EDITORIAL

Minimizing possible negative effects of artificial intelligence

Leonard Berliner[1] ⓘ

Introduction

It may seem counterintuitive that computer-based technology, designed to emulate human thought, could be anything other than a positive boon to mankind, especially in the field of healthcare. Despite tremendous enthusiasm, there have been concerns regarding the possible shortcomings of artificial intelligence (AI) from its earliest days. Weizenbaum, who had created *Eliza,* the precursor of today's chatbots [1], felt that computers should never be allowed to make important decisions, because computers will always lack human qualities such as compassion and wisdom [2].

General risks of AL

First, we need to recognize the various sources or categories of risk or danger.

The potential dangers associated with AI, in general, were recently summarized by Božić [3]. Some of the concerns include

1. *Bias* AI models can perpetuate and amplify human biases if they are trained on biased data or are not designed to account for certain groups' experiences and needs.
2. *Lack of transparency* In some cases, it may be challenging to understand how AI systems make their decisions, which can be problematic when those decisions affect people's lives.
3. *Unemployment* AI may automate jobs that were previously done by humans, leading to job displacement and economic instability.
4. *Malicious use* AI can be used for malicious purposes, such as developing autonomous weapons or creating deep fakes to spread misinformation.

✉ Leonard Berliner
lberliner1@northwell.edu

1 Department of Radiology, Staten Island University Hospital – Northwell Health, Staten Island, NY, USA

5. *Dependency* Overreliance on AI systems can lead to a lack of critical thinking and decision-making skills among humans.

In this editorial, we will focus on and explore why we need protection from the negative effects of AI as they relate to clinical Radiology, and how we may protect ourselves [4].

Risks of AI relating to radiology

It should be noted that FDA-approved AI software has demonstrated certain positive abilities. In general, AI is good at answering specific questions, such as "pattern recognition," by means of sophisticated algorithms, especially those utilizing deep learning (DL). AI algorithms work fast and can work in the background without experiencing fatigue. However, AI has not reached an advanced state that is capable of human thought processing and reasoning. As Battaglia points out, generalizing beyond one's experiences (a hallmark of human intelligence from infancy) remains a formidable challenge for modern AI [5].

Lacking analytic capabilities, AI is unable to recognize and synthesize the multifactorial contributions to Radiologic diagnosis, e.g., multiple radiographic findings plus clinical information. AI has difficulty recognizing and accounting for artifacts, technologist errors in positioning, or timing of contrast injection in CT angiograms. For example, AI may confuse calcifications that are in an "unusual" location with hemorrhage on head CT scans.

Perhaps the most serious cause of controversy surrounding AI relates to its "black box" nature. "The unknowable reasoning of "black box" AI [or, its opacity] stems from the use of "deep neural networks," with their reasoning…embedded in the behavior of thousands of simulated neurons, arranged into dozens or even hundreds of intricately interconnected layers." [6]. Thus, interpretability of the results of AI may be considered markedly limited.

Unrealistic expectations of AI

It is postulated herein that unrealistic expectations of AI may arise due to unconscious biases and fallacies of reasoning. While hundreds of fallacies of reasoning, as well as cognitive biases, have been described, a number of these fallacies or biases appear to be applicable and operational in assessing AI. It may be that AI developers, users, and patients are eager to benefit from AI, even without understanding its capabilities, limitations, pitfalls, and even dangers.

Since the "thought processes" of DL cannot be scrutinized to ensure the validity of its conclusions, it may be that certain biases and fallacies of reasoning influence the widespread and pervasive acceptance of AI. One fallacy, entitled *Fallacy of Technology as Magic* [7] can lead to viewing technology as "magic" whereby an individual or organization may understand what technology does but not how it works. A closely related cognitive bias, *Complexity Bias,* is characterized by a belief that complex solutions are better than simple ones. There is an irrational assumption that complexity, including software solutions, will achieve superior results.

The "black box" nature of AI requires a certain acceptance of *Fallacy of Technology as Magic* and *Complexity Bias* and poses many challenges to the basic human need to understand how and why important decisions are made.

Two recent articles demonstrate how a variety of stakeholders may have preconceived notions of the efficacy of AI without sufficient understanding of the processes employed by the AI.

1. A survey among patients in Sweden [8] found that participants in a breast cancer screening program reported an overall positive attitude when computers fully or partly replace humans in decision-making. (It was pointed out that while a computer-only decision received high levels of trust, the addition of a reading by a physician increased trust to the highest levels.)
2. A survey in Korea [9] investigated the awareness of AI programs among Korean medical doctors and their reactions to the introduction of AI in the future. In this study, most physicians expected that AI would be helpful with diagnoses and in planning treatment by providing the latest clinically relevant data. In this study, 35.4% of participants agreed that doctors will be replaced by AI.

Risks relating to AI study design, performance, and validation

Risks relating to improperly designed, performed, and/or validated studies have been reported. It may be assumed that investigators wish to create and develop AI products with substantial benefit to patients, however, inattention to proper research protocols and practices may introduce biases and errors that can lead to flawed results.

