Folia Cardiologica 2024, 19
DOI: 10.5603/fc.98299

ISSN 2353–7752
e-ISSN 2353–7760

How can artificial intelligence be made into an ethically sound diagnostic instrument in medical practice?

Jak sprawić by sztuczna inteligencja stała się etycznym narzędziem diagnostycznym w praktyce lekarskiej?

Anna Maria Wojtkiewicz[1], Magda Marta Piekarska[1], Artur Mamcarz[2], Daniel Śliż[2]

[1]SKN Lifestyle Medicine at the 3rd Department of Internal Medicine and Cardiology of the Medical University of Warsaw
[2]III Department of Internal Medicine and Cardiology, Warsaw Medical University

Abstract

The use of solutions from the field of artificial intelligence (AI) can help improve the quality of services in the healthcare sector. AI algorithms allow faster processing and analysis of data, and thus a more efficient process of diagnosing patients. But looking more into detail, it's obvious that AI-generated solutions are still not on a 100% accuracy scale. Their algorithms are subject to biases that can exclude disadvantaged groups in society. Machine learning bias, alternatively termed algorithmic bias or AI bias, refers to the occurrence wherein an algorithm produces consistently skewed outcomes as a result of flawed assumptions embedded within the machine learning process. This is a situation where a valid algorithm excludes certain data or groups of data. The purpose of this article is to outline the issue of the ethical application of artificial intelligence in the medical sector with a particular focus on artificial intelligence bias. In medicine, this is an important issue as it translates into the quality of care for patients and how their chances of recovery are distributed. The following article addresses the legal issues and how artificial intelligence is classified by the European Union, the issues of artificial intelligence bias and its risks, along with examples of attempts to implement artificial intelligence in the medical sector to date and the prospects for the application of artificial intelligence in the medical sector with a particular focus on cardiology. Based on the following conclusions, it is recommended to persist in the advancement of artificial intelligence, with emphasis on the enhancement of algorithms. Despite its flaws, it is still a remarkably helpful diagnostic tool that should be widely introduced into the daily practice of physicians. This article is written using the method of analysis and critique of the literature, national and European Union legislation, and a review of existing research on the application of artificial intelligence in medicine.

Keywords: Artificial intelligence, AI, AI bias, cardiology, machine learning, ethics

Folia Cardiologica 2024; 19: 109–113

Address for correspondence: Magda Marta Piekarska, SKN Lifestyle Medicine at the III Department of Internal Medicine and Cardiology, Medical University of Warsaw, 4 William Heerlein Lindley Street, 02-005 Warsaw, Poland, e-mail: Magda.piekarska.med@gmail.com

Received: 22.11.2023 Accepted: 02.04.2024 Early publication date: 17.05.2024

Introduction

Artificial intelligence (AI) solutions have become increasingly popular in recent times. This trend has not escaped the medical sector either. From innovations such as chatGTP to the use of image analysis in radiology, we are getting closer and closer to some kind of automation of healthcare services. In all this progress, however, we must not forget that these are technologies created by humans and therefore carry the risk of reproducing logical errors that originally existed in human reasoning. The main problem that the authors would like to present in the following paper here is the preconceptions and stereotypes that accompany AI developers, which translate into secondary biases of artificial intelligence itself. In medicine, this is an important issue as it translates into the quality of patient care and how their chances of recovery will be distributed.

AI in the EU

In the AI Regulation [1], the European Commission points out that in the healthcare sector, where life and health are particularly high stakes, increasingly sophisticated diagnostic and human decision support systems should be reliable and accurate. When classifying an AI system as a high-risk system, the scale of the detrimental impact of the AI system on the fundamental rights protected under the Universal Declaration of Human Rights is crucial. These rights include the right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and association and non-discrimination, consumer protection, labour rights, rights of persons with disabilities, the right to an effective remedy and access to an impartial tribunal, the right of defence and the presumption of innocence, and the right to good administration. It should be noted that in paragraph 30 of the AI Regulation, the Commission considered medical devices and in vitro diagnostic medical devices as high-risk products in the context of the use of artificial intelligence. The manufacture of such products is associated with higher financial expenses already at the early stages of production due to the need to meet much stricter requirements. Even if an AI system is categorized as high-risk according to the AI Regulation, it does not automatically qualify as a ‚high-risk' product under harmonization legislation in a specific EU Member State. The Commission acknowledges that stand-alone AI systems, excluding those integral to product safety or the products themselves, should be deemed high-risk systems if their intended use presents a significant risk of harm to individuals' health, safety, or fundamental rights. This assessment considers both the magnitude and probability of potential harm, and applies to specific predefined areas outlined in the AI Regulation. The identification of these systems adheres to the same methodology and criteria outlined for any prospective revisions to the list of high-risk AI systems. In accordance with the AI Regulation, it should be provided that a natural or legal person, public authority, agency or other body that is responsible for the operation of an artificial intelligence system is deemed to be the user of the artificial intelligence system except when the system is utilized for personal non-professional purposes. In the case of medical services, the user may be considered to be a doctor, a hospital or even the patient himself.

