

Beggiato A et al. (2017). Gender differences in autism spectrum disorders: divergence among specific core symptoms. Autism Research 10, 680–689.

Berendas BS, Waldstein SM, Schmid-Erfurth U (2018). Screening and Management of Retinal Diseases Using Digital Medicine. The Ophthalmologist 115 (9), 728–736.

Bioethics Commission (2008) Empfehlungen mit Genderbezug für Ethikkommissionen und klinische Studien.

Birch K (2017). Rethinking value in the bio-economy: Finance, assetization, and the management of value. Science, Technology, & Human Values 42/3, 460–490.

Birklbauer A (2019). Die Bedeutung von (medizinischen) Leitlinien im Strafrecht. Journal für Medizin und Gesundheitsrecht 1/2019, 16 et seqq.

Bonham VL, Green ED, Pérez-Stable EJ (2018). Examining how race, ethnicity, and ancestry data are used in biomedical research. JAMA 320 (15), 1533–1534.

Borck C (2016). Medizinphilosophie. Hamburg, Junius Verlag.

Bowlby J (2003). Bindung und menschliche Entwicklung: John Bowlby, Mary Ainsworth und die Grundlagen der Bindungstheorie. Stuttgart, Klett-Cotta.

Calude CS, Longo G (2017). The deluge of spurious correlations in big data. Foundations of science 22 (3), 595–612.

Central Ethics Committee at the German Medical Association (2013). Statement: Ärztliches Handeln zwischen Berufsethos und Ökonomisierung. Das Beispiel der Verträge mit leitenden Klinikärzten und -ärztinnen. Deutsches Ärzteblatt 110 (38).

Caruana R et al. (2015). Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730.

Chin Yee B, Upshur RE (2018). Clinical judgement in the era of big data and predictive analytics. Journal of Evaluation in Clinical Practice, 635–637, https://doi.org/10.1111/jep.12852.

Citron DK, Pasquale FA (2014). The Scored Society: Due Process for Automated Predictions. Washington Law Review 89, 1–32, https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2/.

Cohen IG et al. (2014). The legal and ethical concerns that arise from using complex predictive analytics in health care. Health affairs 33/7, 1139–1147.

Cohen IG, Graver HS (2017). Cops, docs, and code: A dialogue between big data in health care and predictive policing. UC Davis Law Review 51, 437–474.

Comfort N (2018). Genetic determinism redux. Nature 561, 461–463.

Cooper RS, Nadkami GN, Ogedegbe G (2018). Race, ancestry, and reporting in medical journals. JAMA 320 (15), 1531–1532.

Dang BN et al. (2017). Building trust and rapport early in the new doctor-patient-relationship: a longitudinal qualitative study. BMC Med Educ 32,https://doi.org/10.1186/s12909-017-0868-5.

Deo RC (2015). Machine learning in medicine. Circulation 132, 1920–1930.

Dörn S (2016). Programmieren für Ingenieure und Naturwissenschaftler: Grundlagen. Berlin, Vieweg.

Editorial (2017). Clinical decision making: more than just an algorithm. The Lancet Oncology 18/12, 1553.

European Parliament (2017). Resolution with recommendations to the Commission on Civil Law Rules on Robotics. Recommendation No. 59 seq.

European Parliament (2020). Draft Report with recommendations to the Commission on a Civil liability regime for artificial intelligence. 2020/2014(INL), https://www.europarl.europa.eu/doceo/document/JURI-PR-650556_EN.pdf.

Epstein D, ProPublica (2017). When evidence says no, but doctors say yes. Atlantic (1).

Esteva A et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature 542 (7639), 115–118.

European Society of Radiology (ESR) (2019). What the radiologist should know about artificial intelligence – an ESR white paper. European Society of Radiology Insights into Imaging, https://doi.org/10.1186/s13244-019-0738-2.

Expert Group on Liability and New Technologies (2019). Report on liability for Artificial Intelligence and other Emerging Digital Technologies. European Union.

Fagerhaugh SY et al. (1987). Hazards in hospital care: ensuring patient safety. San Francisco, Jossey-Bass Publishers.

Ferryman K, Pitcan M (2018). Fairness in precision medicine. Data & Society.

Fischer F (2017). Ethische Aspekte von E-Health aus der Perspektive von Public Health. In: Müller-Mielitz S, Lux T (eds) E-Health-Ökonomie, 141–151, https://doi.org/10.1007/978-3-658-10788-8_9.

