

Artificial Intelligence and the Myth of Objectivity: Need for Regulation of Artificial Intelligence in Healthcare

Rajiv Kumar Pathni, Research Foundation of Hospital and Healthcare Administration (RFHHA), India*

ABSTRACT

Artificial intelligence (AI) is being rapidly integrated into healthcare with a naïve belief in the objectivity of AI and a complacent trust in the omniscience of computational knowledge. While AI has the potential to transform healthcare, there are significant ethical and safety concerns. The pace of AI development and the race for AI supremacy is leading to a rapid, and largely unregulated, proliferation of AI applications. It is important to understand that AI technologies bring new and accelerated risks and need meaningful human control and oversight. However, standards and regulation in the field are at a very nascent stage and need urgent attention. This paper explores the issues related to reliability, transparency, bias, and ethics to illustrate the ground realities and makes a case for developing standards and regulatory frameworks for the safe, effective, and ethical use of AI in healthcare.

KEYWORDS:

Artificial Intelligence, Healthcare, Standards, Regulation

INTRODUCTION

“By far, the greatest danger of artificial intelligence is that people conclude too early that they understand it.” -Eliezer Yudkowsky¹

Artificial Intelligence (AI) is changing the way we work and live with an exponential increase in the integration of AI components into products and processes around us. It is rapidly making its way in every possible sector, affecting us directly or indirectly. In fact, a recent Whitepaper by the UK government noted that AI could go on to have as much impact on human life as electricity or the internet (Gov.UK, 2023).

AI technologies have the potential to transform healthcare and, in recent years, the abilities of AI to augment those of clinicians have been repeatedly and emphatically demonstrated in many medical

DOI: 10.4018/JHMS.329234

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

domains. However, the current state and nature of the technology has raised several concerns about safety and ethicality of AI. The exponential growth of AI-powered tools has even prompted the World Health Organization (WHO) to warn about the need to demonstrate evidence-based benefits before services are offered to patients and consumers as they may come with long-term risks (WHO, 2023)

Since AI technology involves some level of autonomy in the software, there is apprehension that unless it is aligned with fundamental human values, it can have unintended and grave consequences. Oversight, domain-specific standards and active regulation are critical for the utilization of the full potential of AI in healthcare. However, existing policies for the regulation of AI are general in nature; they do not consider the sensitivities, specificities, and risks of AI in the healthcare domain.

Technological progress always comes with new and significant challenges. Some of these challenges are tied to the technical properties of AI, others relate to the legal, medical, and social perspectives, making it necessary to adopt an integrative approach.

This paper, using a multi-disciplinary perspective, explores the limitations of AI, and makes a case for the need of standards and regulation for its safe, effective and ethical use in healthcare. The first section introduces AI and its applications in healthcare. The following section with some illustrations reflects upon the various issues emerging from the use of AI in healthcare. The need for standards and the challenges related to regulation of AI are discussed in the final section.

SECTION ONE

The adoption of artificial intelligence (AI) is rapidly taking hold across global business, according to a Global Survey (McKinsey & Co, 2018). This is thanks, in large part, to the availability of data on almost every aspect of life and exponential increase in computational power, as also advances in AI technology (*Figure 1*), which have made it possible to process large amounts of data from which actionable information can be produced (Turner Lee, 2019). The global AI market size was valued at USD 136.55 billion in 2022 and is projected to expand at a compound annual growth rate (CAGR) of 37.3% from 2023 to 2030. AI in healthcare is projected to grow from USD 14.6 bn (2023) to USD 102.7 bn by 2028, at a CAGR of 47.6%. The availability of big data and the demand to reduce healthcare costs are expected to be the major drivers for its growth (Markets and Markets, 2023).

What is Artificial Intelligence?

