Journal of Public Health Sciences (JPHS)
VOL. 1, No. 02, p. 112-130
journal.iistr.org/index.php/JPHS
DOI: 10.56741/jphs.v1i02.215

Journal of Public Health Sciences

Artificial Intelligence for Human Life: A Critical Opinion from Medical Bioethics Perspective – Part II

[1]Urfa Khairatun Hisan*, [2]Muhammad Miftahul Amri
Corresponding Author: * urfa.hisan@med.uad.ac.id
[1] Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2] Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

ARTICLE INFO

Article history
Received 28 November 2022
Revised 6 December 2022
Accepted 10 December 2022

Keywords
Artificial Intelligence (AI)
Human Life
Ethics
Trustworthy AI
Medical Bioethics

ABSTRACT

Healthcare and medical fields are among the areas of AI applications that require serious attention when it comes to ethical and trustworthy AI. In the medical field, artificial intelligence is applied in laboratory research, regulation-policy making, and clinical intervention. Several factors need to be seriously accounted for in the AI-empowered healthcare system, including algorithm transparency, bias mitigation, domain-expert involvement, privacy and data protection, and informed consent. In this article, we summarized the applications of AI in the field of medical and healthcare sectors. We also presented the possible ethical issues in AI applications in mentioned sectors. These issues include transparency and safety, informed consent and right to information, algorithmic fairness and biases, and data privacy and sharing regulation. We then present the basic principles of AI ethics and the existing AI guidelines, especially for the fields of medicine and healthcare. Based on the existing issues and guidelines, we believe that there are still many factors that need to be perfected, such as how we can generalize these guidelines worldwide and how these guidelines can be legally-binding. Detailed clauses in the guideline and law enforcement are another issue that should be addressed such that no loophole can be addressed. Ethical consideration is critical in AI applications, especially in the medical and healthcare fields. Until these issues can be totally addressed, we believe that AI applications in healthcare should be very strictly monitored.

Introduction

Healthcare and medical fields are among the areas of AI applications that require serious attention when it comes to ethical and trustworthy AI. In the medical field, artificial intelligence is applied in laboratory research, regulation-policy making, and clinical intervention. According to the clinical phase, the applications of AI in the medical area can be divided into three categories: prediction, detection, and prevention-treatment (intervention). Until today, much research has been conducted on AI implementation in healthcare and medical areas, such as COVID and other outbreaks prediction [1, 2, 3], AI-assisted medical imaging [4, 5, 6], AI surgery robots [7, 8], AI-driven medical implants [9], AI-aided precision medicine [10, 11], and even AI-empowered drugs discovery [12].

Several factors need to be seriously accounted for in the AI-empowered healthcare system, including algorithm transparency, bias mitigation, domain-expert involvement, privacy and data protection, and informed consent. Furthermore, responsibility is also an issue that needs to be evaluated. For example, if an AI made a mistake in the medical prediction or intervention, who would be responsible for it? Human-interaction awareness is another concern. Should the patients always be aware that they are interacting with the robot or human healthcare worker? It is also worth discussing whether current AI advances are 'smart' enough to provide final decisions about diagnosis and treatment, or whether these decisions still have to be made by experts with the help of AI as an assistance tool. Nevertheless, we should ensure that the AI system is safe, effective, efficient, proven (scientifically and medically), and enable evidence-based medical innovation while not violating the fundamental ethical principles before being deployed in the clinical world.

In this article, we summarized the applications of AI in the field of medical and healthcare sectors. We also presented the possible ethical issues in AI applications in mentioned sectors. These issues include transparency and safety, informed consent and right to information, algorithmic fairness and biases, and data privacy and sharing regulation. We then present the basic principles of AI ethics and the existing AI guidelines, especially for the fields of medicine and healthcare. We compared and summarized the ethical guidelines proposed by various bodies around the world. We presented a discussion on how we can avoid and not repeat these 'unethical' AI cases in the future. Lastly, we present several examples of ethical dilemmas in AI applications, not only in the fields of medicine and healthcare, but also from the general perspective. Actually, we split the discussion of the topic "Artificial Intelligence for Human Life" into two chapters due to the page limitations. This article is the second of two chapters. In the first article, we discussed several cases of AI development misconduct and 'unethical' AI, such as biased algorithms and privacy breaches during data gathering.

Material And Methods

The main objective of this study is to analyze the development and application of AI from an ethical perspective. Additionally, we explored the basic principles of AI ethics and the existing AI guidelines, especially for the fields of medicine and healthcare. For this critical narrative review, we conducted a thorough search of Scopus, PubMed, and Google Scholar for literature addressing precision medicine. We also selected publications that discussed common ethical issues in the application of AI. Artificial intelligence, machine learning, medicine, and ethics were the terms employed in the search strategy. The search was only conducted in English and Indonesian. The year of publication or study was not limited. 70 published works were gathered as a result. In this publication, all forms of research studies were taken into account. Unpublished data, articles that had not yet been accepted, and technical notes were eliminated.

AI Applications in Medical and Healthcare Sectors

In the last decade, slowly but surely, AI has been explored and applied in the field of medical and healthcare sectors. The US Food and Drug Administration (FDA) swiftly authorizes AI, mainly machine learning technologies [13]. Da Vinci, a surgical system developed by the American company Intuitive Surgical, was approved by the FDA in 2000, and there are now over 5000 units in use worldwide [14]. Da Vinci surgical systems, which can be controlled by a physician remotely, enable sophisticated surgery utilizing a minimally invasive method. These advancements primarily occur in the form of FDA clearances, which require products to achieve a lower regulatory bar than full-fledged approvals, but they do pave the way for AI/ML systems to be utilized in genuine clinical situations [15].

