

# DIGITAL TECHNOLOGY The automation of doctors and machines: A classification for AI in medicine (ADAM framework)

Author: Fawz Kazzazi<sup>A</sup>

## ABSTRACT

The advances in artificial intelligence (AI) provide an opportunity to expand the frontier of medicine to improve diagnosis, efficiency and management. By extension of being able to perform any task that a human could, a machine that meets the requirements of artificial general intelligence ('strong' AI; AGI) possesses the basic necessities to perform as, or at least qualify to become, a doctor. In this emerging field, this article explores the distinctions between doctors and AGI, and the prerequisites for AGI performing as clinicians. In doing so, it necessitates the requirement for a classification of medical AI and prepares for the development of AGI. With its imminent arrival, it is beneficial to create a framework from which leading institutions can define specific criteria for AGI.

**KEYWORDS:** medical technology, artificial intelligence, AI, clinical governance

**DOI:**10.7861/fhj.2020-0189

## Introduction

The advances in artificial intelligence (AI) provide a great opportunity to expand the frontier of medicine to improve diagnosis, efficiency and management. The term 'artificial intelligence' was coined by John McCarthy in 1956, and describes the ability of machines to 'have beliefs' that make them capable of problem-solving performance.<sup>1</sup> Fundamentally, AI is the ability of non-organic systems to perform tasks that require human-level intelligence, but as these tasks evolve then so does AI's potential. In Tesler's theorem, this is known as the 'AI Effect': an understanding that AI is 'whatever hasn't been done yet'.<sup>2</sup>

When contextualising this to medicine, it is useful to classify technologies into 'augmented', 'narrow' and 'general'. Augmented AI is an adjunct to human intelligence and does not replace it; it is a tool that assists humans towards an end. Narrow AI (also referred to as 'weak' AI) is focused towards a specific task within a limited range of pre-defined functions; it does not exhibit the flexibility to work outside of these parameters. Artificial general intelligence

(AGI; also known as 'strong' or 'true' AI) refers to a machine with the ability to perform any intellectual task that a human could. When John Searle coined this term in 1980, it included attributes of consciousness, sentience and mind; however, the moral legitimacy of machines mimicking characteristics ordinarily attributed to 'personhood' remains an active debate.<sup>3,4</sup>

Yet, by extension of being able to perform any task that a human could, a machine that meets the requirements of AGI possesses the basic necessities to perform as, or at least to qualify to become, a doctor. In this evolving fast-paced field, I explore the distinctions between doctors and AGI, and the emerging prerequisites for AGI performing as clinicians. In doing so, this work necessitates the requirement for a framework of criteria to classify the involvement of AI in medicine and prepare for the advancement of AGI.

## Achieving AGI in medicine

The original test for an operational definition of AGI was the 'imitation game' proposed by Alan Turing in 1950 (colloquially known as the 'Turing Test').<sup>5</sup> This test assesses the ability of a machine to exhibit sufficiently intelligent behaviour that it is indistinguishable from a human in a written conversation to a third party. Since this first proposition, alternate proposals have been put forward. Steve Wozniak, co-founder of Apple, has suggested the 'coffee test' for AGI.<sup>6</sup> This is a functional test with the requirement that a machine enters a home, finds coffee equipment and uses it to make a cup of coffee. Goertzel proposes a more academic challenge and suggests that to achieve AGI, machines need to attend a university and pass exams.<sup>7</sup>

These tests exhibit some key characteristics of doctors: the ability to learn, make decisions and communicate with people. Technology is already complementing clinical practitioners through machine learning in augmented and narrow intelligence applications that enhance diagnosis, understanding and treatment.<sup>8,9</sup> One prominent example is AI-enhanced retinal examination. This was originally designed to encourage general and emergency clinicians to not forgo diagnostic retinal examination nor rely on secondary referrals to ophthalmologists. A deep-learning system used fundus photographs to identify optic disks with papilloedema, normal disks, and disks with non-papilloedema abnormalities to an acceptably high sensitivity.<sup>10</sup> These innovations expand the capability of the

**Author:** <sup>A</sup>core surgical trainee and law student, Mason Institute for Medicine, Life Sciences and Law, Edinburgh, UK

standalone physician. Governments are recognising and rewarding this potential; for example, the UK government has committed over £200 million in funding to AI in the field of prevention, early diagnosis and treatment of chronic diseases.<sup>11</sup>

This transformative effect of AI on medicine is reflected in the 20,000 publications on this topic in the last 10 years.<sup>12</sup> Interestingly, over half of these were produced by teams in the USA (based on the institutional affiliation of the lead author); yet, the UK did not feature in the top 20 countries for medical AI outputs in that literature review.<sup>12</sup> When stratifying these publications, the most common research themes were 'cancer' (~23%), 'heart disease' (~7%), 'vision' (~6.5%), 'stroke' (<5%), 'Alzheimer's' (<5%), depression (<5%), kidney disease (<5%) and diabetes (<5%).<sup>12</sup> The authors noted that there is a scarcity of publications regarding the ethics of AI (0.7%) despite its potential widespread use and implications. This supports the notion that a robust framework that can act as a normative regulation and decision-making tool may be required.