Fiske A, Buyx A, Prainsack B (2019). Health Information Counselors: A New Profession for the Age of Big Data. Academic Medicine 94/1, 37–41.

FitzGerald C, Hurst S (2017). Implicit bias in healthcare professionals: a systematic review. BMC medical ethics 18 (1), 19.

Forbes Insights Team (2019). AI And Healthcare: A Giant Opportunity. 11 Feb 19.

Foucault M (1973). The Birth of the Clinic. Munich, Carl Hanser Verlag.

Gallagher MD, Chen-Plotkin AS (2018). The Post-GWAS Era: From Association to Function. The American Journal of Human Genetics 102, 717–730.

Gandomi A, Haider M (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35.2, 137–144.

Gerendas BS, Waldstein SM, Schmidt-Erfurth U (2018). Screening and Management of Retinal Diseases Using Digital Medicine. The Ophthalmologist 115 (9), 728–736.

German Federal Government's Data Ethics Commission (2019). Opinion of the Data Ethics Commission Berlin, without publisher

Gianfrancesco MA et al. (2019). Potential biases in machine learning algorithms using electronic health record data. JAMA internal medicine 178 (11), 1544–1547.

Greenwald AG, Hamilton Krieger L (2006). Implicit bias: Scientific foundations. California Law Review 94/4, 945–967.

Gründinger W et al. (2019). Mensch Moral Maschine – Digitale Ethik, Algorithmen und künstliche Intelligenz. Berlin, German Association for the Digital Economy (BVDW), 12.

Gundling F et al. (2019). Defizite in der Gesundheitskompetenz stationär behandelter Patienten – eine Querschnittstudie. German Medical Weekly (DMW) 144 (04), 21–29.

Hammarlund N (2018). Racial Treatment Disparities after Machine Learning Surgical-Appropriateness Adjustment. SSRN Electronic Journal 3057607.

Hecker D et al. (2017). Künstliche Intelligenz und die Potentiale des maschinellen Lernens für die Industrie. Wirtschaftsinformatik & Management 9 (5), 26–35.

Hehner S, Biesdorf S, Möller M (2018). Digitizing healthcare – opportunities for Germany. Digital McKinsey.

Hehner S, Liese K et al. (2018). Digitalisierung in deutschen Krankenhäusern – Eine Chance mit Milliardenpotenzial für das Gesundheitssystem. McKinsey & Company.

Hipp R, Schlude C, Göller N (2017). Healthcare of the Future – The digital revolution of the healthcare sector. Porsche Consulting.

Hipp R et al. (2018). Patient im Fokus – Investitionschancen in das digitale Gesundheitswesen. Porsche Consulting.

Hoeyer K (2016). Denmark at a crossroad? Intensified data sourcing in a research radical country. The ethics of Biomedical Big Data, Springer, Cham, 73–93.

Horton RH, Lucassen AM (2019). Recent developments in genetic / genomic medicine. Clinical Science 133 (5), 697–708.

Jackson G (2019). The female problem: how male bias in medical trials ruined women's health. The Guardian 13 Nov 2019, https://www.theguardian.com/lifeandstyle/2019/nov/13/the-female-problem-male-bias-in-medical-trials.

Jannes M et al. (2018). Algorithmen in der digitalen Gesundheitsversorgung – Eine interdisziplinäre Analyse; Gütersloh, on behalf of the Bertelsmann Foundation.

Kaltenbach T, Erharter M et al. (2016). Digital & Disrupted – All change for healthcare. Roland Berger GmbH.

Kautzky-Willer A, Thurner S, Klimek P (2017). Use of statins offsets insulin-related cancer risk. Journal of internal medicine 281 (2), 206–216.

Kleinberg S, Hripcsak G (2011). A review of causal inference for biomedical informatics. Journal of Biomedical Informatics 44(6), 1102–1112.

Klimek P et al. (2015). Quantification of diabetes comorbidity risks across life using nation-wide big claims data. PLOS Computational Biology 11 (4), e1004125.

Klimek P, Aichberger S, Thurner S (2016). Disentangling genetic and environmental risk factors for individual diseases from multiplex comorbidity networks. Scientific reports 6, 39658.

Kopetzki C (2008). "Off-label-use" von Arzneimitteln. In: Ennöckl D, Raschauer N, Schulev-Steindl E, Wessely W (Publisher), Über Struktur und Vielfalt im Öffentlichen Recht. Vienna / New York, Springer, 73 et seqq.