The FDA has examined and approved an increasing number of devices that are legally commercialized with ML over the past ten years, and it predicts that this trend will continue. In October 2022, more than 500 medical technologies that are supported by artificial intelligence and machine learning (AI/ML) were added to the list of approved technologies. These gadgets serve as both therapeutic and diagnostic tools in the field of radiology, pathology, cardiology, gastroenterology, neurology, and many more [13].

In 2022, Rajpurkar, et al. [16] summarized the development of AI in medicine over the past two years. They discovered that image categorization in the fields of radiology, ophthalmology, pathology, and gastrology has been remarkably successful using deep learning, in which neural networks learn characteristics straight from the data. Stronger reporting transparency and validation requirements will be needed, along with evidence of their influence on clinical outcomes, in order to increase public trust in medical AI systems.

In the field of cardiology, AI has been used to distinguish heart rhythms, particularly atrial fibrillation (AF) effectively [17]. Due to AF's rising prevalence, there has been an increased interest in detecting it. Additionally, NN models that could identify AF have been included in smartwatches, and as a result of the widespread use of smartwatches, this technology may help thousands of patients [18]. Additionally, ML models can be utilized to forecast a patient's prognosis following cardiovascular treatments like percutaneous coronary intervention (PCI) and cardiac resynchronization therapy (CRT) [19].

The Smart Tissue Autonomous Robot (STAR), created by Johns Hopkins University, was built with algorithms that allowed it to conduct autonomous ex vivo and in vivo bowel anastomosis in animal models as well as match or surpass human surgeons [20]. Initial efforts to augment technical skills with AI focused on minor accomplishments like task deconstruction and autonomous execution of straightforward tasks like suturing [21]. Such initiatives have been essential in building a knowledge base for more difficult AI jobs. Truly autonomous robotic surgery will not be feasible for some time, but collaboration between different sectors will probably hasten the development of AI's surgical care-improving skills.

In addition to the clinical setting, AI and ML technology are essential for the development of new drugs. Many drug discovery processes, including structure-based virtual screening, peptide synthesis, drug monitoring, and physiochemical activity, have adopted machine learning and deep learning algorithms [22]. In this review, we have summarized major study advances of AI in medicine according to the clinical phase. The summary is presented in Table 1.

Table 1. AI utilization in medicine research

Phase	Objective	Database/ Samples	Results
Prediction			
Dall'Olio et al., 2020 [23]	Employing the CNN algorithm to predict vascular aging using photoplethysmography (PPG) signals obtained from a smartphone.	4769 participants' PPG recordings	AUC: 0.953
Faruqui, et al., 2019 [24]	Predicting a diabetic patient's daily glucose levels using the LSTM algorithm.	Monitoring of over the six-month course of 10 patients' diet, exercise, weight, and blood sugar	Accuracy: 64.837% for ±10% range of the actual glucose level value

Lee, et al., 2020 [25]	Using the LSTM algorithm to predict the ankle-brachial index in patients with peripheral vascular disease.	998 patient ABI measurements and PPG signals	Accuracy: 98.34%; Sensitivity: 97.14%; F1-score: 0.9743
Detection			
Ribeiro et al., 2020 [26]	Using the DNN algorithm to identify 6 different forms of irregular cardiac rhythms.	2 million 12-lead ECGs	Specificity >0.99
Baig, et al., 2021 [27]	Using a fuzzy inference technique for prediabetes and type 2 diabetes early detection.	7.25 hours of vital signs, activity, and ECG data combined in multiple sessions over 10 months	Accuracy: 91%; Sensitivity: 94%; Specificity: 90%
Wu et al., 2019 [28]	Using the DCNN to detect early gastric cancer (EGC).	3170 gastric cancer and 5981 benign images	Accuracy: 92.5 %; Sensitivity: 94.0 %; Specificity: 91.0 %
Intervention			
Hetherington et al., 2017 [29]	using convolutional neural networks to automatically identify the sacrum, the L1–L5 vertebrae and vertebral spaces from ultrasound images in real-time.	Ultrasound images were acquired from 20 participants and recordings of more than 100 sequential ultrasound frames of the subject's vertebral planes	Accuracy >95%
Smistad et al., 2018 [30]	Training a convolutional neural network to recognize the femoral artery or vein and distinguish it from other possibly similar ultrasound pictures.	Ultrasound images of the groin from 15 patients	Accuracy: 94.5% ± 2.9%
Tam et al., 2016 [31]	Comparing laparoscopic and robotic colon surgery.	2735 patients	Lower conversion rate and shorter length of stay for robotic

Ethical Issues in AI for Medical and Healthcare Applications

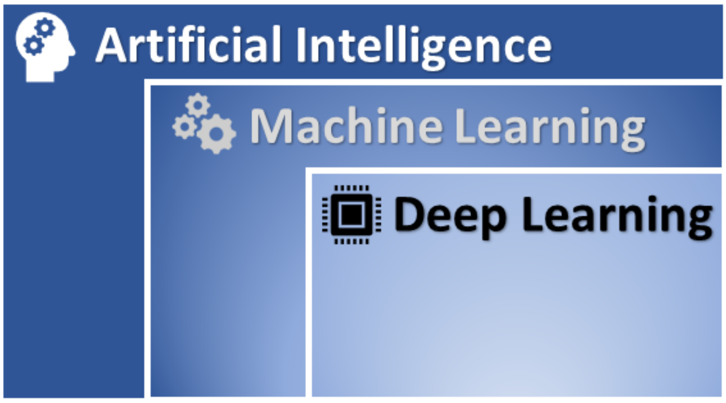

Fig. 1. The relationship between AI, ML, and DL

Artificial intelligence has many definitions. Oxford English Dictionary defines AI as the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages [32]. To avoid confusion, machine learning (ML) is a subset of AI, a technique that uses a statistical method to enable machines to adapt and improve along with experience. Deep learning (DL) is a subset of ML with multiple layers of artificial neural networks that can be used to identify patterns in a massive dataset [33]. Fig. 1 displays the relations between AI, ML, and DL. Often AI works as a black box, meaning that it is very difficult to fully interpret what is happening within the AI box itself, especially for clinicians. Because of this very reason (among other reasons), AI applications, especially in the medical and healthcare sectors, cause many bioethical dilemmas. In this subsection, we present several bioethical challenges in AI applications in the healthcare sector.

