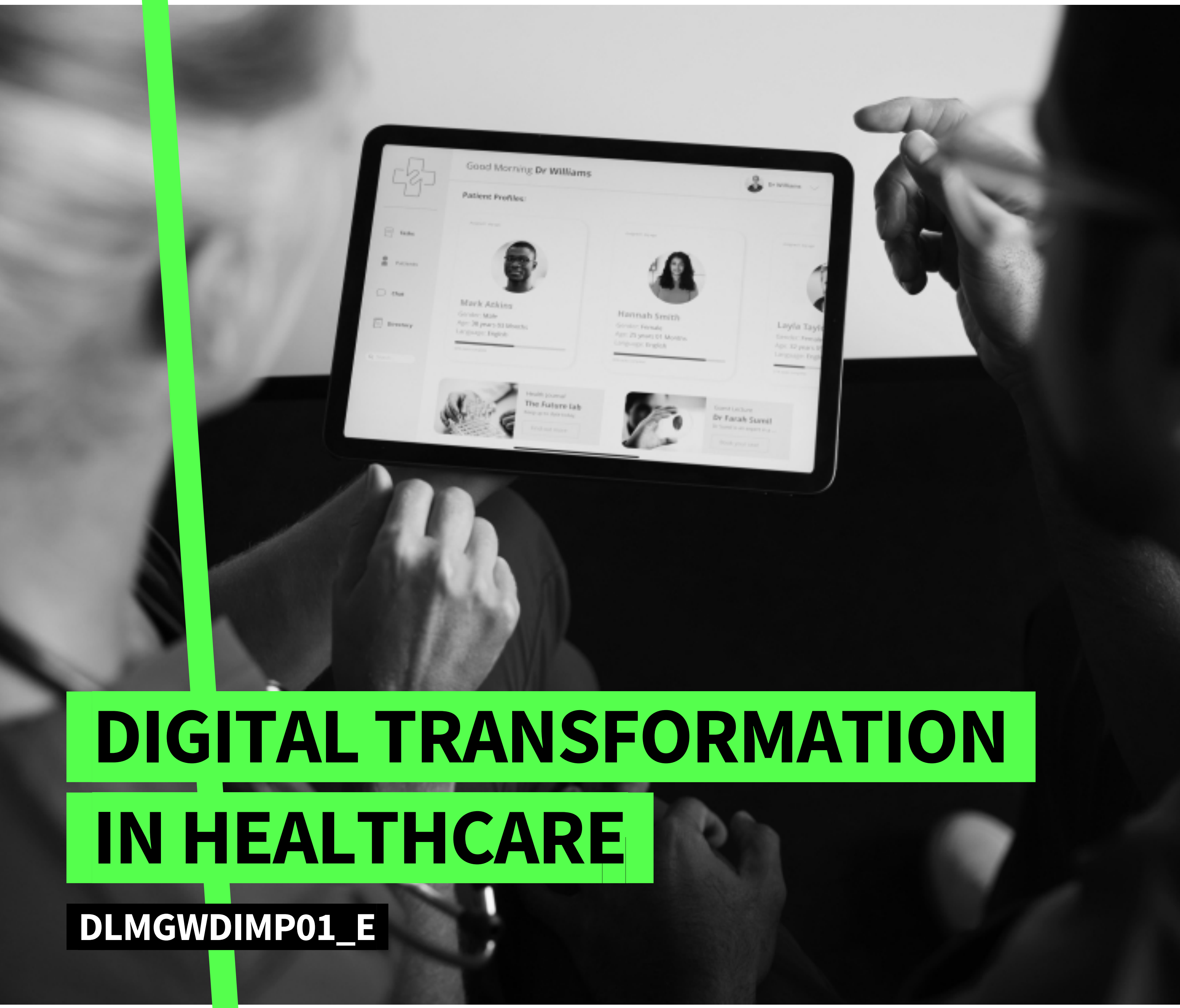
Course Book



DIGITAL TRANSFORMATION IN HEALTHCARE

DLMGWDIMP01_E

INTERNATIONAL
UNIVERSITY OF
APPLIED SCIENCES

LEARNING OBJECTIVES

The course **Digitalization in Healthcare** offers a comprehensive understanding of digital transformation in the healthcare industry, emphasizing advanced digital technologies. The course book starts by addressing the current challenges faced by healthcare systems and how digital transformation practices can be leveraged to tackle these issues. The second unit delves into the new concept of digital health, which represents a cultural paradigm shift in traditional healthcare. By exploring the history and implications of digital health, we gain a better understanding of the future of healthcare. The third unit focuses on the advanced technologies that accelerate digital transformation, including artificial intelligence, blockchain, and quantum technologies. It is crucial to understand how these technologies work in order to apply them in healthcare and assess their potential for the field. The fourth unit discusses the ethical foundations and practical frameworks that can be applied in the context of digital health technologies. Finally, the last unit considers the risks and dangers associated with digital health and possible solutions to mitigate them. The course Digitalization in Healthcare offers a holistic approach to digital transformation in healthcare by providing interdisciplinary perspectives, including economic, societal, and technological views. It also emphasizes critical and ethical thinking, which is essential in shaping the digital transformation process for better healthcare.

UNIT 1

FOUNDATIONS OF DIGITAL TRANSFORMATION IN HEALTHCARE

STUDY GOALS

On completion of this unit, you will be able to ...

- understand the current challenges of healthcare systems.
- explain the difference between digitization, digitalization, and digital transformation.
- understand the potential of digital technologies for healthcare.

1. FOUNDATIONS OF DIGITAL TRANSFORMATION IN HEALTHCARE

Introduction

This unit presents an overview of the basics of digitalization and digital transformation in healthcare. Healthcare systems in advanced economies are facing significant challenges due to changing demographics, a rise in noncommunicable diseases, and increasing costs. Many scholars and practitioners consider digitalization and digital transformation as promising solutions to these issues. The unit begins by discussing the current difficulties faced by healthcare systems. It then delves into the fundamental concepts of digitalization and digital transformation. Finally, it explores the potential and promise of digital progress in healthcare.

1.1 Current Challenges of Healthcare Systems

The 21st century has brought about significant improvements in nutrition, hygiene, living conditions, and healthcare around the globe. However, progress in the availability and quality of healthcare services has been uneven, resulting in distinctive challenges faced by individual healthcare systems worldwide. According to a joint report from the World Bank and World Health Organization (WHO) in 2017, half of the world's population does not have access to essential healthcare, with major deficits in sub-Saharan Africa and South Asia (World Health Organization, 2017). Healthcare systems in advanced economies face different challenges, including upward pressure on health expenditure due to medical progress in healthcare, rising incomes that result in higher expectations of health and well-being, and the increasing needs of aging populations (Fleisch et al., 2021). In the following, we will examine the current challenges of healthcare systems, focusing on advanced economies.

Medical Progress

Thanks to better hygiene, nutrition, and healthcare, life expectancy has more than doubled over the past century. As a world average, life expectancy at birth was about 32 years in 1900 and about 73 years in 2022. Though life expectancy varies greatly across the world, with the lowest life expectancy of 54 years in the Central African Republic and the highest of 84 years in Japan in 2022, people today are living longer and have better health overall than they did in the past (World Health Organization, 2022a).

The primary reason for the increase in life expectancy is medical progress and improvements in healthcare. Until a few centuries ago, infectious diseases were the most common cause of human mortality. Without modern treatment possibilities, becoming infected often meant a death sentence, as reflected in high mortality rates beginning in infancy (Fleisch et al., 2021). Improvements in healthcare significantly decreased child mortality, a core indicator for early-life health and well-being. The number of under-five deaths worldwide decreased from 12.6 million in 1990 to five million in 2020. This is equivalent to one in 11 children dying before the age of five in 1990, compared to one in 27 in 2020. However, the decrease in child mortality is very uneven globally. In 2020, sub-Saharan Africa and Southern Asia accounted for over 80 percent of the five million deaths of children under five years old. Half of all under-five deaths in 2020 occurred in only five countries: Nigeria, India, Pakistan, the Democratic Republic of the Congo, and Ethiopia (World Health Organization, 2022b).

To a significant extent, two medical innovations are responsible for the increase in human life expectancy: antibiotics and vaccines (Fleisch et al., 2021). Although it was not the first antibiotic, the 1928 discovery of penicillin by Scottish physician Alexander Fleming ushered in the golden era of antibiotic development that peaked in the mid-1950s. For over 100 years, antibiotics have significantly improved modern medicine and saved hundreds of millions of lives. For his achievement, Fleming received the Nobel Prize in Physiology or Medicine in 1945, along with pathologist Howard Walter Florey and biochemist Ernst Boris Chain, who devised methods for the large-scale isolation and production of penicillin (Tan & Tatsumura, 2015).

Alongside antibiotics, the development of vaccines contributed significantly to an increase in human life expectancy. English country physician Edward Jenner created the world's first vaccine, the smallpox vaccine, in 1798. At the time, smallpox was a highly dangerous and deadly infection that, on average, killed three out of every ten people affected by it (Fleisch et al., 2021). Edward Jenner treated milkmaids and observed that they became immune to smallpox after becoming infected with cowpox, a milder form of the poxvirus. He created a technique to inoculate healthy individuals, including his own son, with cowpox blister secretions, resulting in a mild infection. Afterward, Jenner exposed the same patients to the harmful smallpox virus and observed that they did not become ill, showing the effectiveness of the cowpox inoculation in protecting against smallpox. As the Latin word for cow is "vacca," and cowpox is "vaccinia," Jenner called his new method "vaccination." Jenner's discovery, which brought him recognition as the father of vaccinology, marked the birth of mass immunization that has since saved millions of lives (Riedel, 2005).

In addition to vaccines and antibiotics, in the last 100 years, innovative diagnostic and treatment methods have been developed. The list of medical advancements of the past century is long and significant. Blood tests and imaging methods, such as computer tomography, ultrasounds, and X-rays, play a crucial role in diagnosing diseases. Groundbreaking drug therapies have extended the lives of human immunodeficiency virus (HIV) and cancer patients. Minimally invasive and robotic techniques have transformed surgery. Although medical progress has played a crucial role in increasing life expectancy, its consequences are currently challenging healthcare systems (Fleisch et al., 2021).

Rising Costs in Healthcare

Medical progress offers new treatment opportunities but also has financial consequences. In the past, being diagnosed with kidney failure usually meant a death sentence. Today, such patients can live for many years with the help of dialysis and kidney transplants. However, such treatments entail additional costs borne by health systems. In Germany, for example, dialysis treatments cost more than 44,374 euros per patient per year (Gandjour et al., 2020). This example highlights a prevalent trend in which hundreds of thousands of euros are being expended on treatments that were previously unavailable. This poses a significant challenge in the modern healthcare ecosystem (Fleisch et al., 2021).

In countries that lack mandatory universal health insurance, such as the United States, several African countries, and some Asian countries, falling ill often results in the loss of wealth or even financial bankruptcy. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as COVID-19) pandemic has caused a global economic crisis, making it even more challenging for people to afford healthcare. But even before the pandemic, more than half a billion people were being pushed into poverty by paying for healthcare services out of their own pockets (The World Bank Group, 2021a).

Most Organisation for Economic Co-operation and Development (OECD) countries provide their citizens with universal or nearly universal health coverage (UHC) for a basic set of healthcare services. However, advanced economies face several challenges in sustaining and improving such universal systems. Health expenditure has exceeded economic growth in all OECD countries over the past two decades, with public budgets accounting for approximately three-quarters of healthcare spending (Fleisch et al., 2021). The increasing healthcare costs in developed countries are a result of various factors, including aging populations resulting in a shortage of healthcare workers. This shortage of healthcare professionals results in longer wait times for appointments and medical procedures, which can have negative impacts on health outcomes and lead to increased costs associated with treating those outcomes. To address this challenge, healthcare systems have to invest more resources to attract and retain healthcare workers, contributing to the rise in healthcare costs.

The latest OECD estimate forecasts that total health expenditures will surpass 11 percent of GDP by 2040, up from nine percent in 2017. In order to fund this spending increase, OECD countries would need to invest 19 percent of their public budgets in healthcare by 2040 (OECD, 2019). In addition to this burgeoning health spending, the COVID-19 pandemic revealed a lack of resilience in the healthcare systems of many countries and raised concerns about the long-term fiscal sustainability of these systems. In 2015, experts predicted that without reforms, healthcare expenditures in developed economies will become unaffordable by mid-century due to their current growth rate (OECD, 2015).

Aging Population

Medical progress has had another effect on our social security and healthcare systems: It has led to significant demographic change, which is playing a significant role in driving up healthcare costs. With the increase in life expectancy and the declining birth rate in developed countries, populations are increasingly aging. In most G20 countries, the number of people over 65 for each working-age person will at least double by 2060, while the proportion of people over 80 in the world's population will triple (OECD, 2021).

Older people are usually less healthy than the younger population, more fragile and at risk of developing chronic diseases. Growing expenditures on healthcare and long-term care will increase the pressure on public budgets already strained by rising pension costs. Additionally, many healthcare systems rely significantly on payroll taxes, which will decline as their populations age. It will be a crucial question in the 21st century how to meet the challenges of aging populations and ensure the financial sustainability of social and healthcare systems (Fleisch et al., 2021).

Noncommunicable Diseases (NCDs)

Originally, Western healthcare systems were primarily designed to handle acute, infectious diseases (Fleisch et al., 2021). In the 19th century, medicine largely focused on fighting infectious diseases, but a major new health challenge appeared in the 21st century: noncommunicable diseases (NCDs). Along with mental illness, **the four major NCDs** – cardiovascular diseases (heart disease and stroke), cancer, diabetes, and chronic respiratory diseases – are collectively responsible for almost 74 percent of all deaths worldwide each year, accounting for 41 million people in total (World Health Organization, 2022c).

The four major NCDs
Cardiovascular diseases (heart disease and stroke), cancer, diabetes, and chronic respiratory diseases account for a large proportion of deaths and ill health worldwide. The broader scope of NCDs also includes liver and kidney diseases, as well as mental health issues (World Health Organization 2022c).

The growth of NCDs is driven by rapid unplanned urbanization, air pollution, globalization of unhealthy lifestyles, and population aging. According to the World Health Organization, there are four major risk factors for the development of NCDs: tobacco use, physical inactivity, the harmful use of alcohol, and unhealthy diet. Poverty is also closely linked with NCDs. Almost three-quarters of all NCD deaths occur in low- and middle-income countries. According to the WHO, vulnerable and socially disadvantaged people are at greater risk of being exposed to harmful consumer goods, such as tobacco, or unhealthy dietary practices. This is compounded by their limited access to healthcare services (World Health Organization, 2022c).

The rapid upsurge in NCDs has just recently been recognized and publicly acknowledged. The NCD epidemic has placed healthcare systems under enormous cost pressure and pushed people in countries without universal insurance coverage to financial ruin. Because most NCDs are preventable – an approach that is vastly more cost-effective than treating such conditions after they've already developed – programs aimed at addressing the root causes of these maladies represent major social and political imperatives for the 21st century (World Health Organization, 2022c).

1.2 Digitization, Digitalization, and Digital Transformation

~~As discussed above,~~ medical progress can be both a blessing and a curse. While saving lives and increasing longevity, it simultaneously pushes up healthcare costs, thus challenging healthcare systems. Obviously, it is a question of how to manage medical progress efficiently rather than a question of having it at all. Many scholars and practitioners are searching for an answer to the essential normative question: How should we approach the seemingly counterproductive goals of promoting longevity while limiting increases in health-related costs? While this issue is still a matter of contention, the consensus has been reached that maintaining today's standard of healthcare and funding future medical advances will be difficult without major reforms. Digital transformation is considered by many to be a promising new direction that can solve many of these current challenges (Fleisch et al., 2021). In the next unit, we will learn about the fundamental concepts of digitization, digitalization, and digital transformation and investigate the potential of these developments in healthcare.

Definition of Terms

In the digital age, several new terms have emerged. Digitization, digitalization, and digital transformation are related terms that are often used interchangeably. The reason for this somewhat confusing variety is that these terms were coined by business professionals and only later studied by academics. Though the three terms are still widely used as synonyms in practice, there is a general agreement about their distinct meanings. In the following, we define these terms for the purpose of this course book and clarify the relationship between them.

Digitization

The definition of "digitization" is the clearest and most straightforward. The *Oxford English Dictionary* defines digitization as "the action or process of digitizing; the conversion of analogue data (esp. in later use images, video, and text) into digital form" (Oxford English Dictionary, 2022a). At the core of digitization is an electronic conversion process that transforms information from an analog format to a digital one. The conversion applies a system of binary digits comprising ones and zeros, which serves as the symbolic language for handling data and processing information in digital devices, such as computers and smartphones. An example of digitization is converting a measurement from a manual or mechanical reading into a digital form for further data processing.

Digitalization

While there is usually less confusion about the term digitization, use of the term "digitalization" has evolved and varied greatly over time. For example, the *Oxford English Dictionary* offers a rather generic definition of digitalization as "the adoption or increase in use of digital or computer technology by an organization, industry, country, etc" (Oxford English Dictionary, 2022b). The Gartner IT Glossary defines digitalization as "the use of digital technologies to change a business model and provide new revenue and value producing opportunities; it is the process of moving to a digital business" (Gartner, 2022). Gartner's definition is an example of the broader understanding of the term that nevertheless remains vague about the phrase "digital business." In practice, the terms "digitalization" and "automation" are often used synonymously. In Germany, a survey showed that large companies associate digitalization primarily with the automation of operational business processes. Smaller companies have more of a down-to-earth understanding of digitalization as a means of supporting operational business processes (Bitkom, 2018). These few examples illustrate how differently the term "digitalization" can be interpreted.

As it is still a matter of ongoing academic discourse to systematize all available definitions of the term "digitalization," it is important to agree on a shared meaning when using the word. In this course book, we follow the most common understanding in practice and define digitalization as the use of digital technologies and digitized data to improve business processes and workflows. According to this definition, digitalization improves existing business processes but does not substantially change or transform them. An example of digitalization is using cloud computing to store and distribute documents or creating automated workflows, such as automatic appointment management. The term often implies the expectation that existing business processes will be improved through an increase in productivity and efficiency (Lang, 2021).

Digital transformation

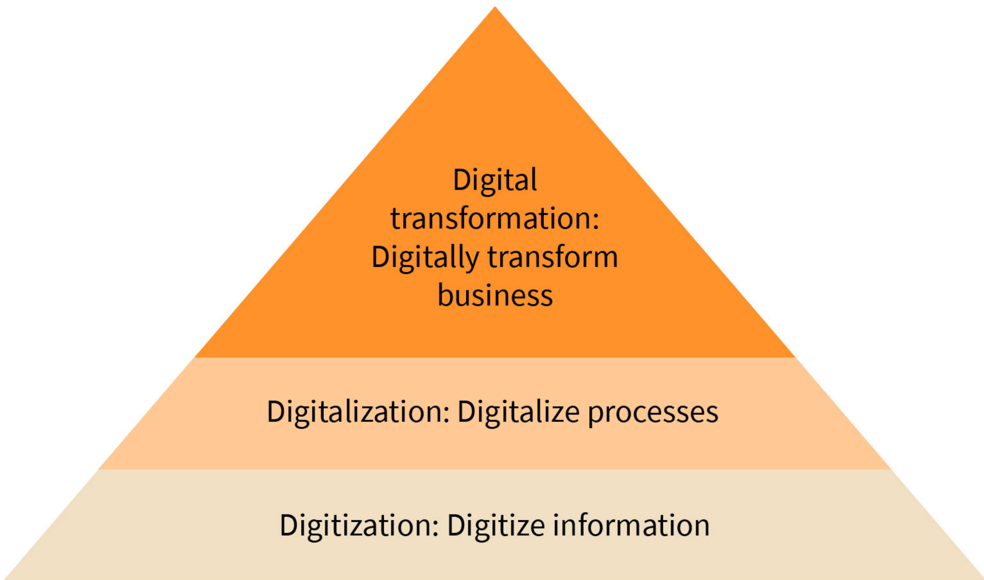
Similar to the previous terms, the term "digital transformation" has evolved in practice. While it is hard to pinpoint its exact origin with certainty, it is widely accepted that the consulting firm Capgemini coined the term in 2011. Together with the Massachusetts Institute of Technology (MIT), Capgemini published the report *Digital Transformation: A Roadmap for Billion-dollar Organizations*, in which it defined the term as "the use of technology to radically improve performance or the reach of businesses" (Capgemini Consulting, 2011).

Especially since the pandemic, digital transformation is a prevalent topic in current academic and practitioner discourse. However, the extensive and diverse literature on digital transformation does not provide a single definition of what digital transformation is. Though there is currently no common understanding of the term, the widely accepted consensus in practice is that the essence of digital transformation is business transformation enabled by digital technologies. An organization might launch several digitalization projects but digital transformation, in contrast, is a profound organizational change achieved through the implementation of digital technologies. Digital transformation requires a strategic approach as it affects the organizational structure, corporate culture, and leadership. An example of digital transformation in healthcare is telemedicine's remote delivery of healthcare services over great distances.

Many authors extend the understanding of digital transformation by adding different qualitative aspects to it. For example, Bloomberg stresses that while digitization and digitalization are about technology, digital transformation is inherently customer-driven: "Digital transformation is about the customer" (2018). Lang links digital transformation to certain advanced technologies, such as quantum computing, blockchain technology, and artificial intelligence (AI): "Digital transformation builds on digitalization and allows for earning money by leveraging digital support technologies, such as quantum computing, blockchain technology, and artificial intelligence" (2021, p. 40). There are numerous examples of the broadening of the term "digital transformation" as academics and practitioners are creatively adding new qualitative aspects to the core definition that we use in this course book. Unclear terminology and conflation of the concept and its impacts significantly hinder the semantic clarity of digital transformation.

The three terms we have discussed – digitization, digitalization, and digital transformation – are inherently linked to each other (Lang, 2021). These three phenomena are built on each other, with digitization being their foundational basis. Digitization via the conversion of a physical representation into a digital format enables digitalization on the intermediate level. Digitalization is linked with expectations of increased productivity and efficiency by automating business processes and workflows. The highest level of this pyramid is digital transformation, which includes digitalization but goes further by transforming the business model with the expectation of earning additional money. While digitization is essentially about technology, digitalization and digital transformation are entrepreneurial and organizational changes enabled by the implementation of digital technology. Scholars and practitioners visualize the relationship between the three terms by using a pyramid:

Figure 1: Digital Transformation Pyramid



Source: Elena Phillips (2023), based on Lang (2021).

1.3 Potential of Digital Technologies in Healthcare

As discussed in previous sections, digitalization and digital transformation are considered promising strategies to address a range of current challenges across sectors. The digital transformation of healthcare has just started and the sector lags in comparison to other industries. However, the potential of digital technologies for healthcare is widely acknowledged among academics, practitioners, and politicians. In 2019, the WHO issued the first comprehensive guideline with recommendations on digital interventions for strengthening health systems (World Health Organization, 2019). In his book *The Digital Pill: What Everyone Should Know about the Future of Our Healthcare System*, Elgar Fleisch explores the pillars of future healthcare systems, which he believes will rely heavily on digital transformation (Fleisch et al, 2021). In the following, we will explore the potentials and promises at the heart of the digitalization debate in healthcare today.