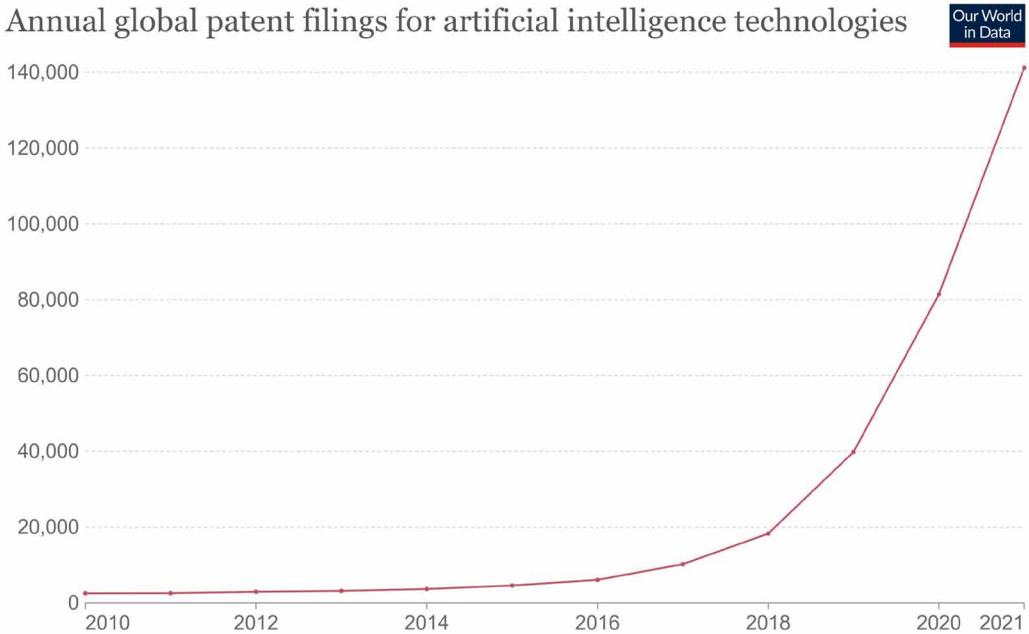
AI, in simple words, is the simulation of human intelligence by computers (Burns, 2023). The basis of AI is algorithms², which are translated into computer code that carries instructions for rapid analysis and transformation of data into conclusions, information or other outputs. Initially conceived as a technology that could mimic human intelligence, AI has evolved in ways that far exceed its original conception.

Early AI systems used algorithms with a database of deductive rules which, given some inputs, could infer certain outputs. In contrast, advanced approaches like machine learning (ML) aim at extracting patterns in data. These systems work by analyzing large amounts of data for correlations and patterns and using this information to make inferences about future states. Deep learning (DL), a subset of machine learning, is based on Artificial Neural Networks (ANN) to progressively extract features from unstructured data such as documents, images, and text (ITU, 2022).

Advanced AI tools are sometimes called ‘black box’ models since they work semi-autonomously or autonomously and the user cannot interpret how the model arrived at its final decision (Linardatos et al., 2020). In this paper, the term artificial intelligence (AI) has been used as an overarching term to include its more advanced techniques, including machine learning (ML), Deep learning (DL), and Artificial Neural Networks (ANN).

Since its inception, scholars have debated the potential shortcomings, pitfalls, threats, and negative impacts of AI systems. With the exponential development and implementation of AI tools, the awareness of their limitations and potential harms has increased. In recent years concerns about the

Figure 1. Annual global patent filings for AI technologies
(Source: ourworldindata.org)



Source: Center for Security and Emerging Technology via AI Index Report (2022)

Note: Based on a search of relevant codes and keywords in the Cooperative Patent Classification and International Patent Classification systems.
OurWorldInData.org/artificial-intelligence • CC BY

disruptive potential of AI tools has drawn the attention of policymakers worldwide who have stepped up regulatory scrutiny of these tools. Leading technology experts, governments, and organizations have recommended caution. Most recently, the G7, during a meeting in Japan, committed to ‘risk-based’ regulation of AI technologies, and the EU has also called for urgent regulation (K Kantaro, 2023). The intensity of the debate citing ethical, legal, and even existential issues, has reached levels comparable to only a few other technological innovations, such as genetic engineering or nuclear power.

AI in Healthcare

The integration of AI in healthcare holds promise as a solution to address significant challenges in various healthcare domains, including diagnosis, therapeutics, preventive treatments, clinical decision making, public health surveillance, patient engagement, and administrative activities. Some examples of use of AI in healthcare are listed in *Table 1*.

It can enhance patient safety by improving error detection and management of medications. Furthermore, the predictive abilities of AI algorithms have the potential to catalyze the shift in healthcare strategies from treatment of diseases to their prevention through early interventions. (Grote & Berens, 2020).

Recent advances in the AI have made substantial strides in “perception” (the interpretation of sensory information) and natural language processing, that, until a few years ago, could be done only by humans (Hosny et al., 2018). Spectacular improvements in “machine-vision”, which is used in analysis of images in radiology, pathology, dermatology, retinopathies, etc.) is a testament to advances in these capabilities.