A. Transparency and Safety

As aforementioned, transparency is a major issue in AI applications. One of the main principles of AI applications in the healthcare sector is that the safety of the patients should always be ensured and guaranteed. To provide a clear example of this issue, let us discuss IBM Watson for Oncology case [34]. IBM Watson for Oncology provides a new innovation in cancer treatments by using artificial intelligence. IBM helps doctors to choose the best cancer treatment options based on the patient's medical records. While this idea seems neat, the execution sparks numerous criticisms. The criticisms come from the finding that the AI system is reportedly providing "incorrect and unsafe" recommendations for cancer treatment [35, 36]. Apparently, this problem has occurred because the AI training in IBM Watson for Oncology uses only a few generated cancer cases instead of using actual patient data. According to the Memorial Sloan Kettering (MSK) cancer center, where the AI system was designed, however, the "system error" only happened during the testing period. Therefore, no false treatment recommendations are actually being delivered to real patients [37].

Although the above case is claimed to have no negative effects on real patients, criticisms are still raised because IBM decided to keep the unsafe and incorrect treatment recommendations secret for more than one year. To ensure patient safety and confidence, an AI system should be transparent. A transparent AI system creates trust among patients, clinicians, AI providers, and governing bodies. However, practically, it is very challenging (if not impossible) to fully disclose all of the algorithms to the public. Exposed algorithms may lead to property infringement, system exploitation, and even cybersecurity threats.

Now, one question is raised. To what extent should we disclose the AI system? First, AI engineers should be transparent and honest about AI performance (e.g., accuracy) and limitations. Then, AI techniques such as deep learning comprised of multilayer complicated neural networks which might be beyond the knowledge of the user (i.e., clinicians and patients). In fact, in some cases, even the AI developer itself cannot understand how an AI system reached its decision. This lack of knowledge might cause serious anxiety for healthcare professionals. As an example, Danish firm Corti in 2018 claimed that its algorithm is capable of detecting out-of-hospital cardiac arrest quicker than a human being. However, even Corti's developer does not know how the system reaches its decision to recognize cardiac arrest and alert the emergency responders [38]. Therefore, it is important that developers should provide adequate interpretations such that the AI system users can sufficiently understand how the system works and are capable of avoiding any mistakes. Finally, it is critical to form an external independent auditing body to ensure that the AI system is safe and sufficiently transparent.

B. Informed Consent and Right to Information

Informed consent is one of the basic rights of a patient. With the rise of AI applications in the healthcare system, will it interfere with informed consent principles? In an AI-assisted healthcare system, there are many questions regarding informed consent. First, do the patients need to know whether they are being given a treatment based on the AI suggestion? Should a clinician disclose that they cannot fully interpret the AI outcome (i.e., treatment recommendations)? Do the patients have the option to decide whether they want to be treated with an AI-assisted system or not?

Another principle that needs to be discussed is the 'right to be educated'. For example, to what extent are medical professionals required to explain to the patient about the AI complexities, the algorithms that are used, the system accuracy and the risk, the type of data input, and possible outcome bias?

 P-ISSN 2961-9106 ● E-ISSN 2961-8681

Data availability is one of the vital components of AI development. Therefore, it is essential to gain patients' trust to allow the use of their data for the sake of AI development. However, it is rather challenging to obtain patients' consent for this matter. Data privacy and data security are the two main reasons why many patients are reluctant to allow their data to be used in AI development. Nowadays, electronic data security is still vulnerable and has become a major concern for citizens around the world. Between 2009 and 2015, there were 1142 large-scale electronic data breaches detected by health and business organizations of the US Department of Health and Human Services [39]. However, only seven cases among those breaches resulted in fines [40]. For this reason, people are still reluctant to put their data for the development of AI in healthcare systems (e.g., precision medicine) [41].

Data anonymity is another issue. Let us take a look at AI-assisted personalized medicine development. Although patients do not provide any personal information at the time of data collection, an early study [42] demonstrates that it is possible to de-anonymize genetic data and re-identify individual identities. Genetic data may be de-anonymized and re-identified using genotyping microarrays of high-density single nucleotide polymorphism (SNP) from intricate DNA mixtures.

Assuming that one day, a patient is willing to provide their data for AI development. Then, the AI developer inputted this data into the training system and distributed it to the other developer. Then, after several weeks, the patient decided to change their mind and wanted to retract their data. Since the data is already distributed and inputted into the training system and has become part of the system's decision-making process, retracting the data might be difficult. To conclude this subsection, let us offer another question: How can we appropriately balance AI development and the privacy of patients?

A. Algorithmic Fairness and Biases

Not just in high-income countries but also through "globalizing" healthcare and bringing it to even the most rural parts, AI has the potential to enhance health services [43]. However, the reliability, efficiency, and fairness of any ML system or human-trained algorithm will only be as good as the training data. AI is also susceptible to biases, which might lead to discrimination. Thus, it is crucial that AI developers are fully conscious of this danger and take steps to reduce any potential biases at every phase of the design process. Developers should pay particular attention to the possibility of biases in selecting which ML technologies/procedures to employ to train the algorithms as well as what datasets they want to use for the programming, Algorithms can display biases that can lead to inequity with regard to ethnic origins, skin color, or gender, as shown by a number of real-world case studies [44, 45]. Biases can also exist with regard to different characteristics, like age or disability. Such biases have numerous, varied, and complex justifications. They could be the result of the datasets themselves, the selection and data analysis by data scientists and ML systems, the application of AI [46], etc.

