

Artificial intelligence in medical education: Typologies and ethical approaches

Agnieszka Pregowska¹ & Mark Perkins²

Abstract

Artificial Intelligence (AI) has an increasing role to play in medical education and has great potential to revolutionize health professional education systems overall. However, this is accompanied by substantial questions concerning technical and ethical risks which are of particular importance because the quality of medical education has a direct effect on physical and psychological health and wellbeing. This article establishes an overarching distinction of AI across two typological dimensions, functional and humanistic. As indispensable foundations, these are then related to medical practice overall, and forms of implementation with examples are described in both general and medical education. Increasingly, the conditions for successful medical education will depend on an understanding of AI and the ethical issues surrounding its implementation, as well as the formulation of appropriate guidelines by regulatory and other authorities. Within that discussion, the limits of both narrow or Routine AI (RAI) and artificial general intelligence or Decision AI (DAI) are examined particularly in view of the ethical need for Trustworthy AI (TAI) as part of the humanistic dimension. All stakeholders, from patients to medical practitioners, managers, and institutions, need to be able to trust AI, and loss of confidence could be catastrophic in some cases.

Keywords: artificial intelligence typology; artificial intelligence in medicine; ethics; bioethics; medical education; health professional education

Introduction

The human brain provides an ideal reference model for the foundations of artificial intelligence, an indispensable IT tool that processes large data sets based on numerical algorithms to obtain solutions to problems in focus (Wang, Sun & Chen, 2023). The study of the foundation of natural intelligence may contribute to both the understanding of the general mechanics of intelligence toward pervasive brain-inspired systems and the formulation of intelligent entities (Wang et al., 2018). While the development of AI, including deep learning where algorithms are becoming more precise, faster, and able to reproduce the processes taking place in the human brain more faithfully, is rapidly permeating virtually all aspects of everyday life, necessary human dimensions such as ethical behavior and learning practice need to be accounted for and integrated. Nowhere is this more important than in medical education where outcomes impact human beings directly both physically and mentally. For example, AI may also be responsible for the introduction of errors in decision-making processes such as those related to bias in algorithms, deriving in particular from disproportionate and over-representation of certain data in electronic databases. A ready-made data set can come with biases incorporated or the data scientist in the pre-filtering step can introduce them, accidentally or deliberately, and these processes can be conscious or unconscious. Such biases can be even reinforced at later points, as in, for example, the US judiciary's probably recidivist AI-based system that inflated scores for African Americans (Zhao et al., 2017; Chouldechova, 2017). Thus, the term Explainable Artificial Intelligence (XAI) has been formulated to guard against the development and mystery of the black-box model (Confalonieri et al., 2021). XAI requires at least two levels of explainability (Solanki, Grundy & Hussain, 2023). The first is described in terms of the processes taking place in AI, its principles of functioning, transparency, and the possibility of interpreting the results obtained from the perspective of developers, while the second includes transparency and explainability from the point of view of end users, involving more general explanations. AI systems based on meeting the needs of both groups are still being

¹ Polish Academy of Sciences (Poland); aprego@ippt.pan.pl; ORCID: 0000-0001-9163-9931

² Collegium Prometricum (Poland); Royal Society of Arts (United Kingdom); mark.perkins@prometricum.eu

developed. One area that is at the forefront in the wholesale adoption of technical innovation is that of medicine, and this trend can be also observed in the case of XAI (Albahri et al., 2023). Indeed, it has been shown that XAI can be successfully applied to the health sector, as in the diagnosis of glaucoma and identification of women with Gestational Diabetes Mellitus (Solanki, Grundy & Hussain, 2023).

Moreover, according to one future vision of AI, the development of so-called Human-Centered AI (HCAI) will see a substantial change in the human position in the whole AI life cycle that will hopefully occur in a safe, reliable, and trustworthy way. In this scenario, AI-based solutions will both need and be able to understand human behavior, intentions, and expectations in some human-like way. However, people do not always say what they think, and this is an important issue in AI-based system development. Implication and sub-texts need to be taken into account, especially in terms of AI comprehension and possibly also of production. Another significant issue in HCAI concerns providing support to non-expert AI users in understanding its essence and capabilities. Six challenges, which include human well-being, respect for privacy, individual and social responsibility, appropriate design processes, respect for cognition in the sense of human-AI interaction, and enabling supervision, have been formulated for HCAI (Graibay et al., 2023), and these are all highly relevant to the medical field. AI-based systems offer great promise in the quest to raise and equalize human well-being as a support tool for learning, monitoring, and decision-making. The key issue is to answer the question of what the nature of AI-based solutions should be in such a way that this ambition may be met. Initial attempts in this field have been carried out, including those of the Wellbeing AI Research Institute which proposed the requirements that artificial intelligence should meet (Wellbeing AI Research Institute, 2024). Features necessary for AI such as simplicity, transparency, respect for human and fundamental rights and freedom, counteraction of social exclusion, and respect for private data were identified. There are also consequences for AI in the implementation of the UN Sustainable Development Goals (SDGs) (United Nations, 2024), which are discussed on the platform AI for Good (Vinueas et al., 2020). While at the current stage of AI development, it is becoming difficult and perhaps even impossible to identify a specific person as susceptible to error, responsibility in the context of AI implies a number of limitations related to legal, ethical, and moral analysis that AI designers need to have in mind when respecting the well-being of potential users of the new technology (Arriete et al., 2020). In other words, this is to be seen as essential good practice in AI design, development, and implementation processes. Thus, legal liability in the context of AI is a complex issue, especially considering that IP protection can be integrated with AI more simply than opportunities for criminal redress and legal principles (Gless, Silverman & Weigend, 2016). Moreover, it is highly advisable that AI-based solutions be certified by relevant authorities and include detailed user instructions.

Ethical and social consequences, therefore, are of critical concern in the field of apparently self-thinking machines, especially in the context of the potential harm that could be suffered by people. Indeed, there arises the question of the moral status of advanced AI-based solutions. These are based on substantial numbers of large data sets whose acquisition is a social process itself, and which may be susceptible to the socio-cultural perspectives of their stakeholders (Slota et al., 2020). For instance, trust in AI and the safety of personal data, including electronic health records, patient and disease registries, pharmacy dispensations, and insurance claims are essential issues to be considered. Indeed, this is especially so in the medical sector, in which the building of trust between the medical health profession and patients is crucial. As artificial intelligence develops and begins to be deployed more widely, trust needs to keep pace, and the notion of Trustworthy AI (TAI) will become prominent. For example, AI has started to change medical procedures, and as a consequence, many debates have arisen on how to integrate it with healthcare and legislative systems (Kumar, Chauhan & Awasthi, 2023). At the same time, this

raises the question of how AI ethics affects the medical education system.

In sum, we face a new challenge appearing before our eyes. A coexisting entity in the form of artificial intelligence is emerging, and it is up to us with which values it will be built. Responsible as we are for the future of new generations, it is today that we must transfer and extend our work in the field of eliminating inequalities to the digital world. The creation of artificial intelligence is a great achievement and represents a major part of our future. It is not only the fruit of the development of science, but also a child that requires proper education at every stage of its development. For it to respect us in the future, we must teach it the meaning of respect today, metaphorically speaking. In this paper, we concentrate on applications of artificial intelligence in the medical sector, in particular health professions education, and point towards further research on ethical risk points. This is of particular importance considering that medical education is distinct from other types of education in that it has a direct impact on the health and wellbeing of people.

Typology of artificial intelligence

Artificial intelligence, as has been emphasized in many technical reports, is of strategic importance (Samoili et al., 2020) but is viewed and described from a variety of perspectives. For example, in Europe, an AI monitoring commission (AI Watch) has been appointed (AI Watch, 2024) and given the goal of monitoring development and producing an inclusive definition of artificial intelligence alongside a taxonomy (Schwalbe & Finzel, 2023). In many papers, explainability and interpretability of XAI are treated as equivalent concepts (Bellucci et al., 2023), although neither concept has been given a formal, strict mathematical definition (Lipton, 2018). Many taxonomies of artificial intelligence take into account different features such as its capabilities, functionalities, similarity to human intelligence, and interpretability. Here we briefly review certain well-known AI taxonomies beginning with the functional (see Figure 1 below), and then concentrate on the humanistic perspective, important from the point of view of ethics.