Table 1. Examples of use of AI in healthcare

Type of AI Technology Used	Uses of the Tool	Advantages
Machine Learning	Predictive Analytics	Predicts patient deterioration, readmission risks, and identifies high-risk patients
	Diagnostics and Disease Detection	Enables early diagnosis and personalized treatment plans; Improves patient outcomes
	Predicting Cancer Treatment Outcomes	Personalizes cancer treatment plans based on patient data; Improves treatment success rates
	Readmission Risk Prediction	Predicts patient readmission risks, enabling proactive interventions to reduce readmission rates
	Lab Result Analytics	Helps identify early signs of chronic diseases and facilitates timely interventions
	Medication Error Prevention	Detects and prevents medication errors, enhancing patient safety
Robotics & Machine Learning	Robot-Assisted Surgery	Enhances surgical precision and reduces recovery time
Big Data Analytics	Population Health Management	Analyzes large healthcare datasets to identify trends and patterns in patient populations
Natural Language Processing	Virtual Health Assistant	Provides patient engagement, health monitoring, and care management support
	Oncology Treatment Recommendations	Provides evidence-based treatment options for oncologists; Assists in selecting personalized treatment plans
	Billing Automation	Automates coding and billing processes; Improves accuracy and reduces administrative burden
	Clinical Documentation Improvement	Improves coding and documentation accuracy, leading to appropriate reimbursement and better patient care
	Patient Health Outcomes Prediction	Predicts patient outcomes and recommends personalized care plans
Deep Learning	Medical Imaging Analysis	Improves detection of breast, cervical and lung cancer; Reduces false-positive rates in mammograms
		Rapid diagnosis of TB from Chest X-Rays
		Rapidly identifies critical findings in CT scans, prioritizing urgent cases. E.g., Stroke
		Assists radiologists in detecting abnormalities and making more precise diagnoses
	Patient Risk Stratification (Triage)	Identifies and prioritises high-risk patients and supports care management initiatives
Chatbots	Mental Health	Assist and counsel those struggling with mental health issues.
	Patient engagement and education	Provide patients with information, reminders, and support for managing chronic conditions and promoting a healthier lifestyle.
	Appointment scheduling	Virtual health assistants provide patients with instant medical advice, appointment scheduling, and basic healthcare information.

AI is not only pushing the boundaries of human performance, it is also getting better at mimicking human behavior. In fact, AI has had a significant impact even in areas typically considered “non-machine” domains like emotional intelligence (Green, 2023) and empathy. A recent study published in

the Journal of American Medical Association (JAMA) demonstrated that responses by an AI chatbot to patient questions were nearly ten times more empathetic than those by physicians (Ayers et al., 2023).

Studies have found that it can even positively contribute to the well-being of physicians by reducing burnout and improving their work-life balance (Awan, 2023). Thus, the convergence of human and artificial intelligence has immense potential in healthcare both at the individual and at the institutional level (Topol, 2019a).

The list applications in healthcare is likely to grow continuously in the future and many applications not yet envisaged will emerge. However, for their widespread adoption in the medical field, regulators need put in place safeguards to ensure that the AI applications are safe, reliable and trustworthy.

SECTION TWO

“Getting diversity in the training of these algorithms is going to be incredibly important, otherwise we will be, in some sense, pouring concrete over whatever current distortions exist.”
-Isaac Kohane³

Reliability of Medical AI: Potential and Reality

A study using EHR data from nearly 30,000 patients in University of Michigan found that Epic Sepsis Model, a proprietary sepsis prediction model implemented at hundreds of US hospitals, missed most instances (67%) of sepsis and produced multiple false alarms. (Wong et al., 2021). Considering that sepsis is a life-threatening emergency and a leading cause of hospital death, these findings raise serious concerns about the reliability of such systems.

Unreliable predictions can cause more harm than benefit in guiding clinical decisions. During the COVID-19 pandemic hundreds of clinical prediction models were developed and deployed. However, the Turing Institute found that none of the models had any valuable impact (The Alan Turing Institute, 2020). Similarly, a systematic review of AI-based skin-checking apps available in the public domain found that they were frequently inaccurate (Freeman et al., 2021).

ChatGPT, a large language model (LLM) which has rapidly gained public acclaim has found place in many healthcare applications too. However, LLMs are known to “hallucinate”, or make up false information (Eysenbach, 2023). The problem is that the output of these tools is so coherent that without a high degree of diligence, it is very difficult to separate the truth from the chaff.

Data Scientists also warn of “Clever Hans⁴ predictors” found in ML models. In the previous section we have seen that ML models work by analyzing large amounts of data for correlations and patterns which are then used for making their predictions and recommendations. However, sometimes they base their inferences on spurious or irrelevant correlations learnt during their training phase. A famous non-medical example is that of an AI model to distinguish between photographs of huskies and wolves. The AI made the right judgment most of the time, but when the data scientists looked more deeply, they found that it wasn’t analyzing the physical attributes of each animal — instead, its decisions were based on whether there was snow in the background (Lapuschkin et al., 2019).

Similar spurious correlations have been found in medical AI models too. An AI tool developed at Mount Sinai Hospital performed very well in identifying high-risk patients based on chest x-rays. However, when it was applied at other hospitals there was a severe decline in its performance. It turned out that the predictions were not based on the x-ray features; but only on identifying hardware related information about the specific x-ray machine that was used in the ICU at Mount Sinai (Zech et al., 2018).

Wallis et al demonstrated that a commonly used brain MRI dataset achieved a high tumor classification accuracy, even when no information regarding the tumor itself was available. Here, the ML model used the position and orientation of the 2D MRI slices as a feature to classify tumors (Wallis & Buvat, 2022).