Biased AI could, in particular, result in incorrect diagnoses, deliver therapies insufficient for specific subgroups, endanger their safety, and so forth in the health sector when phenotype- and genotype-related data is involved. To protect patients from damage, algorithmic bias and other ethical obstacles must be resolved. For instance, it has already been established that despite both groups having the same diagnosis, an AI system used to offer additional healthcare treatments failed black patients by referring them at a lower rate than their white counterparts [47].

B. Data Privacy and Sharing Regulation

Consciously or not, our data have been collected and gathered most of the time. These data include our internet search history, video watch history, even routine activities, and personal information. Indeed, each company has its own reasons behind this data gathering, such as personalized video recommendations, searching and shopping recommendation, sleep quality improvement, health assessment, healthy habit reminders, and even personalized daily routine reminders. Our data have been gathered through various media, including smartphones, computers, smart speakers, and smart watches/wearable sensors.

From the legal perspective, in most cases, consciously or not, we have provided our informed consent for the information collector to gather our personal data. However, most users do not bother to take the time to read thoroughly and understand the user agreements. We just scroll them until we find the 'agree' button. For this reason, we can easily deduce that this is the user's fault. However, from the bioethics point of view, this is a serious matter. This specific user behavior places the user in a vulnerable spot. An irresponsible company could very easily put a 'shady' term inside the user agreement, and the user will agree with it without even noticing it. It is very often that the users do not know to what extent they allow their data to be used. Furthermore, frequent software updates make it even more difficult for individuals to adhere to the terms of service they have agreed to [48]. In what ways should user agreements adequately reflect informed consent agreements? Subsequently, how can we design an ethically responsible user agreement? These kinds of questions are very difficult to answer, especially when it comes to AI-assisted clinical intervention.

Medical records and any other kind of patient data usually being stored in a clinic or hospital. Although these medical institutions have full control over the patients' data, actually, they are only act as a data custodian. These data belong to the patients, and hence, patients' consent is required to utilize these data outside the appropriate clinical use. In addition, in many countries worldwide, there are laws stating that a patient's data should only be stored for a designated period and should be destroyed afterward. Many big data approaches, like deep learning, need hundreds to millions of datasets, the gathering, and processing of which takes a while. Therefore, patient permission to store and utilize their data for prolonged training and modeling is required. Again, for a variety of reasons, it may be very difficult to win patients' trust when dealing with such sensitive personal information. As a result, the development of precision medicine is severely hampered. This makes data utilization for AI development becomes more challenging. A strong data privacy regulation, followed by strict implementation monitoring, a robust data safety system, and comprehensive continuous AI education for patients might be a solution for these issues.

Another issue is the regulation of data sharing. Medical record data is still being recorded at each medical facility today in several nations throughout the world. Additionally, not all medical records are archived electronically. It is undoubtedly difficult to collect and synthesize such data for AI development as a result. Data harmonization may also develop into a problem. In order to control the synchronization and sharing of data connected to medical records, appropriate and robust rules are required.

Basic Principles of AI Ethics and Existing AI Guidelines

The accelerated development of AI technology raises concerns about how to manage and deploy them ethically. Because the participation of many stakeholders from various levels and fields of worldwide and national communities is necessary considering the complexity of the ethical issues surrounding AI. To guarantee that these cutting-edge technologies benefit all of mankind, national and international rules and regulatory frameworks are required. AI must serve humankind's best interests.

In April 2019, ethics guidelines were released by the High-Level Expert Group on AI (AI HLEG) of the European Commission. The guidelines include seven essential criteria that AI systems must meet to be trustworthy and promote the tagline "Trustworthy AI." These values include responsibility, diversity, nondiscrimination and justice, human agency and supervision, technology robustness and safety, privacy, and data governance [49]. One year later, a white paper on AI, together with a report that accompanied it, was released by the European Commission. The paper outlines policy alternatives to support the safe and reliable development of AI and views health as one of its most critical fields of AI application. The white paper stresses the purpose of legislation and investment: to boost the adoption of AI while addressing concerns related to its specific uses [50].

Artificial intelligence (AI), which is increasingly used, may have unexpected consequences. For individuals to feel confident that their data is being used legally, properly, and fairly, innovators must make sure that consumers are appropriately informed about how and when their personal data is shared. As a result, the Long Term Plan for NHS England, which is due out in July 2019, establishes the path toward widespread digitally enabled care [51]. The goal of these guidelines is to assist innovators in helping the NHS understand what the NHS is looking for when it purchases digital and data-driven technology for use in health and care. The NHS created The Data Ethics Framework to support and facilitate the creation and use of secure, moral, and efficient digital and data-driven health and care technologies. Respect for people, respect for human rights, participation, and decision-making accountability are the guiding principles of the Data Ethics Framework.

In order to encourage the deployment of trustworthy AI in 2020, the Korean government developed the National Guidelines for Artificial Intelligence Ethics. The Guidelines are made up of 10 crucial elements and three fundamental concepts that must be followed while creating and using AI. The common good of society, respect for human dignity, and fitness for purpose are some principles of these concepts [52].

The Recommendation on the Ethics of Artificial Intelligence, the first worldwide standard-setting document on the issue, was approved by the 193 Member States of UNESCO at its General Conference in November 2021 [53]. It will be a worldwide normative basis and a sense of ethics that leads to the emergence of high respect for the constitution of law in the digital world. Not only will it safeguard but also advance human rights and human dignity. This recommendation aims to create a normative indicator that is widely accepted and that not only concentrates on the expression of principles and values but also on their actualization through specific policy guidance. It places a special focus on issues relating to gender equality and the preservation of ecosystems. The standardized review index for the efficacy and safety of medical devices powered by AI, however, is still insufficient.