A common basic taxonomy of artificial intelligence based on functional capabilities splits AI into Narrow AI (NAI), General AI (GAI), and Strong or Super AI (SAI) (Miloski, 2023). A narrow AI system can perform a dedicated task, cannot operate outside clear limitations, and depends on supervised learning (datasets are pre-coded). Examples include Siri (Apple), Watson (IBM), chess, speech recognition, and image recognition. GAI models have the ability to take in and classify uncoded datasets, can learn through iterations, and can perform many tasks at least as well as humans. This has led theorists to advance an extended hypothetical concept of Artificial General Intelligence (AGI) whereby attributes similar to reasoning and other human cognitive faculties might be manifested. According to the further notion of Super AI, also known as Artificial Super Intelligence (ASI), it is debated whether systems can surpass human intelligence in many other respects.

Other AI divisions are based on comparison to human intelligence and five types have typically been introduced such as logico-mathematical intelligence, social intelligence, visual-spatial intelligence, verbal-linguistic intelligence, and processing-speed intelligence (Pantano & Scarpi, 2022). The first concerns the ability to solve complex analytical problems, the second is related to the ability of machines to understand human emotions (Caic et al., 2019), the third is connected with spatial perception or awareness and has important implications in the context of the metaverse, the fourth type is connected with human natural language processing and the machine's ability to communicate with humans (McLean, Osei-Frimpong & Barhorst, 2021), and the last type involves the ability to solve tasks quickly (Dong et al., 2020).

Another important aspect of AI taxonomy from the ethical point of view is connected with the definition of Interpretable AI (IAI) which is a non-trivial issue, incorporating as it does an understanding of the task, its metrics, and type of experience. Taking human decision-making

as an exemplar, interpretability will also increasingly become a requirement of AI (Coeckelbergh, 2023). The discussion concerning differences and similarities between interpretable and explainable AI taking into account non-technical terminology is still ongoing. In addition to AI being interpretable and explainable, it needs to be trustworthy, and so Trustworthy AI (TAI) forms the third part of a trilogy of humanistic definitions that should be accessible to the general user (EAI, IAI, and TAI). Figure 2 illustrates the components of humanistic artificial intelligence.

Not all AI systems are highly complex, however. For example, models such as random trees and sparse linear models do not require additional actions connected with interpretability: it can be said that they have inherent interpretability due to the principles of their operation (Molnar, 2023). In general, there is a relation to the trade-off between the complexity of an AI-based system and its effectiveness, which is especially important in medical diagnostics (Petrick et al., 2013). Moreover, it is worth stressing that interpretability is not equivalent to fairness, privacy, and transparency (Wellar, 2019). Moreover, interpretability and explainability are often treated as equivalent terms (Vilone & Longo, 2020), although some papers describe their visible differences (Chromil & Schuessler, 2020). Explainability has been described as the interface where events occur understandable for humans (Vinuesa et al., 2020). This concept has also been used as synonymous with transparency (Belluci et al., 2021). To be able to define the above issue it is necessary to employ social, ethical, and legal points of view (Graziani et al., 2023). In general, IAI can be understood as comprising AI-based systems whose principles of functioning are understandable for humans, as opposed to the treatment of AI as a black-box model. According to EU law regarding IAI, the developer is obliged to provide information on the principles of operation, logic, and expected consequences of use. If we look symbolically, we should offer a symbolic definition, while in the case of sociology, we must incorporate the trust definition between humans and machines. In this sense it is important to shift the burden of creating algorithms to building trust relationships with people, not just automating the work they do.

Artificial intelligence with the ability to effectively process large amounts of data is increasingly used in medicine (Hollon et al., 2023). Medical image diagnostics supported by AI is an important tool that has an impact on medical doctors' decisions (Borys et al., 2023), although effective diagnosis, especially in the early stages of a disease, as well as the determination of disease development trends, is a very difficult issue that requires taking into account many factors and parameters. These examples serve to indicate that the application of AI can limit research to a smaller but significant part of the state space. A key limitation in the large-scale use of this technology is the lack of transparency in decision-making in critical cases and the ethical issues connected with AI (Rudin & Radin, 2019). Important issues in the general understanding of AI in diagnostics include trust and understanding of feature extraction that are crucial in Machine Learning (ML) in particular (Fuhrman et al., 2022). Performing a feature distribution analysis will help increase users' confidence in the algorithm if they know that there may be a case that is difficult to classify. This is also important information for the developers themselves when adapting models to automatically recognize such cases, whereby the interpretability of such algorithms will increase. This contributes to the fact that over time, users will start to have more confidence in input data, especially in clinical trials.

Another issue relates to the fact that doctors have begun to wonder what information to provide to the patient if the use of artificial intelligence is involved. Who is then responsible for the diagnosis, and who is liable if things go wrong? The adoption of XAI in medicine has been proposed on four levels, namely disclosure, intelligibility, interpretability, and explainability (Brady & Neri, 2020). It is known that the use of AI in a clinical procedure requires the informed consent of the patient, as with any other clinical procedure (LG Muchen, 2024). From an ethical point of view, medical doctors should not conceal information that

artificial intelligence supports them in the diagnosis process; they should disclose this fact. The patient should give informed consent to a medical procedure using AI and information about the patient and his health condition. Here the question arises as to what degree of AI explainability should be used to confirm that the patient gave informed consent. When we consider AI intelligibility in medicine, we must inform people about general risks and benefits. Then, AI in medical diagnostics can be beneficial through the high accuracy of diagnostics as in radiology (Canadian Association of Radiologists, 2019). However, such AI may be burdened with various types of biases when algorithms are trained, including algorithmic and automation biases (where there is a tendency for humans to favor machine-generated decisions) (Geis et al., 2019). It should be stressed that most AI-based algorithms, which are applied in medicine, are trained, and validated on race-skewed datasets (mostly dealing with European ancestry). This contributes to the fact that they may give not entirely corresponding or coextensive results for other racial types (Chen et al., 2023): algorithms may pass clinical-grade validation although they contain disparities in representation in minority subpopulations. Moreover, minimum criteria of interpretability may provide information concerning data (implicit grouping and/or conditions) that can lead to inaccurate results (Arbelaez Ossa et al., 2023). Thus, in the medical field, the patient should be informed in an understandable way about features of AI-based diagnostics, including characterization of the data, algorithmic bias, their accuracy, specificity, and sensitivity as well as the degree of involvement of AI in making a diagnosis. While it is highly complex, numerical calculations and parameters of the neural networks are not necessarily connected with biological structures and/or mechanisms, and thus explanation is very difficult (Rosler et al., 2023). In the case of explainability, it is necessary to distinguish between the two issues of explanation from the doctor's perspective and explanation from the patient's point of view. Natural questions arise as to why some algorithms are more effective in diagnosing certain health disorders and why. It is obvious that medical doctors should obtain such information (Kempt & Nagel, 2022), while diagnoses can differ (Heinin & Le Metayer, 2021). The introduction of XAI also has implications for health professions education. Medical students as well as doctors should receive such explanations when learning to make diagnoses based on artificial intelligence. When we consider the decision-making processes, one can also distinguish between Routine AI (RAI) and Decision-making Artificial Intelligence (also called Decision AI (DAI)) (Blackwell, 2024). The first, as a tool for analysis and comparison of large data sets and prediction at least, helps people make decisions in the form of recommendations (weak version), and the second at most makes decisions on their behalf (strong version), thus automating the decision-making processes.