While these examples are not exhaustive, they indicate that such problems are serious practical realities. Building trust in AI solutions is extremely important when dealing with highly critical applications, such as healthcare.

Explainability: The Need for Transparency

Machine Learning technologies, and most AI in healthcare is in this domain, are often called ‘black box’ models. This is because it is difficult, if at all, to understand how they arrived at a particular conclusion. (Beres, 2017). Many AI algorithms in the medical domain– particularly those used for image analysis – are virtually impossible to interpret or explain. Even though the capabilities of AI tools have been demonstrated in certain tasks, when human lives are at stake, it may not be prudent to blindly trust the decision of an AI tool, if one were to be held accountable for it.

For AI models to be effectively utilized in healthcare, they must be able to provide a clear rationale for their decisions. This not only develops trust among medical professionals but also ensures that the patients receive the highest standard of care. It can also aid system developers to fact check their models and prevent ‘Clever Hans’ phenomena (Eversberg, 2023)

Notably, the EU’s General Data Protection Regulation (GDPR) has also envisioned such a requirement and provides the ‘right to explanation’ to the users of AI (Casey et al., 2019).

Bias: Perpetuating Prejudices

There are numerous examples of bias from the use of AI tools in education, hiring, finance, and even the justice system (Baker & Hawn, 2022; Lagioia et al., 2022). Applications of AI in healthcare are similarly susceptible to bias.

Biases come in the way of the equitable provision of and access to healthcare services. A seminal study by Obermeyer et al. found that an algorithm used widely in US hospitals to predict which patients would require additional medical care favored white patients over black patients by a considerable margin. The algorithm used to identify patients in need of “high-risk care management” was far less likely to nominate black patients as they received lower risk scores, even when their needs were greater (Obermeyer et al., 2019).

Studies have shown that AI can reflect the same prejudices that we are striving to overcome in society. And, as the AI systems lack the contextual understanding of humans, they can neither discriminate nor mitigate biases effectively. In addition, being based on algorithms and feedback loops, these embedded biases can get further amplified and perpetuate themselves (A. Cohen, 2023).

AI is not inherently biased. However, biases arise in AI models because they are trained on data collected from healthcare organizations and reflect the disparities in accessing healthcare which can lead to imbalances in the representation of different populations within medical datasets. Societal inequities related to factors like race, gender, and socioeconomic status are ingrained in social structures, reflecting long-standing inequalities. As AI relies only on the available data, excluding realities outside the database, it inadvertently embeds the inequities present in the skewed data into its models.

Data of ethnicity is often not recorded in patient records; and this is with the best of intentions. A review of publicly available dermatological datasets found that subject ethnicity data were available for only 1.3% of the images (Wen et al., 2022). This makes it very difficult to address the bias through any statistical means. On the other hand, a study published in the *New England Journal of Medicine* (NEJM) found that race-corrections in algorithms could also further race-based inequities, causing delay or even denial of healthcare to traditionally marginalized communities. The study was conducted on various tools in use at US hospitals across multiple specialties (Vyas et al., 2020).

Some biases can creep in inadvertently. Images taken by sophisticated CT scan machines, being of better quality, may be preferentially used for training AI models. Although this would be a reasonable thing to do, it would end up ignoring a large segment of representative images from regions that cannot afford to install high-end CT machines (Monga, 2022).

Bias is neither new nor unique to AI and, the National Institute of Standards and Technology (NIST) has noted that it is impossible to achieve zero bias (NIST, 2022). Yet there are factors specific to AI that require new perspectives. Harmful impacts stemming from AI bias can impact society at large, undetected, and at unprecedented scale and speed. As the AI tools become more complex, biases will become even more difficult to identify, let alone control (Racine et al., 2019; Stahl et al., 2021).

Besides the social context, biases could also be clinically dangerous. Optimal recognition and timely management of myocardial infarction (MI), especially for reducing patient delay in seeking acute medical care, is critical. However, as cited in study in BMJ, a self-administered symptom-checker app was found to infer that a 60-year-old man experiencing chest pain was likely having a heart attack, whereas a 60-year-old woman with similar symptoms was likely experiencing a panic attack (Salisbury & Oxford, 2020).

It is critical to appreciate that AI systems are not objective; they reflect existing inequalities, and can also augment them through data cascades which, in turn, cause negative downstream effects (London, 2022). While these aberrations may enter the models without any malicious intent by the developers, their unintended consequences would certainly challenge public trust in AI (Turner Lee, 2018).

Ethical Aspects of AI in Healthcare

Medical ethics describes the moral principles by which a doctor must conduct themselves. The four principles of biomedical ethics described by Beauchamp and Childress - autonomy, non-maleficence, beneficence, and justice - have been highly influential in the field of medical ethics. Confidentiality is sometimes considered the fifth pillar of medical ethics. However, with the increasing penetration of AI systems in healthcare, we need to go beyond the conventional ethical frameworks to address the novel complex challenges posed by them (Hagendorff, 2020).