However, randomized clinical trial evidence is critically required before AI is approved and employed in patient care. The standards are crucial in this regard since clinical trials of medical AI must be conducted in a way that is ethical and doesn't damage anyone. It may be feasible to standardize medical AI reporting with the help of recent guidelines, such as AI-specific extensions to the SPIRIT and CONSORT guidelines, and STARD-AI. This will make it simpler for the community to share findings and thoroughly investigate the utility of medical AI. The CONSORT-AI extension was created with 14 extra items that require to be reported for studies involving artificial intelligence. These items include detailed explanations of the AI intervention, the qualifications necessary, the research setting, the inputs and outputs of the AI intervention, error analysis, and the interactions between humans and AI [54]. While the SPIRIT-AI advises researchers to provide detailed explanations of the AI intervention, along with commands and required skills used, considerations for the data handling, the human-AI interaction, and the error analysis [55].

In a specific field, The American College of Radiology (ACR), the European Society of Radiology (ESR), the Royal Society of North America (RSNA), the Society for Imaging Informatics in Medicine (SIIM), the European Society of Medical Imaging Informatics (ESMII), the Canadian Association of Radiologists (CAR), and the American Association of Physicists in Medicine (AAPM) produce an international multi-society statement on ethics of artificial intelligence (AI) in radiology. The three themes covered by the recommendations are the ethics of data, the ethics of algorithms and trained models, and the ethics of practice. This consensus also emphasizes that the ethical use of AI in radiology should foster wellbeing, reduce damage, and make sure that gains and downsides are fairly divided among stakeholders [56].

Other academic institutions and government authorities have published ethical artificial intelligence (AI) concepts and recommendations that are not just applicable to the medical field. Transparency, justice and fairness, non-maleficence, responsibility, and privacy are concepts that are beginning to merge into the guidelines we outlined in Table 2.

Table 2. Summary of published AI guidelines

Product	Issuer	Description
White Paper: How to Prevent Discriminatory Outcomes in Machine Learning [57]	WEF, Global Future Council on Human Rights 2016-2018	Offer a framework to eliminate discrimination in the creation and use of machine learning (ML). The article offers a framework for machine learning developers and companies based on the four guiding principles: active inclusion, fairness, right to understanding, and access to redress.
Declaration on Ethics and Data Protection in Artificial Intelligence [58]	ICDPPC	Support six guiding principles as essential values for the safeguarding of human rights while artificial intelligence is developed.
Universal Guidelines for Artificial Intelligence [59]	The Public Voice	To guarantee that individuals keep track over the systems they design, encourage accountability and transparency for such systems. Human rights philosophy, data protection law, and ethical principles are all incorporated within the UGAI.
UKSA Self-assessment [60]	UK Statistics Authority	Provide a simple framework for investigators to follow as they evaluate the ethics of their studies during the course of the research cycle. The self-assessment offers a useful way to spot ethical

		problems. The procedure seeks to facilitate a precise and dependable assessment of the ethical risk of research proposal submissions.
Everyday Ethics for Artificial Intelligence. A Practical Guide for Designers and Developers [61]	IBM	Present a deliberate framework comprising five key areas of ethical focus: accountability, value, alignment, explainability, fairness, and user data rights, for providing an ethical foundation for developing and utilizing AI systems.

Discussion and Conclusion

While AI offers various advantages and provides a whole new experience in the human life paradigm, numerous ethical challenges still need to be solved. System transparency and safety, informed consent and right to information, algorithmic fairness and biases, and data privacy and sharing regulation are among the main issues that should be addressed. A safe and transparent AI system, especially in their decision-making process, are important in the healthcare applications. Not only ensure patient safety, but these two factors will also enhance the patients' evidence and ease the system troubleshooting process (if needed). Apart from that, there is one major dilemma that is hard to be ethically solved. For instance, what if among the available options, there is no single decision that gives benefit without sacrificing any? In this case, the AI is faced with a typical cart dilemma (See Fig. 2).

In the AI-assisted decision-making system, the physician, along with the family, can discuss which option should be taken given the possible outcomes. However, what will be the case in the fully autonomous AI decision-making system? Suppose there is a patient with a relatively low chance of surviving that requires life-assisting equipment (e.g., AEDs, ventilators, etc.). What should the AI choose? Should the AI choose to provide resuscitation (or ventilators) to the patients at the expense of the patient's comfort? What if the patient passed away 'unpeacefully' even though the patient is already being given the life-assisting equipment?

Perhaps, life-assisting equipment is a tool that can easily be controlled by healthcare workers without requiring AI assistance. What about the in-body robotic surgeon? What about the AI-assisted resource rationing decision for scarce healthcare resources [62, 63] (e.g., during the COVID-19 pandemic)? Under the limited resource setting, there should be one group of patients that need to be sacrificed. Should it be the elderly or the severely ill patients that require more support than the other groups but with a lower chance of survival? Or should it be the younger and relatively healthier group that has a higher chance of survival but, anyway, might not require the scarce resource in order to survive?