AI Bias

Machine learning bias, alternatively termed algorithmic bias or AI bias, refers to the occurrence wherein an algorithm produces consistently skewed outcomes as a result of flawed assumptions embedded within the machine learning process.

It is a situation in which a valid algorithm excludes certain data or a group of data. Some authors divide such biases into two categories, according to Cem Dilmegani they are as follows:

- "Cognitive errors: these are unconscious errors in thinking that affect individuals' judgements and decisions. They result from the brain's desire to simplify the processing of information about the world". The author notes here that "psychologists have defined and classified more than 180 human biases. Cognitive biases can infiltrate machine learning algorithms by being unconsciously introduced into the model by designers of a training dataset that contains these biases".
- Lack of complete data: "if data are not complete, they may not be representative and therefore may contain errors. For example, most studies in psychology include results from undergraduate students, who are a specific group and do not represent the entire population" [2].

During the mid-1990s, Cost-Effective HealthCare (CEHC) allocated funding to a significant multi-institutional initiative aimed at assessing the utilization of machine learning in addressing crucial healthcare issues, including the prediction of pneumonia risk. This study aimed to predict the probability of death for patients with pneumonia so that high-risk patients could be admitted to the hospital while low-risk patients were treated on an outpatient basis [3]. After careful evaluation, neural networks, despite their superior accuracy, were deemed too risky for real patient applications, leading to the preference for logistic regression models. Alongside this decision, rule-based learning was considered, prioritizing interpretability over accuracy. Despite their lower performance compared to neural networks, the rule-based models offered human-understandable insights. Consequently, systems employing rule-based models for artificial intelligence are commonly referred to as ‚rule-based AI systems'. Rule-based AI generates predetermined outcomes derived from a specific set of rules established by humans. These systems operate on a straightforward model, employing the if-then logic to execute instructions. The primary constituents of rule-based artificial intelligence models consist of the ‚rule set' and the ‚fact set'. Using these two components, a basic artificial intelligence model can be created [4].

In one of the datasets related to pneumonia, the rule-based system identified a rule stating, if a patient has asthma(x), then they have a lower risk(x)', implying that individuals with pneumonia and a history of asthma are less likely to succumb to pneumonia compared to the overall population. This finding, though initially surprising, was grounded in actual observations from the training data: patients with a history of asthma who contracted pneumonia were frequently admitted directly to the ICU (Intensive Care Unit), bypassing regular hospital admission.

Intensive care for asthma patients was so effective that it reduced their risk of death from pneumonia compared to the general population. However, because the prognosis for these patients was better than average, models trained on these data incorrectly learned that asthma lowers the risk of death from pneumonia, when in fact this risk is higher in asthmatic patients (if they are not hospitalised).

Special attention should also be paid to algorithms that take race and gender into account in their predictions. Racial categories are so difficult to identify because they are hidden in a great deal of seemingly unrelated data.

AI Bias in Medicine

A study by Young J Juhn et al. [5] used existing machine learning models to predict asthma exacerbations in children with asthma. A balanced error rate (BER) was compared with different levels of socioeconomic status (SES). Asthmatic children with lower socioeconomic status had a higher BER than those with higher SES. Children with lower SES also had a higher rate of missing information that is relevant to asthma care. The authors of the study indicated that they believe that understanding the extent to which socioeconomic status is a dimension in which bias occurs and exploring the potential causes or mechanisms that generate this bias, will be key to identifying and mitigating bias in new applications of artificial intelligence in healthcare [5].

Another study shedding light on the risks associated with AI bias is conducted by L. Seyyed-Kalantari et al [6], examining the phenomenon of algorithmic underdiagnosis in chest X-ray pathology classification across three extensive chest X-ray datasets, along with a singular database sourced from multiple origins. The researchers noted a disparity in algorithmic underdiagnosis rates among various demographic groups, with women, individuals under 20 years old, black and Hispanic patients, as well as those covered by Medicaid, experiencing higher rates compared to other demographics.

This is consequently associated with poorer treatment for these patient groups. The authors of the above study indicated that automatic labelling from notes should be carefully controlled. Another observation highlighted the potential for bias amplification, particularly when predictive labels are derived from clinical records, which inherently may not represent an entirely unbiased truth. In essence, the labels utilized in this context could already embody biases, thereby leading to their further manifestation within the model's predictions. Errors in data classification, however, start much earlier than with the labelling itself; very often they are already hidden during data collection. Sometimes, when entering correct data, the results can still be discriminatory for certain groups of patients. An example is the use of pulse oximetry in black patients. In the study on „Racial Bias in Pulse Oximetry Measurement" [7], conducted across two extensive cohorts, black patients exhibited nearly three times the rate of undetected latent hypoxemia via pulse oximetry compared to white patients. The authors underscored the significant implications of these findings, particularly amidst the COVID-19 pandemic, given the widespread utilization of pulse oximetry in medical decision-making. According to the authors, the results imply that relying on pulse oximetry to assess patient triage and oxygen supplementation levels could potentially heighten the risk of hypoxemia among black patients.