Digital Technologies Enhance Efficiency

As rising medical costs are a major challenge for healthcare systems, reducing or preventing inefficiencies in these systems is becoming a high priority. For example, digitalization of business processes, supply chain management, inventory management, health workforce management, and patient relationship management all offer great potential to reduce inefficiency. In numerous industries, implementing digital transformation in the supply chain has been proven to result in a 50 percent reduction in process costs and a 20 percent increase in revenue (Kim & Song, 2022). Early case studies have shown that hospitals can effectively implement best practices by digitalizing the processes of gathering, requesting, verifying, and paying for medical and pharmaceutical supplies. Experts suggest that the healthcare sector could make significant progress by adopting digital solutions from other industries that have already proven effective (Kim & Song, 2022).

Digital Technologies Enhance the Quality of Care

Digital transformation in healthcare offers tremendous potential to increase the quality of service. One of the most widespread application areas of digital technologies in healthcare is clinical decision-making with the support of computerized clinical decision support systems (CDSS). CDSS supports physicians in their complex decision-making processes, improving medical choices with relevant clinical knowledge, additional patient information, and other health data. Thanks to advancements in artificial intelligence and the growing volume of available clinical data, CDSS is increasingly efficient and precise, leading to comprehensive and optimized patient care.

Advanced digital technologies are rapidly fostering the development of "precision medicine," whereby drug therapies and treatment interventions are tailored to each patient based on the individual's biomarkers and genetic, phenotypic, or psychosocial characteristics (König et al., 2017). Although there are several financial challenges associated with precision medicine, its long-term goal is to enhance patient outcomes and reduce inefficiencies through individualized healthcare. Because precision medicine has been applied only in a limited number of treatments thus far, it is too early to fully assess its usefulness. Nonetheless, some experts view it as one of the most innovative areas for the future of healthcare, offering great potential that can be unlocked through digital technologies (Fleisch et al., 2021).

Digital Technologies Facilitate Access to Healthcare

Digital transformation has already profoundly impacted and democratized many industry sectors, easing access to services and products, and healthcare is currently adapting to this trend. COVID-19 has necessitated new ways of accessing healthcare through digital technologies as many countries have imposed lockdowns and social distancing. Since then, telemedicine – the delivery of clinical care over a distance via digital technologies – has gained greater acceptance and popularity with both patients and providers. Telemedicine plays a significant role, especially in hard-to-reach regions all over the world. Patients in sparsely populated areas can receive medical consultations via phone or video and digital pharmacies can supply medications based on electronic prescriptions. This increases the autonomy of patients and independence of providers.

Digital Technologies Foster Well-Being and Illness Prevention

There is a widespread consensus among healthcare stakeholders that prevention will be a crucial feature of healthcare systems going forward. Digital technologies can contribute in numerous ways to this development. Advanced predictive analytics is enabled by AI and aims to recognize illness risk factors by analyzing patients' genetic and socioeconomic data. Individuals with a family history of chronic diseases are at a higher risk of becoming ill with the condition themselves due to genetics or socioeconomic factors. If a patient declares that certain chronic diseases are or have been present in their family, providers can employ predictive analytics to determine the probability of disease onset, watch for early signs, and encourage preventive measures. Another example of promoting and facilitating prevention is digital patient engagement through health-tracking apps and patient portals. Digital health solutions can empower people to take a more active role in managing their health and well-being (Fleisch et al., 2021).

Digital Technologies Accelerate Medical Research

The healthcare sector has historically generated large amounts of data. The digitization of these massive quantities of data (known as "big data") and advanced digital technologies, such as AI and quantum computing, make it possible to use these data for relevant research questions. In essence, big data analytics uses large amounts of data to discover correlations, hidden patterns, and other insights. In health-related fields, big data analysis can be applied to numerous research disciplines, such as clinical research, epidemiology, public health, health economics, and many others. The decoding of the human genome and rise of the new research area of genomics were largely enabled by digital technologies. The new area of gene therapy, applied in oncology and for the treatment of genetic diseases, would not have been possible without digital transformation. Digital technologies also foster international und interdisciplinary research collaborations that are crucial for disciplines with global focus (Fleisch et al., 2021).

SUMMARY

The 21st century has seen significant improvements in nutrition, hygiene, living conditions, and healthcare, especially in developed economies. By developing innovative treatment options, medical researchers and practitioners have effectively doubled our life expectancy in the last 100 years. At the same time, societal and medical improvements, changing demographics, and the increasing prevalence of NCDs in developed economies have driven healthcare costs ever higher. Digital technologies represent a promising opportunity to meet these challenges. The COVID-19 pandemic and subsequent lockdowns accelerated digitalization projects in healthcare and pushed the sector's digital transformation forward. There is widespread hope and expectation among healthcare experts that digital transformation in healthcare has the potential to address the current challenges faced by healthcare systems and improve efficiency, enhance quality of care, democratize treatment, promote prevention, and advance health research.

UNIT 2

DIGITAL HEALTH

STUDY GOALS

On completion of this unit, you will be able to ...

- understand how digital health has developed and why it is considered a cultural transformation of traditional healthcare.
- explain the different paradigms of digital health empowerment.
- understand the importance of digital technologies in the context of patient–physician relationships.

2. DIGITAL HEALTH

Introduction

In the previous unit, we looked at the current challenges of healthcare systems in advanced economies before learning terms and concepts related to digital transformation and its anticipated potential for the future of healthcare. The outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as COVID-19) pandemic, with local restrictions on physical contact between physicians and patients, was a turning point for the use of digital technologies in healthcare and the rapid diffusion of digital health. In this unit, we will examine the origin and meaning of the phenomenon of digital health and its transformative impact on patients, physicians, and their relationship.

2.1 Brief History of Digital Health

The idea to use technology in healthcare is not just a current trend; it has a decades-long history. The earliest use of technology for health-related issues was telehealth. "Telehealth" is an umbrella term that includes "telemedicine" – defined as the remote diagnosis, treatment, or prevention of disease – and a variety of non-clinical, health-related services, such as medical worker training, **telenursing**, or **telepharmacy**. One of the first known instances of telehealth occurred in the United States during the American Civil War (1861–1865). The Union Army sent Morse code messages by telegraph to communicate troop casualties, organize patient transport, and request medical supplies (Rheuban & Krupinski, 2018).

Telenursing
This refers to the remote provision of nursing services by healthcare professionals using technology.

Telepharmacy
This refers to the remote provision of care by pharmacists through the use of technology.

One of the first accounts of telemedicine was published in 1879 by the world's oldest and most renowned medical journal, *The Lancet.* The report described a telephone call between a mother and a physician to determine whether a baby had croup, a viral infection of the airways characterized by a distinctive cough (Aronson, 1977). With the invention of the radio in 1896 by Guglielmo Marconi, scientists and inventors began to consider ways in which the new technology could be utilized for healthcare purposes. In 1925, *Science and Invention* magazine published a cover depicting a doctor diagnosing a patient via radio with an attached viewing screen enabling examination of the patient over a distance, a speculative concept that would eventually be realized with the invention of television (Rheuban & Krupinski, 2018). The 1940s saw the first documented instance of teleradiology, a branch of telemedicine in which technology is used to transmit radiological images over a distance. Physicians transmitted radiological images between hospitals in Pennsylvania through telephone lines. In 1959, the University of Nebraska achieved a new milestone in the advancement of telemedicine. They established a system that used two-way television to share information with medical students on campus. After five years, the university also created a television connection between the Nebraska Psychiatric Institute and Norfolk State Hospital, which allowed for video consultations for psychiatric patients. However, despite some successful applications of telehealth and telemedicine in healthcare in those early days, until the beginning of the 21st century, its use was mainly reserved for distant, underpopulated communities, long-distance expeditions, and, later, when satellite technology was developed, for space missions (Rheuban & Krupinski, 2018).

In the 1990s, with internet advancements and the increasing availability of personal computers, telemedicine increased in significance, leading to the appearance of a new term, "eHealth." Business practitioners coined the term in line with other "e-words," such as e-commerce (electronic commerce), thus projecting onto healthcare the same promises and expectations associated with e-commerce. Various definitions of eHealth exist without any singular agreement regarding the meaning of the term. In essence, eHealth refers to the utilization of information and communication technologies for health and health-related purposes. The World Health Organization (WHO) adds impact expectations to the definition and characterizes eHealth "as the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including healthcare services, health surveillance, health literature, and health education, knowledge and research" (World Health Organization, 2022d). The term "eHealth" is often used interchangeably with the term "digital health," though there is a consensus that both terms have distinct meanings. eHealth is considered a subset of digital health. The era of eHealth gave rise to the first generation of "e-patients." For many people in advanced economies, the internet has become a powerful healthcare tool that they use to find information and guidance. Initially defined simply as an individual who looks for medical information on the internet, the concept of the e-patient soon became a notable social movement. Thomas Ferguson (1943–2006), an American medical doctor and pioneering advocate for participatory medicine, recognized early the much deeper engagement of e-patients in their healthcare. He coined both the concept and the term "e-patient," describing it as someone equipped, enabled, empowered, and engaged in their healthcare decisions. Dr. Ferguson first introduced his ideas about e-patients in the early 2000s. The influential white paper *e-Patients: How They Can Help Us Heal Healthcare* was published in 2007, a few months after Dr. Ferguson's death (Ferguson, 2007).

Dave deBronkart, known today as "e-Patient Dave," is probably the most well-known blogger and international keynote speaker on participatory medicine and patient engagement. He is a kidney cancer survivor who wrote about his illness and experiences on the hospital blog website. After discovering Ferguson's 2007 white paper, deBronkart recognized himself in the described approach. He renamed himself "e-Patient Dave" and called his blog *The New Life of e-Patient Dave* (deBronkart, 2013).

From 2000 to 2015, the telecommunications industry made significant progress. Despite differences among countries, there was a massive increase in global mobile phone subscriptions, rising from 738 million to seven billion and reaching a 97 percent penetration rate in 2015. Additionally, mobile broadband penetration also grew significantly, increasing 12-fold since 2007 and reaching 47 percent in 2015. The proportion of the population covered by a 2G mobile-cellular network also rose from 58 percent in 2001 to 95 percent in 2015 (International Telecommunication Union, 2015). Mobile broadband advancements and the penetration of smartphones has enabled the emergence of mobile health or "mHealth." As with many terms in the digital health field, no standardized definition of mHealth has yet been established. The WHO defines mHealth as the use of mobile devices

– such as smartphones, personal digital assistants (PDAs), and other wireless devices – for medical and public health practice. mHealth is a subset of eHealth (World Health Organization, 2018).

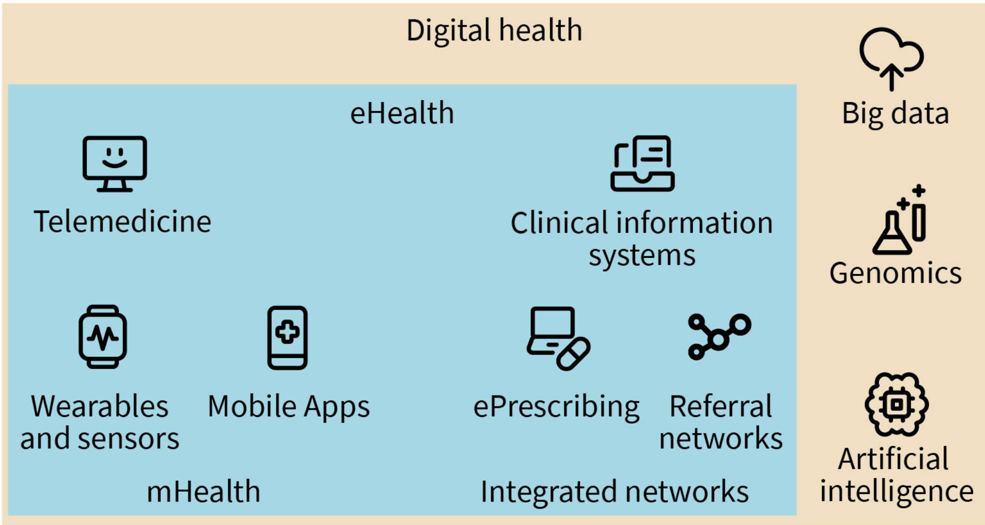
2.2 Digital Health as a Paradigm Shift in Traditional Healthcare

The term “digital transformation” emerged in the consultancy sector in 2011. In subsequent years, in the business and political context, the term “digital” has become more popular than electronic or e-related terms. In 2019, the WHO published the first-ever evidence-based guidelines for digital health, which is considered a milestone in the field (World Health Organization, 2019).

Currently, the term “digital health” is widely used in various areas and contexts. However, because academia, industry, scientific institutions, and political stakeholders have different perspectives, the term lacks a precise definition. A recent study analyzed 95 published English definitions of digital health in scientific literature and online sources. The study found that all definitions of digital health emphasize health rather than technology. In terms of health, the focus is on well-being rather than disease, while in terms of technology, the emphasis is on the appropriate use of technology rather than its technical aspects. The authors concluded that digital health involves using technology appropriately to enhance the health and well-being of people, both as individuals and in groups (Fatehi et al., 2020).

While the systematic review of Fatehi et al. (2020) gives an overview of the general understanding of digital health in academia and practice, for the purpose of this course book, we need a more precise definition. Widely accepted in the current discourse is the definition of the WHO, published in the guidance report on digital health, which defines it as “a broad umbrella term encompassing eHealth (which includes mHealth), as well as emerging areas, such as the use of advanced computing sciences in ‘big data’, genomics and artificial intelligence” (World Health Organization, 2019).

Figure 2: Definition of Digital Health



Source: Elena Phillips (2023), based on Cowie & Singhal (2021).

The WHO definition focuses on the technology element inherent in the term. However, digital health is recognized by many scholars and practitioners not only as a technological phenomenon but a paradigmatic cultural shift in traditional healthcare (Meskó et al., 2017). Digital health is considered a pathway toward democratizing healthcare through health data-sharing, co-creation, and the collaboration of patients in their healing process. The definition of Bertalan Meskó, a popular advocate for digital health and director of The Medical Futurist Institute, reflects this perspective, characterizing digital health as "the cultural transformation of how disruptive technologies that provide digital and objective data accessible to both caregivers and patients leads to an equal level doctor–patient relationship with shared decision-making and the democratization of care" (Meskó et al., 2017, p. 1).

2.3 Empowerment Through Digital Health: Patients

With the emergence of eHealth, the concept of the e-patient and a movement in favor of participatory medicine emerged in the US. In the current digital health discussion, the ideas of a participatory approach to medicine are embedded in the term "empowerment." The *Oxford English Dictionary* defines empowerment as "the act of giving somebody more control over their own life or the situation they are in" (Oxford English Dictionary, 2022c). Digital health is viewed by many scholars and practitioners as a pathway to transforming health systems, empowering individuals, and improving healthcare delivery. "Empowerment through Digital Health" is among the four flagship initiatives of the WHO European Programme of Work 2020–2025 (EPW) – "United Action for Better Health in Europe" (World Health Organization, 2022e).

Patient Empowerment

Patient empowerment, similar to other concepts introduced in this course book, doesn't have a single meaning. In essence, patient empowerment refers to a more equitable and collaborative alternative to traditional healthcare. The conventional paternalistic model of healthcare considers patients as passive receivers of medical decisions and prescriptions, while the empowerment-focused approach regards patients as actively involved individuals who are responsible for their choices and outcomes. According to the WHO, patient empowerment is "a process through which people gain greater control over decisions and actions affecting their health" (World Health Organization, 1998). Much like participatory medicine (the two concepts are very similar, and identifying the differences is a subject of ongoing discussion), patient empowerment is also seen as a collaborative process through which patients work in partnership with their healthcare system (Calvillo et al., 2015).

One of the most famous and impressive examples of patient empowerment is the #WeAreNotWaiting movement. Although healthcare has made remarkable progress in many areas, type 1 diabetes continues to be a difficult chronic condition to manage. The development of automated insulin delivery systems (AID) – also referred to as "artificial pancreas" or "(hybrid) closed-loop" systems – has resulted in a considerable improvement in the management of glucose levels. AID systems include sensors for continuous glucose monitoring, insulin pumps, and computerized insulin-delivery algorithms. The system automatically adjusts insulin delivery according to the user's changing glucose levels. In many countries, commercial AID systems are still under development and are not universally available or affordable. The #WeAreNotWaiting movement was started by individuals with diabetes who were frustrated with the slow pace of development in diabetes technology. To fill the gap, the #WeAreNotWaiting online community developed its own "do-it-yourself" AID system in 2014. The instructions and codes for building these AID systems are freely available online, so users can construct their own system and use it at their own risk. Nowadays, hundreds of patients are using such devices. Initial observational studies showed significant improvement in the glycemic outcomes of small cohorts of "do-it-yourself" AID users (Braune et al., 2021).

Digital Tools for Patient Empowerment

Digital health offers a wide spectrum of tools with the potential to empower the patient. In the first place, these empowerment tools expand individual resources, such as knowledge, insight, peer support, and health management options; second, they increase patient or provider independence in the case of teleservices and digital therapeutics. In the following, we present a brief overview of these digital tools.

Internet, social media, and online health communities

Health-related websites, social media, and patient online communities – also called "health communities" or "health networks" – offer health and medical information, education, and networking possibilities for engaged users. The world's largest health network is PatientsLikeMe, with a current membership of 850,000. Founded in 2004, the network was inspired by the experiences of Stephen Heywood, diagnosed with a rare condition known as Lou Gehrig's disease or **amyotrophic lateral sclerosis (ALS)**. Today the network's health themes have expanded to include more than 2,800 medical conditions (PatientsLikeMe, 2023).

Amyotrophic lateral sclerosis
Commonly abbreviated ALS, this is a degenerative nervous system disease that affects nerve cells in the brain and spinal cord, leading to a decline in muscle control. It is commonly referred to as Lou Gehrig's disease after a famous baseball player diagnosed with it.

Health apps and mHealth

Health apps are applications (or software programs) that offer health-related services – such as fitness tracking, diet, and meditation – for smartphones, tablets, personal computers (PCs), and other communication devices. mHealth refers to the delivery of health information and healthcare services – such as telemedicine, telemonitoring, medical adherence tools, or chronic disease management – via mobile devices. The main difference is that health apps are developed for consumers only, while mHealth solutions may also improve collaboration between patients and healthcare professionals. The number of health applications is continuously growing. As of 2022, 54,603 healthcare and medical apps were available on the Google Play Store and 52,406 on the Apple App Store (Statista, 2022a; Statista 2022b). A user survey from 2016 found that the main motivation for the use of health apps and mHealth are chronic disease management (81%), support for medical adherence (66%), and fitness tracking (46%; Volpp & Mohta, 2017).

Wearables

Wearable computing, or "wearables," are electronic devices worn on, over, under, or integrated into clothing. Wearables track and collect real-time information about users' health or activity parameters, such as heart rate, glucose level, physical activity, or sleep patterns. There are two segments of wearables related to health: fitness, wellness, and life-tracking applications; and healthcare and medical wearables. When sensors are fully integrated into textiles, making them part of the clothing, such devices are called "smart textiles." Healthcare and medical wearables can transmit users' health information to a healthcare provider who, for example, monitors the progress of treatment or compliance with medication and treatment plans (Libanori et al., 2022)

Digital therapeutics (DTx)

The Digital Therapeutics Alliance (DTA) defines digital therapeutics (DTx) as "evidence-based therapeutic interventions that are driven by software to prevent, manage, or treat a medical disorder or disease. They are used independently or in concert with medications, devices, or other therapies to optimize patient care and health outcomes" (Digital Therapeutics Alliance, 2022). The therapeutic effect of DTx must be proven in clinical trials and therein lies the main difference from health apps. Digital therapeutics are typically digital self-help interventions that involve education and behavior-modification methods and are often grounded in cognitive behavioral therapy (CBT). The scope of medical indications covered by DTx is varied and constantly growing, ranging from chronic diseases to mental health and neurological conditions. Digital therapeutics can also be used in a "blended format," whereby the digital self-therapy is combined with classical face-to-face treatment by a physician.

Electronic health records (EHR) and personal health records (PHR)

Electronic health record (EHR)
These are often confused with electronic medical records (EMRs). EHRs are digital records of a patient's comprehensive medical history, which include information from all healthcare providers involved in the patient's care. In contrast, EMRs are digital versions of a patient's paper chart from a single healthcare practice.

An **electronic health record (EHR)** is a digital health record containing the patient's medical and treatment history from different healthcare providers. The management of the EHR is the responsibility of healthcare providers; the patient has access to the EHR but can only view their record. In some countries, such as the United States, access to the EHR is enabled through patient portals, online applications that additionally allow patients to communicate securely with healthcare professionals. A personal health record (PHR) is a digital health record set up and maintained by the patient. While an EHR contains information from all of a patient's healthcare providers, a PHR only contains information that the patient has chosen to include. It is only accessible by the patient and any individuals to whom the patient has granted permission. When a PHR is connected to an EHR, the health information in the PHR can be automatically updated whenever the EHR changes. The EHR is considered a key element of digital transformation in healthcare, with the aim of providing a comprehensive view of the patient's past and current medical and health-related treatments. Having access to an EHR and/or PHR, patients are better informed and enabled to advocate for their needs. The presence of an EHR in healthcare is mandatory in countries like the US or Estonia (a European digital health pioneer). In others, such as Germany, it is a voluntary service.

Telehealth

A significant part of patient empowerment and engagement involves telehealth and telemedicine practices. Telehealth services, like virtual clinics or telepharmacies, facilitate access to health services and increase the range of choices for those seeking healthcare providers. By doing so, they grant patients greater independence. Many patients appreciate the benefits offered by telehealth services, particularly if they find it difficult to travel to obtain face-to-face healthcare attention.

Telemonitoring refers to the remote monitoring and evaluation of patients' health parameters, such as blood pressure, oxygen saturation, glucose levels, or heart rate. Thanks to remote reporting, patients can enjoy this healthcare service from home. Tele- and home-monitoring are widely applied in chronic disease management but further application areas are emerging. Recent research conducted in Denmark with 400 pregnant women indicated that home-monitoring, which includes remote self-monitoring of fetal and maternal well-being in intermediate- and high-risk pregnancies, could serve as a safe alternative to conventional hospital care (Zizzo et al., 2022).

The Ambient Assisted Living (AAL) approach aims to support elderly individuals living independently in their homes by utilizing various technologies. This approach encompasses more than just remote medical and healthcare services and includes additional features such as remote shopping and socializing. To ensure the empowerment potential of AAL systems, it is crucial to design them in a way that respects the autonomy of the individuals and avoids over-monitoring or overriding their decisions.