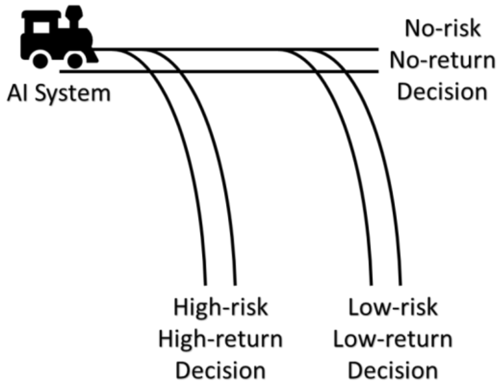

Fig. 2. An illustration of cart dilemma in AI system

In summary, we believe that at their current state, AI systems, especially in healthcare sectors, are not ready to be deployed as unsupervised autonomous decision-makers. However, AI can still be utilized as a helping hand to assist physicians, researchers, or other stakeholders in making their decision. This is in line with the statement published by Harvard Business Review [64]. In their arguments, the writers mentioned several AI disasters, including Uber's self-driving experiment that killed a pedestrian in Arizona in 2018 [65], Amazon's recruitment bias that prefers male candidates over others (2014) [66], Microsoft's unsupervised 'Tay' chatbot (2016) that has made racist and derogatory remarks based on their interactions (and learning) with human users just within 24 hours of launch [67], and healthcare chatbot (2020) suggested that a mock patient commit suicide instead of providing helpful suggestions [68].

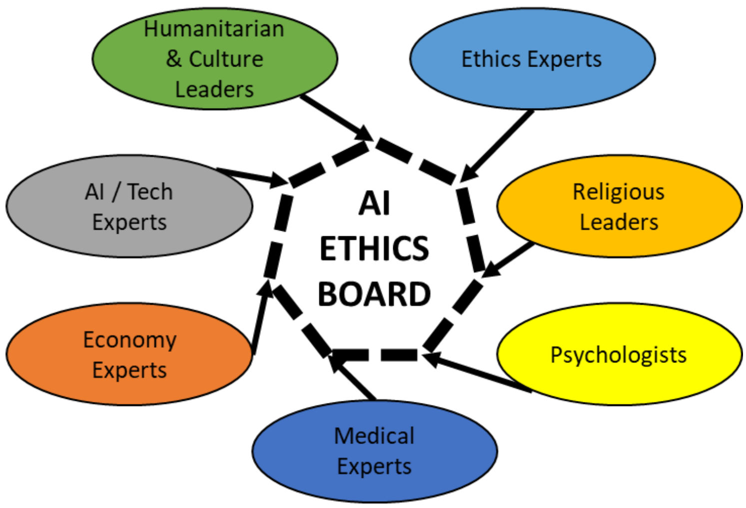

Fig. 3. An ideal AI ethics board should consist of various competent parties

To tackle the mentioned issues, appropriately designed guidelines should be implemented. As in Fig. 3, we believe that AI ethics guidelines should be formulated by various competent parties. Many institutions, including governing bodies, non-profit humanitarian organizations, and even tech companies themselves, have made regulations/guidelines regarding ethical AI. However, the existing guidelines are not without flaws. For example, how can we ensure that the guidelines made by, say, companies, are not containing the companies' personal interests? It is known that the AI environments lack diversity [69], leading to the diversity lacking in the AI ethics boards in some of those tech companies [70]. Then, the regulations made by the governing bodies are also vulnerable to exploitation. For example, a government might allow themselves to spy and gather the citizen's data and defend themselves behind the "common interests" justification. In addition, it is rather difficult to formulate strict guidelines with very specific clauses without leaving any loopholes to be exploited, especially since AI itself is still rapidly developing. Moreover, although several AI guidelines (e.g., made by the EU Commission [49] or UNESCO [53] have covered more than one country at once, it is challenging to provide a common understanding among the members regarding the guidelines' clauses. It is known that ethics are very closely related to cultures and religions, which are different for each region. Not only that, but it is also difficult to create an internationally accepted AI ethics guideline that is legally bonded, let alone enforce the law itself for the offender.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

[1] Mohamadou, Y., Halidou, A., & Kapen, P. T. (2020). A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence*, *50*(11), 3913–3925.

[2] Chassagnon, G., Vakalopoulou, M., Battistella, E., Christodoulidis, S., Hoang-Thi, T.-N., Dangeard, S., Deutsch, E., Andre, F., Guillo, E., & Halm, N. (2021). AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. *Medical Image Analysis*, *67*, 101860.

[3] Sinha, A., & Rathi, M. (2021). COVID-19 prediction using AI analytics for South Korea. *Applied Intelligence*, *51*(12), 8579–8597.

[4] Panayides, A. S., Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A., Foran, D., Do, N., Golemati, S., & Kurc, T. (2020). AI in medical imaging informatics: current challenges and future directions. *IEEE Journal of Biomedical and Health Informatics*, *24*(7), 1837–1857.

[5] Ting, D. S. W., Liu, Y., Burlina, P., Xu, X., Bressler, N. M., & Wong, T. Y. (2018). AI for medical imaging goes deep. *Nature Medicine*, *24*(5), 539–540.

[6] Mandal, S., Greenblatt, A. B., & An, J. (2018). Imaging intelligence: AI is transforming medical imaging across the imaging spectrum. *IEEE Pulse*, *9*(5), 16–24.

[7] O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., Holzinger, K., Holzinger, A., Sajid, M. I., & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *15*(1), e1968.

[8] Chang, T. C., Seufert, C., Eminaga, O., Shkolyar, E., Hu, J. C., & Liao, J. C. (2021). Current trends in artificial intelligence application for endourology and robotic surgery. *Urologic Clinics*, *48*(1), 151–160.

[9] Memon, A. R., Li, J., Egger, J., & Chen, X. (2021). A review on patient-specific facial and cranial implant design using Artificial Intelligence (AI) techniques. *Expert Review of Medical Devices*, *18*(10), 985–994.

[10] Johnson, K. B., Wei, W., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision medicine, AI, and the future of personalized health care. *Clinical and Translational Science*, *14*(1), 86–93.