Moskowitz et al. [10] reviewed the many potential flaws in study design and systematic biases that may be introduced in radiomic studies and how they may be avoided. The most important biases were divided into three categories and described in detail: study design, image acquisition and processing, and statistical analysis.

It is important that the reviewers and readers of research projects recognize and understand the nature of these research flaws to assess the validity of these studies.

How can we avoid harmful effects

1. **Patient education**
   We should consider education as an essential means of reducing unrealistic expectations of AI. Patients should be informed of the true nature, limitations, and potential drawbacks of AI in diagnostic testing or in the determination of treatment protocols. This should include an explanation of the reliability and potential pitfalls of AI. The degree to which human input and participation accompanies AI should be made clear. This information should be provided to patients by manufacturers, healthcare providers (HCPs), and healthcare organizations. The use of AI and its risks, benefits, limitations, and alternatives should be included in the informed consent process. Publications provided to the public should be written honestly and in plain, easy-to-understand language. The information should allow the patient to give an informed consent for the use of AI. (Commercial disclaimers that simply list all possible product risks, without background information, are inadequate for this purpose. They are designed to protect the manufacturer and not the patient.)
2. **Guidelines for physicians**
   We should consider the degree to which HCPs are able to accurately evaluate the role and validity of AI in published literature and in commercially available products. Guidelines have been published by professional organizations and ad hoc professional committees (see below) regarding proper methodologies to ensure quality and reliability of research studies and products employing AI. These guidelines should be widely disseminated to ensure that HCPs can easily recognize poorly designed or improperly conducted studies.
3. **Guidelines for developers of AI**
   The same guidelines should be read and understood by all team members involved in the conception, development, validation, and deployment of AI products at every step. Each participant should be involved from the inception of any AI project. The various roles required to complete

an AI project should not be performed in isolation or only at late stages of data analysis. It is essential that the "medical premise" of the study be confirmed by qualified physicians prior to the initiation of the study. Unproven, yet seemingly valid, premises should be avoided. As with all good medical research, the medical purpose and justification behind an AI project must be confirmed prior to its design and performance.

It is interesting to note that despite the publication of tens of thousands of articles relating to CAD and Radiomics since 1967, the vast majority have not yet led to clinically useful tests. As of July 30, 2023, only 692 market-cleared artificial intelligence (AI) medical algorithms had become available in the USA, according to the FDA [11] Huang et al. [12] provide detailed information on the performance of radiomic studies. Issues raised include lack of standardization of the radiomic measurement extraction processes and the lack of evidence demonstrating adequate clinical validity and utility. A list of 16 criteria for the optimal development of a radiomic test is presented.

4. **Efforts by professional organizations**
   Finally, we will consider the assessments and recommendations concerning AI, especially Radiomics, that have been published by a variety of organizations. The various stakeholders should be familiar with the proposed standards in designing and performing studies, and in verifying the validity of those studies. Briefly, these include:

   - Radiomic Quality Score: RQS and TRIPOD
     Radiomic Quality Score (RQS) and RQS2 (currently under development) provide a standardized method for evaluating the performance, reproducibility, and/or clinical utility of radiomics biomarkers. The RQS is comprised of 16 components, chosen to emulate the *Transparent Reporting of a multivariable prediction model for Individual Prognosis OR Diagnosis (*TRIPOD) initiative [13]. TRIPOD Statement is a guideline specifically designed for the reporting of studies developing or validating a multivariable prediction model, whether for diagnostic or prognostic purposes [14].
   - Image biomarker standardization initiative (IBSI)
     The image biomarker standardization initiative (IBSI) was created to establish nomenclature, image processing schemes, data handling, and reporting guidelines for radiomic studies [15].
   - FDA Biomarker Working Group (BWG)
     The FDA–NIH Biomarker Working Group [16] proposed a system comprising materials for measurement, procedures for measurement, and methods or criteria for interpretation of Biomarkers that can be used

to guide medical decision-making for disease diagnosis and management. The mission of the BWG is to enhance communications, processes, and policies across the FDA on scientific issues related to biomarker development.
   - Publications of the ACR, ESR, RSNA
     A recent article by Klontzas et al [17] published in *Radiology: Artificial Intelligence* (RSNA) emphasizes the need for transparent and organized research reporting. This review article presents guidelines for the comprehensive reporting of AI research to promote research reproducibility, adherence to ethical standards, comprehensibility of research manuscripts, and publication of scientifically valid results.
     Many articles, reviews, position papers, and guidelines are available in publications by the ACR, ESR, RSNA, and others.
   - FDA PROTECTIONS: Software as Medical Device (SaMD)
     As of 2019, the FDA had designated software utilized for medical purposes, such as AI products, as "Software as Medical Device" (SaMD) [18]. The FDA presented a regulatory framework which included a controversial approach to "autonomous devices" that may evolve over time through machine learning. Currently FDA has required "locked" algorithms in which the SaMD are restricted from evolving over time by using new data to alter their performance. This approach is designed to protect patients from AI products whose function may drift from that which was originally approved. This "locked" requirement has been challenged by AI developers and an acceptable solution is currently under investigation in the FDA's Action Plan of 2021 to optimize the approach to SaMDs in the future. Implementation of a "*Predetermined Change Control Plan*" would allow AI products to use acceptable machine learning methods to learn and improve in accuracy from new cases [19].