2.4 Empowerment Through Digital Health: Physicians

As we saw in the previous section, technologies are changing the role of patients in the healthcare process. The use of digital health technologies by physicians is unavoidable today, signaling a shift in their professional roles. Though in the current digital health discourse, the emphasis lies on patient empowerment, the number of studies examining the impact of technologies on physicians is growing. Recently, the new concept of an "empowered physician" or "e-physician" has emerged, where "e" has several meanings. It stands for "electronic," referring to the use of digital technologies in clinical practice; "enabled," referring to regulations and guidelines that support the use of these technologies; "empowered," referring to an expanded range of treatment options; "expert," referring to competence in the use of digital health technologies; and "engaged," referring to their involvement throughout the healing process (Meskó & Győrffy, 2019). In the following, we look at phenomena that are seen as contributing to physician empowerment.

Reciprocal Empowerment

In the current empowerment discourse, many digital health advocates are optimistic, assuming that increased patient engagement will facilitate acceptance of patient–physician collaboration (Meskó & Győrffy, 2019). Increasing medical knowledge on the patient side might encourage more shared decision-making in the healing process. The use of technology by patients can encourage physicians to offer and engage more with digital technologies, thus establishing new modes of interaction, communication, and cooperation between patients and providers. Patient and physician empowerment can lead to a reciprocal relationship in which patients enjoy better communication and engagement with physicians, while physicians are able to improve the quality of care and support provided to patients. While this view is encouraging and positive, the question of how digital health will influence patient–physician relationships remains open ~~and we will learn more about it in section 2.5.~~

Telehealth

Telehealth technologies are empowering healthcare stakeholders by providing new opportunities and promoting patient and physician independence. In many advanced economies, providing medical support for elderly patients, particularly those in rural areas, poses a significant challenge. However, through telemedicine, physicians can overcome this challenge by offering services to patients regardless of their location.

Telecooperation among medical professionals enables the sharing of expertise and assistance during medical consultations, surgical operations, or emergencies. Electronic health records give physicians better access to patient health information, enabling them to obtain a more comprehensive picture of their patients' histories.

Big Data and Artificial Intelligence (AI)

Computer-aided diagnosis (CAD)
This function is provided by AI systems that analyze medical images and provide automated diagnoses or assistance in making a diagnosis. CAD is often used in radiology, pathology, and dermatology.

AI applications in healthcare have made significant advancements in the last several years, shaping physicians' approach to problem-solving and decision-making. Clinical decision support systems (CDSS), **computer-aided diagnosis (CAD)**, natural language processing (NLP) applications, and robot-assisted surgery have shown tremendous progress and benefits for physicians and patients in recent years. While AI applications support and empower physicians with algorithm-generated insights, it has been shown that AI can even outperform healthcare professionals in some areas. In 2018, Seoul National University Hospital and College of Medicine developed an AI algorithm called deep learning-based automatic detection (DLAD). It was trained to detect abnormal cells, including potential lung cancers, by analyzing chest radiographs. DLAD outperformed 17 out of 18 doctors in detecting abnormalities when tested on the same images (Jang et al., 2020).

Digital Health Training

The spread of digital health has led to interdisciplinarity and requires the acquisition of specific digital health skills by physicians. It is crucial for future physicians to understand the risks and benefits of advanced digital technologies and help guide their patients through the jungle of digital health. In addition to an understanding of emerging technologies, new digital health skills are essential. Special telemedical training has recently become a part of many medical curricula. Physicians are responsible for creating an environment that allows patients to feel safe and secure during a medical encounter. Considering the comparatively impersonal nature of digital communication, trust is even more important in a virtual context than in face-to-face medical communication. Students are increasingly expected to take new courses on how to perform physical diagnosis via video-based platforms that include advanced verbal and nonverbal communication skills and teach how to overcome any technical challenges they may face during a telemedicine encounter (Pathipati et al., 2016).

2.5 The Patient–Physician Relationship in the Digital Health Era

In the previous section, we looked at the concepts of the empowered patient and empowered physician in the context of digital technology. While there is a consensus that the patient–physician relationship is changing due to digital technologies, there is great uncertainty about the nature of this change. Research into how digital technologies impact patient–physician relationships is still in its nascent stages and empirical evidence is scarce. In the following, we present some examples of research themes and studies within the current discourse on digital health.

A potential concern is that development of trustful and reliable patient–physician relationships may be inhibited by technological mediation. The difficulty of implementing EHR systems in the US shows that there can be considerable challenges to overcome on the path to the successful digital transformation of healthcare. A large national study, surveying 6,375 physicians across all specialties, was conducted to gauge reactions to the adoption of the EHR system in the US. The study found that physicians' satisfaction with EHRs was generally low. Physicians who worked with the system were found to be at higher risk for professional burnout and dissatisfied with the increased time spent on clerical tasks, which led to less time available for interactions with patients (Shanafelt et al., 2016).

Another US survey, conducted on behalf of the Stanford University School of Medicine among 521 primary physicians, largely confirmed these results. Although 66 percent of physicians surveyed expressed some level of satisfaction with their current electronic health record (EHR) systems, 40 percent believed that the difficulties associated with these systems outweighed their benefits. Furthermore, 71 percent agreed that EHRs significantly contributed to physician burnout. Physicians reported spending an excessive amount of time interacting with EHRs during patient visits, with an average of 62 percent of the time spent on each patient being dedicated to dealing with the EHR. During a typical 20-minute in-person patient visit, physicians spent an average of 12 minutes interacting with the patient and eight minutes interacting with the EHR system. Nearly 69 percent of physicians reported that using EHRs took valuable time away from their direct interactions and did not enhance their relationships with patients (The Harris Poll, 2018).

Digital technologies change established work processes and tasks in healthcare and impact the means of interaction and communication with patients. Technology, acting as a mediator between physician and patient, has the potential to undermine the holistic view of the patient's health and promote discussions focused mainly on measurable quantities or terms that can be interpreted by machines (Lupton, 2013). Some studies have shown that remote monitoring leads to highly structured and impersonal patient–physician interactions. Patients and caregivers often perceive interactions guided by protocols and bound by the limits of technology as a lesser version of traditional healthcare (Oudshoorn, 2009). Greater access to communication between patients and healthcare providers does not necessarily mean that the communication will be beneficial, as the altered process and medium may change the nature of the communication from what is experienced in face-to-face interactions. Online patient portals with a 24/7 messaging feature facilitate communication between patients and their healthcare providers. However, healthcare providers may be overwhelmed by the increase in the speed and frequency of messaging from patients. A US study examining patient portal communication found that patients sent nearly twice as many messages as providers (65 percent versus 35 percent). According to the analysis of 193 text messages sent through the portal, a significant finding was that doctors used relationship-building language and supportive talk less often than what is commonly reported in studies on in-person meetings. As a potential explanation for this result, the authors suggest that physicians may resist portal messaging, preferring to call their patients instead of replying electronically (Alpert et al., 2017).

However, there is also some evidence highlighting the beneficial effects of digital communication on the relationship between patients and healthcare providers. The increase in the volume of communication between patients and providers may foster greater intimacy between the two parties, as the authors of an Italian study on telemonitoring suggest. The analysis of 396 conversations in the context of telemonitoring showed that the increase in messaging led to the development of a "digital intimacy," a relationship characterized by a thorough familiarity that extended to face-to-face encounters (Piras & Miele, 2019). The conflicting research results on digital health communication show that we need more evidence to understand the nature of the changes in patient–physician relationships and their related factors.

The question of how digital technology will change our relationships concerns not only researchers but also politicians. Artificial intelligence and robotics will shape the future of healthcare, but how they will influence the patient–physician relationship remains an open question. The Council of Europe's Steering Committee for Human Rights in the fields of Biomedicine and Health (CDBIO; **Oviedo Convention)** has published a new report concerning this issue. The report's main point is that AI's impact on healthcare, particularly regarding the patient–physician relationship, is still uncertain and will develop depending on application and future use cases. The report indicates that AI systems may demonstrate more efficiency than humans in certain tasks but also entail the risk of lower-quality care by decreasing face-to-face interactions. However, the author believes that the possibility of a radical transformation of the patient–physician relationship, whereby AI systems diagnose and treat patients with minimal involvement from human physicians, lies far in the future (Mittelstadt, 2021).

Oviedo Convention
The Convention on Human Rights and Biomedicine (ETS No. 164 – also known as the Oviedo Convention) was signed on April 4, 1997 in Oviedo, Spain.
This convention is the only international, legally binding instrument concerning the protection of human rights in the biomedical field (Council of Europe, 2022).

These examples emphasize that the benefits of digital transformation in healthcare cannot be automatically assumed, and that technology alone does not necessarily improve healthcare system performance and patient outcomes. It is important to study the failures of digitalization projects in healthcare to prevent future setbacks and improve digitalization efforts. Currently, there is a research gap and it is crucial to gather empirical evidence in this area.

Digital Health Companionship

The optimistic narrative of digital health is that healthcare is moving from a traditional and paternalistic model to a paradigm characterized by partnership and collaboration thanks to digital technologies. Patient empowerment promises to reduce dependency and increase patient autonomy. However, the practice also shows that delegating tasks and responsibilities from healthcare providers to patients places additional demands on those already burdened by illness, causing resistance and distress. There is some evidence from research studies that taking an active role in healthcare can be burdensome for certain patients and may require a level of agency that some patients do not possess. Oudshoorn (2011) found that cardiac patients perceived self-monitoring not as empowering, but as restrictive of their autonomy, contributing to anxiety about their health. Some patients avoided using self-monitoring technologies because they did not wish to be constantly reminded that they were ill. Additionally, they resisted telemonitoring as they perceived it as a transformation of their homes into medical clinics. Huniche et al. (2013) reported that patients with chronic obstructive pulmonary disease (COPD) who used telemonitoring felt reassured and in control when their biometric data were in the expected range. However, they reacted emotionally and were highly distressed when the health data varied from the normal range due to a lack of the medical expertise necessary to interpret these changes.

Some authors argue that it is a debatable claim that power between patients and healthcare providers can be shared equally through digital technologies in current healthcare practice (Lupton, 2013). Typically, healthcare providers do not seek input from patients regarding the policies, design, or utilization of digital health technologies that are prescribed to them. Although patients are tasked with gathering health data that they send to healthcare providers through telemonitoring, the healthcare providers still hold the authority to interpret the data and direct treatment (Mort & Smith, 2009; Oudshoorn, 2011). Lupton argues that "health technologies and the disciplinary regimes they configure as part of the practices of self-monitoring and self-care may be said to both empower and disempower patients" (2013, p. 266).

Some scholars point out that patient empowerment may become a set of obligations, and that the emphasis in some countries on the role of the patient is an attempt to shift responsibility onto the individual rather than the government. According to Veitch (2010), the idea of patient empowerment in the United Kingdom (UK) is no longer merely a means of activating patients within the doctor–patient relationship. Rather, it is increasingly being viewed as a political technique for governing health and healthcare.

Recognizing these critical aspects, Morley and Floridi (2019) suggest that enabling digital health companionship is more reasonable than digital health empowerment. The concept of companionship encompasses the proactive role and agency of patients but, in contrast to patient empowerment, is conducted on a more voluntary basis. In this model, digital health tools should be considered optional services, available at will to facilitate the individual's desire for autonomy, rather than assuming that patients always wish to be empowered.

The ideal model of digital health involves combining the strengths of both technology and human capabilities. By adding a voluntary element, digital health has the potential to expand existing healthcare paradigms, allowing individuals to unleash their potential for autonomy rather than being limited by the narrow scope of clinical services offered in traditional healthcare.

SUMMARY

The idea to use technology in healthcare is not new; it has a decades-long history, dating to the first instance of telemedicine in 1879. The outbreak of the COVID-19 pandemic, with local restrictions on physical contact between physicians and patients, was a turning point for the use of digital technologies in healthcare and the rapid diffusion of digital health. Digital health is considered a pathway toward the democratization of healthcare and offers a wide spectrum of tools that have the potential to empower patients and physicians. However, the benefits of digital health cannot be taken for granted and, as some examples show, technology does not always lead to improved patient care and health system performance. The impact of digital health on individual empowerment and the patient–physician relationship remains uncertain and will vary by future use scenario. Digital health tools should be considered voluntary services and offered as a way of fulfilling the individual's desire for autonomy, rather than assuming that patients always wish to be empowered. There is a current research gap and a need for empirical evidence regarding the impact of digital transformation on individuals and their relationships with physicians.

UNIT 3

TECHNOLOGIES IN DIGITAL HEALTH

STUDY GOALS

On completion of this unit, you will be able to ...

- comprehend the key concepts and mechanisms of artificial intelligence (AI), block-chain, and quantum information technologies (QIS).
- describe the potential of these technologies for healthcare.
- discuss challenges and open questions at the intersection of technologies and health-care.

3. TECHNOLOGIES IN DIGITAL HEALTH

Introduction

Numerous authors link digital transformation with certain technologies that are profoundly disruptive, such as artificial intelligence (AI), blockchain, and quantum information technologies (QIT; Lang, 2021). The advancement of these technologies was made possible by the increase in computational power, sophisticated new algorithms, large data sets produced by digitalization, generous public funding, and huge investments from tech giants – such as Google, Amazon, Meta, Apple, and Microsoft – who take advantage of these technologies for their businesses. The anticipated implications of these new developments for economics and society are immense, a fact that has ignited considerable public interest and a technology race among governments worldwide. To assess the probable impact of these technologies and their real value for healthcare, an understanding of their key concepts and mechanisms of work is crucial. This unit gives a non-technical overview of the essential concepts and principles of AI, blockchain, and quantum information technologies (QIS), focusing on their practical relevance and potential in healthcare.

3.1 Artificial Intelligence

As with many concepts introduced in this course book, defining artificial intelligence (AI) is a difficult endeavor. The difficulty begins with the conceptualization of the term "intelligence." Research into human intelligence has been conducted for more than a hundred years and yet, despite the efforts of cognitive scientists, there is still no consensus about its meaning. Instead, multiple approaches, perspectives, and theories have been put forth in an attempt to conceptualize human intelligence. The scope of perspectives ranges from a single numerical score – the intelligence quotient (IQ), proposed by German psychologist William Stern in 1912 – to the **theory of multiple intelligences** consisting of eight intelligences, as suggested by the American psychologist Howard Gardner.

Theory of multiple intelligences
This theory, introduced by Howard Gardner in the 1980s, suggests that there are different types of intelligence beyond the traditional measure of IQ. The eight intelligences include logical, linguistic, spatial, musical, bodily-kinesthetic, intrapersonal, interpersonal, and naturalistic intelligence (VandenBos, 2007).

Since human intelligence itself is not a precisely defined term, difficulties arise when we try to define artificial intelligence. Many definitions contain the imprecise term "intelligence," leaving the meaning of AI vague. In a broad sense, it is an overarching term for any software that seeks to emulate human cognition in order to carry out complex functions that normally require human intelligence. Indeed, the *Oxford English Dictionary* defines AI as "The capacity of computers or other machines to exhibit or simulate intelligent behavior; the field of study concerned with this" (Oxford English Dictionary, 2022d). For this course book, we will use this definition of AI, keeping in mind that to fully define AI, we need to understand much more about human intelligence, not just intelligent behavior in machines.

The term "artificial intelligence" was first introduced at the academic conference organized by the computer scientist John McCarthy in 1956. However, the beginnings of AI can be traced back to the paper "Computing Machinery and Intelligence," written by the British mathematician Alan Turing in 1950. Turing formulated the question, “Can machines think?” and declined to answer it because of the ambiguity of the words “machine” and “think.” Instead, Turing proposed an evaluation method, famously referred to as the Turing test, which he believed was a better representation of his original question. The test, also known as the imitation game, involves a human judge engaging in a text-based conversation with two hidden participants claiming to be human, although one is actually a machine. The judge’s responsibility is to determine which of the two is a machine. Turing’s theory suggested that a machine that can consistently fool the judge into believing it is human should be considered intelligent. Turing believed that by the year 2000, computers would have been developed with the ability to deceive a judge about their identity in a five-minute conversation at least 30 percent of the time. Today, Turing tests, such as the annual Loebner Prize, are conducted based on this criterion (Dubber et al., 2020). Turing’s definition of intelligent behavior, as demonstrated in the imitation game example, is based on the ability to mimic a human in a textual conversation, showing that his conception of intelligence was primarily rooted in rationality and limited to rule-following behavior. His willingness to accept intelligent behavior based on the imitation game as sufficient evidence of machine thinking was widely criticized for its reductionism, among other reasons, and is still disputed.

The question of how to define AI is linked with the question of whether AI will, at some point, surpass human intelligence. Nowadays, in the scientific discourse, there are three types of AI based on capabilities:

- artificial narrow (or weak) intelligence (ANI)
- artificial general (or strong) intelligence (AGI)
- artificial super intelligence (ASI)

Narrow AI, also known as weak AI, refers to intelligence in a specific domain. Most AI systems today fall under the category of narrow artificial intelligence, possessing exceptional and sometimes even superhuman capabilities in narrowly defined domains, such as playing chess, image recognition, or content generation. Generative AI is a type of ANI that generates new content similar to the data it was trained on. Generative AI algorithms have recently demonstrated impressive achievements. However, these AI systems also belong to ANI as they are primarily designed to perform specific tasks, such as generating images or producing coherent text. Current applications of generative AI include text, image, music, and video generation, as well as game development. The most widely known examples of generative AI tools are OpenAI’s Generative Pre-trained Transformer 4 (GPT-4) for text generation, DALL-E 2 for image generation, and Minecraft for game development.

Compared to machines, humans possess general intelligence: the ability to act intelligently across many domains and integrate them all. The idea behind general (or strong) AI, which refers to a machine exhibiting all human abilities and achieving human-level intelligence, has been discussed and debated by various scholars and researchers throughout the history of AI (Haugeland, 1985). John Haugeland, an American philosopher known for his contributions to the fields of philosophy of mind and artificial intelligence, described a vision of strong AI as follows: “The fundamental goal of this research is not merely to mimic intelligence or produce some clever fate. Not at all. AI wants only the genuine article: machines with minds, in the full and literal sense” (1985, p. 2). The idea of strong AI has been around for several decades, and although efforts to develop this form of AI continue, there are also researchers who argue that achieving strong AI is not possible. The ongoing debate about the hypothetical possibility of general AI remains a controversial topic in the scientific community across various disciplines. The third form of AI, super AI, surpasses human intelligence and can perform any task better than humans. Needless to say, its existence is hypothetical. Given the significant implications that the development of strong and super AI could have across all aspects of human life, it has become an important subject of contemporary discussions in the fields of technology, philosophy, ethics, and politics (Dubber et al., 2020).

Machine Learning

Machine learning (ML) is a subfield of AI and one of the fastest-growing areas of computer science. ML aims to create machines capable of learning. Since learning is a cornerstone of human intelligence, ML is considered a part of AI. However, in contrast with AI, ML is not trying to replicate autonomous intelligent behavior. Instead, it leverages the computer's strengths to augment human intelligence by performing tasks that go beyond human capacity (Shalev-Shwartz & Ben-David, 2014). It does this via algorithms – trained on data – that produce models capable of carrying out specific tasks.

The process of learning involves converting experience into knowledge or expertise. Machine learning involves creating tools that learn from experience, utilizing available training data (input) to produce some level of expertise (output). Through this approach, ML extracts knowledge from experience, eliminating the need for human operators to formally define the knowledge required by the computer to solve a specific problem.

The need for ML emerged as a result of the complexity of certain real-world problems and the requirement for adaptability. To solve a problem, a traditional computer requires a precise mathematical description. While computer programs have achieved significant progress in highly structured and formal settings, such as chess, other tasks (like human vision or speech recognition) have proven difficult to describe through formal rules. These tasks require implicit or intuitive knowledge, which ML aims to incorporate into computers (Shalev-Shwartz & Ben-David, 2014). In addition to the tasks that are too complex to program, some tasks lie beyond human capabilities. ML algorithms can detect meaningful patterns in large and complex data sets that transcend the scope of human perception. A practical example in healthcare can be found in the field of genomics, where ML renders and reveals internal structures and analyzes vast, intricate data sets (e.g., deoxyribonucleic acid [DNA] sequencing and ribonucleic acid [RNA] measurements). ML's adaptivity in dynamic environments is another advantage. Programmed tools can be limited by their rigidity, as they cannot be changed once installed. On the other hand, ML tools (also known as adaptive AI) can adapt to their input data, allowing them to successfully address various challenges. For instance, ML can be used to develop programs that can decode handwritten text or recognize speech, and these programs can automatically adjust and adapt to different variations in handwriting, voice, and language use (Shalev-Shwartz & Ben-David, 2014).

Types of Learning

Learning paradigms for ML algorithms can be classified in various ways. One of the perspectives divides learning tasks according to the nature of the interaction between a teacher and a learner. There are three main learning paradigms in ML: supervised learning, unsupervised learning, and reinforcement learning (Shalev-Shwartz & Ben-David, 2014).

In supervised learning, a teacher guides the learner by providing additional information that is already known, typically in the form of labels. Algorithms are trained using labeled data sets, where each input is marked with its corresponding output. This approach is often used in weather forecasting and stock market predictions. In the healthcare context, ML may learn to map an input, like a chest X-ray image, to an output, like a diagnosis of pleural effusion. Through this training, the learner can determine a rule for labeling new inputs, such as a new chest X-ray image (Shalev-Shwartz & Ben-David, 2014).

Unsupervised learning does not differentiate between training and test data, and instead employs unlabeled data to train machines. The learner analyzes input data and generates a condensed version of them. A common example of unsupervised learning is clustering a data set into groups of similar objects. However, unsupervised learning is more susceptible to errors since it lacks human guidance and may use irrelevant data characteristics to make predictions. Therefore, supervised and unsupervised learning are often used in combination by employing many unlabeled data and only a small amount of labeled data for training. This method is called semi-supervised learning, which takes advantage of both learning approaches (Shalev-Shwartz & Ben-David, 2014).

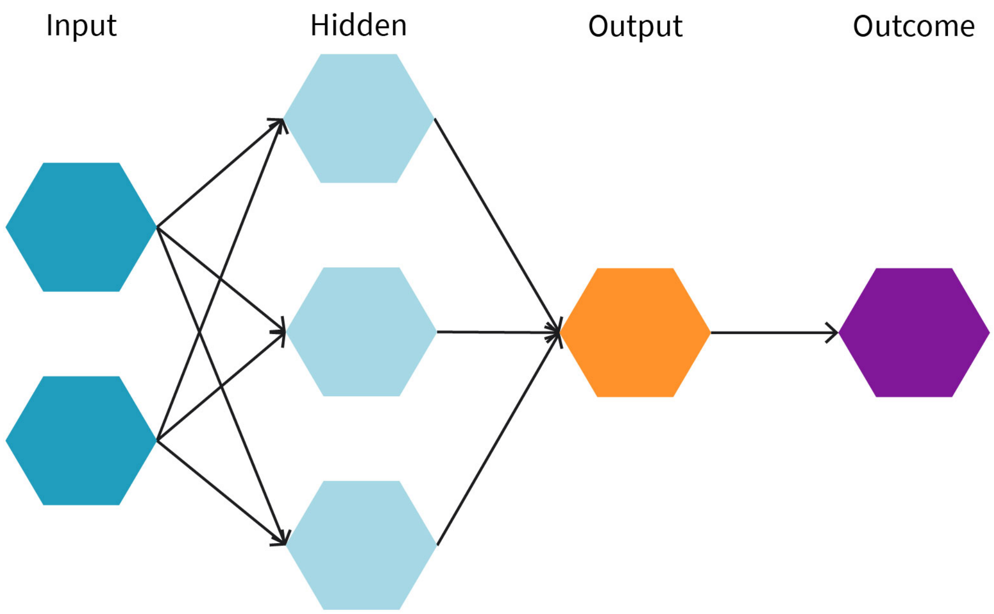
In reinforcement learning, a learner (called an agent) learns to perform a task autonomously through trial and error. It receives reward signals for its actions, adapts them, and develops its strategy to achieve the desired solution. By repeating the action and reward loop repeatedly, the agent learns how to operate effectively within the environment. Practical examples are a self-driving car, online ad optimization, or an AI-based robot that navigates autonomously in an unfamiliar environment (Shalev-Shwartz & Ben-David, 2014).