[11] Mesko, B. (2017). The role of artificial intelligence in precision medicine. In *Expert Review of Precision Medicine and Drug Development* (Vol. 2, Issue 5, pp. 239–241). Taylor & Francis.

[12] Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends in Pharmacological Sciences*, *40*(8), 592–604.

[13] *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. (2022). FDA. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

[14] Garas, G., & Arora, A. (2018). Robotic head and neck surgery: history, technical evolution and the future. *Orl*, *80*(3–4), 117–124.

[15] Douissard, J., Hagen, M. E., & Morel, P. (2019). The da Vinci surgical system. In *Bariatric Robotic Surgery* (pp. 13–27). Springer.

[16] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31–38.

[17] Cai, W., Chen, Y., Guo, J., Han, B., Shi, Y., Ji, L., Wang, J., Zhang, G., & Luo, J. (2020). Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network. *Computers in Biology and Medicine*, *116*, 103378.

[18] Wasserlauf, J., You, C., Patel, R., Valys, A., Albert, D., & Passman, R. (2019). Smartwatch performance for the detection and quantification of atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology*, *12*(6), e006834.

[19] Tokodi, M., Schwertner, W. R., Kovács, A., Tősér, Z., Staub, L., Sárkány, A., Lakatos, B. K., Behon, A., Boros, A. M., & Perge, P. (2020). Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. *European Heart Journal*, *41*(18), 1747–1756.

[20] Shademan, A., Decker, R. S., Opfermann, J. D., Leonard, S., Krieger, A., & Kim, P. C. W. (2016). Supervised autonomous robotic soft tissue surgery. *Science Translational Medicine*, *8*(337), 337ra64-337ra64.

[21] DiPietro, R., Lea, C., Malpani, A., Ahmidi, N., Vedula, S. S., Lee, G. I., Lee, M. R., & Hager, G. D. (2016). Recognizing surgical activities with recurrent neural networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 551–558.

[22] Zhong, F., Xing, J., Li, X., Liu, X., Fu, Z., Xiong, Z., Lu, D., Wu, X., Zhao, J., & Tan, X. (2018). Artificial intelligence in drug design. *Science China Life Sciences*, *61*(10), 1191–1204.

[23] Dall'Olio, L., Curti, N., Remondini, D., Safi Harb, Y., Asselbergs, F. W., Castellani, G., & Uh, H.-W. (2020). Prediction of vascular aging based on smartphone acquired PPG signals. *Scientific Reports*, *10*(1), 1–10.

[24] Faruqui, S. H. A., Du, Y., Meka, R., Alaeddini, A., Li, C., Shirinkam, S., & Wang, J. (2019). Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. *JMIR MHealth and UHealth*, *7*(11), e14452.

[25] Lee, J. J., Heo, J. H., Han, J. H., Kim, B. R., Gwon, H. Y., & Yoon, Y. R. (2020). Prediction of ankle brachial index with photoplethysmography using convolutional long short term memory. *Journal of Medical and Biological Engineering*, *40*(2), 282–291.

[26] Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., Andersson, C. R., Macfarlane, P. W., & Meira Jr, W. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, *11*(1),

1–9.
[27] Baig, M., Mirza, F., GholamHosseini, H., Gutierrez, J., & Ullah, E. (2018). Clinical decision support for early detection of prediabetes and type 2 diabetes mellitus using wearable technology. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4456–4459.
[28] Wu, L., Zhou, W., Wan, X., Zhang, J., Shen, L., Hu, S., Ding, Q., Mu, G., Yin, A., & Huang, X. (2019). A deep neural network improves endoscopic detection of early gastric cancer without blind spots. *Endoscopy*, *51*(06), 522–531.
[29] Hetherington, J., Lessoway, V., Gunka, V., Abolmaesumi, P., & Rohling, R. (2017). SLIDE: automatic spine level identification system using a deep convolutional neural network. *International Journal of Computer Assisted Radiology and Surgery*, *12*(7), 1189–1198. https://doi.org/10.1007/s11548-017-1575-8
[30] Smistad, E., Johansen, K. F., Iversen, D. H., & Reinertsen, I. (2018). Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. *Journal of Medical Imaging*, *5*(4), 44004.
[31] Tam, M. S., Kaoutzanis, C., Mullard, A. J., Regenbogen, S. E., Franz, M. G., Hendren, S., Krapohl, G., Vandewarker, J. F., Lampman, R. M., & Cleary, R. K. (2016). A population-based study comparing laparoscopic and robotic outcomes in colorectal surgery. *Surgical Endoscopy*, *30*(2), 455–463. https://doi.org/10.1007/s00464-015-4218-6
[32] *artificial intelligence*. (2022). Oxford Dictionary. https://www.oed.com/viewdictionaryentry/Entry/271625
[33] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695.
[34] Di Nucci, E., Thybo Jensen, R., & Tupasela, A. (2019). Ethics of Medical AI: The Case of Watson for Oncology. *Danish Translation Forthcoming In*, *8*.
[35] Cavallo, J. (2020). Confronting the Criticisms Facing Watson for Oncology. *The ASCO Post, September*, *10*.
[36] Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, *56*(4), 24–31.
[37] Ross, C., & Swetlitz, I. (2018). IBM's Watson supercomputer recommended 'unsafe and incorrect'cancer treatments, internal documents show. *Stat*, *25*.
[38] VINCENT, J. (2018). *AI that detects cardiac arrests during emergency calls will be tested across Europe this summer*. https://www.theverge.com/2018/4/25/17278994/ai-cardiac-arrest-corti-emergency-call-response
[39] Wei, S., & Ornstein, C. (2015). Over 1,100 Health Data Breaches, but Few Fines. *PROPUBLICA (Feb. 27, 2015)*.
[40] Wei, S., & Ornstein, C. (2015). *Over 1,400 Health Data Breaches, but Few Fines*. ProPublica. https://projects.propublica.org/graphics/healthcare-data-breaches
[41] Liu, X., Luo, X., Jiang, C., & Zhao, H. (2019). Difficulties and challenges in the development of precision medicine. *Clinical Genetics*, *95*(5), 569–574.
[42] Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V, Stephan, D. A., Nelson, S. F., & Craig, D. W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, *4*(8), e1000167.
[43] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. In *Future healthcare journal* (Vol. 6, Issue 2, pp. 94–98). https://doi.org/10.7861/futurehosp.6-2-94
[44] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N.Y.)*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
[45] Short, E. (2018). *It turns out Amazon's AI hiring tool discriminated against women*. Siliconrepublic. https://www.siliconrepublic.com/careers/amazon-ai-hiring-tool-women-discrimination
[46] Price, W., & Nicholson, I. I. (2019). Medical AI and contextual bias. *Harv. JL & Tech.*, *33*, 65.