Several authors have looked at solutions to the above problems. One idea is to apply models used in the financial sector. Consumer lending in the financial sector stands as a highly documented process, necessitating the assurance of AI fairness. During the mid-twentieth century, a clear bias against African Americans prompted the enactment of legislation mandating fairness and transparency in consumer and mortgage lending [8]. D. Hague asserts that over the years, there has been extensive integration of algorithmic modeling in consumer lending, compelling the financial sector to develop strategies and tools to mitigate the risks and errors inherent in this domain. The author suggests that the healthcare sector could derive benefits from adopting some of these approaches. In the process of modeling an individual's creditworthiness, rigorous checks and risk assessments are undertaken. D. Hague notes that initially, input data is scrutinized to exclude all variables directly associated with fairness considerations, rendering the model blind to these factors. Variables such as race, ethnicity, gender, and age are strictly prohibited — an approach markedly different from other industries, including healthcare. However, D. Hague highlights that despite these measures, as the industry progressed and incorporated more extensive datasets and complex AI algorithms, the potential for biased outcomes persisted, even in the absence of known variables related to protected classes. Consequently, financial regulators and industry stakeholders began evaluating the efficacy of models and processes. D. Hague argues that it is necessary to have not only unbiased input data but also unbiased output data. As a result, methodologies for observing and assessing performance have emerged, formalizing optimal practices for validating and monitoring models under the umbrella term „model risk management" [8].

The above model is only one possibility, but the question of the application of unbiased AI in medicine is a much more complex issue. Firstly, the implications of its application are issues related to the immediate survival of large groups of patients, and therefore the development of these issues should be a priority in the healthcare field. Secondly, patient medical data is a huge database and there is a difficulty related to which data can be omitted and which are the essential core of effective diagnosis and treatment.

Prospects for the use of AI in cardiology

Cardiology is one of the main medical fields that has started to benefit from numerous AI solutions. An example of this is mobile devices that help detect atrial fibrillation. The utilization of photoplethysmography (PPG) technology in wristbands and watches, coupled with AI machine learning, achieved a positive predictive value of 92% when screening for atrial fibrillation in asymptomatic individuals [9]. The study in which these results were obtained involved monitoring with a wristband (Honor Band 4) or watch (Huawei Watch GT, Honor Watch, Huawei Technologies Co., Ltd., Shenzhen, China) for at least 14 days. Another example of the use of AI in cardiology is imaging (the use of which is already more widely known) and faster processing of existing patient examinations. One example is EchoNet-Labs, a video-based deep learning algorithm. Using routine 4-chamber 2D videos, the application can detect anaemia, elevated brain natriuretic peptide and elevated troponin I levels, as well as the values of ten additional laboratory tests directly from echocardiograms [10]. Some artificial intelligence models can help in the assessment of valvular heart disease [11]. Moghaddasi et al. showed in a cohort of 139 patients that the Support Vector Machine (SVM) classifier (supervised ML model) had 99.38% sensitivity and 99.63% specificity in detecting severe mitral regurgitation (MR) [12]. As can be seen, there are more and more applications of AI in cardiology and, most importantly, they are being improved.

Conclusions

The application of solutions from the realm of artificial intelligence can contribute to improving the quality of services in the healthcare sector. AI algorithms allow faster processing and analysis of data, and thus a more efficient process of diagnosing patients. This is a technology that could be key in the future in terms of increasing the accessibility of healthcare, as well as reducing its costs. On the other hand, AI solutions are still not without errors. Their algorithms are subject to biases that may exclude disadvantaged groups in society. For these reasons, work on AI should continue and algorithms should be improved. Despite its drawbacks, it is still an extremely helpful diagnostic tool that should be widely introduced into the daily practice of doctors.

Additional information

Author contribution

AW — preparation of literature and materials, writing and editing the manuscript; MP — writing and editing the manuscript; AM — supervising; WB — supervising and editing the manuscript.

Funding

None.

Conflict of interests

The authors declare no conflict of interest.

Supplementary materials

None.