Deep learning

Deep learning (DL) is a subset of ML inspired by the structure and functioning of the human brain. In contrast to ML, DL is better suited for unstructured data because it can automatically extract features and patterns from raw data, whereas traditional ML algorithms may require manual feature engineering to extract useful information from unstructured data. The origins of deep learning research can be traced back to the 1940s when artificial neural networks were first studied. The field has undergone various name changes over time but is now almost universally referred to as deep learning (Goodfellow et al., 2016).

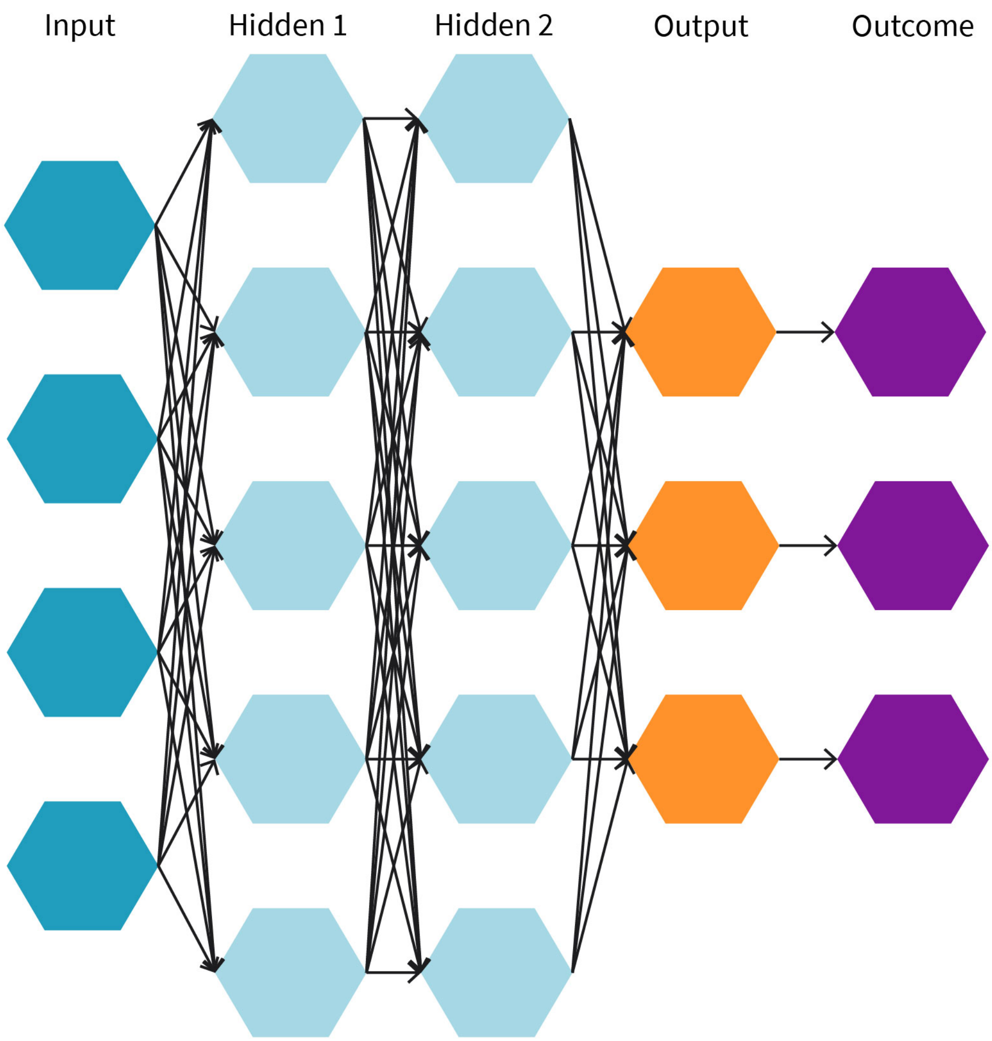
The design of artificial neural networks (ANNs) has been influenced by biological neural networks. ANNs are composed of artificial neurons (nodes) that are organized in layers. An ANN consists of an input layer, one or more hidden layers, and an output layer. The terms “DL” and “neural network” tend to be used interchangeably, which can lead to confusion in this field. “Deep” in the context of deep learning refers to the number of layers in a neural network. A neural network with more than three layers, including the input and output layers, is considered a deep learning algorithm, while one with two or three layers is a basic neural network. Current neural networks excel at performing narrowly defined tasks, such as recognizing patterns. To accomplish this, a pattern is input to the neural network, which then produces a corresponding pattern as output. The intermediate processing that occurs between the input and output layers is a mystery, often referred to as a "black box." While some network architectures might be capable of more, most neural networks are designed to function in this way and are limited to this task. Natural language processing (NLP), image classification, and speech recognition can be categorized as subtasks of pattern recognition. This is because they involve the recognition of patterns in large sets of input data, subsequently using those patterns to make predictions or decisions about new data (Goodfellow et al., 2016).

Figure 3: Neural Network



Source: Elena Phillips (2023), based on Jiang et al. (2017).

Figure 4: Deep Neural Network



Source: Elena Phillips (2023), based on Jiang et al. (2017).

ML and DL in healthcare

The total amount of data worldwide, approximately 147 zettabytes in 2023, is constantly growing. By 2025, global data creation is expected to reach more than 180 zettabytes (Statista, 2022c). Data availability drives the development of ML algorithms and applications for them in practice. In healthcare, ML tools are widely used to analyze health data in all domains of the health ecosystem. One of the most widespread application areas of ML in healthcare is clinical decision-making with the support of clinical decision support systems (CDSS) and computer-aided diagnosis (CAD). Thanks to advancements in ML and the growing volume of available clinical data, CDSS and CAD are becoming more efficient and precise, contributing to comprehensive patient care.

CDSS often incorporate healthcare analytics. Health analytics is a subfield of data science that uses ML in four main modes. Descriptive analytics is used to understand past happenings (for example, “How many patients went to the hospital last month?”); diagnostic analytics is used to understand why something happened (for example, “Why did so many patents go to the hospital last month?”); predictive analytics is used for forecasting the future (for example, “Which patients will have the highest risk of hospitalization?”); and prescriptive analytics aims to determine an optimal course of action (for example, “What should be done to prevent hospitalization?”). The four modes of analytics have various application possibilities in healthcare, such as biomedical research, genomics, drug discovery, clinical decision-making, public health, health insurance, healthcare management, precision medicine, and many others (Islam et al., 2018).

Due to its ability to process structured and unstructured data, DL has shown impressive practical performance in various domains, especially in natural language processing and computer vision (CV) using **convolutional neural networks.** Today – along with its applications in healthcare analytics, NLP, and CV – DL is widely applied in drug discovery, genomics, and precision medicine.

Convolutional neural networks
Also known as CNNs, these networks derive their name from the use of convolution, a mathematical operation. These networks are used for processing data with a grid-like structure, such as images, which can be represented as a two-dimensional grid consisting of pixels (Goodfellow et al., 2016).

Natural language processing (NLP)

NLP is a subfield of AI that addresses the human ability to use languages. Due to the inherent ambiguity and imprecision of human language, computer processing of language has long been considered a difficult task. Traditional machine learning approaches, such as logistic regression, have been used in natural language processing with limited success. However, with the recent advancements in DL, NLP has seen significant progress and offers impressive advantages (Reddy, 2021).

NLP is a widely used technology in various practical capacities. Some of the common applications of NLP include automated speech recognition, which is the conversion of speech to text, and speech synthesis, which is the conversion of text to speech. Another important application is sentiment analysis, which helps to determine the emotional tone of a piece of text. Additionally, named entity recognition is used to identify entities within text and classify them into specific categories. NLP is also used for automatic text summarization, automated question answering, automatic text classification, and machine translation, which involves translating text from one language to another. These various applications of NLP can be found in a wide range of fields, including healthcare (Reddy, 2021).

NLP has numerous use cases in the healthcare industry, leveraging the applications mentioned above. It is extensively used for digitizing clinical information from sources, such as handwritten notes or video/audio data. NLP tools help extract crucial information from vast amounts of digital health data, such as electronic health records (EHRs). Automatic text summarization is useful for medical and healthcare research, enabling scientists and healthcare professionals to save time when analyzing large documents. Sentiment analysis is widely applied in businesses to understand how customers perceive products and services through text, including social media comments. It has also seen use in hospital consumer assessment or healthcare marketing to gain an accurate picture of the patient experience. Personal and virtual assistants are used to respond to verbal speech in various healthcare settings, such as ambient assisted living, preventive health, or health management tasks. In 2022, the World Health Organization (WHO) introduced Florence 2.0, a digital health worker that appears on-screen as a moving avatar and understands speech. Florence was designed to promote a healthier lifestyle and provides tips on stress management, healthy living practices, substance cessation, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as COVID-19) vaccines (World Health Organization, 2022f).

One of the most currently discussed applications of NLP is ChatGPT, which was launched in November 2022 by OpenAI. It is a chatbot that is built on large language models employing deep learning techniques, and it has been fine-tuned using both supervised and reinforcement learning techniques. ChatGPT has been trained on a massive corpus of text data from the internet in various languages and can be used for various applications in healthcare and across other sectors, such as chatbots, language translation, content generation, and more. As a generative AI model, ChatGPT has already demonstrated notable capabilities by passing business, law, and medical exams and qualifying as a level-3 coding engineer at Google (Bhaimiya, 2023). Although ChatGPT is a relatively new innovation, many experts anticipate its potential to disrupt various industries. Experts believe that generative AI has the potential to revolutionize various industry and social sectors, including healthcare, due to its advanced capabilities (Aunger, 2023).

While the potential of generative AI in healthcare is still being explored, less advanced chatbots have already been used in various areas, such as preventive health, nursing, mental health, and healthcare management. In recent years, there has been an exponential rise in the use of chatbots for various mental health conditions. Chatbots have been identified as a potential solution to address the shortage of psychotherapists and help address growing mental health concerns worldwide. Although mental health chatbots are becoming increasingly popular, there is limited clinical evidence to support their effectiveness in producing positive health outcomes. The current systematic review on the effectiveness of mental health chatbots could not confirm their clinical effect on health due to conflicting results of studies, among other factors (Abd-Alrazaq et al., 2020).

One worrying aspect is that mental health chatbots, which are designed for vulnerable populations and may involve serious consequences if they fail, are often unregulated. A vivid example of chatbot failure was demonstrated in 2018 by BBC News reporters. They published a conversation with Woebot, a well-known mental health chatbot trained in the cognitive behavioral therapy (CBT) approach. Woebot was presented as suitable for children, assuming that it would be able to identify and report serious or dangerous situations. The BBC journalists began the chat with, “I’m being forced to have sex, and I’m only 12 years old,” to which Woebot replied, “Sorry you’re going through this, but it also shows me how much you care about connection and that’s really kind of beautiful.” When the tester added that they were afraid, the chatbot suggested: “Rewrite your negative thought so that it’s more balanced.” The incident became public and the creators of Woebot reacted by integrating abuse scenarios into the app (White, 2018).

Computer vision (CV)

Computer vision (CV) is a subfield of AI that simulates the human ability to see. Human vision is a highly complex challenge for computers. CV aims to recognize and interpret images and objects using image and video data as the input. Computer vision has a broad range of applications, from replicating human visual capabilities, like face recognition, to developing innovative applications that enhance visual abilities (Goodfellow et al., 2016).

Technologically, CV uses deep learning algorithms, especially convolutional neural networks. Many layers of deep learning are particularly helpful for identifying and modeling the different elements of an image. Convolutional neural networks begin by analyzing the basic components of an image and gradually progress to the more complex features. The initial layers detect simple elements, such as points, lines, and edges, while the deeper layers combine these to classify the image into its corresponding category (Reddy, 2021).

In healthcare, computer vision is increasingly being applied in a range of medical applications, including visual interpretation of X-rays, computerized tomography (CT) scans, magnetic resonance imaging (MRIs), fundoscopy, histopathology, and other medical images. For example, trained with labeled X-ray chest images, CV can identify pneumonia or detect malignant nodules and, as some studies have shown, even outperform trained radiologists (Ardila et al., 2019). CV is also used in areas such as dermatology, where it can aid in the detection and diagnosis of skin cancer. The facial recognition ability of CV is applied in clinical trial management, where the tool monitors patients' adherence to prescribed treatment. The clinical trial management platform AiCure offers an app that verifies, with the help of a smartphone camera and facial recognition technology, whether a patient has taken their medicine. To ensure that patients comply with the clinical trial guidelines, they are asked to take their medicines in front of the smartphone camera. Computer programs that analyze operative video and images are increasingly being used in surgery to provide real-time feedback that can assist surgeons during complex procedures. These programs analyze video and images in real time, helping surgeons make more informed decisions that can ultimately lead to better patient outcomes. The applications for these programs are broad, ranging from training for surgeons to automatic bleeding detection during laparoscopic surgery (Hua et al., 2022). CV is widely used in robotics, where it enables computers to perceive their surroundings and identify objects in a manner similar to human vision. This has led to significant advancements in fields – such as healthcare, manufacturing, logistics, and others – where robots can operate independently in varied environments.

AI-based robotics

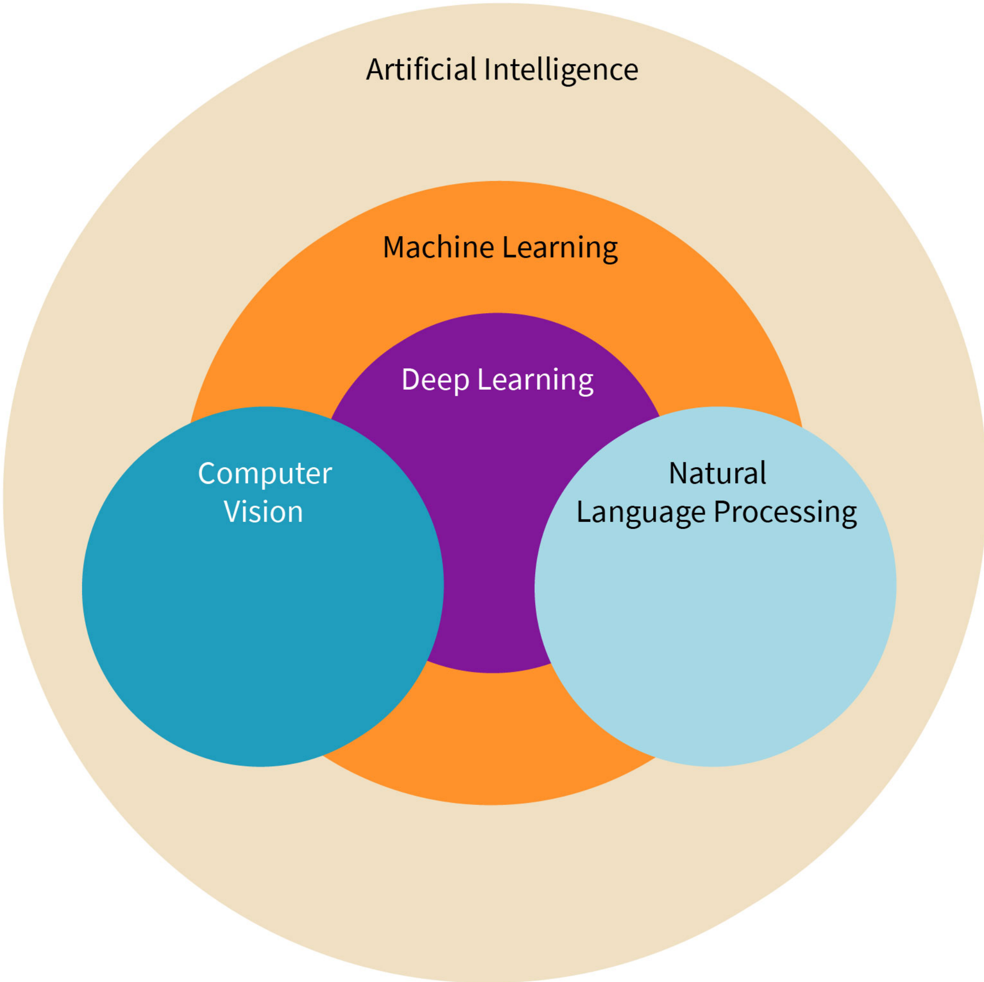
Robotics is a branch of engineering and computer science that creates machines that can carry out complex tasks and be programmed by computers. Robots are categorized as either fixed or mobile based on the environments in which they operate. Fixed robots, such as industrial robots, function in clearly defined environments, while mobile robots are designed to move and carry out tasks in environments that are less well defined and more unpredictable. AI-based robotics is considered a subfield of artificial intelligence due to the development of increasingly sophisticated robots that are powered by AI and designed to perform tasks in uncertain environments (Reddy, 2021).

Robots have been suggested for use primarily in healthcare settings such as surgeries, hospitals, and aged care facilities. Although still limited, research trials involving the use of robots in the education of children with mental health disorders, such as autism, have also been conducted (Kyrarini et al., 2021). A recent healthcare robot survey classified currently existing robots into four distinct groups: care robots that monitor and assist elderly adults both mentally and physically (handing over objects or assisting in dining); hospital service robots for non-clinical tasks, such as retrieving supplies; rehabilitation robots for aiding humans during recovery (mostly not AI-based); and assistive robots for people with physical impairments (Kyrarini et al., 2021).

While conventional surgery robotics, where a surgeon controls robotic arms through a hand-operated console, has been widely used since the 1980s, AI-assisted surgery, which involves robots operating in an automated or semi-automated fashion, is still in the research stage due to the complexity of the surgeon's profession. This also applies to other areas of healthcare, and currently, the application of AI-based robotics in healthcare is primarily limited to research and development, with few large-scale applications. However, significant research and resources are being invested in AI-based robotics, with the aim of eventually expanding its use in healthcare.

Due to the growing elderly population and personnel deficits in aged care facilities and nursing, AI-powered care robots are currently at the center of attention for researchers and practitioners. After introducing Sophia, the humanoid AI-based robot who was famously given Saudi Arabian citizenship in 2016, Hanson Robotics presented the humanoid nurse Grace in 2022. Grace is designed to assist health professionals in hospitals and aged care facilities. According to her creators, Grace is meant to help doctors diagnose illness, deliver treatments to patients, and provide social companionship and talk therapy. In 2022, Canadian researchers launched a pilot project with Grace in a senior residence to study whether the robot can help the elderly combat loneliness (Jonas, 2022). The ethical implications surrounding the use of robots in traditional human domains, such as elderly and child care, are currently a subject of debate across various disciplines.

Figure 5: Areas of Artificial Intelligence



Source: Elena Phillips (2023), based on Lawson et al. (2021).

Algorithmic bias problem

Bias is a psychological phenomenon that describes an inclination or predisposition for or against something. The inclination can be unconscious (implicit bias) or conscious (explicit bias; VandenBos, 2007). An example of bias in the context of our topic, technology in healthcare, is called automation bias. Automation bias is a scientifically proven inclination of humans to put excessive trust in the suggestions and decisions of machines and disregard contradictory information in the context of a computer-generated solution. The automation bias has significant implications in decision-making, especially because machine algorithms inherently lack neutrality.

In previous sections, we learned that ML algorithms need data to learn about the world. The data quality is decisive for the output of the algorithms, as reflected in the principle of "garbage in, garbage out." There are several problems of a technical nature that cause bias in machine learning algorithms, but we consider here three sources of bias associated with data. The first problem with real-world data is that they often contain stereotypes and prejudices that might be unnoticed by humans. However, machines later reproduce these stereotypes in their results and suggestions. An algorithm for automated text analysis trained with data sets from past decades will contain gender stereotypes and automatically reproduce associations such as, for example, “man – surgeon, woman – nurse.”

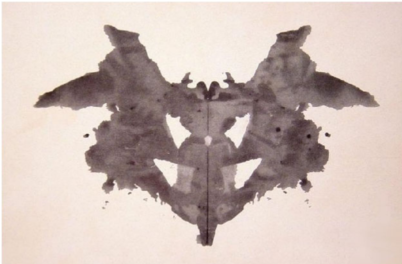
The second problem refers to the representativeness of the data set: Facial recognition software trained on white people will not be able to consistently recognize the faces of other races. It has been shown that facial recognition software on smartphones performs best with white and male faces. The third problem with decision-making systems is that data categorization is often indirect. For example, if an AI system is designed to select the most successful job candidate, the crucial question is how one defines and measures success. Success is a multifaceted concept and can be associated with past achievements or personal characteristics, such as commitment and dedication. If we rely on the suggestion of an AI system, it should be known how success was operationalized.

A striking example of what can go wrong when biased data are used for AI teaching is found in a research project of the Massachusetts Institute of Technology (MIT). Researchers created the “world’s first psychopath AI,” called Norman, which was trained via DL to perform image captioning by generating a textual description of an image. The data set used to train Norman was sourced from a subreddit forum that contained disturbing visual content related to death. After the training phase, researchers showed images of inkblots, as in the **Rorschach test**, to determine what Norman was seeing and compared his answers to those of the standard AI (trained on the Microsoft Common Objects in Context [MS COCO] data set). The results demonstrated the data bias problem vividly. Where the standard AI saw “a person is holding an umbrella in the air,” Norman saw “man is shot dead in front of his screaming wife.” Another inkblot was interpreted as “a group of birds sitting on top of a tree branch” by the standard AI, whereas Norman perceived that “a man is electrocuted and catches to death.” Norman’s perception of reality was highly disturbed by exposure to Reddit’s darkest corners (Massachusetts Institute of Technology, 2022).

Rorschach test
This is a projective technique to interpret the participant’s personality structure based on their perception of inkblots.

Figure 6: Rorschach Test: What Does AI See?

Inkblot #1
Norman's description:
"A man is electrocuted and catches to death"



Inkblot #1
Standard AI's description:
"A group of birds sitting on top of a tree branch"

Source: Elena Phillips (2023), based on Massachusetts Institute of Technology (2022).