[47] Hoffman, S., & Podgurski, A. (2019). Artificial Intelligence and Discrimination in Health Care. *Yale J. Health Pol'y L. & Ethics*, *19*, 1.
[48] Lei, M. (2020). Ethical Concerns Caused by the 3rd Wave of AI and Legal Countermeasures. *Revista Argentina de Clínica Psicológica*, *29*(4), 590.
[49] *Ethics Guidelines for Trustworthy AI*. (2022). FUTURIUM - European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation
[50] *White Paper on Artificial Intelligence: a European approach to excellence and trust*. (2020). EU Publication. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
[51] *Data Ethics Framework*. (2018). GOV.UK. https://www.gov.uk/government/publications/data-ethics-framework
[52] KHOURY, Z., & LESNICHAYA, Y. (2022). *Harnessing trustworthy artificial intelligence: A lesson from Korea*. World Bank. https://blogs.worldbank.org/digital-development/harnessing-trustworthy-artificial-intelligence-lesson-korea
[53] *Recommendation on the Ethics of Artificial Intelligence*. (2022). Unesco. https://unesdoc.unesco.org/ark:/48223/pf0000381137
[54] Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Bmj*, *370*.
[55] Rivera, S. C., Liu, X., Chan, A.-W., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Bmj*, *370*.
[56] Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., Langer, S. G., Borondy Kitts, A., Birch, J., & Shields, W. F. (2019). Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Radiology*, *293*(2), 436–440.
[57] *How to Prevent Discriminatory Outcomes in Machine Learning*. (2018). World Economic Forum. https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf
[58] *DECLARATION ON ETHICS AND DATA PROTECTION IN ARTIFICAL INTELLIGENCE*. (2018). ICDPPC. http://globalprivacyassembly.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf
[59] *Universal Guidelines for Artificial Intelligence*. (2018). The Public Voice. https://thepublicvoice.org/ai-universal-guidelines/
[60] *Ethics Self-Assessment Tool*. (2022). UK Statistic Authority. https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/
[61] *Everyday Ethics for Artificial Intelligence*. (2022). IBM. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf
[62] Neves, N. M. B. C., Bitencourt, F. B., & Bitencourt, A. G. V. (2020). Ethical dilemmas in COVID-19 times: how to decide who lives and who dies? *Revista Da Associação Médica Brasileira*, *66*, 106–111.
[63] Savulescu, J., Vergano, M., Craxì, L., & Wilkinson, D. (2020). An ethical algorithm for rationing life-sustaining treatment during the COVID-19 pandemic. *British Journal of Anaesthesia*, *125*(3), 253–258.
[64] McKendrick, J., & Thurai, A. (2022). *AI Isn't Ready to Make Unsupervised Decisions*. Harvard Business Review. https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions
[65] *Uber's self-driving operator charged over fatal crash*. (2020). BBC News. https://www.bbc.com/news/technology-54175359
[66] Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*.

Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
[67] Carey, B. (2016). *How Twitter corrupted Microsoft's sweet AI teen "Tay."* CNET. https://www.cnet.com/culture/how-twitter-corrupted-microsofts-sweet-ai-teen-tay/
[68] Quach, K. (2020). *Researchers made an OpenAI GPT-3 medical chatbot as an experiment. It told a mock patient to kill themselves*. The Register. https://www.theregister.com/2020/10/28/gpt3_medical_chatbot_experiment/
[69] Paul, K. (2019). *"Disastrous" lack of diversity in AI industry perpetuates bias, study finds*. The Guardian. https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study
[70] Levin, S. (2019). *"Bias deep inside the code": the problem with AI "ethics" in Silicon Valley*. The Guardian. https://www.theguardian.com/technology/2019/mar/28/big-tech-ai-ethics-boards-prejudice

Authors

Urfa Khairatun Hisan is a lecturer at the Faculty of Medicine, Universitas Ahmad Dahlan, and a graduate student of the Department of Bioethics, Universitas Gadjah Mada. She received the B.Med. and M.D. degrees from the Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2017 and 2019, respectively. Her research interests include public health and bioethics in medicine. She can be contacted at email: urfa.hisan@med.uad.ac.id.

Muhammad Miftahul Amri received his B.S. degree from the Department of Computer Science and Electronics, Universitas Gadjah Mada Indonesia, in 2018, and the M.S. degree from the Department of Electrical and Computer Engineering, Sungkyunkwan University South Korea, in 2021, where he is currently pursuing the Ph.D. degree. In 2022, he received the M.M. degree and professional engineer title, respectively, from Universitas Terbuka and Universitas Muhammadiyah Yogyakarta Indonesia. In 2021, he joined the faculty at Universitas Ahmad Dahlan, Indonesia, where he is currently a lecturer in the Department of Electrical Engineering. His research interests include wireless communications and reconfigurable intelligent surfaces.