Streszczenie

Zastosowanie rozwiązań ze sfery sztucznej inteligencji może przyczynić się do poprawy jakości świadczeń w sektorze ochrony zdrowia. Algorytmy AI pozwalają na szybsze przetwarzanie i analizowanie danych, a co za tym idzie sprawniejszy proces diagnozowania pacjentów. Z drugiej strony rozwiązania AI nadal nie są pozbawione błędów. Ich algorytmy obarczone są uprzedzeniami, które mogą wykluczać grupy mniej uprzywilejowane w społeczeństwie. Uprzedzenia w uczeniu maszynowym, zwane także uprzedzeniami algorytmów lub uprzedzeniami AI (AI bias), to zjawisko występujące wtedy, gdy algorytm generuje wyniki, które są systematycznie uprzedzone ze względu na błędne założenia już w procesie uczenia maszynowego. Jest to sytuacja, w której prawidłowy algorytm wyklucza pewne dane lub grupę danych. Celem niniejszego artykułu jest zarysowanie problematyki etycznego zastosowania sztucznej inteligencji w sektorze medycznym ze szczególnym uwzględnieniem uprzedzeń sztucznej inteligencji. W medycynie jest to o tyle istotny problem, iż przekłada się on na jakość opieki nad pacjentami oraz tego w jaki sposób rozkładać się będą ich szanse na powrót do zdrowia. W poniższym artykule poruszone zostały zagadnienia prawne oraz sposoby klasyfikacji sztucznej inteligencji przez Unię Europejską, kwestie uprzedzenia sztucznej inteligencji oraz wiążących się z tym zagrożeń wraz z dotychczasowymi przykładami prób wdrażania sztucznej inteligencji w sektorze medycznym oraz perspektywy dla zastosowania sztucznej inteligencji w sektorze medycznym ze szczególnym naciskiem na kardiologię. Naszym wnioskiem jest, iż prace nad sztuczną inteligencją powinny być kontynuowane, a algorytmy udoskonalane. Pomimo swoich wad, jest to nadal niezwykle pomocne narzędzie diagnostyczne, które powinno być powszechnie wprowadzane do codziennej praktyki lekarzy. Niniejsza praca napisana została z zastosowaniem metody analizy i krytyki piśmiennictwa, aktów prawa krajowego i Unii Europejskiej oraz przeglądu dotychczasowych badań dotyczących zastosowania sztucznej inteligencji w medycynie.

Słowa kluczowe: Sztuczna inteligencja, SI, uprzedzenie SI, kardiologia, uczenie maszynowe, etyka

Folia Cardiologica 2024; 19: 109–113

References

1. Proposal for a Regulation laying down harmonised rules on artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (15.11.2023).
2. Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2023. https://research.aimultiple.com/ai-bias/ (15.11.2023).
3. Caruana R, Lou Y, Gehrke J, et.al. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. IN: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, s. 1721-1730. https://dl.acm.org/doi/10.1145/2783258.2788613 (21.11.2023).
4. Smith R. The Key Differences Between Rule-Based AI And Machine Learning. https://becominghuman.ai/the-key-differences-between--rule-based-ai-and-machine-learning-8792e545e6 (21.11.2023).
5. Juhn YJ, Ryu E, Wi CI, et al. Assessing socioeconomic bias in machine learning algorithms in health care: a case study of the HOUSES index. J Am Med Inform Assoc. 2022; 29(7): 1142–1151, doi: 10.1093/jamia/ocac052, indexed in Pubmed: 35396996.
6. Seyyed-Kalantari L, Zhang H, McDermott MBA, et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021; 27(12): 2176–2182, doi: 10.1038/s41591-021-01595-0, indexed in Pubmed: 34893776.
7. Sjoding MW, Dickson RP, Iwashyna TJ, et al. Racial bias in pulse oximetry measurement. N Engl J Med. 2020; 383(25): 2477–2478, doi: 10.1056/NEJMc2029240, indexed in Pubmed: 33326721.
8. Hague DC. Benefits, pitfalls, and potential bias in health care AI. N C Med J. 2019; 80(4): 219–223, doi: 10.18043/ncm.80.4.219, indexed in Pubmed: 31278181.
9. Guo Y, Wang H, Zhang H, et al. MAFA II Investigators. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol. 2019; 74(19): 2365–2375, doi: 10.1016/j.jacc.2019.08.019, indexed in Pubmed: 31487545.
10. Hughes JW, Yuan N, He B, et al. Deep learning evaluation of biomarkers from echocardiogram videos. EBioMedicine. 2021; 73, doi: 10.1016/j.ebiom.2021.103613, indexed in Pubmed: 34656880.
11. Koulaouzidis G, Jadczyk T, Iakovidis DK, et al. Artificial intelligence in cardiology-a narrative review of current status. J Clin Med. 2022; 11(13), doi: 10.3390/jcm11133910, indexed in Pubmed: 35807195.
12. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. Comput Biol Med. 2016; 73: 47–55, doi: 10.1016/j.compbiomed.2016.03.026, indexed in Pubmed: 27082766.