Artificial intelligence in education in general

Artificial intelligence in one form or another has been used in education for almost forty years (Aleven & Koedinger, 2022). Recently AI applications in the field of higher education have grown very rapidly (Chu, Tu & Yang, 2022). However, ever-accelerating development has meant that people and institutions are struggling to keep abreast of new applications and their potential use in educational settings. While UNESCO, in its substantial report on technology in education, recognizes that, coupled with the fact that large numbers of teachers globally are ill-prepared to meet technological challenges, the fast pace of change puts strains on education systems to adapt, only three pages are dedicated to artificial intelligence (UNESCO, 2019, pp. 251–253). On the other hand, an active discussion is to be found in Europe as to the impact of AI on health care including ethics and education (United Nations Educational, Scientific and Cultural Organization, 2019). Indeed, the European Network for Academic Integrity (ENAI) has formulated recommendations on Artificial Intelligence in Education (AIEd), in particular regarding AI ethics (Spatharou, Hieronimus & Jenkins, 2020). One recommendation states that an AI-supported system cannot be the author of a publication due to its inability to take

responsibility for the published content, while a human being or legal entity always assumes responsibility for its actions. According to another recommendation, AI users should be aware of the fact that the input data may have biased or incorrect content. Ethical issues are important in the context of AIED because both students and teachers need to be taught how to use AI ethically to prevent misuse of AI tools and thereby prevent academic misconduct such as hiding the fact that AI has been used in their work.

AIED facilitates the combination of different disciplines and technologies and the enrichment of learning and teaching processes. It can be a supportive tool for both teachers and students, in particular in distance education (Foltynek et al., 2023) with many potential benefits stemming from its implementation. For example, AIED can be applied to automated assessment (Garlinska et al., 2023). This allows evaluators to shorten both the evaluation time and waiting time for results (Crompton & Burke, 2023). Students can also receive feedback on their skills faster, which directly translates into faster improvement of the same. When AI is applied in Intelligent Tutoring Systems (ITS) it enables the personalization of learning processes taking into account individual needs. Stat-Knowlab, for instance, is a solution that offers students only those classes for which they are prepared at the moment (Rutner & Scott, 2022). In turn, the ITS LabTutor enables profiling access to classes and laboratories for engineering students (de Chiusole et al., 2020). An AI-based system can both support students and find changes in their engagement, as in foreign language training (Khalfallah & Slama, 2018). An interesting use of AIED is to predict future student academic performance (Huang et al., 2020). Krause and Lowe (2014) used predictions to obtain information about which students would be at risk of dropping out of online courses in the future.

Studies have shown that AI has a positive impact on students (Wang, Sun & Chen, 2023). When students are able to integrate and apply AI-based technologies, their levels of self-confidence, and creativity increase (Krause & Lowe, 2014). Thus, AI capabilities can be split into three categories, namely: resources (data, technical resources), skills (technical, teaching application, the ability to cooperate), and consciousness (Wang, Sun & Chen, 2023). Then, a different viewpoint from that of user application has been proposed by Creely (2022) whereby the relation between AI and creativity can only be seen as an AI-human co-creative approach rather than viewing generative machines only as self-contained.

However, AIED not only benefits, but also threatens and challenges, and this is addressed by UNESCO in terms of issues of comprehensive public policy, inclusion and equality, preparing academic/teaching staff to use AI in the teaching process, supervision of the quality and scope of data (counteracting bias in algorithms, privacy) taking into account ethics and transparency in data collection, use, and dissemination (United Nations Educational, Scientific and Cultural Organization, 2019). Further principles in AIED have been proposed such as governance and stewardship, and human-centered AIED (Mikalef & Gupta, 2021). Unsurprisingly, the massive collection, storage, and processing of sensitive information has attracted the attention of human rights defenders (Nguyen et al., 2023). Considering the principles of biomedical ethics together with artificial intelligence in education it is typically believed that practice should be informed by the following principles: autonomy, beneficence, non-maleficence, justice, and explicability (Anshari et al., 2023). However, this may not always be borne out in practice. In a study by Kim and Kim (2020), the demand for the use of AI in education in the case of Korean teachers was shown to be high. However, it was worrying that their interest in understanding AI principles and AI ethics was low. In another piece of research, it was found that more than 60 percent of elementary school teachers and pre-service teachers do not know how to use AI in the teaching process (Lee, 2021).

In addition, a further important aspect is the development of AI skills among young children: limitations, ethical consequences, and a basic understanding of this technology are crucial (Su & Yang, 2023) and this goes for AI-based toys such as robots, drawing tools, and chatbots (Su

& Yang, 2022). It can be claimed that these toys help children improve their technical and communication skills, including writing and digital storytelling. However, not much is known about how a teacher should support children at an early stage of development in interacting with artificial intelligence (Kewalramani et al., 2021). While there are a number of curriculum types and materials aimed in particular at elementary and high school students, there is no material to help young non-programmers gain and expand their knowledge of AI and the ability to implement it in their education (Yang, 2022). Of course, nothing can replace a child's contact with a living person, although getting accustomed to AI at an early stage of development will counteract digital exclusion in the future. On the other hand, while there is some evidence of the improvement in numeracy, literacy, and science due to AI in particular, however, there is no overall study of the impact of AI on education in general. Moreover, public and private partnerships proposed for the management of AIEd have not provided much evidence for reducing inequalities in the provision and access to public services, including education (Nemorin et al., 2023).

Artificial intelligence in medical education

The three main categories of AI are all relevant to health professions education. These are Artificial Narrow Intelligence (ANI) focussed on specific applications, Artificial General Intelligence (AGI) whose features emulate human capabilities, and Artificial Super Intelligence (ASI) where AI may be able to go further than the human being and surpass it in many aspects of all kinds. All of these conceptions of AI are relevant to health professions education and have their own ethical implications.

Health professions education has always been closely related to ethical issues and the four principles of autonomy, non-maleficence, beneficence, and justice established by (Beauchamp & Childress, 2019) continue to be highly influential. This has been primarily due to the need to handle sensitive patient information appropriately, and to respect medical confidentiality when treating patients in the future. Consequently, the introduction of artificial intelligence into health profession education should cover all the principles of Bioethics supplemented with the general principles of AI ethics used in other areas.

There have been comparatively few studies on the use, actual or potential, of artificial intelligence in medical education (AIMEd). In their scoping review of literature on the ethics of AI in health care, Murphy et al. (2021) found that while there were many recommendations as to how AI could be used responsibly, there were few guides as to how practical implementation could be achieved. Some specific mentions have been made of the potential in terms of the use of Virtual Reality (Mittermaier, Raza & Kvedar, 2023), whereas Wartman and Combs (2019) focus on information overload, issues of confidence in using AI, and curriculum reform, and Katznelson and Gerke (2021) highlight six AI-related key ethical issues that need to be accounted for in medical school education: informed consent, bias, safety, transparency, patient privacy, and allocation of resources. Paranjape et al. (2019) comment on the lack of standards for the use of AI in patient care and concomitant liability questions, associated as they are with the need for better medical training. Masters (2023) offers a guide to the ethical issues that HPE teachers and administrators are likely to face as AI enters the teaching environment, focussing on traditional areas of ethical concern in health such as data collection, consent and ownership, security, privacy and transparency, responsibility and beneficence (for whose benefit are data used?). However, the account of algorithmic bias is basic, and the brief discussion of AI sentience and AI rights could be said to assume possible conclusions on their probable development that are by no means in widespread acceptance (Gillet, 2016). Any notion of sentience, for example, would rest on a theory of emergent properties: at some point increasing complexity in machine systems will result in the emergence of consciousness. However, there is a substantial historical discussion on this across disciplines which is ongoing

[80], although at least the main principles of the debate need to be considered, especially where an AI medical trainer may appear.