Beneficence and non-maleficence can be traced back to the times of Hippocrates and are fundamental to defining the responsibility of a practicing physician towards a patient. While beneficence indicates that medical actions will have a reasonable expectation of benefit to the person, non-maleficence implies that they would not harm the patient. Justice implies having systems to assure the fair distribution of benefits and risks across all impacted populations. Autonomy means that each person has the right to make decisions related to their own body and their own personal information.

Earlier in this section we have explored concerns relating to reliability, bias and lack of transparency in AI and how this can well prejudice the safe and equitable delivery of healthcare contravening the basic ethical tenets. AI is supposed to be value-neutral, but it is not. AI-based decisions are susceptible to inaccuracies and bias.

Informed consent, which derives from autonomy, plays a key role in medical practice. Informing the patient about the role of AI in clinical decision-making will be a very important ethical and legal consideration for informed consent. This is, however, quite challenging for the physicians as it requires them to have full knowledge of the decision-making process of the AI system allowing them to reflect on, even query, the system's output.

SECTION: THREE

“AI is only as good as the humans programming it and the system in which it operates. If we are not careful, AI could unintentionally exacerbate many of the worst aspects of our current healthcare system.”
-Bob Kocher⁵

The relationship between humans and technology is not as straightforward as it might seem. The very meaning of technology itself changes through the way it is used. Sometimes the way new technology functions in the real world goes beyond all its intended purposes and the way we make sense of it. Social media is an excellent example of this.

AI is mistakenly considered to be more objective than our cognitive abilities. This is due to ‘cognitive complacency’, whereby users tend to put more trust in the computational power of AI systems without mindfully analyzing the outputs (Jarrahi et al., 2023). Misplaced confidence in AI could be very persuasive and increase the uptake of incorrect outputs.

Developing a clinical prediction model is a science and an art. It is important to find a balance between statistical performance and clinical applications because unreliable predictions could cause more harm than benefit. Despite all efforts, it is basically impossible to develop a model without any errors. Being based on assumptions, probabilities and statistics, a hundred percent accuracy is not a realistic goal in any model. And so it is essential that healthcare institutions, as well as regulatory bodies, establish structures to monitor key issues to limit negative effects.

AI tools, being adaptive, change through interactions with their environment and are also liable to decay in performance over time. Gravely though, these changes may be difficult to detect due to the intrinsic opacity of the tools. Consequently, from a medical point of view, not only regulatory oversight, but also constant human oversight is essential. It is important to ensure that clinicians have control over the decision-making process and that mechanisms are in place to verify the accuracy of AI decisions.-

The adoption of AI tools in healthcare has lagged mainly because of ethical and safety considerations as discussed in the previous section. It would be natural for clinicians to be wary of AI systems because it is difficult to detect or fix aberrant behavior, or even trust them when it comes to safety and ethics. The absence of standard guidelines for the use of AI in healthcare has only served to worsen the situation. Robust clinical evaluation, using metrics that are intuitive to clinicians, and appropriately regulated systems are needed for trust in the technology. However, scientific, peer-reviewed evidence of efficacy is lacking for most commercially available AI products (Daneshjou et al., 2021; Voter et al., 2021).

Pace and Race

Globally there is an ongoing race for AI dominance between many countries. When such race dynamics take over there is a real risk that ethics would be sidelined (I. G. Cohen, 2023). Furthermore, mutual mistrust makes any open mature discussion on regulation of the technology very difficult.

A separate race is parallely playing out in the industry. In the last decade there has been an exponential increase in the annual global patent filings for AI technologies (*Figure 1*). The pace of AI growth has made the race to market one for survival as the innovators need to fully exploit their technological head start. As they sprint from the lab to the market, companies often cut corners, compromising testing, validation and safety (Skelton, 2023).

Transformative technologies like AI require comprehensive adoption strategies, including national-level strategies to contribute to the larger public good (NITI, 2018). The prospects of economic progress, and the quest for AI dominance, would thus influence strategies for regulation too.

Building an Ecosystem of Trust

A 2021 study by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems found that only 26% of the general public trust AI, while 73% of respondents believe that AI should be regulated. Two years later, a survey of 17000 people by KPMG in 17 countries in 2023 showed similar findings (KPMG Australia, 2023). Evidently, despite widespread availability and awareness of AI tools, trust in the technology has not improved.

Such ethical dilemmas, and the implications on accountability and liability, unless resolved through appropriate regulations, could hinder the larger adoption of AI in medical practice.

It is difficult to regulate AI due to its complexity and pace of development. Robust safeguards in terms of validation, effective governance, regulatory and policy frameworks are critical for safe and ethical adoption of AI. Managing these concerns without stifling innovation, and the potential of AI, is, however, a challenge for regulators. While most countries have issued guidelines, the US

and UK have both adopted a ‘pro-innovation approach’ by enabling current regulators to determine how best to adapt existing policies to the field (Skelton, 2023).