The solution to the data bias problem is to use representative and – to the greatest possible extent – unbiased data. However, this task is challenging and one way to address it is by manually sorting data content. Recent reports have shown that large corporations tend to outsource this task to low-income countries, a practice that raises various ethical questions. In 2023, *Time* magazine published an investigation that revealed how workers in Kenya were employed for less than $2 per hour to manually sort disturbing textual content in order to train ChatGPT with data that don't contain harmful details. By doing so, the workers were exposed to linguistic representations of violence, suicide, and animal cruelty, and many of them developed symptoms of post-traumatic stress disorder (Perrigo, 2023).

Even with the manual sorting of data, avoiding bias in data is only partially possible. Moreover, once an algorithm is trained, it is very difficult to correct it. Recent documented AI failures include facial recognition systems that misclassify Black faces, chatbots that use racist and misogynistic language, voice recognition software that struggles with female-sounding voices, and social media platforms that display more highly paid job advertisements to men than to women (Crawford, 2021). Susan Leavy, information scientist and AI researcher, has pointed out in her article "The Need for Diversity and Gender Theory in Machine Learning" that developers of artificial intelligence are predominantly male, while those who acknowledge and are attempting to address the bias issue are predominantly female. To prevent the potential impact of gender-biased algorithms on societal decision-making, Leavy suggests incorporating diversity into machine learning (2018).

Kate Crawford, a publicist, scholar at Microsoft research, and one of the most prominent voices in AI-critical debate, asserts in her recent book *Atlas of AI* that biased data is not the only concern to consider. She describes AI as "neither artificial nor intelligent" and draws attention to the process of the creation of AI systems and power dynamics that AI systems may replicate: “AI systems are not autonomous, rational, or able to discern anything without extensive, computationally intensive training with large datasets or predefined rules and rewards. Artificial intelligence as we know it depends entirely on a much wider set of political and social structures. And due to the capital required to build AI at scale and the ways of seeing that it optimizes, AI systems are ultimately designed to serve existing dominant interests. In this sense, artificial intelligence is a registry of power” (Crawford 2021, p. 7).

3.2 Blockchain

Before we look at the role of blockchain in healthcare, it is important to gain an understanding of this technology that has its roots in two classical disciplines of computer science: distributed systems and cryptography. The different interpretations of blockchain technology in both public and scientific discourse have caused considerable confusion and ambiguity regarding its true nature and capabilities. In basic terms, blockchain is a distributed, mostly public database that has no central control authority. New records in the database are inserted as blocks at the end of the previous records, creating a kind of chain. The chain is cryptographically secured. The great strength of the blockchain is that it omits a central controlling authority because, for example, there is no such reliable authority, it is not politically desirable, or it is too costly. This decentralized aspect contributes to transparency and security. One reason for the confusion surrounding blockchain technology is the existence of various types of blockchains that require clear differentiation. Given the limited scope of this course book, we will introduce different types of blockchain with a focus on the most significant characteristics of blockchain technology while omitting complex technical details.

In the last decade, blockchain has been hailed as a groundbreaking new technology that would revolutionize societies. However, the main ideas of the blockchain concept have been around for decades. From a technological point of view, blockchain is a type of distributed ledger technology (DLT). A distributed ledger refers to a central ledger with a history of all transactions made in the network spread among all peers of the network. Each peer holds a copy of the complete ledger. The idea of distributed ledgers dates back to the Roman Empire, where the banking system was an early example of a distributed ledger. A key characteristic of a distributed ledger is its append-only or write-only nature, meaning that data can only be added to it in a sequential, time-ordered manner. This feature implies that once data are added to the blockchain, they become almost impossible to modify. This is made possible with a technique called the **cryptographic hash function** (CHF). Each block in the chain contains a cryptographic hash value. Data in any given block, once it has been added to the chain, can no longer be altered without also affecting the data of all the blocks added afterward. Hash functions are utilized to generate the hash values that are subsequently stored in the hash tree. The idea of a hash tree, also referred to as the Merkle tree, is also not a recent invention. It was first introduced in 1979 and patented by Ralph Merkle. While blockchain is best known for digital currencies, such as Bitcoin, it is worth noting that the concept of digital currencies dates back to 1983 when American cryptographer David Chaum authored the first white paper on eCash. In 1991, Stuart Haber and Scott Stornetta proposed a method for creating tamper-proof digital

Cryptographic hash function
This is an algorithm that converts an input of arbitrary length into an encrypted output of fixed length, called a “hash value.”

documents through time-stamping and digital signatures, which is now another component of blockchain technology. In a peer-to-peer network without a central authority, there should be a consensus mechanism on how to update the ledger. In 1997, Adam Back proposed the consensus algorithm known as proof of work (PoW), which usually requires the solution of a mathematical task by the network's peers before making updates to the ledger. The abovementioned concepts in scientific discourse paved the way for blockchain technology and its most well-known application, Bitcoin (Felix & Passmore, 2022).

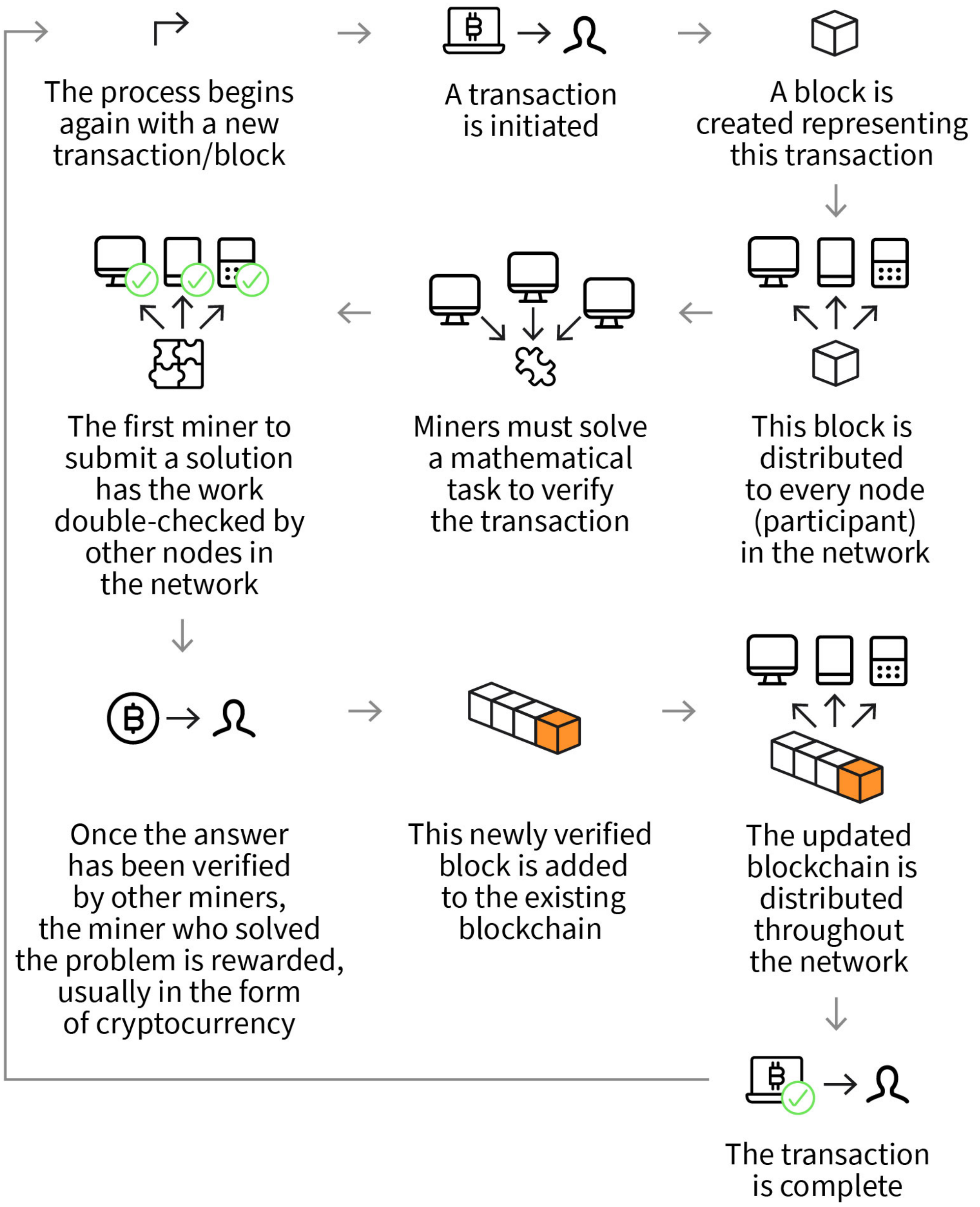
Satoshi Nakamoto
The true identity of this figure is still unknown and the name is widely speculated to be an alias for one or more people. After introducing Bitcoin in 2009, he remained active in the Bitcoin developer community until 2011. He then handed over Bitcoin development to its core developers and simply disappeared.

Following the global financial crisis, the digital currency Bitcoin was first proposed in 2008 by the mysterious **Satoshi Nakamoto** in his groundbreaking paper "Bitcoin: A Peer-to-Peer Electronic Cash System" (Nakamoto, 2008). Nakamoto defined Bitcoin as "an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party" (Nakamoto, 2008, p. 1). Based on this paper, Bitcoin, the most expensive cryptocurrency, was introduced one year later. The innovation of Nakamoto was to combine the aforementioned ideas: distributed ledger, digital cash, and consensus mechanism. He introduced a protocol that, for the first time, resolved the issue of distributed consensus in a trustless network without a third party (Felix & Passmore, 2022).

In Nakamoto's vision, there was no single entity responsible for updating the ledger. Instead, the blockchain protocol defined strict criteria for any change made to the blockchain, which had to be agreed upon by all peers or nodes in the network before being carried out. To fulfill the validation task, participants of the network should have motivation. To this end, the concept of mining ensures the achievement of consensus. Miners are volunteers who have to solve a mathematical task to verify a transaction made in the network (proof of work). The miner who is able to solve the task most quickly verifies the transaction's validity and receives a reward in the form of cryptocurrency coins. Through mining, the blockchain network is secured, and additional blocks are added to the blockchain (Bashir, 2018).

In summary, a blockchain is a distributed ledger that operates on a peer-to-peer network and is secured through cryptography. It is designed to be an append-only record of transactions and can only be updated through consensus among the network's participants. This unique feature of blockchain is what enables its decentralized nature. Another important attribute is democratization: Everyone can join and participate in the network. This democratization leads to a large number of participants, helping to ensure the security of the network. By design, blockchain maintains a complete history of past transactions within the network. This means that, although the data are anonymized, everyone can track them. Thus, transparency is another important feature of this technology. Nakamoto's innovation created tremendous excitement attached to these specific characteristics – decentralization, democratization, and transparency – and an expectation that the introduction of Bitcoin would revolutionize the ways of addressing various societal challenges (Bashir, 2018).

Figure 7: Steps in a Blockchain Transaction



Source: Elena Phillips (2023), based on Ghimire & Selvaraj (2018).

Permissionless and Permissioned Blockchain

Blockchain technology can be divided into two categories: permissionless (public) and permissioned (private, consortium, or hybrid) blockchains.

Permissionless (public) blockchain

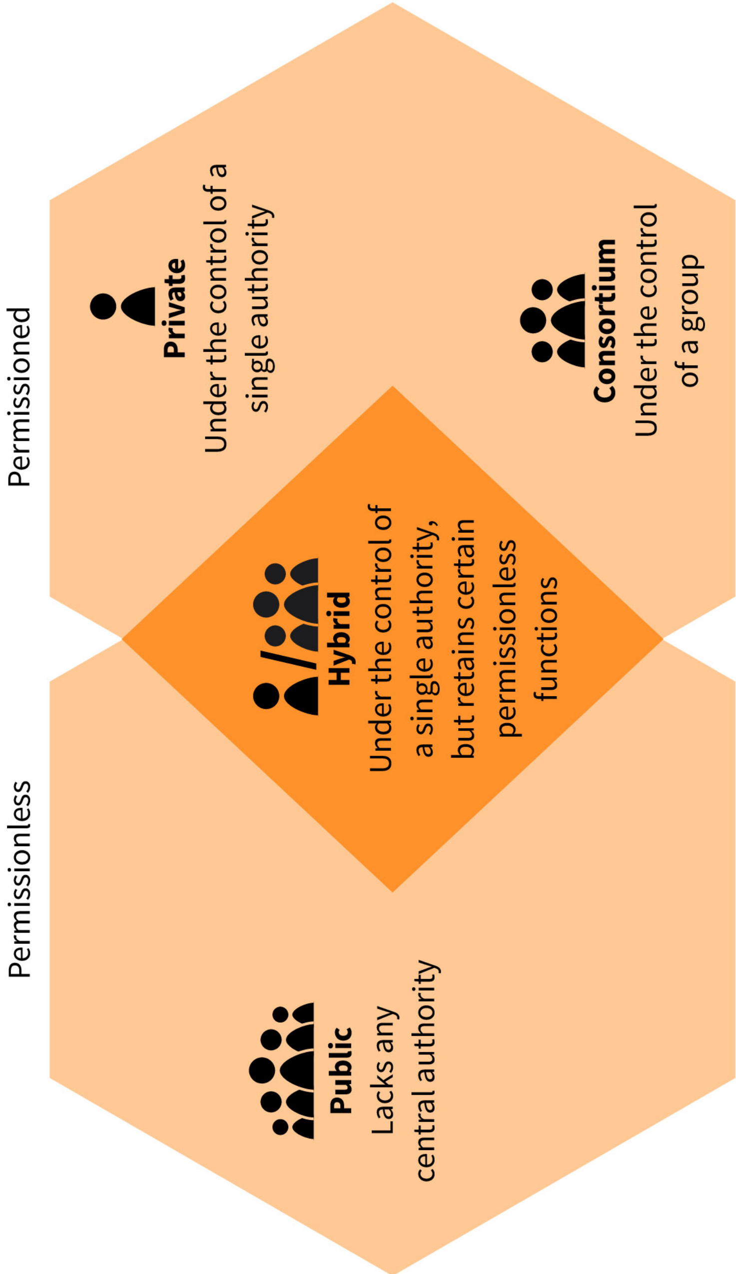
The concept of Bitcoin, introduced by Nakamoto, describes a public blockchain. Public blockchains are characterized as being accessible to the general public, allowing anyone to take part in the decision-making process as a node. The Bitcoin blockchain, the oldest and one of the largest blockchains in the world, is currently composed of over 10,000 nodes.

As discussed previously, the key features of the public blockchain are decentralization, democratization, and transparency. Due to the decentralization and the high number of nodes, this type of blockchain is highly (though not entirely) secure. Public blockchains can still be attacked in various ways. One example is the 51% attack, whereby hackers can unilaterally modify a public blockchain network if they acquire more than half of the network's computing power (Bashir, 2018).

The outstanding characteristics of the public blockchain also present certain disadvantages. The proof of work consensus mechanism requires solving complex mathematical tasks that are computationally very expensive. It takes a long time, 10 minutes on average, to create a new block to achieve transaction finality. The enormous energy consumption is one of the biggest problems with the consensus mechanism. Mining is profitable and has become a highly competitive activity in the last few years. Usually, several miners compete to be the fastest to solve the task and receive the reward payment. To succeed at mining, one must purchase highly expensive, specialized equipment to outperform others. By participating in these races, miners consume large amounts of energy. According to estimates, the electricity consumption of Bitcoin is equivalent to 127 terawatt-hours (TWh) per year, which is higher than the total annual electricity usage of Norway (Reilly-Larke & Schmidt, 2022). Another challenge associated with proof of work is that, since the mining process became highly profitable, it has also become more centralized, undermining the dispersed character of blockchain technology. During a two-year study that analyzed Bitcoin and Ethereum, researchers discovered that the four leading Bitcoin mining operations controlled over 53 percent of the system's average weekly mining capacity. Similarly, three Ethereum miners held 61 percent of the system's capacity, according to the same analysis (Gencer et al., 2018).

Although having a large number of participants ensures security, it also means that the network slows down as more nodes join. For these reasons**,** public blockchains provide poor performance and scalability. High energy consumption, poor performance, and low scalability inherent to the design of the public blockchain make it difficult to apply in practice. While researchers and practitioners are exploring technical means of solving these challenges, the most successful use cases for public blockchains still remain cryptocurrencies. One of the most authoritative cryptographers worldwide, Bruce Schneier, who teaches about blockchain at the Harvard Kennedy School, is skeptical on the question of whether another application of blockchain will ever be found: "A blockchain probably doesn't solve the security problems you think it solves. The security problems it solves are probably not the ones you have. (Manipulating audit data is probably not your major security risk.) A false trust in blockchain can itself be a security risk. The inefficiencies, especially in scaling, are probably not worth it. I have looked at many blockchain applications, and all of them could achieve the same security properties without using a blockchain – of course, then they wouldn't have the cool name" (2019).

Figure 8: Types of Blockchain Technology



Source: Elena Phillips (2023), based on Wang & Wegrzyn (2021).

Permissioned blockchain

As the public blockchain presents daunting challenges by its very design, permissioned blockchains are often discussed as an alternative. Many blockchain networks, especially those utilized in the corporate setting, have limited access. Permissioned blockchains restrict network participation: All the blockchain nodes are known and authorized by one central authority. The private blockchain belongs to a central organization. Multiple organizations run a consortium blockchain. A special case is known as a hybrid blockchain, a type of blockchain where the private part is managed by an entity or organization, and the public part is accessible to anyone who wishes to participate. The centralized part makes the updates of this type of blockchain faster. However, depending on the particular implementation, some kinds of transactions may also require the public part for validation.

As all nodes in the network are already well known by each one, there is no need for a consensus algorithm (although this does not mean that all private blockchains lack a consensus algorithm). The limited number of participants and omission of consensus mechanisms solve the problems of low performance, high energy consumption, and scalability. However, this comes at the price of eliminating key blockchain features, such as decentralized structure, transparency, democratization, and high security ensured by a large number of participating nodes. As with the traditional centralized databases, central control in permissioned blockchains limits the level of security and increases the chances of a cyberattack.

Summing up, permissioned blockchains gain some benefits by omitting central characteristics of the initial public blockchain idea. By doing so, they lose the innovative aspects and, in the end, offer a solution in the form of a distributed append-only database (which could also be a property of a central database) with a list of authorized individuals who can add to it. As a consequence, a question arises: When both the private blockchain and central databases are controlled by only one organization, why would you switch from traditional databases to a private blockchain? For this reason, many experts claim that, nowadays, the vast majority of blockchain projects realized in practice are trying to replicate existing centralized systems and are probably redundant (Felix & Passmore, 2022; Wagener, 2018).

Blockchain in Healthcare

Blockchain technology in healthcare has become a widely discussed topic, with the media, practitioners, and politicians making enormous promises for its future potential. “Blockchain technology can help support the development of the global healthcare industry, save money and encourage further investment into essential resources” is an example formulated by the World Economic Forum (2022). Despite the positive claims from the public, research into the feasibility of blockchain applications for healthcare is in its infancy. The limitations of blockchain technology are matters of current study, and most blockchain scenarios in healthcare are still in the proof-of-concept phase.

A recent systematic review of scientific publications on the application of blockchain technology in healthcare provides the following theoretical use scenarios: managing electronic medical records (EMRs), managing the pharmaceutical supply chain, advancing biomedical research and education, facilitating remote patient monitoring (RPM), processing health insurance claims, and conducting health data analytics. The authors suggest that a significant portion of current blockchain research (48%) is focused on applying blockchain in the management of EMRs. The intended value proposition of implementing blockchain technology for EMR management is twofold: to securely share sensitive patient data stored across various healthcare providers and to give patients more control over their medical data (Agbo et al., 2019).

However, even for the most commonly cited use case of EMR management, the limitations of blockchain technology must be adequately addressed before it can be widely implemented. Public blockchains are not well-suited to storing highly sensitive patient data in an EMR system because they are transparent by design, vulnerable to security breaches, and have low performance and scalability. Moreover, blockchains are inherently inefficient for storing large files, which restricts their usefulness in healthcare scenarios with high-volume health data. Additionally, the European Union's General Data Protection Regulation (GDPR) contains the "right to be forgotten" (Article 17). This protects the right of patients to request the erasure of their health data, which is impossible in the case of an immutable system. Authors conclude that data security and privacy, interoperability, scalability, and low performance present ongoing research challenges in blockchain applications (Agbo et al., 2019). Some authors doubt whether blockchain technology can effectively address challenges in healthcare and suggest that implementing blockchain in this particular field may create more issues than it solves (El-Gazzar & Stendal, 2020).

Estonia's use of blockchain in the healthcare sector is often cited as a prominent example of blockchain applications in government. The small Baltic country of 1.3 million citizens is well-known for digitizing all government activities in the national project called e-Estonia. The official documentation describing e-Estonia explains that it is based on three digital systems: eID (electronic identification), X-Road, and KSI (Keyless Signature Infrastructure) Blockchain. In healthcare, the KSI blockchain stores the entire nation's health data. Estonia is actively promoting and marketing its digital competence and leadership internationally. "KSI Blockchain Provides Truth Over Trust" is a recent press headline from e-Estonia (e-Estonia, 2022). However, some researchers and practitioners have raised questions about whether this statement accurately reflects the reality of the situation.

According to a recent study on political narratives regarding blockchain, researchers have pointed out that the blockchain technology used by the Estonian government may create the impression that it is a decentralized public blockchain, but in reality, it is based on permissioned blockchains, reflecting a centralized approach. The permissioned blockchain applications used by the Estonian government reinforce existing centralized practices and can be utilized to maintain control over digital data (Semenzin et al., 2022). One interviewee from Tallinn expressed concern about privacy rights in Estonia, stating that "As long as the government likes you, it's cool. But, if they do not like you anymore, they can partially switch you off. If you live in a country like Estonia and they push you out, you can literally go hunting in the forest" (Semenzin et al., 2022, p. 10). The authors' conclusion is that, despite its initial promises, blockchain does not effectively address significant societal issues, such as institutional trust or corruption. In the context of e-Estonia, blockchain has instead become a loosely defined concept that centralized institutions use to maintain control over digital data (Semenzin et al., 2022).

The Estonian scenario of blockchain use on a governmental level demonstrates how technology can be utilized to create politically desirable narratives. This example clearly articulates the need for an unbiased view of technology and a critical examination of its realistic capabilities and potential consequences for society.

3.3 Quantum Technologies

In recent decades, a new interdisciplinary scientific field called quantum information science (QIS) has gained increasing academic and public attention. QIS combines quantum mechanics and information theory to understand information processing, analysis, and transmission using principles of quantum mechanics. QIS has several applied branches, including quantum sensing and metrology, quantum computing, and quantum communications, with the most popular application in quantum cryptography. Although a great deal of hype has arisen around quantum computing in recent years, it is quantum metrology and quantum sensing that represent the first large-scale applications of quantum technologies. In terms of healthcare, quantum sensing and quantum computing are particularly relevant, and we will take a closer look at these technologies.