Among practical cases, artificial intelligence is currently used in surgery as in systems evaluating the effectiveness of surgeons' work based on videos. Often the explanations generated by artificial intelligence do not coincide with the explanations given by surgeons. Kiyasseh et al (2023) propose human explanations as supervising the AI-based algorithm (supervised learning) so that the algorithm correctly assesses the importance of individual frames in the video system TWIX. TWIX also takes into account the level of experience of the person performing the operation. The proposed solution, used together with the surgical artificial intelligence system (SAIS), would improve the skills of surgeons. Not only surgeons stand to benefit from the introduction of AI into the educational process. In surgery, repeatability is an important element. AI can be used as a tool to analyze operational techniques and structure feedback to improve operational efficiency. Guerrero et al. found that, as a consequence, the efficiency of operating rooms increased overall (2023).

Another medical area in which AIMed can be beneficial is radiology. Shah et al. (2023) showed that AI enables interns to receive real-time feedback when interpreting clinical and simulation brain MRI scans. An important limitation of this study from the statistical point of view was the size of the sample, but it can be treated as illustrative. In turn, Schwendicke et al. (2023) posed a set of questions that educators should ask students when using AI in dentistry. It can be also used to create more realistic virtual medical courses in the form of AI-enabled Virtual Reality Simulation (VRS), as in medical training (tr.AI.ning). VRS uses three-dimensional (3D) immersive technology to allow students to enter a virtual world where they can experience classes from any geographical location, while AI provides the tool for the creation of more realistic 3D models and scenes (Mergen et al., 2023). Another significant technology, which can be drawn upon by AIMed is that of digital twins (DTs) (von Ende et al., 2023). DTs provide a digital replica of a physical object or process: this is a virtual model with special features that combine the physical and digital worlds (Peshkova et al., 2023). It enables the transfer of human physical characteristics, including changes and disorders in the body to the digital environment. Thus, many different areas of medical education can benefit from the application of artificial intelligence through different technological applications.

Apart from AI solutions designed to analyze datasets such as clinical and patient data, generative AI (as in ChatGPT) is likely to have great potential in AIMed. Generative AI has become quite prominent through the common use of the benchmark application ChatGPT which is now recognized as a distinct form of AI in itself with the arrival of the name gen AI (Kamel Boulos & Zhang, 2021), and it is so powerful that in one study it was found that although ChatGPT may have a positive influence on health professions education, its output can fool scientists (Else, 2023). Gilson et al. (2023) demonstrated that ChatGPT is able to formulate the answers at the level of a third-year medical student. Eysenbach decided to check this statement and had a conversation (interview) with ChatGPT (Chui, 2023). While ChatGPT's knowledge and ideas on how to introduce this AI tool in health professions education were impressive, the disturbing issue was that this tool came up with literature that does not exist in reality. This phenomenon is called hallucination in AI and occurs in large language models in the form of false facts or information (Eysenbach, 2023). Another study focussed on the application of ChatGPT in the field of diabetes education (Alkaissi & Mcfarlane, 2023). It turned out that it provided accurate answers, although of a quite general nature. This is a consequence of training the language model on limited datasets that do not cover the nuances of the education of the diabetic patient. As datasets improve in quality and quantity, however, the accuracy of answers will improve. One disadvantage of ChatGPT is the fact that it is not able to assess the reliability of output datasets. ChatGPT responses are formulated based on the probability of occurrence of phrases and sentences. The model as it is

does not include references to specific knowledge bases in answering processes. Corsello & Santangelo (Sng et al., 2023) demonstrate the positive sides of ChatGPT in pediatrics where the analysis of huge amounts of text data generates new insights, and Abdelhady et al. (2023) showed that ChatGPT can work as a tool for creating accurate operating notes in plastic surgery. This is in a sense a logical contradiction to the fact that ChatGPT has a tendency to produce phenomena such as AI hallucination, where generative AI produces information and scenarios that are far from correct. However, it is important to encourage students to use ChatGPT-type technologies effectively, otherwise, without supervision, they may violate the integrity of the academic space, leading to well-known issues such as plagiarism and cheating (Abdelhady & Davis, 2023). On the other hand, it appears that ChatGPT can be used for proofreading on a basic level and some papers contain an acknowledgment of this (Malinka et al., 2023). Additionally, it has been claimed that ChatGPT, used in this form, can be conducive to constructive writing (Kitamura, 2023). With regard to nursing, Liu & Liu (2023) point out the positive advantages of using generative text AI such as ChatGPT in nursing education in terms of the ability to personalize learning, provide immediate feedback, and suggest simulations. Indeed, ChatGPT may be introduced in medical education in several ways. It can be applied as an automatic scoring tool for the evaluation of student essays and papers or as a teaching assistance tool to generate exercises, quizzes, and scenario descriptions (Khan et al., 2023). This solution may be also beneficial to students as a help in answering questions and a tool for creating bibliographies (although these should be checked in view of the phenomenon of hallucination). However, a certain level of knowledge is required to be able to judge their correctness. It has also great potential in personalizing learning content. On the other hand, disadvantages include variable accuracy and discrepancies, questions over privacy and security, types of bias, lack of human interaction, and the temptation to over-rely, plagiarize, and cheat (Alkhaqani, 2023). Additionally, the question has even been raised as to whether ChatGPT should be listed as an author in scientific publications (Stokel-Walker, 2023). Overall, it would appear that ChatGPT has a great deal to offer when it comes to health profession education but the nature of the relationship between the AI thus manifested and the human needs to be examined carefully from the ethical point of view.

In order to score and teach, a gen AI must have suitable knowledge. One way of testing that is to require the AI itself to take tests. Kung et al. (2023) tested the performance of ChatGPT on all three tests of the United States Medical Licensing Exam (USMLE). Achieving a pass mark of 60% overall, it was judged good enough to educate 2nd-year students (Step 1) and was able to produce significant insights in its answers. Li et al. (2023) required ChatGPT to take a virtual objective structured clinical examination in obstetrics and gynecology. A panel of 14 qualified examiners scored the results and found that it outscored the average human scores by 3.5% (77.2% as against 73.7%). Giannos & Delardas (2023) evaluated ChatGPT's performance in the UK BMAT, TMUA, LNAT, and TSA examinations with an eye to its potential in setting tests. They concluded that it has value as a supplemental educational tool, although its performance was uneven across test sections (its scientific and mathematical knowledge was found to be poor, for example). On the other hand, Das et al. (2023) found that ChatGPT scored approximately 80% on both first-order questions and second-order knowledge questions based on the Indian competency-based medical education (CBME) microbiology curriculum and concluded that it could serve as a useful automated testing system. In sum, if gen AI can give the right answers to tests it will be capable of creating tests, and these studies also indicate that a gen AI is also likely to have the capability to be a teacher given improvements in its knowledge and performance.

Thus, large language models (LLMs) such as Generative Pre-trained Transformer (GPT) systems may potentially change the whole experience of medical content (van de Ridder et al., 2023). Theoretically, they can provide a tool to create or/and support the creation process of the

learning content, personalized learning plans, and transform teaching methodologies (Abd-Alrazaq et al., 2023). Indeed, instead of banning the use of AI, medical units have developed regulations and mechanisms to ensure ethical and common sense applications, and it is certain that medical staff are also being encouraged to increase their competence in this area. Indeed, medical education programs should include education on the proper use of AI, including LLM-based gen text solutions. This is especially important in the context of ethical behavior and the danger of student cheating. However, it is also worth pointing out that LLM-based solutions can increase student writing quality (Marzuki et al., 2023). Such solutions can be used, although it is not acceptable that one gen text solution be used by the lecturer to check output from another such solution used by the student. A separate issue arises of whether students are prepared to use the achievements of technology wisely in practice, perhaps in contradiction to their declared intentions, as well as the perception and preparation of doctors to conduct classes with AI participation (Kedar & Khazanchi, 2023).