Whichever approach is adopted, a new regulatory mindset will be required to keep up with the pace of change.

Need for Standards and Regulation

One area where AI poses a significant challenge is the approaches traditionally used to regulate software and other technologies. Although AI tools consist of software, the existing standards for software are not compatible with typical AI methods because of their autonomous and complex nature and lack of transparency (Zielke, 2020).

A basic approach to standardization would involve the definition of common terms. When looking at terms relating to AI, the term artificial intelligence itself is a subject of discussion. European Commission’s Rolling plan for ICT standardisation notes that there is no generally accepted definition of Artificial Intelligence (EU, 2023).

It is important to appreciate that AI is not just one technology but a group of capabilities being applied to diverse domains. For instance early AI was knowledge-based, while today’s AI is data-driven, and uses a multitude of methods and models. Further, in most cases AI tools are not stand-alone devices, but involve the integration of AI components into other products and processes to augment their capabilities.

Standardization brings consensus which can support scientific and commercial exchange, It can also establish confidence amongst users and facilitate regulation.

Most guidelines for the regulation of AI, and the same are used for AI applications in healthcare, are general in nature. These cannot meet the sensitivities, specificities, and risks of AI in the healthcare domain. US FDA considers AI applications in healthcare under the category of “Software as a Medical Device” (SaMD) (US FDA, 2019). However, regulations for medical devices, being based on legacy models and traditional approaches, lack mechanisms for addressing the specific issues related to AI technologies.

AI technologies bring new and accelerated risks. They are evolving at a pace that exceeds the speed at which regulatory agencies can keep up because the existing models of regulation of healthcare are designed for ‘locked’ solutions, whereas AI is ‘adaptive’ i.e. it is flexible and evolves and changes over time (Dettling et al., 2021). Accordingly, it is essential for the regulatory environment to adapt to this fast-evolving field to anticipate and prevent potential risks and unintended harm.

CONCLUSION

“In healthcare, artificial intelligence needs to pass the implementation game, not the imitation game.”
- John Powell⁶

Alan Turing, widely considered the father of computer science and artificial intelligence, developed a test in 1950 called the “Imitation Game” to test a machine’s ability to exhibit intelligent behavior. The test requires that a human examiner should be unable to distinguish the machine from another human being from replies to questions in natural language. We have come a long way from the times when the test was very challenging, and many AI applications available now can successfully pass the test. The implementation of AI, in the absence of clear standards and robust regulation, is however, a challenge.

As discussed in previous sections, AI technology holds immense potential but the same is perhaps yet to be realized. Individual studies showing fantastic results sometimes fail when subjected to rigorous scientific scrutiny and meta-analysis, underlining the need for robust testing, validation, and regulation.

AI works in the digital world, but it impacts us in the real world. So, in the glare of possibilities we should not lose sight of its limitations. Perhaps the greatest source of risk is the myth of objectivity of AI, an illusion of neutrality of algorithms, and complacent trust in the omniscience of computational knowledge. AI, in its present state, is not worthy of any of these epithets.

Commercial interests and the race for AI supremacy can, and will affect the design, use, and long-term impact of the technology, which, unless regulated through appropriate oversight mechanisms can have serious consequences. The ethical and legal implications of AI in healthcare are not only quite significant but also challenging. Managing these concerns, and building regulatory safeguards, without stifling the potential of AI is a key challenge for regulators.

AI is not infallible. Over time, marked improvements in accuracy, productivity, and workflow will likely be actualized, and it will ultimately become a mature and effective tool for the healthcare sector (Topol, 2019b). For now, AI technologies need meaningful human control and regulations.