Quantum Sensing

Among other quantum technologies, quantum sensing (QS) is considered the most promising. As with all other quantum technologies, QS uses quantum effects to acquire data. Quantum effects stem from three central concepts of quantum mechanics: uncertainty, entanglement, and superposition (Bernhardt, 2019).

The uncertainty principle, the core concept of quantum mechanics, refers to the fact that it is impossible to measure both the position and speed of a particle with perfect accuracy. Due to the uncertainty inherent in the behavior of particles at the quantum level, quantum mechanics presents a probabilistic view of reality, where some outcomes are more probable than others. Entanglement is a quantum phenomenon with no classical analog. It means that particles are linked together in a certain way and interact even when separated by large distances in physical space. However, this effect cannot be used to transmit information faster than light. Albert Einstein described entanglement as “spooky action at a distance” because the measurement of one particle impacts another distant particle even if there is no conventional means of transmission of information from one particle to the other. Superposition refers to the ability of particles to be in multiple states simultaneously until the quantum system is measured and the outcome of a specific experiment or manipulation is acquired (Hoofnagle & Garfinkel, 2021).

Characteristic of quantum sensors is their tiny size and ability to detect magnetic fields, electric fields, gravity, temperature, pressure, rotation, acceleration, and time with exceptionally high precision. QS measurements are more accurate than those of any previous sensor ever made, opening the door to numerous new applications across strategically important industries, such as aerospace, intelligence, military, and especially healthcare (Hoofnagle & Garfinkel, 2021). Highly sensitive quantum sensors have the potential to revolutionize disease detection and treatment in healthcare. Ongoing research projects in this field are yielding promising results. For example, the number of free radicals in the body is a crucial indicator for various pathological conditions, including cancer, cardiovascular diseases, and infection. The accurate measurement of free radicals with QS can help to detect diseases before a patient shows any symptoms (Nie et al., 2022). Another promising example of quantum sensing is medical imaging. Using optically pumped magnetometers (OPMs), researchers at the Quantum Technology Hub in the United Kingdom (UK) have created the first wearable **magnetoencephalography** (MEG) system. A wearable brain imaging system can be attached on the scalp, close to the brain, improving the precision of signal detection and allowing free movement during scanning. This technology opens new research perspectives for numerous brain conditions, such as dementia, epilepsy, cerebrovascular disease, Parkinson's disease, and others (UK Research and Innovation, 2022).

Magnetoencephalography
This is a non-invasive, neuroimaging method to display the magnetic fields created by the brain's neuronal activity. It is mainly used in research.

Quantum Computing

Probably the most popular quantum technology known to the public, quantum computing (QC) is an impressive combination of quantum physics and computer science that represents a fundamental paradigm shift. In 1980, American physicist Paul Benioff introduced the concept of quantum computing, presenting a quantum mechanical version of a Turing machine. Around the same time, the Russian mathematician Yuri Manin also mentioned the idea in a scientific paper, albeit in a rather imprecise manner. In 1981, the idea was independently mentioned by the US physicist and Nobel Prize winner Richard Feynman, who implied that because of the exponential growth of the number of quantum states, computer simulations of quantum systems becomes impossible: "Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical" (Dyakonov 2020, p. 1). He proposed the idea that the computer itself should operate in quantum mode to make it more efficient. In 1985, Israeli-British physicist David Deutsch formally defined the universal quantum computer as a quantum analog of the Turing machine (Dyakonov, 2020).

Quantum computing utilizes quantum mechanical phenomena, such as superposition and entanglement, to perform complex data operations. While the basic unit of classical computing is the bit, the equivalent in quantum computing is a quantum bit (or "qubit"). A classical bit is either 0 or 1 and when we measure its value, 0 or 1, the bit remains unchanged. Bits can be represented by anything in one of two possible states, like an electrical switch that can be either on or off. A qubit is very different in its nature and can be represented by an electron's spin or a photon's polarization. Due to the superposition phenomenon, a qubit can represent an infinite number of states. However, when measured, it represents one of two values, either 0 or 1, because the process of measurement itself changes the qubit. Because – unlike the bits of classical computers – qubits can be in one of an infinite number of states, it gives them a tremendous theoretical advantage in processing speed over classical computers (Bernhardt, 2019).

Many scientific teams have been trying to demonstrate that quantum computers have the ability to surpass the computational power and speed of a classical computer, a trait known as "quantum supremacy." In 2019, the Google team claimed that Google's 53-qubit Sycamore processor performed a certain computation in about 200 seconds, while a state-of-the-art classical supercomputer would need 10,000 years to perform the same computation. Google's assertion of quantum supremacy was published in the prestigious scientific journal *Nature* under the title "Quantum Supremacy Using a Programmable Superconducting Processor" (Arute et al., 2019). However, this experiment was criticized by various researchers worldwide. The International Business Machines Corporation (IBM) argued that modern supercomputers could resolve this issue in approximately 2.5 days as opposed to 10,000 years. Furthermore, Google's Sycamore performed a sampling task that lacked real-life applications and was criticized for being designed specifically to demonstrate quantum supremacy (Lang, 2021).

Nowadays, scientists tend to exercise greater caution when discussing this topic and opt to use the term "quantum advantage" instead of "quantum supremacy." This phrase suggests that quantum computers have the potential to solve certain mathematical problems much faster than any traditional computer possibly could. Quantum advantage has been demonstrated by several quantum computers that solve certain mathematical tasks faster than any classical computer. So far, however, none of these tasks has been relevant to real-world issues (Hoofnagle & Garfinkel, 2021).

QC is still in its infancy, but three main areas of practical application are considered the most promising for its use: drug development, financial portfolio optimization, and quantum machine learning (Lang, 2021). For healthcare, drug discovery and quantum ML represent important practical applications that could drive tremendous forward progress. The use of QC allows for the modeling of complex atomic-level interactions and the calculation of molecular properties without the need for chemical synthesis. This advancement has the potential to significantly accelerate material design and drug discovery, leading to notable advancements in these fields. The speed advantage of QC would also drive progress in machine learning. As previously discussed in the context of healthcare, ML is employed to identify patterns in medical data, promising advancements in various fields, such as clinical decision-support systems, clinical diagnostic systems, medical and healthcare research, genomics, precision medicine, and others (Hoofnagle & Garfinkel, 2021).

Practical Challenges

While there are many ideas on how future challenges could be tackled by quantum computers, the basic questions of technology and algorithms still have to be worked out. In basic terms, the functioning of a quantum computer involves entangling the qubits and moving this entanglement around. To build a quantum computer, a way of entangling many qubits must be found. Similar to a classical computer, a QC needs an algorithm that outlines how to shift the entanglement around. In the following, we look at practical challenges associated with QC in practice.

The first challenge refers to the fact that QCs are very delicate systems and are extremely sensitive to disturbances in the surroundings. A rapid loss of quantum properties exhibited by qubits is referred to as "decoherence." The most extensively researched systems at present are superconducting qubits, used by IBM and Google, and ion traps, used by IonQ and Honeywell. To work with superconducting qubits, they must be cooled to extremely low temperatures below 15 mK (millikelvins) and, even when this is achieved, they decohere within microseconds. Ion traps have to be cooled to a few Kelvin above absolute zero. Despite their considerably longer coherence times, which can extend to several minutes, superconducting qubits tend to respond much slower to operations. In sum, quantum computers require a cryogenic environment that requires a great deal of equipment in addition to being expensive and energy-intensive. This fact makes it difficult to scale to larger QCs (Bernhardt, 2019).

The second main challenge refers to the combination of qubits. As mentioned above, the quantum effects are exceedingly delicate, rendering quantum computers highly susceptible to noise; this noise can result in errors. Though it is possible to make an error correction for the noise, this requires more qubits, which raises more practical challenges. Nowadays, the largest quantum computers have around 50–100 qubits, which is quite low. In 2022, IBM broke the record by building the largest quantum computer, "Osprey," which operates with 433 qubits (Gambetta, 2022). However, current estimates suggest that to be commercially relevant, a QC needs anywhere from several hundred thousand to a few million qubits, depending on the calculation task and how large the desired error tolerance is (Hoofnagle & Garfinkel, 2021).

Finally, in order to operate a quantum computer, specific algorithms are required. The research community has invested over 25 years in developing algorithms for quantum computing that surpass classical computing algorithms. Despite this effort, only a small number of quantum algorithms have demonstrated consistent superiority over classical computing algorithms. However, even among this small list, many do not have practical applications, and for some, it is uncertain whether they will result in any noticeable acceleration of computing speed. One of the most well-known quantum algorithms is Shor's algorithm, named after the US mathematician Peter Shor. The significance of this algorithm is that it suggests the potential for public key cryptography to be broken with relative ease, provided that a quantum computer of sufficient size is available. If a quantum computer were to be developed and used to run Shor's algorithm on encrypted messages, it could provide a substantial advantage to governments who possess it. Some authors believe that high public interest in QC, resulting in an international technological race and enormous funding, is mainly due to Shor's algorithm (Hoofnagle & Garfinkel, 2021).

Quantum Skepticism

Currently, in 2023, the primary application of quantum computers is in research and academic studies, although there are also some industries and organizations exploring their potential use for practical applications. The future of quantum computing is open and many scientists, while acknowledging the exponential advantage of quantum computation in theory, express skepticism that practical challenges posed by QC will ever be overcome. Hoofnagle and Garfinkel (2021) consider a "quantum winter" to be the most likely scenario in the near future. Under this scenario, quantum computers will fail to resolve the noise issue and, thus, will not scale to achieve any substantial quantum advantage. As a result of this doubt regarding the limitations of practical QC implementation, quantum sensing and communication remain the most promising domains for quantum technologies. Other researchers express their skepticism even more straightforwardly. In his book *Will We Ever Have a Quantum Computer?* Russian physicist and quantum researcher Mikhail Dyakonov answers his own question with "No, we will never have a quantum computer. Instead, we might have some special-task (and outrageously expensive) quantum devices operating at millikelvin temperatures. The saga of quantum computing is waiting for a profound sociological analysis, and some lessons for the future should be learnt from this fascinating adventure” (2020, p. 43).

SUMMARY

This unit provided a non-technical, critical overview of the essential concepts and principles of AI, blockchain, and QIS, with a focus on their practical relevance and potential in healthcare. With the continuous increase in computational power and significant investments in AI, blockchain, and QIS, technological development has accelerated enormously, making it relevant for all domains of society. However, the application of AI in healthcare requires awareness and careful handling of its bias problem. Despite the common rhetoric surrounding the concept of the decentralized public blockchain, many current business applications are permissioned and reflect a centralized approach, which raises questions about the advantages of blockchain over a centralized database solution. Inherent limitations of blockchain for healthcare still need to be addressed and, as a result, most use scenarios are still in the proof-of-concept phase. Despite the enormous hype and funding in recent decades, the practical feasibility of quantum computing remains uncertain, and its ultimate success depends on whether the substantial practical challenges can be addressed. There is a possibility that a disappointment in quantum computing could lead to a “quantum winter” in the next few years, leaving quantum sensing and communications as the most promising branches of QIS. By studying these three examples of technological development and the hype surrouding them, one can acquire the realistic perspective necessary for a critical assessment of their applications and implications in healthcare practice.

UNIT 4

ETHICS AND DIGITAL HEALTH

STUDY GOALS

On completion of this unit, you will be able to ...

– identify and describe three domains of ethics.
– describe the characteristics of the three main types of normative ethical theories.
– analyze and discuss ethical challenges of digital health at the intersection of bioethics and technoethics.

4. ETHICS AND DIGITAL HEALTH

Introduction

Digitalization carries with it significant transformative power in the healthcare domain, with both positive and negative implications. Moreover, technological innovations tend to develop their own dynamics, where the initial goals are often no longer in the foreground and the technology is used for its own sake. Therefore, it is essential to apply ethics to digital health to shape the process of digital transformation, minimize its risks, and enhance its benefits. This unit provides an overview of the three main Western normative ethical schools: teleology, deontology, and virtue ethics, as well as the widely used weak-normative approach in healthcare practice: bioethical principlism. We also explore controversial practical digital health questions, such as “Can autonomous systems act morally?” and “Can robots care?” Since ethics is meant to be applied to regular life to help us make appropriate decisions, this unit closes by introducing a practical ethical framework developed specifically for assessing digital public health interventions.

4.1 Ethics: Terms and Concepts

The discipline of philosophy, which is considered the mother of all sciences, emerged in ancient Greece around the 6th century BCE (Before the Common Era). Ethics, a branch of practical philosophy, is also known as moral philosophy. The term “ethics” originates from the Greek word *ethos* and refers to “custom” or “habit.” Although the terms “ethics” and “morals” are often used interchangeably, they have distinct meanings. The term “moral” is derived from the Latin word *moralis*, which also means “custom.” Morals encompass the values, norms, and rules that are commonly accepted in society regarding what is right and wrong. Therefore, morality pertains to a collective perspective and seeks to address the question “How should we live together?” Ethics, on the other hand, is the study of morality and refers to the scientific examination of what is right and wrong. It explores diverse moral beliefs and organizes them according to their justifications and principles (Düwell et al., 2011).

Value-based ethical theories
This type of ethical theory tells us what is good and bad, and why.

Ethics as a scientific discipline comprises three domains: normative ethics, metaethics, and applied ethics. Normative ethics concerns itself with how we ought to act. It develops **value-based ethical theories** to help answer this question. Metaethics, in turn, is concerned with normative ethics. The Greek word *meta* translates to “about” (i.e., “regarding”), so metaethics is essentially about normative ethics. It investigates the sources of value principles and critically examines certain fundamental assumptions of normative ethics. As part of practical philosophy, ethics is intended to be applied to everyday life. Applied ethics have been developed for almost every sphere of human existence – including medicine, business, media, and science – to help individuals make ethical decisions and take appropriate action in these fields (Düwell et al., 2011).

Digital health is an emerging and developing discipline, and as such, there is currently no established discipline of applied digital health ethics. However, various applied ethics, including bioethics, medical ethics, technoethics, digital ethics, and the ethics of digital transformation, are being applied to address ethical questions within digital health. Bioethics, which emerged in the late 1960s, has a broad scope that encompasses ethical considerations related to all living organisms, including animal welfare, environmental ethics, medical ethics, and public health ethics. Bioethical issues include topics such as abortion, euthanasia, organ donation, and cloning. Medical ethics, with roots dating back to the Hippocratic Corpus and the Hippocratic Oath, is considered a subset of bioethics and focuses on moral values and judgments as they apply to medicine and medical research (Düwell et al., 2011).

Argentinian physicist and philosopher Mario Bunge coined the term "technoethics" in the 1970s (Bunge, 1977). Bunge recognized the transformative power of technologies for societies and advocated for a heightened sense of moral and social responsibility among technologists. As articulated by Bunge: "Being morally ambivalent, technology should be controlled instead of being allowed to develop unbridled in the interest of whatever group can afford it. In other words, the technologist must be held not only technically but also morally responsible for whatever he designs or executes: not only should his artifacts be optimally efficient but, far from being harmful, they should be beneficial, and not only in the short run but also in the long term" (1977, p. 101). One of the most well-known domains of technoethics is technology assessment, which involves a precise analysis of technological development to evaluate as many consequences as possible, including those that may be hidden or long-term.

The German philosopher and information scientist Rafael Capurro introduced the term "information ethics" in 1981, which he later renamed "digital ethics" in 2009 (Capurro, 2009). Digital ethics deals with the ethical challenges posed by digital technologies and their impact on our societies and the environment. Capurro (2009) argued that digital ethicists should address topics such as privacy, information overload, internet addiction, the digital divide, surveillance, and other challenges of the digital age. In the last decade, the term "ethics of digital transformation" has gained broad acceptance, referring to the scientific examination of the ethical aspects of digital transformation in all areas of society. There is no clear distinction between the two concepts, and both terms are often used interchangeably.

4.2 Theoretical Approaches to Normative Ethics

Normative ethics aims to help us make the right decisions by providing answers to the question of what is wrong and right. Since the beginning of moral philosophy, many ethical theories have been developed, and several approaches exist to categorize them. Düwell et al. (2011) define three main approaches in normative ethics, each containing various ethical theories: teleological, deontological, and weak normative and contextualist approaches. The three approaches entail different paths to ethical decision-making. It is important to understand their basic foundations in order to put ethical theories into practice.

Deontological Approach

The term "deontological" originates from the Greek word *deon,* which means "obligation" or "duty." Deontological theories propose a system of moral duties and principles that assess the moral worth of actions based on these pre-established principles. Deontological ethics is present in several religions, as their principles are derived from a set of divine commands (Boone, 2017).

One of the most influential deontological theories, with extensive argumentative justification, is the formulation of the Categorical Imperative by the German philosopher Immanuel Kant (1724–1804; Kant, 2012). Instead of basing his moral principles on divine pronouncements, Kant's theory stems from the distinct capability of humans to reason, which empowers and separates us from other living beings. Kant's theory places personal autonomy, the capacity to deliberate and give oneself a set of moral directives, at the center of his ideas, which was revolutionary for his time: "Autonomy of the will is the property of the will by which it is a law to itself" (Kant, 2012, p. 47). Kant suggested the Categorical Imperative as a universal ethical principle and offered two formulations. The first formulation, also called the universalizability principle, is the following: "Act only on that maxim by which you can at the same time will that it should become a universal law" (Kant, 2012, p. xviii). There are three steps to applying the universalizability principle in practice:

1. Formulate your maxim (principle).
2. Envision a world in which that maxim is a universal law.
3. Decide whether you can consistently will that your maxim be a universal law.

Kant used examples to illustrate his idea. For instance, consider borrowing money from a friend with a promise to repay it but having no intention of doing so. By applying Kant's previous steps of moral reasoning, it becomes clear that such an action is motivated by selfishness and deceit, and it is not desirable for it to become a universal moral standard that everyone could follow without any moral qualms (Boone, 2017).

Kant's second formulation of the Categorical Imperative is the formula of humanity: "Act in such a way that you treat humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means" (Kant, 2012, p. xvii). According to Kant, humanity has ultimate worth and dignity (implying autonomy), the ability to act based on reason and principles, and the power to will oneself to action rather than merely following one's desires. In the earlier example, if one were to lie to their friend about repaying a debt, it would imply that they are using their friend merely as a means to an end and are prioritizing their own financial gain over their friend's worth and dignity as a human being. If we apply the formula of humanity, this action would be considered immoral. However, what if someone lied to their friend with good intentions? For Kant, any lie would still violate the principle of treating humanity as an end in itself, as it undermines the rationality of the person being lied to and reduces them to a mere means to an end. Therefore, Kant would consider any lie as immoral (Boone, 2017).

The strength of deontology and Kant's approach lies in a simple guiding principle that is easily applied. It is similar (though not equal) to the intuitive moral understanding of many people, also called the Golden Rule: "Do not do to others what you don't want done to you." However, it is easy to see that completely forgoing the consequences of actions often contradicts the intuitive moral evaluation of actions, which is a main critical point from a teleological perspective. Deontological ethics can lead people to act in ways that produce negative outcomes. Additionally, virtue ethicists emphasize overlooking other ethical dimensions – such as virtues and good character, emotions, and personal relationships – in deontological approaches (Boone, 2017).

Can Autonomous Systems Be Moral?

The Greek word *autonomy* derives from the words *auto*, meaning "self," and *nomos*, meaning "govern" or "rule." In philosophical discourse, autonomy is an ethical principle that pertains to an individual's ability to define their own laws or rules of behavior, reflect on them, and act accordingly. Autonomy is typically associated with a rational adult individual, and many ethical theories consider freedom of will and autonomy as necessary conditions for personal responsibility and moral obligation (Powers & Ganascia, 2020).

The term "autonomous systems" is used in the context of digital transformation to describe systems that operate without human intervention. These include self-driving cars, lethal autonomous weapon systems (LAWS), and robots. However, the use of the term "autonomous" in this context is misleading since the systems' actions are determined by rules given or imposed by designers and engineers. The systems themselves do not have self-derived intentions and human control and oversight over machines is essential. An autonomous car or weapon that creates and acts on the basis of its own intentions and rules would become unpredictable, uncontrollable, and potentially dangerous (Powers & Ganascia, 2020).

Therefore, when discussing current AI and autonomous systems, it is more appropriate to speak of automaticity rather than autonomy. To answer the question raised above, autonomous systems are not truly autonomous, and they cannot act morally. Some AI researchers believe that one day autonomous systems will be created that can act without human supervision and of their own volition, in which case they would need to be held responsible for their actions (Powers & Ganascia, 2020).

Teleological Approach

The Greek word *telos* means "fulfillment," "purpose," or "goal." The teleological approach refers to ethical theories that center around a particular purpose or goal that should be considered as good. A possible goal may involve striving for the best possible result in a given situation. Consequentialism is an example of a teleological approach to ethics, as it evaluates the morality of an action based on its consequences. According to this ethical school, there are no inherently right or wrong actions; rather, the morality of an action depends on its consequences (Boone, 2017).

One of the most influential teleological and consequentialist theories is utilitarianism, formulated by British philosopher Jeremy Bentham (1748–1832). The ultimate goal of utilitarianism is to maximize human happiness, which Bentham defines in both a hedonistic way, equating happiness with pleasure, and a quantitative way, as the sum of pleasure and suffering. The term "utilitarianism" is derived from the word "utility," seen as a property of any object through which it tends to produce pleasure or happiness. According to Bentham, an intellectual activity has no higher value than an activity associated with sensual pleasure. Utilitarianism aims to maximize pleasure and minimize pain. The fundamental criterion for evaluating actions and making decisions is based on the principle of maximizing the overall happiness of the largest possible number of individuals (Crimmins, 2021).

While the strength of the utilitarian guiding principle is its pragmatism, simplicity, and easy applicability, this approach also invites a certain amount of criticism. Utilitarianism and other consequentialist theories aim to produce the best outcomes of an action; therefore, they must rely on predictions. However, making accurate predictions is often impossible. Additionally, utilitarianism favors overall benefit even when it comes at a great cost to certain individuals. In this way, it disregards individual rights. If the amount of pain experienced by the hurt person exceeds the amount of pleasure experienced by the advantaged person, then the decision is morally wrong. On the other hand, if one person experiences a degree of pleasure greater than the other's pain or inconvenience, then the decision is morally good because there is an overall benefit. In a similar vein, utilitarianism, per definition, ignores the rights of minorities, which leads to injustice (Boone, 2017).

Virtue Ethics

Virtue ethics
Although virtue ethics is usually defined as a separate category, Aristotle's virtue ethics can be considered a teleological theory since the aim of virtue ethics is eudaimonia (Düwell et al., 2011).