Krismer F et al. (2019). Morphometric MRI profiles of multiple system atrophy variants and implications for differential diagnosis. Movement Disorders 34(7), 1041–1048, https://doi.org/10.1002/mds.27669.

Laney D (2001). 3D data management: Controlling data volume, velocity and variety. META Group.

Leonelli S (2016). Data-centric biology: A philosophical study. University of Chicago Press.

Looker AC et al. (1997). Prevalence of low femoral bone density in older U.S. adults from NHANES III. Journal of Bone and Mineral Research 12 (11), 1761–1768, https://doi.org/10.1359/jbmr.1997.12.11.1761.

Lown B, Rodriguez D (2012). Commentary: Lost in translation? How electronic health records structure communication, relationships, and meaning. Academic Medicine 87 (4), 392–394.

Mayrhofer (2014). Off-label-use von Analgetika in der perioperativen Kinderschmerztherapie aus rechtlicher Sicht. Der Schmerz, 65–66.

Manyika J et al. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Mitchell A, Vaze A, Rao S (2009). Clinical diagnosis of depression in primary care: A meta analysis. The Lancet 374 (9690), 609–619.

Müller-Mielitz S (2017). E-Health-Ökonomie – Begriff und Abgrenzung. in: Müller-Mielitz S, Lux T, E-Health-Ökonomie, Wiesbaden, Springer Gabler, 35–49.

Nanji KC et al. (2018). Medication-related clinical decision support alert overrides in inpatients. Journal of the American Medical Informatics Association 25 (5), 476–481.

National Research Council (2011). Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease. Washington (DC), National Academies Press.

Need AC, Goldstein DB (2009). Next generation disparities in human genomics: concerns and remedies. Trends in Genetics 25 (11), 489–494.

Obermeyer Z, Emanuel EJ (2016). Predicting the future – big data, machine learning, and clinical medicine. The New England journal of medicine 375 (13), 1216.

Obermeyer Z, Lee TH (2017). Lost in Thought – The Limits of the Human Mind and the Future of Medicine. New England Journal of Medicine 377 (13), 1209–1211, https://doi.org/10.1056/NEJMp1705348.

Open Letter to the European Commission, Artificial Intelligence and Robotics (2018), http://www.robotics-openletter.eu/.

Osoba O, William W IV (2017). An intelligence in our image: The risks of bias and errors in artificial intelligence. Santa Monica (CA), RAND Corporation.

Parasuraman R, Sheridan TB, Wickens CD (2000). A model of types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part A 30 (3), 286–297.

Parikh RB, Obermeyer Z, Navathe AS (2019). Regulation of predictive analytics in medicine. Science 363 (6429), 810–812.

Pearl J (2019). The Seven Tools of Causal Inference, with Reflections on Machine Learning. Communications of the ACM 62 (3), 54–60.

Pine K et al. (2018). Data Work in Healthcare: Challenges for Patients, Clinicians and Administrators. Conference Paper, 433–439, https://doi.org/10.1145/3272973.3273017.

Pot M, Spahl W, Prainsack B (2019). The Gender of Biomedical Data: Challenges for Personalised and Precision Medicine. Soma 9:2–3, 170–187.

Prainsack B (2015). Is personalized medicine different? (Reinscription: the sequel) A response to Troy Duster. The British journal of sociology 66 (1), 28–35.

Prainsack B (2019). Precision Medicine Needs a Cure for Inequality. Current History 118 (804), 11–16.

Rajkomar A et al. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine 1, 18–28.

Rieder G, Simon J (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. Big Data & Society Vol 3, issue 1, https://doi.org/10.1177/2053951716649398.

Rimscha M (2014). Algorithmen kompakt und verständlich: Lösungsstrategien am Computer. Wiesbaden, Springer Vieweg.

Ropohl G (1987). Neue Wege, die Technik zu verantworten. In: Lenk H, Ropohl G: Technik und Ethik, Stuttgart, Reclam, 149–176.

Rosenberg NA et al. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. Evolution, Medicine, and Public Health (1), 26–34, https://doi.org/10.1093/emph/eoy036.

Russell C et al. (2017). When Worlds Collide: Integrating Different Counterfactual Assumptions in Fairness. Advances in Neural Information Processing Systems 30 (NIPS).

Samulowitz A et al. (2018). "Brave men" and "emotional women": a theory-guided literature review on gender bias in health care and gendered norms towards patients with chronic pain. Pain Research and Management, 14.