Basic barriers to the effective implementation of artificial intelligence in medical education include the skills of academic staff in teaching using AI and developing curricula (Foster & Tasnim, 2020) as well as the lack of infrastructure for teaching students using AI (where the level is dependent on the country) (Rogerson et al., 2022). However, over 80 percent of radiologists believe that AI-enabled classes should be included in residency programs (Salastekar et al., 2023).

Discussion

It seems that in the future, a great deal of routine work will be done by artificial intelligence, including in the sphere of medical education. AI offers huge promise, but even here the development of systems implies huge responsibility, as we stand at a turning point in AI development. AI needs to be guided by decisions for the good of future generations and ethical, social, and legal aspects must be integral to its progress in the quest to create Trustworthy Artificial Intelligence (TAI) (Arrieta et al., 2020), especially important in medicine and medical education. Trust is a core principle of this area, especially as medical education, as distinct from other types of education, has a direct impact on health and well-being. In that AI-based solutions will take over a great number of tasks performed by people, they should be developed according to the observance of ethical principles that guide people and, in the case of medicine and medical education, follow the principles of bioethics (Kumar & Choudhury, 2023), even or especially where factors along the way may distort those principles as reflected in practical solutions.

Other issues are connected with human trust in AI-based systems when it comes to decision-making as to what to teach, when, and how, in a medical curriculum, and it is imperative that these be considered in the design of AI-powered medical education programs. The general principles of designing TAI are built on three pillars: it should be lawful, ethical, and robust. There is a growing awareness of this amongst regulators and authorities. For example, the European Union has formulated a tool called the Assessment List for Trustworthy Artificial Intelligence (ALTAI) (Radclyffe, Ribeiro & Wortham, 2023). Designed to enable institutions to align their AI-based systems, this comprises a set of seven principles, including human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal wellbeing, and accountability. Thanks to the scalability of the ALTAI assessment system, the organization has the ability to see where its AI-based system is in relation to the reference standard. The disadvantage of this approach is the lack of consideration of the size of organizations which may vary between huge corporations and start-ups, and so the method of application to small clinics or large hospitals may be an issue. Another weakness of ALTAI is that it does not take into account relative risk issues, of great importance in the field of medical AI.

A further recommendation proposed is that AI-based systems be evaluated according to a Fundamental Rights Impact Assessment (FRIA), which is intended to assess how algorithms meet human rights. This tool is commonly used in new legislative processes. It is also worth stressing that these principles and recommendations are not sufficient: government and regulatory bodies must use them in the creation of structures and regulations to monitor the implementation and functioning of AI-based solutions, especially in medical education. The first official attempt at AI regulation is likely to be the EU AI Act (European Union Agency for Fundamental Rights, 2020). According to this, AI-based systems with an unacceptable level of risk to people's safety will be prohibited. These include post-remote biometric identification systems (apart from when law enforcement authorities are given judicial approval), biometric categorization systems using sensitive data, predictive policing systems, emotion recognition systems in law enforcement and public institutions, and untargeted scraping of facial images for the development of audiovisual content. Thus, limited levels of risk are connected with the minimal transparency requirements of AI-based systems, and the aim is to ensure that the user is aware that he/she is using a system with such characteristics.

Additionally, some countries such as Germany and the United Kingdom have already established government bodies to promote the adoption of new technologies in industry, notably the Office for AI (OfAI) and the Centre for Data Ethics and Innovation (CDEI) in the UK. AI regulations are also beginning to be implemented in the European Union, where, for example, in the area of data privacy laws, online platforms, and online e-commerce regulations, although no equivalent or similar regulations have been introduced in the United States. In contrast with regulatory action designed to promote societal trust, it has been demonstrated in a study by Kleizen et al. (2023) that short-term campaigns do not increase public trust in AI, as opposed to long-term campaigns. This point should also be taken into account in the field of ALTAI development as crucial to understanding public responses. On the operational level, robustness is a critical factor influencing the efficiency of algorithms that should be taken into account in the design of any AI-based system (Li et al., 2022). This can be accomplished with a robustness test with a special emphasis on adversarial attacks bearing in mind ethical concerns such as the proper functioning of AI-based systems regarding privacy issues. A mathematical formulation for TAI-based system robustness in the case of adversarial attacks by deep learning algorithms has been proposed by Huang et al. (2020).

Conclusions

Two basic divisions of AI are considered here: functional and humanistic, and routine AI (RAI) and decision AI (DIA). At the current stage of development, narrow AI is being deployed to conduct precise and well-defined tasks for routine AI purposes. In future steps, forms of human-like artificial general intelligence (AGI) may be developed, at least where the machine is permitted to make decisions at appropriate levels, and there also exists a vision of Super AI (strong AGI) that surpasses human intelligence even regarding the unique aspects of being human. In this future vision, it has to be assumed that human-centered AI (HCAI) will put humans as a priority, otherwise a dystopia could ensue. The functional typology comprises one of the possible future AI taxonomies. A second and related operational typology is based on the comparison of the taxonomic components of human intelligence, namely logico-mathematical intelligence (as in decision trees), social intelligence, visual-spatial intelligence, verbal-linguistic intelligence, and processing-speed intelligence. However, taking into account the influence of AI on human beings, inextricably linked as it is with the ethical dimension, a further humanistic typology including explainable AI (XAI), interpretable AI (IAI), and trustworthy AI (TAI) would seem to be essential in terms of human understanding. Of these three, the first assumes that transparent and accountable AI-based solutions can be built, while the second is concerned with explaining the exact principles of AI by analogy with the human brain. The

third and last one is connected with countering disinformation and building awareness and trust in AI-based solutions. However, these three are not subject to strict mathematical definitions, and it is also to be stressed that ethical concepts themselves are not defined mathematically and are best expressed as part of human discourse. In many sources, they are imprecise and overlap, although attempts at systematization in international organizations are in progress. The establishment of good typologies and taxonomies, underpinned by ethical principles, is essential in the use of AI in medicine overall and medical education in particular. All stakeholders need to have confidence in AI when it comes to the selection of datasets, algorithm construction, and potential bias, disclosure, and liability. While there are many considerations in common between education overall and medical education as a significant subset, medical education is special since its outputs impact the health and well-being of people. In future work, we will describe and define the ethical dimension of artificial intelligence in medical education in terms of its role, management, and risks.

References

- ABD-ALRAZAQ, A. et al. (2023): Large language models in medical education: Opportunities, challenges, and future directions. In: *JMIR Med Educ*, 9, e48291.
- ABDELHADY, A. M. & DAVIS, C. R. (2023): Plastic surgery and artificial intelligence: How ChatGPT improved operation note accuracy, time, and education. In: *Mayo Clinic Proceedings: Digital Health*, 1(3), pp. 299–308.
- AI WATCH (2023): [online] [Retrieved January 16, 2024] Available at: European Commission https://ai-watch.ec.europa.eu/index_en
- ALBAHRI, A. S. et al. (2023): A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. In: *Information Fusion*, 96, pp. 156–191.
- ALEVEN, V. A. W. M. M. & KOEDINGER, K. R. (2002): An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. In: *Cognitive Science*, 26(2), pp. 147–179.
- ALKAISSI, H. & MCFARLANE, S. I. (2023): Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. In: *Cureus*, 15(2), e35179.
- ALKHAQANI, A. L. (2023): ChatGPT and Nursing Education: Challenges and Opportunities. In: *Al-Rafidain Journal of Medical Sciences*, 4, pp. 50–51.
- ANSHARI, M. et al. COVID-19. (2023): Artificial intelligence, ethical challenges and policy implications. In: *AI & Soc*, 38, pp. 707–720.
- ARBELAEZ OSSA, L. et al. (2022): Re-focusing explainability in medicine. In: *Digit Health*, 8, 20552076221074488.
- ARRIETA, A. B. et al. (2020): Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. In: *Information Fusion*, 58, pp. 82–115.
- ASHKAN, F. (2021): There is no I in AI. In *AI & Society*, 36(3), pp. 1035–1044.
- BEAUCHAMP, T. L. & CHILDRESS, J. F. (2019): *Principles of Biomedical Ethics*, 8th ed. New York: Oxford University Press.
- BELLUCCI, M. et al (2021): Towards a terminology for a fully contextualized. In: *XAI Proc. Comput. Sci.*, pp. 241–250.
- BLACKWELL, A. F. (forthcoming, 2024): *Moral Codes: Designing alternatives to AI*. MIT Press. [online] [Retrieved January 16, 2024] Available at: <https://moralcodes.pubpub.org/>
- BORYS, K. et al. (2023): Explainable AI in medical imaging: An overview for clinical practitioners – Saliency-based XAI approaches. In: *European Journal of Radiology*, 162, 110787.