REFERENCES

- Awan, O. (2023). Artificial Intelligence And Its Potential To Combat Physician Burnout. *Forbes*. <https://www.forbes.com/sites/omerawan/2023/04/12/artificial-intelligence-and-its-potential-to-combat-physician-burnout/?sh=39d0a4937734>
- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine*, 1–8. doi:10.1001/jamainternmed.2023.1838
- Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. doi:10.1007/s40593-021-00285-9
- Beres, D. (2017). Even AI Creators Don't Understand How Complex AI Works. *Big Think*. <https://bigthink.com/the-future/black-box-ai/>
- Burns, E. (2023). What is artificial intelligence (AI)? - AI definition and how it works. Tech Target. <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence>
- Casey, B., Farhangi, A., & Vogl, R. (2019). Rethinking explainable machines: The GDPR's "right to explanation" debate and the rise of algorithmic audits in enterprise. *Berkeley Technology Law Journal*, 34, 143. doi:10.15779/Z38M32N986
- Choudhury, A., & Asan, O. (2020). Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review. *JMIR Medical Informatics*, 8(7), e18599. doi:10.2196/18599 PMID:32706688
- Cohen, A. (2023). Unveiling the Inevitable Bias: Why Human Data Remains Biased in AI Applications? *Medium*. <https://iamitcohen.medium.com/unveiling-the-inevitable-bias-why-human-data-remains-biased-in-ai-applications-f7ca5c4fad9e>
- Cohen, I. G. (2023). What Should ChatGPT Mean for Bioethics? SSRN *Electronic Journal*, 1–9. 10.2139/ssrn.4430100
- Daneshjou, R., Smith, M. P., Sun, M. D., Rotemberg, V., & Zou, J. (2021). Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. *JAMA Dermatology*, 157(11), 1362–1369. doi:10.1001/jamadermatol.2021.3129 PMID:34550305
- EU. (2023). Artificial Intelligence. *Joinup*. <https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/artificial-intelligence>
- Eversberg, L. (2023). As a Data Scientist, You Should Know About a Clever Horse Called Hans. *Medium*. <https://pub.towardsai.net/as-a-data-scientist-you-should-know-about-a-clever-horse-called-hans-51eab1135a61>
- Eysenbach, G. (2023). The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Medical Education*, 9, e46885. doi:10.2196/46885 PMID:36863937
- Freeman, K., Geppert, J., Stinton, C., Todkill, D., Johnson, S., Clarke, A., & Taylor-Phillips, S. (2021). Use of artificial intelligence for image analysis in breast cancer screening programmes: Systematic review of test accuracy. *BMJ (Clinical Research Ed.)*, 374, n1872. doi:10.1136/bmj.n1872 PMID:34470740
- Gov.UK. (2023). A pro-innovation approach to AI regulation - GOV.UK. In *Whitepaper*. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#part-1-introduction>
- Green, F. M. (2023). How AI Can Help You Develop Emotional Intelligence. *Forbes*. <https://www.forbes.com/sites/forbescoachescouncil/2023/03/24/how-ai-can-help-you-develop-emotional-intelligence/?sh=73da0e5d6340>
- Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46(3), 205–211. doi:10.1136/medethics-2019-105586 PMID:31748206
- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120. doi:10.1007/s11023-020-09517-8

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews. Cancer*, 18(8), 500–510. doi:10.1038/s41568-018-0016-5 PMID:29777175

ITU. (2022). *ITU-T FG-AI4H Deliverable*. Issue September.

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*, 66(1), 87–99. doi:10.1016/j.bushor.2022.03.002

K. P. M. G. (2023). *Trust in Artificial Intelligence*. KPMG Australia. <https://kpmg.com/au/en/home/insights/2023/02/trust-in-ai-global-insights-2023.html>

Kantaro, K. (2023). *G7 should adopt “risk-based” AI regulation, ministers say*. Reuters. <https://www.reuters.com/markets/europe/g7-should-adopt-risk-based-ai-regulation-ministers-say-2023-04-30/>

Lagioia, F., Rovatti, R., & Sartor, G. (2022). Algorithmic fairness through group parities? The case of COMPAS-SAPMOC. *AI & Society*, 38(2), 459–478. doi:10.1007/s00146-022-01441-y

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K. R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, 10(1), 1–8. doi:10.1038/s41467-019-08987-4 PMID:30858366

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy 2021*, 23(1), 18. 10.3390/e23010018

London, A. J. (2022). Artificial intelligence in medicine: Overcoming or recapitulating structural challenges to improving patient care? In *Cell Reports Medicine*, 3(5). doi:10.1016/j.xcrm.2022.100622

Markets and Markets. (2023). *Artificial Intelligence (AI) in Healthcare Market Industry Report, Size, Growth Report, 2031*. Report SE5225. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-healthcare-market-54679303.html?gclid=EAIaIqobChMIn9TThMmIgAMVknKLCh2tjQ0KEAMYASAAEgLzb_D_BwE

McKinsey & Co. (2018). *Adoption of AI advances, but foundational barriers remain*. McKinsey. <https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain>

Monga, K. (2022). The Machine Learning Conundrum in the Care Delivery Ecosystem. *Journal of American Health Information Management Association*. <https://journal.ahima.org/page/the-machine-learning-conundrum-in-the-care-delivery-ecosystem>

NIST. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. In *Natl* (Vol. 1270). Inst. Stand. Technol. Spec. Publ. doi:10.6028/NIST.SP.1270

NITI. (2018). National Strategy for Artificial Intelligence. *Enhanced Reader*.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. doi:10.1126/science.aax2342 PMID:31649194

Pfungst, O. (1911). Clever Hans (the Horse of Mr. Von Osten): Contribution to experimental animal and human psychology. *The Journal of Philosophy, Psychology, and Scientific Methods*, 8(24), 663–666. www.archive.org/details/cu31924024783973. doi:10.2307/2012691

Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics*, 2023(1), 1–18. doi:10.1007/s43681-023-00258-9