Sauter SK et al. (2014). Analyzing Healthcare provider Centric networks through secondary use of health chaines data. IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 522–525.

Scherfler C (2018). Automatisierte Magnetresonanztomographie Analyse bei Parkinson-Syndromen. Newsletter of the Austrian Parkinson Society (1).

Scherfler C et al. (2016). Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism. Neurology 86 (13), 1242–1249.

Sennaar K (2019). Machine Learning for Medical Diagnostics – 4 Current Applications. Emerj Artificial Intelligence Research.

Shadowen AN (2017). Ethics and Bias in Machine Learning: A Technical Study of What Makes Us "Good". New York, John Jay College of Criminal Justice, Student Thesis.

Strauss A et al. (1985). Social Organization of Medical Work. Chicago, The University of Chicago Press.

Street RL Jr, Gordon H, Haidet P (2007). Physicians' communication and perceptions of patients: Is it how they look, how they talk, or is it just the doctor?. Soc Sci Med 65, 586–98.

Street RL Jr et al. (2009). How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient education and counseling 74 (3), 295–301.

Street RL Jr, Haidet P (2011). How well do doctors know their patients? Factors affecting physician understanding of patients' health beliefs. J Gen Intern Med (1), 21–27.

Sturmberg JP (2018). Embracing complexity in health and health care – Translating a way of thinking into a way of acting. Journal of Evaluation in Clinical Practice 24 (3), 598–599.

Sundermann EE et al. (2017). Does the female advantage in verbal memory contribute to underestimating Alzheimer's disease pathology in women versus men?. Journal of Alzheimer's Disease 56 (3), 947–957.

Taylor K (2015). Connected Health: How digital technology is transforming health and social care. Deloitte Centre for Health Solutions.

Thaler RH, Sunstein CR (2008). Nudge: improving decisions about health, wealth, and happiness. New Haven & London, Yale University Press.

Topol E (2019), Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, Basic Books.

Torkamani A, Weininger NE, Topol EJ (2018). The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics 19 (9), 581–590.

Tschandl P, Rosendahl C, Kittler H (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data (5), 180161.

Tschandl P et al. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. The Lancet Oncology 20 (7), 938–947.

Tversky A, Kahneman D (1974). Judgment under uncertainty: Heuristics and biases. Science (185) 4157, 1124–1131.

Ubel PA, Scherr KA, Fagerlin A (2017). Empowerment failure: How shortcomings in physician communication unwittingly undermine patient autonomy. The American Journal of Bioethics 17 (11), 31–39.

Überall M, Werner-Felmayer G (2019). Integrative Biology and Big-Data-Centrism: Mapping out a Bioscience Ethics Perspective with a S.W.O.T. Matrix. OMICS: A Journal of Integrative Biology 23(8), 371–379.

Vayena E et al. (2018). Policy implications of big data in the health sector. Bulletin of the World Health Organization 96 (1), 66–68.

Vellido A (2019). Societal issues concerning the application of artificial intelligence in medicine. Kidney Diseases 5 (1), 27–33.

Vellido A et al. (2018). Machine learning in critical care: state-of-the-art and a sepsis case study. BioMed Eng OnLine 17(Suppl 1), 135, https://doi.org/10.1186/s12938-018-0569-2.

Ventres W et al. (2006). Physicians, patients, and the electronic health record: an ethnographic analysis. The Annals of Family Medicine 4 (2), 124–131.

Vogl WD et al. (2019). Automatic segmentation and classification of breast lesions through identification of informative multiparametric PET / MRI features. European Radiology Experimental 3 (1), 18.

Wachter S, Mittelstadt B, Russell C (2018). Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harvard Journal of Law & Technology 31 (2), 841–887.

Wickens CD et al. (2010). Stages and levels of automation: An integrated meta-analysis. Human Factors and Ergonomics Society Annual Meeting Proceedings 54 (4), 389–393.

Wilm S et al. (2004). At Which Point Does the General Practitioner Interrupt his Patients at the Beginning of a Consultation?. Z Allg Med 80, 53–57. http://doi.org/10.1055/s-2004-44933.

Wirtz V, Cribb A, Barber N (2006). Patient-doctor decision-making about treatment within the consultation – A critical analysis of models. Social science & medicine 62.1, 116–124.

Wooden B et al. (2017). Using big data to discover diagnostics and therapeutics for gastrointestinal and liver diseases. Gastroenterology 152 (1), 53–67.