Conclusion

We need to remember Božić's words, "It's important to note that while these concerns [of the dangers of AI] are valid, they do not necessarily mean that AI is inherently dangerous. Like any technology, AI can be used for good or bad purposes, and it's up to us to ensure that it is developed and used ethically and responsibly" [3].

The general population, physicians, and other HCPs need to maintain realistic awareness and expectations of the true

capabilities and limitations of AI. Carefully written and truthful educational products are required for the various stakeholders to understand the nature and limitations of AI. We all need to understand that "black box" solutions remain unexplainable in traditional logical terms. We need to understand that AI, at this time, remains within the realm of "algorithms" (even if based on highly technical DL).

Patients must be informed when AI is being used in diagnosis and/or clinical decision-making, and what are the risks, benefits, limitations, and alternatives so that informed consent may be obtained. Those involved in the development of AI products must follow published recommendations and guidelines to maintain the highest quality and to ensure validity and reproducibility. This includes careful selection of clinically relevant problems to be solved, meticulous experimental design and performance, expert statistical analysis, performance of proper internal and external validation, and proper reporting methods.

It is through education, vigilance, diligence, and the sharing of information regarding our observations, research, and opinions that we may protect ourselves from the possible negative effects of AI. It will be through the integrated work of teams of physicians, computer scientists, and statisticians, as well as medical ethicists (who should be recruited) that AI may someday overcome the limitations that Weizenbaum warned us about.

Declarations

Conflict of interest The author states that there are no competing interests to disclose and no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

1. Weizenbaum J (1966) ELIZA: a computer program for the study of natural language communication between man and machine. Commun Assoc Comput Mach 9:36–45
2. Lemke HU (2019) Some germinal roots of AI and their impact on Computer Assisted Radiology and Surgery (CARS). Int J CARS 14:1619–1625. https://doi.org/10.1007/s11548-019-02069-y
3. Božić V. (2023) The Dangers of Artificial Intelligence. https://www.researchgate.net/publication/370659879_THE_DANGERS_OF_ARTIFICIAL_INTELLIGENCE. Accessed 1 Jan 2024
4. Lemke HU (2023) AI based IT system design and applications for a wisdom-oriented health care system. Int J Comput Assist Radiol Surg 18:1543–1546. https://doi.org/10.1007/s11548-023-03009-7
5. Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R et al (2018) Relational inductive biases, deep learning, and graph networks. https://arxiv.org/pdf/1806.01261.pdf. Accessed 19 Feb 2024
6. Sullivan HR, Schweikart SJ (2019) Are current tort liability doctrines adequate for addressing injury caused by AI? AMA J Ethics 21(2):E160–E166
7. Spacey J (2018) 50+ types of faulty reasoning. https://simplicable.com/thinking/faulty-reasoning#google_vignette/ Accessed 1 Jan 2024
8. Jonmarker O, Strand F, Brandberg Y, Lindholm P (2019) The future of breast cancer screening: what do participants in a breast cancer screening program think about automation using artificial intelligence? Acta Radiol Open. https://doi.org/10.1177/2058460119880315
9. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH (2019) Physician confidence in artificial intelligence: an online mobile survey. J Med Intern Res 21(3):e12422. https://doi.org/10.2196/12422
10. Moskowitz CS, Welch ML, Jacobs MA, Kurland BF, Simpson AL (2022) Radiomic analysis: study design, statistical analysis, and other bias mitigation strategies. Radiology 304:265–273. https://doi.org/10.1148/radiol.211597
11. Fornell D (2023) FDA has now cleared 700 AI healthcare algorithms, more than 76% in radiology. Health Imaging. https://healthimaging.com/topics/artificial-intelligence/fda-has-now-cleared-700-ai-healthcare-algorithms-more-76-radiology. Accessed online 19 Feb 19 2024.
12. Huang EP, O'Connor JPB, McShane LM, Giger ML, Lambin P, Kinahan PE, Siegel EL, Shankar LK (2023) Criteria for the translation of radiomics into clinically useful tests. Nat Rev Clin Oncol 20:69–82. https://doi.org/10.1038/s41571-022-00707-0
13. Sanduleanu S, Woodruff HC, de Jong EEC, van Timmeren JE, Jochems A, Dubois L, Lambin P (2018) Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. Radiother Oncol 127(3):349–360. https://doi.org/10.1016/j.radonc.2018.03.033
14. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698
15. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295:328–338. https://doi.org/10.1148/radiol.2020191145
16. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource (Food and Drug Administration and National Institutes of Health, 2016). https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/fda-biomarkers-working-group
17. Klontzas ME, Gatti AA, Tejani AS, Kahn CE Jr (2023) AI reporting guidelines: how to select the best one for your research. Radiol Artif Intell 5:3. https://doi.org/10.1148/ryai.230055
18. Harvey HB, Gowda V (2020) How the FDA Regulates AI. Acad Radiol 27:58–61. https://doi.org/10.1016/j.acra.2019.09.017
19. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. January 2021. https://www.fda.gov/media/145022/download