Racine, E., Boehlen, W., & Sample, M. (2019). Healthcare uses of artificial intelligence: Challenges and opportunities for growth. *Healthcare Management Forum*, 32(5), 272–275. doi:10.1177/0840470419843831 PMID:31234654

Salisbury, H., & Oxford, G. P. (2020). Helen Salisbury: Prestidigitation. *BMJ (Clinical Research Ed.)*, 368, m648. doi:10.1136/bmj.m648 PMID:32098837

Skelton, S. K. (2023). MPs warned of AI arms race to the bottom. *Computer Weekly*. <https://www.computerweekly.com/news/365529793/MPs-warned-of-AI-arms-race-to-the-bottom>

- Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhé Shaelou, S., Patel, A., Ryan, M., & Wright, D. (2021). Artificial intelligence for human flourishing – Beyond principles for machine learning. *Journal of Business Research*, *124*, 374–388. doi:10.1016/j.jbusres.2020.11.030
- The Alan Turing Institute. (2020). *Data science and AI in the age of COVID-19 – report*. The Alan Turing Institute. <https://www.turing.ac.uk/news/publications/data-science-and-ai-age-covid-19-report>
- Topol, E. J. (2019a). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. doi:10.1038/s41591-018-0300-7 PMID:30617339
- Topol, E. J. (2019b). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–47. <https://go.gale.com/ps/i.do?p=HRC&sw=w&issn=10788956&v=2.1&it=r&id=GALE%7CA573404401&sid=googleScholar&linkaccess=fulltext>. doi:10.1038/s41591-018-0300-7 PMID:30617339
- Turner Lee, N. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information. Communication and Ethics in Society*, *16*(3), 252–260. doi:10.1108/JICES-06-2018-0056
- US FDA. (2019). Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning (AI / ML) -Based Software as a Medical Device (SaMD). *U.S Food & Drug Administration*, 1–20.
- Voter, A. F., Larson, M. E., Garrett, J. W., & Yu, J. P. J. (2021). Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Cervical Spine Fractures. *AJNR. American Journal of Neuroradiology*, *42*(8), 1550–1556. doi:10.3174/ajnr.A7179 PMID:34117018
- Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *The New England Journal of Medicine*, *383*(9), 874–882. doi:10.1056/NEJMms2004740 PMID:32853499
- Wallis, D., & Buvat, I. (2022). Clever Hans effect found in a widely used brain tumour MRI dataset. *Medical Image Analysis*, *77*, 102368. doi:10.1016/j.media.2022.102368 PMID:35063892
- Wen, D., Khan, S. M., Xu, A. J., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., de Blas Perez, C., Denniston, A. K., Liu, X., & Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: A systematic review. *The Lancet. Digital Health*, *4*(1), e64–e74. doi:10.1016/S2589-7500(21)00252-1 PMID:34772649
- WHO. (2023). *WHO calls for safe and ethical AI for health*. World Health Organization. <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health>
- Wong, A., Otlés, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., Phillips, M., Konye, J., Penozo, C., Ghous, M., & Singh, K. (2021). External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Internal Medicine*, *181*(8), 1065–1070. doi:10.1001/jamainternmed.2021.2626 PMID:34152373
- Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Medicine*, *15*(11), e1002683. doi:10.1371/journal.pmed.1002683 PMID:30399157
- Zielke, T. (2020). Is Artificial Intelligence Ready for Standardization? *Communications in Computer and Information Science*, *1251 CCIS*(May), 259–274. doi:10.1007/978-3-030-56441-4_19

ENDNOTES

- ¹ Decision theorist from the USA who is widely cited for his writings on the long-term future of artificial intelligence. He leads research at the Machine Intelligence Research Institute (MIRI), Berkeley, California
- ² A series of well-defined steps for solving a given type of problem
- ³ Chair of the Department of Biomedical Informatics, Harvard Medical School
- ⁴ ‘Clever Hans’ was a famous horse at the beginning of the 20th century in Germany which was believed by his trainer to be capable of performing arithmetic calculations. Subsequent analyses revealed that the

horse was not performing any calculations but gave answers by detecting subtle cues on the face of his trainer (Pfungst, 1911).

⁵ Adjunct Professor at Stanford University, School of Medicine. He also serves on the Advisory Board of the National Institute of Healthcare Management (NIHCM).

⁶ Nuffield Department of Primary Care Health Sciences, Medical Sciences Division, University of Oxford, UK

Rajiv Kumar Pathni, an Air Force Veteran, is a Physician with over three decades of healthcare leadership experience. He is a member of the Executive Committee of Research Foundation of Hospital & Healthcare Administration (RFHHA) and is an elected member of the Board of Executives of the Academy of Hospital Administration (AHA). He has served as an expert on several government and industry panels and continues to provide consultancy and pro bono support as a domain expert in healthcare policy, strategy, quality and technology.