Yudell M et al. (2016). Taking race out of human genetics. Science 351, 564–565.

Zarsky T (2018). Correlation versus Causation in Health-Related Big Data Analysis. In Cohen I, Lynch H, Vayena E, Gasser U (Eds.), Big Data, Health Law, and Bioethics, Cambridge, Cambidge Univeristy Press 42–55.

Zimmermann-Rittereiser M, Schaper H (2016). Big Data – An Efficiency Boost in the Health-care Sector. In: Big Data in Medical Science and Healthcare Management – Diagnosis, Therapy, Side Effects (11), 131–138.

Websites

Gartner Information Technology (IT) Glossary: Big Data – www.gartner.com/en/information-technology/glossary/big-data

Price L. The Digital Health Hype Cycle – www.healthcare.digital/single-post/2019/01/12/The-Digital-Health-Hype-Cycle-2019

Abbreviations

AAL	Ambient Assisted Living
ABGB	Austrian Civil Code
ADS	Attention deficit syndrome
ÄrzteG	Austrian Physicians Act [Ärztegesetz]
CDL-AM	Christian Doppler Laboratory for Applied Metabolomics
DSG	Austrian Data Protection Act [Datenschutzgesetz]
GDPR	General Data Protection Regulation
E-	Electronic
E-Health	Electronic Health
EC	European Community
EGE	European Group on Ethics in Science and New Technologies
EHR	Electronic health record
EKG	Electrocardiogram
ELGA	Electronic Health Record system
ESR	European Society of Radiology
et al.	and others
EU	European Union
EEC	European Economic Community
FFG	Austrian Research Promotion Agency [Forschungsförderungsgesellschaft]
FOG	Austrian Research Organization Act [Forschungsorganisationsgesetz]
GTelG	Austrian Health Telematics Act [Gesundheitstelematikgesetz]
GWAS	Genome-wide association studies
HIC	Health Information Counselors
IIH	Information Infrastructures in Healthcare
ICT	Information and communication technologies
IT	Information Technology
AI	Artificial Intelligence
HIS	Hospital Information System
MDR	EU Medical Device Regulation
MPG	Austrian Medical Devices Act [Medizinproduktegesetz]
MPVO	Austrian Medical Device Regulation [Freie Medizinprodukteverordnung]
NEC	National Ethics Councils
NIST	National Institute of Standards and Technology
No.	Number
OCT	Optical coherence tomography
PET / MRI	Positron emission tomography / magnetic resonance imaging
PHG	Austrian Product Liability Act [Produkthaftungsgesetz]
PSG	Austrian Product Safety Act [Produktsicherheitsgesetz]

RFID	Radio Frequency Identification
SNPs	Single Nucleotide Polymorphisms
StGB	Austrian Penal Code [Strafgesetzbuch]
UNESCO	United Nations Educational, Scientific and Cultural Organization
VbVG	Austrian Corporate Criminal Liability Act [Verbandsverantwortlichkeitsgesetz]

Members of the Austrian Bioethics Commission for the 2017–2020 term

Chair
Dr. Christiane Druml

First Vice Chair
Univ. Prof. Mag. Dr. Markus Hengstschläger

Second Vice Chair
Univ.-Prof. Dr. h.c. Dr. Peter Kampits

Univ.-Prof. DDr. Matthias Beck
Univ.-Prof. Dr. Alois Birklbauer
Dr. Andrea Bronner
Univ.-Prof. Dr. Christian Egarter
Dr. Thomas Frühwald
Dr. Ludwig Kaspar
Univ.-Prof. Dr. Lukas Kenner
Dr. Maria Kletecka-Pulker
Univ.-Prof. Dr. Ursula Köller MPH
Univ.-Prof. Mag. Dr. Michael Mayrhofer
Univ.-Prof. Dr. Johannes Gobertus Meran MA
Dr. Stephanie Merckens
Univ.-Prof. Dr. Siegfried Meryn
Univ.-Prof. Dr. Christina Peters
Univ.-Prof. Mag. Dr. Barbara Prainsack
Univ.-Prof. DDr. Walter Schaupp
Univ.-Prof. Dr. Andreas Valentin MBA
Dr. Klaus Voget
Univ.-Prof. Dr. Ina Wagner
Priv.-Doz. Dr. Jürgen Wallner MBA
Univ.-Prof. Dr. Christiane Wendehorst LL.M
Univ.-Prof. Dr. Gabriele Werner-Felmayer