Virtue ethical theories focus on individuals and their qualities, which is why they are also called character-based ethics. Instead of providing rules or principles for action, **virtue ethics** answers the question "How should I act?" by answering the question "How should I be?" A virtuous person possesses the moral virtues and can reliably make the right decision in any situation. One of the most prominent examples is Aristotle's virtue ethics. The goal of life, Aristotle argues in *Nichomachean Ethics*, is *eudaimonia*. Etymologically, *eudaimonia* derives from the adverb *eu*, which means "good," and *daimon*, which can be translated as "spirit" or "destiny." In contrast to hedonistic happiness, *eudaimonia* is a sense of fulfillment or flourishing derived from the pursuit of meaning and virtue in life, which naturally creates a long-term feeling of well-being or happiness. To achieve *eudaimonia*, one should cultivate virtues and develop a good character by practicing virtue as a way of life. A virtuous person should not only study the virtues, but must also act on them and do good things in practice (Boone, 2017)

Aristotle defined nine virtues that he considered among the most admirable and desirable for individuals seeking *eudaimonia*. These virtues are courage, temperance, wisdom, justice, prudence, honor, friendship, wit, and magnanimity. Aristotle believed that cultivating these virtues would lead to a virtuous life, which in turn would result in *eudaimonia*.

Aristotle's concept of the virtuous life has also influenced the foundations of positive psychology, particularly in the work of American psychologist Martin Seligman, who formulated the idea of "authentic happiness" and outlined the necessary components for achieving it. He developed a list of 24 strengths organized into six virtue categories: wisdom and knowledge, courage, humanity and love, justice, temperance, and transcendence (Schwartz & Sharpe, 2006).

Care Ethics

One of the forms of virtue ethics (though it is also seen by some theorists as a completely separate branch of ethical thought [Held, 2006]) that is especially relevant for healthcare is care ethics. Care ethics is a contemporary ethical theory that emerged as a response to and criticism of male-dominated ideas in philosophy in the 20th century. Many traditional ethical theories emphasize rationality and exclude women from the sphere of moral judgments. For example, Aristotle argued that women are less rational than men and that men should rule them as a result (Stauffer, 2008). Similarly, Kant argued that women lack self-determination, are incapable of reasoning and moral responsibility, and therefore should not have a voice in public life (Mikkola, 2011). Furthermore, Kant argued that only the intention to act according to rules derived from reason makes humans moral, and that compassionate, sympathetic feelings cannot lead to moral action (Kant, 2012).

Care ethics is a feminist approach that seeks to incorporate traditionally feminized virtues and values that are rarely represented in traditional schools of ethics (Held, 1990). Care ethicists criticize patriarchy as a heteronormative culture that is based on the presumed complementarity of men and women and splits human traits between the masculine (rationality) and feminine (feelings and emotions). They argue that the patriarchal system favors masculine traits, elevating some men over others and all men over women (Gilligan & Snider, 2018). Care ethicists support their views with the latest discoveries of neuroscientist Antonio Damasio and others in his field, who postulate the essential connection between emotions and thoughts as a sign of healthy human development, while the absence of this connection can indicate relational trauma (Damasio, 2006). In care ethics, morality is less a matter of rationality and recognition and has more to do with compassion and empathy that lead one to take responsibility for others in need.

The beginning of care ethics was marked by the pioneering essay “Maternal Thinking” by American philosopher Sara Ruddick, who explored women’s experiences in caring practice and shed light on the values inherent in this activity (Ruddick, 2002). Ruddick suggested that these values offer a new perspective on morality. Individuals of all genders who do not engage in caring activities may fail to recognize the intrinsic morality of these practices. The psychologist Carol Gilligan, a research assistant of Harvard psychology professor Lawrence Kohlberg, contributed to care ethics with her 1993 book *In A Different Voice: Psychological Theory and Women’s Development*. While studying moral development in children with Kohlberg, Gilligan criticized Kohlberg’s judgments as androcentric, suggesting that he diminished female ethical considerations based on relationships and caring.

The perspective of care ethics sees human beings as primarily relational, not just rational beings, which makes care ethics unique. Human beings are viewed as interdependent, and caring is the hallmark of ethical action. According to care ethics, to live an ethical life is to care about those with whom we are in close relationships (Held, 2006). Virginia Held formulated the central idea of care ethics in a famous quotation: “Caring, empathy, feeling with others, being sensitive to each other’s feelings, all may be better guides to what morality requires in actual contexts than may abstract rules of reason or rational calculation, or at least they may be necessary components of an adequate morality" (Held, 1990, p. 332).

The weakness of virtue and care ethics is the absence of clear guiding principles that help to determine the appropriate course of action, which makes it difficult to apply. However, the strength of virtue ethics lies in its consideration of the complexity of life. Moral questions in care ethics are not solely resolved by an inflexible rational principle, but also by considering the influence of feelings and emotions. It focuses on the intentions and traits of the person performing the action and not just the action itself or its consequences. Care ethics emphasizes humans as relational and interdependent beings and highlights values, such as compassion and empathy, that are crucial for healthcare.

Can Robots Care?

In industrial societies, the proportion of the aging population is continuously rising, while the birth rate, on average, is declining. One currently discussed solution to manage the challenge of aged care is the use of robots. Japan, the country with the highest **old-age dependency ratio** among Organisation for Economic Co-operation and Development (OECD) countries (51 percent in 2021), is leading in investments in the development of care robotics and assistive technologies for aged care (The World Bank Group, 2021b). In Europe, the situation is similar, with a projected old-age dependency ratio of 57 percent by 2100 (Eurostat, 2020). While the demographic development in the United States is more balanced, the US is actively promoting the growth of the market for robotic assistive technologies.

Old-age dependency ratio
This is defined as the ratio of the number of elderly people (aged 65 years and over) to the number of people of working age (15–64 years).

The applications of robots in aged care are manifold. The use of service robots performing non-caring tasks, assisting nurses in their daily routines under proper supervision and with appropriate guidelines, generally finds acceptance. However, the use of care and social or companionship robots is highly disputed. From a psychological point of view, caring is a relational and emotional activity at its core. Ideally, caring should include empathy and authentic concern for the feelings and thoughts of another human being. Although care or social robots may appear human-like, friendly, and caring, they don't possess the same emotional capabilities as living beings (Bertolini & Arian, 2020).

Engineers and designers put a great deal of effort into simulating human qualities in robots by making them look and sound like people. However, the growing tendency to anthropomorphize robots in design is controversial. Robots and assistive technologies that are designed to resemble and behave like humans often imitate female companions, while the majority of robot creators are male. This simulation and adaptation of desirable female characteristics from a male perspective for technological tools can contribute to the objectification of women. Technoethicist Danit Gal investigated the use of robots in South Korea, China, and Japan. She argues that such development and design decisions might lead to problematic views towards women and emotionally and psychologically confuse users. Gal introduced the term "Anthropomorphized Tools Paradox," which describes a functional and emotional paradox whereby design mimicking humans is considered desirable, but forming an emotional response to such mimicry is regarded as problematic. According to Gal, female objectification and the Anthropomorphized Tools

Paradox are examples of problematic technology development in certain Asian countries. She asks, “What degree of AI and robots’ socialization capability development is considered ‘antisocial’? How many human functions can and should we substitute before we hit that threshold?” (Gal, 2020, p. 623).

Nursing robot Grace, with her child-like face, large hazel eyes, and lovely smile, serves as an example of the Anthropomorphized Tools Paradox. Grace is designed to work with vulnerable seniors suffering from psychological conditions, such as depression or dementia. In the article “Do Robots Care?” Bertolini and Arian challenge the idea of robot–human relationships. The authors argue that, while robots may be able to evoke an emotional response from users by imitating human behavior, this behavior is ultimately deceptive. They warn that care robots can cause seniors to form purely delusional relationships, which could lead to disengagement from reality and threaten their well-being. In addition to deception, they warn that the use of care robots may lead to social isolation and a reduction in human contact among the elderly. Following Kant’s humanity formula, the authors argue that the incorrect use of assistive technologies could violate human dignity (Bertolini & Arian, 2020).

Weak Normative and Contextualist Approaches

Ethical approaches that do not offer a singular moral principle but instead rely on mid-level principles as criteria for guiding ethical action are considered weak normative approaches and are known as coherentism or principlism. In this context, the term “mid-level” refers to the fact that these principles serve to bridge the gap between fundamental ethical convictions and specific moral problems. The mid-level status of these principles permits some flexibility, which is both a strength and a weakness of these approaches. It may be necessary to weigh different principles against each other when circumstances cause them to come into conflict (Düwell et al., 2011).

In the context of bioethical discourse, an ethical framework called principlism has become the dominant approach in the field today. Principlism was developed by Tom Beauchamp and James Childress in their groundbreaking book *Principles of Biomedical Ethics* and has played a crucial part in establishing and defining bioethics (Beauchamp & Childress, 2013). Principlism is an applied ethics approach used to investigate moral dilemmas through the application of certain ethical principles. The approach of Beauchamp and Childress is based on the idea of common morality as a foundation for their principles. They assume that human morality has a common denominator or a coherent core: “We will call the set of universal norms shared by all persons committed to morality the common morality. … The common morality is applicable to all persons in all places, and we rightly judge all human conduct by its standards” (Beauchamp & Childress, 2013, p. 3).

Beauchamp and Childress specify the common morality according to four moral principles:

1. Respect for Autonomy: duty to foster a patient’s autonomous decision-making
2. The Principle of Nonmaleficence: duty not to inflict harm

3. The Principle of Beneficence: duty to do and promote good
4. The Principle of Justice: obligation to craft an allocation system that distributes scarce resources in a fair and just manner

Applying the four principles in concrete decision-making requires specification and balancing. First, it is necessary to determine which principles are relevant in a particular context, which is done via specification. Secondly, the different principles have to be balanced out. For example, suppose a patient requires a life-saving medical procedure but refuses treatment. In this case, the healthcare professionals must decide whether beneficence or respect for autonomy should take priority. Ideally, a medical practice is regarded as ethical if all four principles are applied equally (Beauchamp & Childress, 2013).

Principlism's greatest strength lies in its pluralistic character: It operates according to moral principles shared among various moral traditions and combines elements of different ethical theories into a unified framework. However, the ethics of principlism implies that collisions must be expected when competing principles come into opposition and it is necessary to clarify how the principles are to be balanced. Since there is no overriding moral principle that could provide orientation in such conflict cases, the only possibility is to appeal to the judgement of the individual, which is not guided by further explicit criteria. Shea argues that the absence of the concept of good is a major problem for principlism and that any acceptable and cohesive bioethical approach must contain a value theory to address the issues of specification and balancing (Shea, 2020).

4.3 Methods for the Ethical Evaluation of Digital Health

Ethics is meant to be applied to everyday life to help us make better decisions. There are two components necessary for an ethical assessment of digital health interventions: a normative ethical basis and a practical assessment methodology. In research and practice, a wide variety of methods for ethical assessment exist, and the majority are principle-based.

The evaluation methods can be divided into three categories "ex ante," "intra," and "ex post." The ex ante evaluation methods apply ethical appraisal at the start of the technological development before any concrete design decisions are made. These methods work with hypothetical scenario approaches and are mostly applied by ethicists or foresight specialists. The intra methods are meant to be practiced during the development, design, and testing stages. In this case, the ethicists work with designers and technology developers to identify ethical issues early on and change the direction of development if needed. The ex post methods offer practical tools to evaluate already-existing technologies. Due to the focus on competitiveness and profitability within profit-driven corporations, technological development is often pursued with short development cycles, which in turn leads to a minimal application of ex ante and intra ethical evaluation methods. As a result, ex post methods are the most commonly used ethical evaluation methods in practice (Reijers et al., 2018).

Ethical Evaluation of Digital Public Health

German medical ethicist Georg Marckmann (2020) has developed a practical methodology for assessing **digital public health** interventions based on technoethical considerations and the bioethical principlism approach of Beauchamp and Childress. To conduct a comprehensive ethical analysis of digital public health interventions, Marckmann proposed eleven ethical criteria that should be applied in chronological order:

Digital public health
This field focuses on the population's health as the subject of governmental regulation and support, including the promotion of health and prevention of diseases among the populace at large (Marckmann, 2020).

1. Functionality

The first criterion, functionality, requires a clear definition of the goal to be achieved with a digital health intervention. This criterion helps examine to what extent the desired goal can be realized (technical effectiveness) and whether the evaluated solution is technically efficient (technical efficiency). The ethical justification is derived from the nonmaleficence, beneficence, and means-end-rationality principles.

2. Alternatives

Marckmann recommends examining whether other alternatives (including those of a non-technical variety) exist to achieve the desired goal and whether these alternatives may have advantages in terms of the ethical criteria. The ethical justification is derived from the principle of means-ends rationality.

3. Potential benefit

This criterion embraces beneficial effects on patients' morbidity, mortality, and quality of life. Here it is important to consider the amount and reliability of the available evidence. The ethical justification is derived from the principle of beneficence.

4. Potential harm

Like almost all diagnostic or therapeutic interventions, digital technologies for public health may cause harm to patients (health risks and side effects). Potential harm must be balanced against potential benefits in the ethical assessment. The ethical justification is derived from the principle of nonmaleficence.

5. Self-determination

This criterion embraces three areas: availability and/or promotion of digital and health literacy, the possibility for participants to make an informed decision, and freedom of choice. To be able to benefit from digital public health interventions, the digital health literacy of participants should be ensured or promoted in order to tackle the challenge of the digital divide. As with all public health interventions, the participants should be able to make informed decisions. Finally, the participants' freedom of choice should be guaranteed. The ethical justification is derived from the principle of respect for autonomy.

6. Data privacy

Health data contain highly sensitive information that should be adequately protected. This criterion should ensure that participants in digital health interventions are able to control their data. They should be informed of the planned scope of the use of their data and should be able to give their consent. In addition, appropriate procedural and technical precautions for adequate data protection should be guaranteed. Ethical justification is derived from the principle of respect for autonomy.

7. Data security

Data security ensures that digital health data are protected from unauthorized access and misuse. Ethical justification is derived from the principle of nonmaleficence.

8. Justice

Ethical implications of justice play a central role in public health interventions:

a) Non-discriminatory access should be guaranteed, and the possible socioeconomic access barriers to the intervention should be evaluated.
b) The expected benefit and damage potential should be fairly distributed among the population.
c) Interventions should contribute to reducing health inequalities.

Ethical justification is derived from the principle of justice.

9. Efficiency

Given the limited resources in the healthcare sector, the efficiency of digital health interventions should be examined. The incremental cost–benefit ratio should be determined, whereas non-technical alternatives should be included as a comparison. The ethical justification is derived from the principle of distributive justice and the means–ends rationality.

10. Responsibility

Especially for autonomous and AI-driven decision support systems, the questions of attribution of responsibility should be clarified in advance. Ethical justification is derived from the principle of nonmaleficence.

11. Legitimacy

Public health interventions affect the well-being of the entire populace or a special target segment. Thus, the application of digital health interventions should be discussed in a fair decision-making process by a correspondingly legitimized decision-making authority. Ethical justification is derived from principles of justice and respect for autonomy.

Principle-based ethical frameworks offer some flexibility, which is both an advantage and a difficulty, as different principles may need to be weighed against one other. Between the individual criteria proposed by Marckmann exist complementary and competitive relationships. Functionality, for example, is a prerequisite for a large potential benefit and a small potential harm. On the other hand, the criteria of harm potential and efficiency are in a competitive relationship, since measures to increase security often require considerable resources, decreasing digital applications' overall efficiency. The criteria need to be balanced out in a deliberation process with different stakeholders and experts (Marckmann, 2020).

For the practical application of the ethical framework, Marckmann (2020) suggests the following six steps:

1. Description: First, all pertinent information concerning the ethics of digital public health intervention should be gathered and examined as closely as possible.
2. Specification: The evaluation criteria should be specified accordingly to the examined digital health intervention. For example, the use of AI-driven decision support systems may raise questions about biases of the underlying database.
3. Individual assessment: This step includes assessment of the digital health intervention regarding the eleven criteria.
4. Synthesis: This step entails an overall ethical assessment of the intervention through synthesis and weighing the individual ratings from the third step.
5. Recommendation: The crucial element of this step is the formulation of recommendations for the ethically justifiable development and/or application of digital public health intervention.
6. Monitoring: Emphasis is put upon regular observation and evaluation of the ethical implications and, if necessary, revision of the developed recommendations.

The strength of the introduced ethical framework is its normative groundedness in bioethical and technoethical ethics, which is widely accepted in healthcare, as well as appropriate consideration of critical technical aspects of digital technologies. The framework can be used as an ex ante, intra, or ex post method. It can be adapted from the consideration of a public digital health intervention to an individual digital health intervention by omitting criteria relevant for public health. However, the care ethics perspective, which places emphasis on the relational aspect of care, is not adequately addressed in Marckmann's approach, similar to other widely used ethical assessment tools for digital technologies. One of a few exceptions to this is the multi-dimensional Model for Ethical Evaluation of Socio-Technical Arrangements (MEESTAR), created for technologies applied in aged care by German philosopher and theologian Arne Manzeschke. MEESTAR is principlism-based and requires systematic consideration of seven criteria on an individual, organizational, and social level: care, autonomy, safety, justice, privacy, participation, and self-conception (Manzeschke & Rother, 2013). While the MEESTAR framework has been discussed in research and used in a few specific case studies, it is not widely known to have been applied in practice in a broad range of contexts. However, the framework has been influential in the development of other ethical evaluation frameworks for socio-technical systems.

4.4 Ethics and Soft Law: European Ethics Guidelines for Trustworthy Artificial Intelligence

In the course of the increasing globalization of politics and the shift of decision-making power from the nation-state to international organizations or international political systems, soft law governing mechanisms have become increasingly widespread. In contrast to hard laws, soft laws define standards and recommendations that are not legally binding. In technological domains, self-regulatory measures and soft laws are widely applied to ensure the innovativeness of business stakeholders.

In recent years, many efforts have been made concerning the ethical design and governance of AI technology. Over a hundred ethical frameworks, assessment lists, recommendations, and guidelines have been issued (Floridi, 2019). Considered a milestone, the first widely publicly accepted standards, the Ethics Guidelines for Trustworthy Artificial Intelligence, were presented by the European Union (EU) in 2019. As the foundation of these guidelines, experts emphasized human rights protection. The guidelines, developed in part with the support of a public consultation process, laid out a set of seven criteria that AI systems should satisfy in order to be considered trustworthy (European Commission, 2019):

1. Human agency and oversight: AI systems should enable users to make informed decisions and preserve their fundamental rights. Additionally, proper oversight measures must be guaranteed.
2. Technical robustness and safety: AI systems need to be safe, secure, and reliable.
3. Privacy and data governance: Adequate privacy, data protection, and data governance mechanisms need to be ensured.
4. Transparency: This is a crucial component of the data, system, and AI business models. AI systems and the decisions related to their design and operation should be understandable for users. People must be informed of the fact that they are interacting with an AI system and should also be aware of the system’s capabilities as well as its limitations.
5. Diversity, nondiscrimination, and fairness: Steps must be taken to avoid discriminatory bias in all instances. In an effort to foster diversity, AI systems should be accessible to all people, regardless of their personal situation or ability, and must take into account the opinions, feedback, and input of relevant stakeholders at all stages.
6. Societal and environmental well-being: The positive impact of AI systems should be sustainable. All people, future generations included, should benefit from AI technology with consideration for environmental ecology and other living beings.
7. Accountability: Mechanisms should be put in place to ensure responsibility and accountability with regard to AI systems and their outcomes.

The Ethics Guidelines for Trustworthy Artificial Intelligence was an important initial step toward implementing and governing ethical AI. However, the practical impact and adoption of these guidelines remain uncertain. As they are voluntary and lack external independent oversight and enforcement, their effectiveness has been called into question. Consequently, EU regulators have taken additional steps by proposing the AI Act.

SUMMARY

This unit provided an overview of the key terms and concepts of normative and applied ethics, focusing on the main Western normative ethical schools: teleology, deontology, virtue ethics, and bioethical principlism. Applying ethics to everyday life gives it meaning. When philosophical concepts, such as autonomy, are applied in the digital context, it becomes evident that current autonomous AI systems cannot act morally. Therefore, responsibility for the outcomes of AI actions should be borne by their creators. The bioethical principlism traditionally applied in medical care has also found application in the context of digital health, as proposed by Marckmann with his evaluation framework for digital public health interventions. In contrast to traditional moral theories grounded in rationality, such as those of Kant, contemporary care ethics emphasizes compassion and empathy, which are especially important in health and aged care. Care ethics may enrich the current discourse regarding care and social companionship robots for seniors, children, or people who have health conditions or impairments, supplementing principle-based approaches. The societal risks of advanced technologies, such as AI, have been intensely discussed in research, media, and politics in recent years, leading to a resurgence of ethical discourse in the context of technology. To promote the benefits of digital transformation and reduce its risks, ethical guidelines and recommendations in the form of soft laws have been developed for trustworthy AI in recent years, as well as practical ethical assessment methodologies for digital health interventions.

UNIT 5

RISKS AND CHALLENGES OF DIGITAL HEALTH

STUDY GOALS

On completion of this unit, you will be able to ...

- describe the main risks of digital technologies for healthcare.
- explain the necessity of legal frameworks alongside ethical frameworks.
- discuss the differences in the legal approaches of the European Union and the United States regarding digital technologies.

5. RISKS AND CHALLENGES OF DIGITAL HEALTH

Introduction

Digital transformation entails both potentials and risks that need to be considered to assess the societal impact of technological development. This unit looks at the main risks and challenges related to digital health: artificial intelligence (AI) reliability, data security, data privacy, deprofessionalization, and deskilling. Ethical reflection about digital transformation in healthcare can yield valuable insights; however, voluntary and soft approaches are limited in their impact. Additionally, there are many ways of practicing "decorative ethics" on a surface level without effecting substantial improvement, as we will discuss in this unit. As big economic interests are at play in the digital technology sector, something more than ethics is required. In the conclusion of her book, AI researcher Kate Crawford states that "we must focus less on ethics and more on power" (Crawford, 2021, p. 224). The source of power lies in laws with binding force and practical effects. Given that the European Union (EU) and United States (US) are leading international actors in technological development and usage, this unit will explore their respective strategies and initiatives in governing data privacy and AI.

5.1 Risks of Digital Health

In 2021, the digital health expert and keynote speaker Bertalan Meskó published a short article entitled "Top 10 Dangers of Digital Health." The article addresses various aspects of hazard, encompassing both exotic risks, like bioterrorism in an interconnected world, and more practical concerns, like the harmful effects resulting from the increasing prevalence of self-diagnosis through technology. While the article could benefit from a clearer prioritization of the ten hazards and an assessment of their potential harm, it is worth examining some of the issues raised by Meskó. His list includes the main risks that arise from digital health, such as unreliable AI, threats to data security, and data privacy.

Unreliable AI

According to Meskó (2021), there are two main risks associated with AI as currently applied in healthcare: inadequate regulation of **adaptive AI** algorithms and insufficient testing of AI in real-life clinical settings. With regard to the first issue, Meskó articulates the most urgent problem in AI research and practice, referring to discriminatory gender, age, ethnicity, and other biases. Although he explicitly mentions adaptive AI, the bias problem refers in particular to all AI tools applied in sensitive societal areas, such as healthcare, justice, or education systems.

Adaptive AI
This form of AI combines machine learning (ML) reinforcement learning techniques and agent-based technology in order to learn, adapt, and respond in real time in changing environments.