BRADY, A. P. & NERI, E. (2020): Artificial intelligence in radiology—ethical considerations. In: *Diagnostics*, 10, 231.

ČAIĆ, M., MAHR, D. & ODERKERKEN-SCHRÖDER, G. (2019): Value of social robots in services. In: *Journal of Services Marketing*, 33(4), pp. 463–478.

CANADIAN ASSOCIATION OF RADIOLOGISTS (2019): White paper on ethical and legal issues related to artificial intelligence in radiology. In: *Can Assoc Radiol J*, 70(2), pp. 107–118.

CHEN, R. J. et al. (2023): Algorithmic fairness in artificial intelligence for medicine and healthcare. In: *Nat. Biomed. Eng*, 7, pp. 719–742.

DAS, D. et al. (2023): Assessing the capability of ChatGPT in answering first- and second-order knowledge questions on microbiology as per competency-based medical education curriculum. In: *Cureus*, 15(3), e36034.

DONG, Y. et al. (2020): Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. In: *Complexity*, 1680845, 10.

DU, Y. et al. (2022): An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. In: *Sci Rep*, 12, 1170.

GARIBAY, O. O. et al. (2023): Six human-centered artificial intelligence grand challenges. In: *International Journal of Human-Computer Interaction*, 39(3), pp. 391–437.

GLESS, S., SILVERMAN, E. & WEIGEND, T. (2016): If robots cause harm, who is to blame? Self-driving cars and criminal liability. In: *New Criminal Law Review*, 19(3), pp. 412–436.

DE CHIUSOLE, D. et al. (2020): Stat-Knowlab. Assessment and learning of statistics with competence-based knowledge space theory. In: *International Journal of Artificial Intelligence in Education*, 30, pp. 668–700.

CHOULDECHOVA, A. (2017): Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. In: *Big Data*, 5(2), pp. 153–163.

CHROMIK, M. & SCHUESSLER, M. (2020): A taxonomy for human subject evaluation of black-box explanations in XAI. In: *ExSS-ATEC@ IUI*, p 1. [online] [Retrieved December 4, 2023] Available at: <https://ceur-ws.org/Vol-2582/paper9.pdf>.

CHUI, M. (2023): *The state of AI in 2023: Generative AI's breakout year. McKinsey Insights*. [online] [Retrieved December 4, 2023] Available at: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year#/>.

CHU, H., TU, Y. & YANG, K. (2022): Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. In: *Australasian Journal of Educational Technology*, 38(3), pp. 22–42.

COECKELBERGH, M. (2023): Artificial intelligence, responsibility attribution, and a relational justification of explainability. In: *Sci Eng Ethics*, 26, pp. 2051–2068.

CONFALONIERI, R. et al. (2021): A historical perspective of explainable Artificial Intelligence. In: *WIREs Data Mining Knowl Discov*, 11, e1391.

CORSELLO, A. & SANTANGELO, A. (2023): May artificial intelligence influence future pediatric research?—The case of ChatGPT. In: *Children*, 10, 757.

CREELY, E. (2022): Conceiving creativity and learning in a world of artificial intelligence: A thinking model. In: D. Henriksen & P. Mishra (eds): *Creative provocations: Speculations on the future of creativity, technology & learning*. Cham: Springer, pp. 35–50.

CROMPTON, H. & BURKE, D. (2023): Artificial intelligence in higher education: the state of the field. In: *Int J Educ Technol High Educ*, 20, 22.

ELSE, H. (2023): Abstracts written by ChatGPT fool scientists. In: *Nature*, 613(7944), 423.

EMORIN, S. et al. (2023): AI hyped? A horizon scan of discourse on artificial intelligence in education (AIED) and development. In: *Learning, Media and Technology*, 48(1), pp. 38–51

EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS 2020. Getting the future right: Artificial intelligence and fundamental rights. [online] [Retrieved January 16, 2024] Available at: <https://doi.org/10.2811/774118>.

EYSENBACH, G. (2023): The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation With ChatGPT and a Call for papers. In: *JMIR Med Educ*, 9, e46885.

FOLTYNEK, T. et al. (2023): ENAI recommendations on the ethical use of artificial intelligence in education. In: *Int J Educ Integr*, 19, 12.

FOSTER, M. & TASNIM, Z. (2020): Data science and graduate nursing education: A critical literature review. In: *Clin Nurse Spec*, 34(3), pp. 124–131.

FUHRMAN, J. D. et al. (2022): A review of explainable and interpretable AI with applications in COVID-19 imaging. In: *Med. Phys.*, 49, pp. 1–14.

GARLINSKA, M. et al. (2023): The influence of emerging technologies on distance education. In: *Electronics*, 12(7), 1550.

GIANNOS, P. & DELARDAS, O. (2023): Performance of ChatGPT on UK Standardized Admission tests: Insights From the BMAT, TMUA, LNAT, and TSA Examinations. In: *JMIR Med Educ*, 9, e47737.

GEIS, J. R. et al. (2019) *Ethics of AI in radiology: Joint European and North American Multisociety statement*. [online] [Retrieved December 4, 2023] Available at: <https://www.acr.org/-/media/ACR/Files/Informatics/Ethics-of-AI-in-Radiology-European-and-North-American-Multisociety-Statement--6-13-2019.pdf>.

GILLET, C. (2016): *Reduction and emergence in science and philosophy*. Cambridge: Cambridge University Press.

GILSON, A. et al. (2023): How does ChatGPT perform on the United States Medical licensing examination? The implications of large language models for Medical education and knowledge assessment. In: *JMIR Med Educ*, 9, e45312.

GRAZIANI, M. et al. (2023): A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences. In: *Artif Intell Rev*, 56, pp. 3473–3504.

GUERRERO, D. T. et al. (2023): Advancing surgical education: The use of artificial intelligence in surgical training. In: *The American Surgeon TM*, 89(1), pp. 49–54.

HENIN, C. & LE MÉTAYER, D. (2021): Beyond explainability: justifiability and contestability of algorithmic decision systems. In: *AI & Society*, 37(4), pp. 1397–1410.

HOLLON, T. et al. (2023): Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging. In: *Nat Med*, 29, pp. 828–832.

HUANG, X. et al. (2020) A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. In: *Comput Sci Rev*, 37, 100270.

KAMEL BOULOS, M. N. & ZHANG, P. (2021): Digital twins: From personalised medicine to precision public health. In: *J Pers Med.*, 11.

KATZNELSON, G. & GERKE, S. (2021): The need for health AI ethics in medical school education. In: *Adv in Health Sci Educ*, 26, pp. 1447–1458.

KEDAR, S. & KHAZANCHI, D. (2023): Neurology education in the era of artificial intelligence. In: *Curr Opin Neurol.*, 36(1), pp. 51–58.

KEMPT, H. & NAGEL, S. K. (2022): Responsibility, second opinions and peer-disagreement: ethical and epistemological challenges of using AI in clinical diagnostic contexts. In: *J Med Ethics*, 48(4), pp. 222–229.

KEWALRAMANI, S. et al. (2021): Using robotic toys in early childhood education to support children's social and emotional competencies. In: *Australasian Journal of Early Childhood*, 46(4), pp. 355–369.