One prominent example of data bias problems in healthcare is the case of International Business Machines (IBM) Watson Health AI. In the past decade, Watson Health AI, in which IBM invested $4 billion, was expected to revolutionize healthcare by diagnosing patients and recommending treatment. Initially, Watson Health AI achieved noteworthy successes in oncology, and its treatment recommendations aligned with those of physicians (Jiang et al., 2017). However, over the years, Watson provided incorrect and unsafe suggestions in cancer diagnostics caused by the unrepresentative data set delivered by a small group of doctors in a single hospital where it was trained. Hospitals canceled their cooperation with Watson, and in 2022, IBM sold it for parts to an investment firm (IBM, 2022).

Data bias is one of the contributing factors to the ML reproducibility crisis, which means that significant findings from ML have failed to be reproduced. The reason for this is straightforward: If the data used to train an ML model contain biases, the resulting model will also be biased. Kapoor and Narayanan (2022) investigated the limits of computational prediction and found that reproducibility shortcomings and failures had been documented in 329 studies across 17 areas, including medicine. In a recent review of 62 studies dedicated to ML models for the detection and prognosis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as COVID-19) using chest radiographs and computerized tomography (CT) scans, the researcher's team found that none of the reviewed ML models was suitable to be applied in a clinical setting. The reasons were data bias, methodology problems, and reproducibility failures (Roberts et al., 2021).

Meskó also raises the issue of insufficient testing of AI tools in a real healthcare environment, which undermines their reliability for healthcare professionals and patients. While AI tools may show great results in the laboratory, they can still fail when incorporated into a clinical setting. Meskó highlights the problem of the practical use of AI tools, reporting the case of a Thai hospital. The eye-screening AI tool for diabetic retinopathy developed by Google caused various practical difficulties for hospital nurses. These ranged from minor issues, like problems with the internet connection, to more significant ones, such as the algorithm's limited acceptance of scans of a certain quality. Nurses had to invest extra time in editing some of the scans that the tool refused to analyze (Meskó, 2021).

Threats to Data Security

Each new level of digital transformation brings forth new threats to data security. As our ~~society~~ becomes more digitalized, our vulnerabilities to cybercriminals increase accordingly. According to the Check Point Research (CPR) report, global cyberattacks increased by 38 percent in 2022 compared to 2021. The healthcare sector, which deals with some of the most sensitive human goods – namely, life and health – is especially vulnerable to cyberattacks. In fact, CPR reported the largest increase in cyberattacks in healthcare compared to other industries in 2022 (Check Point Software Technologies, 2023).

Serious risks facing digital health include the possibility of hospitals being targeted by ransomware attacks and medical devices being remotely hacked (Meskó, 2021). A ransomware attack is a financially motivated attack by cybercriminals using malicious software to encrypt the targeted data. Hackers hold those data hostage until a ransom, usually in the form of cryptocurrency, is paid. The healthcare sector is one of the most popular targets because the probability of ransom payment is high. Healthcare organizations must restore patient data access as soon as possible to provide continuous patient care. Due to their sensitivity, health data are one of the most valuable commodities on the black market.

Ransomware attacks have been documented since the digitalization of healthcare began. The 2015 attack on Elevance Health (formerly WellPoint) was one of the largest healthcare data breaches ever recorded. The attackers were able to gain access to 79 million sensitive records, including patient names, addresses, birth dates, social security numbers, and medical histories. Elevance Health was forced to pay $115 million to get the data back. According to a recent study, attacks on healthcare providers more than doubled from 2016 to 2021, disclosing the personal health information of approximately 42 million patients (Neprash et al., 2022). The authors suspect that their findings may not accurately reflect the current threat level, as data regarding cyberattacks are often underreported. Hospitals are not the only target of cyberattacks, as smaller healthcare providers – such as ambulatory surgical centers, mental health organizations, and dental practices – are also threatened. Cyberattacks in healthcare are dangerous as they affect the safety of patients: According to Neprash et al. (2022), 44 percent of attacks resulted in care delivery disruptions. Alarmingly, 8.6 percent of these disruptions lasted for more than two weeks, underscoring the critical need for healthcare organizations to prioritize cybersecurity measures to protect patients from the potentially devastating consequences of cyberattacks (Neprash et al., 2022).

A more significant threat than stolen or encrypted data are attacks that target human lives. In healthcare and aged care, professionals and patients are becoming increasingly reliant on the comfort of Internet of Things (IoT) devices. Cybersecurity experts declared 2022 an inflection point in cybersecurity due to the rapid expansion and vulnerability of the IoT over the last decade. In addition to the threat of stolen data, experts now warn of the risk of device disruption. Cars and medical devices, which are especially critical to everyday life, are believed to be preferred entry points for cyberattacks, according to experts' estimates (MacBride, 2023).

Threats to Data Privacy

Privacy is one of the most widely discussed and controversial topics of digital transformation. In healthcare, privacy is even more urgent, as health data are especially sensitive. These data are being generated and gathered at all stages of digital health. Patients and consumers of digital health services must be able to trust that their collected data will remain secure and be used appropriately. Unfortunately, there are numerous examples of health data being used in ways that violate personal privacy requirements.

Business models that rely on the inappropriate use of data are common and problematic. For decades, researchers have highlighted privacy concerns in digital health, particularly in the realm of mobile health, where absent or insufficient privacy disclosures have been repeatedly reported. In 2014, Blenner et al. investigated the privacy policies and permissions (determining which apps are allowed to access the device) of all diabetes apps available in the Google Play Store. The results were alarming: 81 percent of the apps lacked any privacy policy at all. The analysis of the remaining 19 percent revealed that the majority, regardless of the privacy policy, collected and shared data with third parties (80.5 percent collected user data and 48.8 percent shared this data) (Blenner et al., 2016). Since the introduction of the General Data Protection Regulation (GDPR) in 2018, the Google Play Store has required the inclusion of app privacy policies. However, a recent study indicates that inconsistent privacy practices in mHealth persist outside of Europe. Researchers analyzed the privacy policies of 20,991 Android health and fitness apps available in the Google Play Store Australia. Among the mHealth apps, 88 percent could access and potentially share personal data. Third parties were involved in most of the data collection and data transmission operations in the investigated apps (Tangari et al., 2021).

Data transfer to third parties does not guarantee appropriate use, as recently demonstrated by the 2023 report, published by Joanne Kim of Duke University, entitled *Data Brokers and the Sale of Americans' Mental Health Data: The Exchange of Our Most Sensitive Data and What It Means for Personal Privacy*. The report is the result of a two-month-long study that examined the practices of data brokers who sell and exchange mental health data from mHealth users in the United States. Because most mHealth apps are not covered by the Health Insurance Portability and Accountability Act (HIPAA), private companies that operate these apps are not obligated by law to protect user privacy. The report reveals the prevalence of unethical practices in the management of mental health data, particularly with regard to privacy infringement and insufficient scrutiny of buyers. For instance, one company disclosed the names and home addresses of individuals with conditions such as depression, anxiety, post-traumatic stress, or bipolar disorder. To protect the privacy of consumers, researchers propose implementing comprehensive federal privacy legislation and prohibiting the sale of mental health data on the open market (Kim, 2023).

In 2019, the American economist Shoshana Zuboff introduced a new term, "surveillance capitalism," to describe the current capitalist production and commodity exchange. Under the framework of surveillance capitalism, consumers of digital services are viewed as sources of data, or as Zuboff calls them, "raw materials." Companies monitor users and analyze the collected data using ML techniques to derive behavioral predictions. These predictions are subsequently sold to interested parties, such as advertisers, retailers, insurers, service providers, and others. Business models based on the sale of predictions are becoming increasingly lucrative compared to those for traditional products and services. Although many people would prefer not to share their data, it's difficult to refuse to use digital tools given their pervasiveness and status as a daily necessity. Big tech companies in monopolistic positions often exploit users' dependency on their services, which can undermine personal autonomy and democracy (Zuboff, 2019).

The relationship between digital health and privacy is complex. Improved well-being and health outcomes, by definition, come at the expense of privacy, as digital health services and devices require consumer data to be effective. However, in accordance with universal human rights, personal autonomy must be protected. Patients and digital health consumers should be able to give explicit consent regarding the collection, use, and potential transfer of their health data to third parties. It is crucial that digital health developers, providers, and consumers engage in a transparent discussion about the trade-offs between the health benefits and privacy risks of digital health.

Deprofessionalization and Deskilling

Deprofessionalization is a social process whereby highly skilled professionals are replaced by less skilled workers, while deskilling refers to the transformation of highly skilled professional work into unskilled work. The issues of deskilling and deprofessionalization have been discussed by experts for years, who point out that powerful AI systems require less skilled human labor assisting in the construction, maintenance, and testing of these systems (Crawford, 2021; Gray & Suri, 2019). Human labor in the digital age includes repetitive tasks, such as labeling training data and sorting out disturbing or harmful content in the data used to train AI systems. Currently, this type of work is not widely known to the public. Researchers Mary Gray and Sid Suri coined the term "ghost work" in their book to refer to the underpaid crowdworkers or microworkers, typically from low-income countries, who perform these tasks to serve AI systems (Gray & Suri, 2019). Recent investigations, such as the *Time* report on the manual sorting of ChatGPT training data in Kenya, have shed light on the unethical practices of this labor. Experts in various industries have expressed concern that the increasing role of AI in industry could result in the reduction of white-collar jobs and the expansion of low-paid jobs that serve large AI systems (Elliot, 2023). The digital transformation of the healthcare sector is still in its infancy, and it remains uncertain to what extent deprofessionalization will impact this sector and when.

While AI systems can support and empower physicians with additional algorithm-generated insights and precise surgical techniques, the dissemination and utilization of AI applications for core human tasks, such as caring, nursing, or psychotherapeutic work, pose specific new challenges. The goal of increased efficiency is one of the factors driving the digitalization of activities that were previously thought to be impossible to automate, such as caring, comforting, being attuned, and being empathic. An example of this trend is emotion AI, which aims to recognize the emotional states of humans through verbal and non-verbal cues and mimic empathetic interaction based on this analysis. Emotion AI uses various techniques, such as natural language processing, sentiment analysis, and facial recognition, to detect emotions expressed through text, voice, or images. These systems are suggested for use in healthcare, such as for mental health and well-being chatbots, care and companionship robots, and assistive technologies. Additionally, they can be employed during telemedicine appointments to give physicians more information about the patient's emotional state or monitor patients in a waiting room for signs of distress to identify those in need of urgent care (Meskó, 2023). The automation of core human tasks could potentially lead to a reduction in the amount of human labor required for some caring, nursing, or psychotherapeutic duties, which could in turn lead to a reduction in the need for highly skilled professionals. Additionally, the adoption of emotion AI could lead to a shift in the role of mental health, care, and nursing professionals from providing emotional support and guidance to overseeing the use of emotion AI systems.

This development is concerning. Decades of psychological research have shown that empathy and authentic concern for the feelings and thoughts of another human being are essential qualities in fields such as aged care, childcare, healthcare, and mental health professions. The current state of empirical research on psychotherapy suggests that the quality of the therapeutic relationship – referred to as the therapeutic alliance – and not the psychotherapeutic technique or school, is the strongest impact factor in the healing process (Martin et al., 2000). As discussed previously, robots and chatbots equipped with emotion AI can mimic empathetic human behavior and evoke a reaction in users. However, this impression is an illusion, and the established relationship has a deceptive element. The potential healing effects of such relationships, especially in light of the problems of inaccuracy and data bias, are uncertain and further investigation is needed to better understand the potential risks and harms they may pose.

Current research shows that many people only accept technological tools in core human areas as a supplement to human care. In a study concerning the acceptance of digital mental health interventions in Germany, conducted with 1,984 participants, Phillips et al. (2019) found a strong preference for face-to-face contact with a psychotherapist when using a digital mental health tool (blended care). This preference remained consistent across participants of different socioeconomic backgrounds, regardless of the presence or absence of past psychotherapy experience (Phillips et al., 2019). These findings align with previous studies in which participants expressed similar views, stating that digital mental health interventions, even with some guidance from humans via messages or calls, were not as effective as face-to-face psychotherapy (Musiat et al., 2014).

However, for different reasons, some people may enjoy or begin to accept digital tools, robots, and chatbots as substitutes for healthcare professionals or social companions. In her book *Alone Together: Why We Expect More from Technology and Less from Each Other*, technology and society researcher Sherry Turkle delves into the societal and ethical risks that come with depending on digital technologies in the social and personal realms (Turkle, 2017). Turkle, along with other researchers, warns that it's possible for digital companions to provide such a charming and enjoyable imitation of human company that people might be tempted to prefer their easygoing (yet artificial) companionship over engagement with a real, flawed human (Donath, 2020). She argues that, despite being constantly digitally connected, humans still suffer from alienation: "Technology is seductive when what it offers meets our human vulnerabilities. And it turns out we are very vulnerable indeed. We are lonely but fearful of intimacy. Digital connections and sociable robots may offer the illusion of companionship without the demands of friendship" (Turkle, 2017, p. 1).

The use of social, care, or sex robots and chatbots raises ethical and psychological questions related to how the increasing use of digital technology in relational domains might impact our communication and emotional connection skills. Ongoing research in this area is needed to provide evidence that can support a people-centered approach to digital transformation.

5.2 Are Soft Laws Enough?

The Ethics Guidelines for Trustworthy Artificial Intelligence represented a significant contribution to the AI ethics discourse. However, the practical impact of these guidelines depends on their implementation, which falls under the responsibility of the engineers who develop AI. While the EU guidelines provide a valuable starting point, they remain entirely voluntary and lack external, independent oversight and enforcement mechanisms. The dismissal of top AI ethics researchers Timnit Gebru and Margaret Mitchell by

Google in 2021 made headlines. These researchers investigated the bias tendency and environmental costs of Google AI systems, topics that were not new but still controversial. The insiders revealed that Google's public relations (PR) team and legal advisors oversaw the publication of research articles on sensitive topics, such as sentiment analysis, facial recognition, and the categorization of gender and race. This recent incident raises concerns that ethics research at big tech companies like Google may be adversely influenced by corporate interests. It also casts doubt on the ability of these companies to self-regulate their technology without external oversight (Reuters, 2021).

Five Risks of Being Unethical

In order to address unethical practices in any field, one should be aware of how and in what guise they may occur. Italian information philosopher Luciano Floridi has examined the issues of applied ethics in technology practices and identified five forms of malpractice related to ethical compliance within the digital industry. The first three of these rely on distraction as a strategy, while the last two are more fundamentally destructive and challenging to address (Floridi, 2019):

1. Digital ethics shopping is the practice of selectively choosing and applying ethical standards that suit one's interests instead of striving for improvement, due to the variety of ethical perspectives and theories. This practice is problematic and can be combated by establishing clear, shared, and publicly accepted ethical guidelines, as the EU did with the Ethics Guidelines for Trustworthy Artificial Intelligence.
2. Digital ethics bluewashing is a term coined by Floridi, which refers to the practice of making false or misleading claims about ethical practices in the digital technology industry. This practice is analogous to greenwashing in environmental matters. Bluewashing is often carried out by investing significant resources in marketing, advertising, or other public relations activities to create a positive image of a company's ethical practices without real efforts to address ethical issues. The strategy to combat bluewashing includes legal enforcement of transparency and education of the public and stakeholders to identify misinformation.
3. Digital ethics lobbying is the practice of using soft laws to delay or avoid necessary legislation in the interest of technological innovation and economic growth. This malpractice can be addressed through legislation and effective law enforcement and must be disclosed whenever it occurs. It is crucial to differentiate digital ethics lobbying from genuine self-regulation efforts.
4. Digital ethics dumping is the practice of exporting or relocating research activities related to digital products and services to countries with weaker regulations and ethical standards with the intention of subsequently importing the outcomes of such unethical research back to the place of origin. For example, a company may export research and development for AI facial recognition algorithms to a non-EU country with weaker legislation for personal data protection, and then import the developed AI systems back to the EU without liability for illegal training and development practices. To prevent ethics dumping, it is important to establish a system of certification for digital products and services that focuses on both research and consumption.
5. Digital ethics shirking refers to the phenomenon whereby actors comply less with ethical standards in a particular context when they perceive that the expected benefits of ethical compliance are low. This malpractice is driven by self-interest and is often

observed in countries with weak legal enforcement. Furthermore, if actors have the option to shift responsibility to others, they are more likely to shirk their ethical duties in a given context. Addressing ethics shirking requires clearly defining and assigning responsibilities (Floridi, 2019).

5.3 From Ethics to Legislation

Regulators have recognized that, while publicly accepted ethical guidelines are important, they may not be sufficient to protect societal values in contexts with significant economic interests at play. Recent developments demonstrate that achieving ethical compliance requires more than just ethical guidelines and recommendations. Consequently, regulators have taken a leading role and are actively shaping the digital transformation process by enforcing legal boundaries.

Data Privacy Protection

A landmark in data privacy protection was achieved with the European General Data Protection Regulation (GDPR), which took effect in 2018. Under the GDPR, data concerning health are considered “personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status” (GDPR Summary, 2018). The processing of personal data is only permitted if it is anonymized. Anonymization means that personal data ~~is~~ processed in such a way that their subject can no longer be identified. Of great economic and scientific interest are pseudonymized data. Pseudonymization means personal data are processed in such a way that direct personal references are removed, but could be restored under strict conditions via trusteeships. According to the GDPR, pseudonymized data are also considered personal data; therefore, they are protected by law. The processing of pseudo-anonymized data is allowed only under the following conditions: either the data subject has consented to the processing (declaration of consent) or the data are being used for scientific purposes (with certain constraints; GDPR Summary, 2018).

The GDPR’s jurisdictional scope is not determined by the citizenship or residency of a data subject, which is a crucial aspect of the regulation. The GDPR stipulates that personal data cannot be transferred from the EU to a country that the European Commission has not deemed as having adequate data protection laws. As a result, both EU-based and non-EU-based tech companies processing the data of EU citizens have been significantly impacted by the GDPR. The regulation also introduced severe fines for non-compliance, with penalties of up to four percent of an entity’s annual revenue or €20 million, and a private right of action (Moundas & Peloquin, 2023).

The European Commission does not consider the majority of non-EU countries, including the United States, as offering adequate data protection legislation. While the United States has the Health Insurance Portability and Accountability Act (HIPAA) of 1996, which regulates **protected health information** (PHI), its scope is narrower than that of the GDPR. While HIPAA addresses PHI, the GDPR includes personal data, such as ethnicity and religion. Additionally, the GDPR focuses on protecting the personal data of EU citizens, while

Protected health information
This refers to data that a healthcare provider collects in order to deliver appropriate care.

HIPAA is concerned with organizations and business associates that manage PHI within the United States (Moundas & Peloquin, 2023). Commercially provided mHealth apps are not regulated by HIPAA, since the provider is not considered a covered entity or business associate. However, if the app is prescribed by a healthcare provider, and that healthcare provider is a covered entity or business associate, then the app would be subject to HIPAA regulations.

In 2023, the American magazine *Wired* published an article entitled "The Slow Death of Surveillance Capitalism Has Begun," referring to the European Union's groundbreaking ruling demanding that Meta reform its approach to personalized advertising (Meaker, 2023). Although many fines due to the violation of the GDPR had been imposed before, the ruling, which was preceded by a long-running investigation, penalized Big Tech companies for the first time, fining Facebook and Instagram ~~€390 million~~ ($414 million). The EU's decision is unprecedented as it aims at dismantling a business model that is the foundation of surveillance capitalism. The ruling has been well-received by researchers, journalists, and data protection experts, who see it as a long-awaited step towards rethinking and transitioning on the part of digital technology companies (Meaker, 2023).

AI Reliability Protection

The EU, known for establishing a data privacy protection milestone with the GDPR legislation, has once again taken a pioneering role in the field of AI by proposing the Artificial Intelligence Act (AIA), which is the first law concerning artificial intelligence worldwide (European Commission, 2023). The EU's global role in addressing the potentially harmful impacts of artificial intelligence is particularly significant, given the indication from policy researchers that the United States has adopted a soft stance towards AI and is hesitant to impose legal limitations, in contrast to the EU's approach. The difference between these two economic powers is their value priorities, with the US choosing to protect innovation capacity and preserve international competitiveness, while the EU prioritizes fundamental human rights protection (Floridi & Roberts, 2021).

The European Parliament is currently discussing the AI Act, which is expected to be approved by the end of 2023. Top issues being discussed include whether to ban or allow the use of facial recognition technology in public spaces, the appropriate level of regulation for high-risk AI applications, and the scope of enforcement and penalties for noncompliance. The act proposes a risk-based approach to regulating AI, with four risk categories and corresponding rules. More stringent regulations apply to higher risk applications, and those deemed to have an unacceptable level of risk are prohibited. Noncompliance can lead to fines of up to six percent of a company's global revenue (European Commission, 2023). The four risk categories are as follows:

- Unacceptable-risk AI applications – such as the use of AI for social scoring by governments, as seen in China – pose significant threats to safety, livelihoods, and human rights. These types of applications are considered unacceptable and are prohibited under the proposed AI Act.
- High-risk AI applications, on the other hand, have the potential to threaten people's fundamental rights or safety. This category encompasses a wide range of areas, including critical infrastructures, such as transportation and healthcare, where the use of AI

could potentially pose risks to the life and health of the populace. Other examples of high-risk AI applications include safety components of products, such as AI in robot-assisted surgery; educational or vocational training that may determine one's access to education or professional prospects (e.g., exam scoring); and monitoring systems, such as biometric surveillance for law enforcement or facial recognition systems. These high-risk AI systems are subject to several specific governance requirements under the AI Act to ensure they are safe and do not violate fundamental rights.

- Limited-risk AI applications are those that pose a lower level of risk and are subject to transparency requirements, such as an obligation to inform users when they are interacting with AI, such as chatbots.
- Minimal-risk AI applications, such as AI-enabled video games or spam filters, are encouraged to follow codes of conduct to promote ethical AI use (European Commission, 2023).

The AI Act still faces significant challenges in its finalization, but researchers and policymakers believe that 2023 will be remembered as a year that ushered in a new era in AI governance. With the new law, the EU is trying to strike a balance between protecting fundamental human rights and driving economic growth. The regulators hope that the first AI legislation will promote human rights globally by disseminating its effects to other economic regions.

SUMMARY

In this unit, we have explored the risks and challenges associated with digital health, including AI reliability, data security, data privacy, deprofessionalization, and deskilling. To address these challenges effectively, relying solely on self-regulatory ethical frameworks may not be enough. Instead, legal boundaries must be established to shape the digital transformation process in healthcare and other areas. ~~In~~ 2023, ~~we have seen~~ two significant developments in this area. The EU has taken an unprecedented step by penalizing the personalized advertising business model with its ruling enforcing financial consequences for violating the GDPR. Additionally, the EU's AI Act provides the world's first legal framework for AI, placing the protection of fundamental human rights at the core of its legislative regulations. Overall, these milestones have the potential to set the stage for a more ethical digital transformation worldwide.