KHALFALLAH, J. & SLAMA, J. B. H. (2018): The effect of emotional analysis on the improvement of experimental e-learning systems. In: *Computer Applications in Engineering Education*, 27(2), pp. 303–318.

KHAN, R. A. et al. (2023): ChatGPT - Reshaping medical education and clinical management. In: *Pak J Med Sci.*, 39(2), pp. 605–607.

KIM, N. J. & KIM, M. K. (2022): Teacher's perceptions of using an artificial intelligence-Based educational tool for scientific writing. In: *Frontiers in Education*, 7, 755914.

KIYASSEH, D. et al. (2023): A multi-institutional study using artificial intelligence to provide reliable and fair feedback to surgeons. In: *Commun Med*, 3, 42.

KITAMURA, F. C. (2023): ChatGPT is shaping the future of medical writing but still requires human judgment. In: *Radiology*, 230171.

KLEIZEN, B. et al. (2023): Do citizens trust trustworthy artificial intelligence? Experimental evidence on the limits of ethical AI measures in government. In: *Government Information Quarterly*, 40(4), 101834.

KRAUSE, A. & LOWE, P. J. (2014): Visualization and communication of pharmacometric models with Berkeley Madonna. In: *CPT Pharmacometrics Syst Pharmacol.*, 28, 3(5), e116.

KUMAR, P., CHAUHAN, S. & AWASTHI, L. K. (2023): Artificial intelligence in healthcare: Review, ethics, trust challenges & Future research directions. In: *Engineering Applications of Artificial Intelligence*, 120, 105894.

KUMAR, S. & CHOUDHURY, S. (2023): Normative ethics, human rights, and artificial intelligence. In: *AI Ethics*, 3, pp. 441–450.

KUNG, T. H. et al. (2023): Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. In: *PLOS Digit Health.*, 2(2), e0000198.

LEE, S. (2021). Convergence education of elementary school teachers and pre-service teachers. In: *Journal of Korean Practical Arts Education*, 34(1), pp. 1–17.

LI, X. H., et al. (2022): A Survey of Data-Driven and Knowledge-Aware eXplainable AI. In: *IEEE Transactions on Knowledge and Data Engineering*, 34(1), pp. 29–49.

LIU, J., LIU, F. & LIU, S. (2023): *The application of ChatGPT in nursing education*. [online] [Retrieved January 16, 2024] Available at: <https://doi.org/10.35542/osf.io/2x7a8>.

LG MÜNCHEN II (2022): Urt. v. 4.5.2021 – 1 O 2667/19 Hei. Informationspflicht und Traumaverarbeitung: BGB 278, 280 Abs. 1, 630a Abs. 1, 630c Abs. 2 S. 2, 630h Abs. 5 S. 1. In: *Medizinrecht*, 40(8), pp. 681–685. German. [online] [Retrieved December 4, 2023] Available at: <https://doi.org/10.1007/s00350-022-6278-7>.

LI, S. W. et al. (2023) ChatGPT outscored human candidates in a virtual objective structured clinical examination in obstetrics and gynecology. In: *Am J Obstet Gynecol*, 229 (172), e1-12.

LIPTON, Z. C. (2018): The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. In: *Queue*, 16, 3, pp. 31–57.

MALINKA, K. et al. (2023): On the educational impact of ChatGPT: Is artificial intelligence ready to obtain a university degree? In Proceedings of the 2023 Conference on innovation and technology in Computer science education V. 1 (ITiCSE 2023). In: *Association for Computing Machinery*, New York, NY, USA, pp. 47–53.

MARZUKI, M. et al. (2023): The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. In: *Cogent Education*, 10, 2.

MASTERS, K. (2023): Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. In: *Medical Teacher*, 45(6), pp. 574–584.

MCLEAN, G., OSEI-FRIMPONG K. & BARHORST J. (2021): Alexa, Do voice assistants influence consumer brand engagement? – Examining the role of AI powered voice assistants in influencing consumer brand engagement. In: *Journal of Business Research*, 124, pp. 312–328.

MERGEN, M. et al. (2023): Medical tr.AI.ning consortium. Immersive training of clinical decision making with AI driven virtual patients - a new VR platform called medical tr.AI.ning. In: *GMS J Med Educ.*, 40(2), 18.

MIKALEF, P. & GUPTA, M. (2021): Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. In: *Information & Management*, 58(3), 103434.

MIŁOSKI, B. (2023): Opportunities for artificial intelligence in healthcare and in vitro fertilization. In: *Fertility and Sterility*, 120(1), 3.

MITTERMAIER, M., RAZA, M. & KVEDAR, J. C. (2023): Collaborative strategies for deploying AI-based physician decision support systems: challenges and deployment approaches. In: *Digit. Med.*, 6, 137.

MOLNAR, C. (2023): *A Guide for making black box models explainable* [online] [Retrieved December 4, 2023] Available at: <https://christophm.github.io/interpretable-ml-book/>

MURPHY, K. et al. (2021). Artificial intelligence for good health: a scoping review of the ethics literature. In: *BMC Med Ethics*, 22, 14.

NGUYEN, A. et al. (2023): Ethical principles for artificial intelligence in education. In: *Educ Inf Technol*, 28, pp. 4221–4241.

PANTANO, E. & SCARPI, D. (2022): I, robot, You, consumer: Measuring artificial intelligence types and their effect on consumers emotions in service. In: *Journal of Service Research*, 25(4), pp. 583–600.

PARANJAPPE, K. et al. (2019): Introducing artificial intelligence training in medical education. In: *JMIR Med Educ.*, 3, 5(2), e16048.

PETRICK, N. et al. (2013): Evaluation of computer-aided detection and diagnosis systems. In: *Med Phys.*, 40(8), 087001.

PESHKOVA, M. et al. (2023): Digital twin concept: Healthcare, education, research. In: *J Pathol Inform.*, 16, 14, 100313.

RADCLYFFE, C., RIBEIRO, M. & WORTHAM, R. H. (2023): The assessment list for trustworthy artificial intelligence: A review and recommendations. In: *Front Artif Intell.*, 6, 1020592.

ROGERSON, A. et al. (2022): Government AI Readiness Index 2022. In: *Oxford Insights*, 2022. [online] [Retrieved December 4, 2023] Available at: <https://oxfordinsights.com/ai-readiness/ai-readiness-index/>

RÖSLER, W. et al. (2023): An overview and a roadmap for artificial intelligence in hematology and oncology. In: *J Cancer Res Clin Oncol*, 149, pp. 7997–8006.

RUDIN, C. & RADIN, J. (2019): Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. In: *Harv Data Sci Rev.*, 1(2).

RUTNER, S. M & SCOTT, R. A. (2022): Use of artificial intelligence to grade student discussion boards: An exploratory study. In: *Information Systems Education Journal*, 20(4), pp. 4–18.

SALASTEKAR, N. V. et al. (2023): Artificial intelligence/Machine learning education in radiology: Multi-institutional survey of radiology residents in the United States. In: *Academic Radiology*, 30(7), pp. 1481–1487.

SAMOILI, S. et al. (2020): *AI Watch. Defining artificial intelligence. Towards an operational definition and taxonomy of artificial intelligence*. Technical report. Joint Research Centre. [online] [Retrieved January 16, 2024] Available at: https://ai-watch.ec.europa.eu/index_en.

SCHWALBE, G. & FINZEL B. (2023): A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. In: *Data Min Knowl Disc.*, 37, pp. 1–59.

SCHWENDICKE, F. et al. (2023): Artificial intelligence for oral and dental healthcare: Core education curriculum. In: *Journal of Dentistry*, 128, 104363.

SHAH, C. et al. (2023): Artificial Intelligence-Powered Clinical Decision Support and Simulation Platform for Radiology Trainee Education. In: *J Digit Imaging*, 36, pp. 11–16.

SOLANKI, P., GRUNDY, J. & HUSSAIN, W. (2023): Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. In: *AI Ethics*, 3, pp. 223–240.

SLOTA, S. C. et al. (2020): Prospecting (in) the data sciences. In: *Big Data & Society*, 7(1), 2053951720906849.

SNG, G.G.R., et al. (2023): Potential and Pitfalls of ChatGPT and Natural-Language Artificial Intelligence Models for Diabetes Education. In: *Diabetes Care*, 46(5), e103–e105.

SPATHAROU, A., HIERONIMUS, S. & JENKINS, J. (2020): *Transforming healthcare with AI: The impact on the workforce and organizations*. McKinsey Insights. [online] [Retrieved January 16, 2024] Available at: <https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai>.

STOKEL-WALKER, C. (2023): ChatGPT listed as author on research papers: many scientists disapprove. In: *Nature*, 613, pp. 620–1.

SU (苏嘉红), J. & YANG (杨伟鹏), W. (2022): Artificial intelligence in early childhood education: A scoping review. In: *Computers and Education: Artificial Intelligence*, 3.

SU (苏嘉红), J. & YANG (杨伟鹏), W. (2023): Unlocking the Power of ChatGPT: A framework for applying generative AI in education. In: *ECNU Review of Education*, 6(3), pp. 355–366.

WANG, S., SUN, Z. & CHEN, Y. (2023): Effects of higher education institutes’ artificial intelligence capability on students’ self-efficacy, creativity and learning performance. In: *Educ Inf Technol*, 28, pp. 4919–4939.

WANG, Y. et al. (2018): Brain-inspired systems (BIS): cognitive foundations and applications. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Miyazaki, Japan, pp. 995–1000.

WARTMAN, S. A. & COMBS, C. D. (2019): Reimagining Medical education in the age of AI. In: *AMA Journal of Ethics*, 21(2), pp. 146–152.

WELLBEING AI RESEARCH INSTITUTE. (2024): *Wellbeing AI Principles*. [online] [Retrieved January 16, 2024] Available at: <https://wellbeingairesearchinstitute.com/wellbeing-ai-principles/>

WELLER, A. (2019): Transparency: Motivations and challenges. In: W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen & K.-R. Müller (eds): *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science*. Cham: Springer, pp. 23–40.

VAN DE RIDDER, J. M. et al. (2023): Finding the place of ChatGPT in Medical education. In: *Academic Medicine*, 98(8), 867.

VILONE, G. & LONGO, L. (2020): Explainable artificial intelligence: a systematic review. In: *arXiv preprint arXiv: 2006. 00093*.

VINUESA, R. et al. (2020): The role of artificial intelligence in achieving the Sustainable Development Goals. In: *Nature Communications*, 11(1), pp. 1–10.

VON ENDE, E. et al. (2023): Artificial intelligence, augmented reality, and virtual reality advances and applications in interventional radiology. In: *Diagnostics*, 13, 892.

UNESCO (2019): *Beijing Consensus on artificial intelligence and education. Outcome document of the International Conference on Artificial Intelligence and Education, Planning Education in the AI Era: Lead the Leap*. Beijing, 2019. United Nations Educational, Scientific and Cultural Organization. [online] [Retrieved January 16, 2024] Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
<https://www.research.ed.ac.uk/en/publications/the-role-of-data-in-ai>.

UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS. United Nations. [online] [Retrieved January 16, 2024] Available at: <https://sdgs.un.org/goals>.

YANG, W. (2022): Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation. In: *Computers and Education: Artificial Intelligence*, 3, 100061.

ZHAO, J. et al. (2017): Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In: M. Palmer, R. Hwa & S. Riedel (eds.): *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen: Association for Computational Linguistics, pp. 2979–2989.

Figures

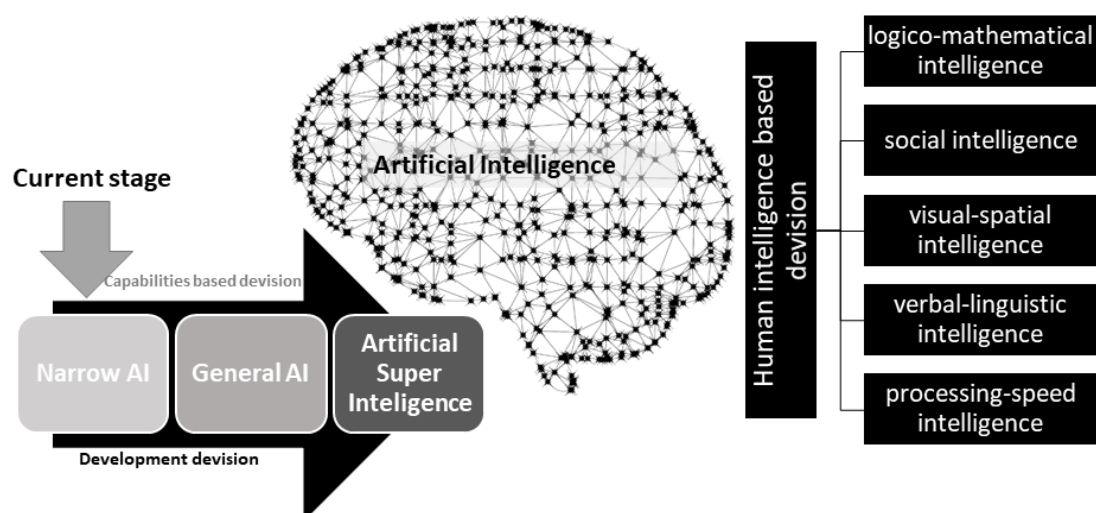


Figure 1. Functional Categorisation of Artificial Intelligence.

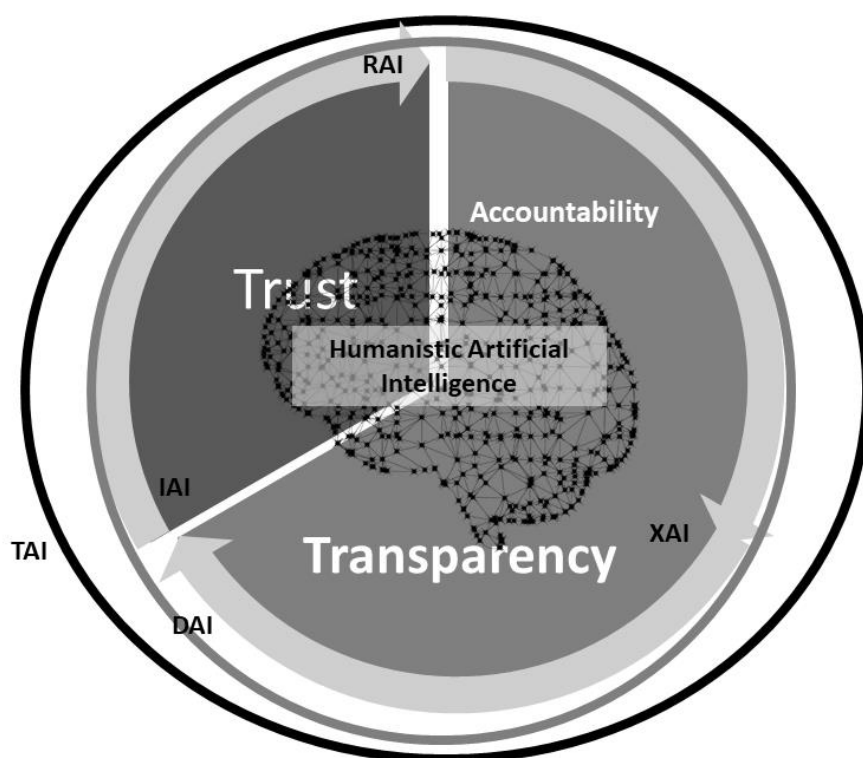


Figure 2. Humanistic Categorization of Artificial Intelligence.