The potential applications of AI are immense and include tools that are broadly classified as for prognosis, diagnosis, treatment, clinical workflow and expertise.<sup>8</sup> More specifically, tools have been designed to improve online consultations, optimise medications in chronic diseases, match tumour genomics to clinical therapy trials and design new drugs.<sup>9</sup> When looking to the data driving these changes, the majority has been from patient records, imaging, genetics and electrodiagnosis (eg electromyography).<sup>13</sup> Notable outputs have included its applications in diagnosing skin lesions, predicting suicide risk and detecting sepsis.<sup>14–16</sup> Extraordinarily, tools are also being designed to help clinicians with clinical processes; one narrow AI tool to help physicians provides real-time instructions/guidance for performing cardiac ultrasounds to optimise views/images.<sup>9</sup> As innovation continues to grow, the conversation is shifting to the safety of these interventions. While the development of automated clinical algorithms becomes more easily achievable, the prospect of a reproducibly safe and accurate automated tool or clinician is inevitable.

## Defining doctorhood

A doctor is more than just a diagnostic tool. In 2017, a Chinese robot (named Xiaoyi) passed a written medical licensing exam comfortably; had it also passed an objective structure clinical examination (OSCE) then this machine would have met the minimum expectation of graduating medical students worldwide.<sup>17</sup> In a Goertzel-inspired manner, this would provide the simplest test of qualification: having the knowledge and minimum social interaction required to pass written medical examinations and OSCEs.

Despite this, a recent survey of general practitioners (GPs) in the UK found that they believe that communication and empathy are exclusively human competencies, and that the scope of AI in primary care was limited to efficiency savings.<sup>18</sup> Technology giant, Intel, completed a study in the USA that showed doctors believe that AI will not replace them, rather it will increase their availability for patient care.<sup>19</sup> However, they raised concerns about the ambiguity in the algorithm (known as the 'black box' dilemma) and the potential for fatal errors. The 'black box' issue that clinicians will be unable to understand or rationalise why an AI reaches a decision is one that has been discussed significantly in the literature as a fundamental downfall of AI. Despite its potential, AI also brings new problems; the greatest barriers to AGI include algorithm bias, reward hacking, insensitivity and automation bias (Table 1).<sup>20–23</sup>

An additional difficulty, first outlined in 1966 by Austin Bradford Hill regarding his novel protocol for randomised controlled trials (RCT) was that determining treatments based on averages does not 'answer the practising doctor's question: what is the most likely outcome when this particular drug is given to this particular patient?'.<sup>24</sup> This incongruence between single-outcome algorithms and heterogeneous treatment effects has been explored in great detail when concerning applicability of RCT results.<sup>25</sup> However, when considering complex decision-making, competing diagnoses and balancing clinical risk, then the literature is surprisingly lacking; perhaps, this is due to the assumption that algorithms will eventually learn best outcomes and superiority as they have

**Table 1. List of common barriers to the successful implementation of artificial general intelligence**

Barrier to artificial general intelligence	Definition	Impact
Algorithmic bias <sup>20,21</sup>	Outcomes reflective of the quality of the input and the knowledge of the designer.	This has the potential to reduce applicability of output in diverse populations. The learning algorithms may also adopt poor practice from historical data that humans would inadvertently avoid.
Reward hacking <sup>22</sup>	Focus in algorithms on outcomes that are seen as 'successful' with poor consideration of the means, process or long-term goals.	The automated system finds methods to achieve short-term goals eg dosing heparin prior to coagulation level measurements.
Insensitivity to impact <sup>22</sup>	Inconsideration of the consequence of the outcome in the decision-making process.	Human doctors tend to err on the side of caution with diagnoses surrounding malignancy, leading to a high false positive but also a high sensitivity (few false negatives). If machine learning systems focus on accuracy, the implications for higher positive predictive value and specificity (higher true positive, lower false positive) may not be optimised for care of the holistic patient.
Automation bias <sup>23</sup>	Reliance on the output of automated systems.	This occurs where clinicians rely on the results of their tools despite evidence to the contrary. Furthermore, eagerness and funding pressure may lead to the premature or inappropriate application of under-developed systems.

demonstrated with risk-stratification and management of patients with upper gastrointestinal bleeds (outpatient vs inpatient intervention).<sup>26</sup> Decisions to operate are archetypical examples of medical deliberations that require hypothetical-deductive reasoning, individual judgment and heuristics. Innovators have recognised that while AI may not provide a definitive answer, it can be utilised to optimise conditions that favour a specific outcome (eg to operate) through the reduction of risk factors.<sup>27</sup> In a study on complex breast cancer patients, the AI decision-making tool was compared with oncologists with the long-term aim of incorporating it in a fashion analogous to a multidisciplinary team (MDT) meeting to aid lone physicians.<sup>28</sup> An alternative option to selecting between therapies in complex patients would be utilising historical patient decision making, particularly in the context of patients who lack capacity.<sup>29,30</sup> While patient-led decisions have been explored previously, particularly in the context of medical versus surgical management of prostate cancer, it risks perpetuating established inequalities in health outcomes due to ethnicity and socio-economics prejudiced decision-making.<sup>31</sup>

These drawbacks need to be acknowledged in the design and implementation of a safe medical AI framework through an iterative learning process that recognises their influence on decision making and aims to mitigate their effect. In this sense, it is similar to unconscious bias training in that working towards widespread recognition and training on the issue helps reduce its influence. However, AGI has a powerful value proposition to medical services though a non-fatigable decision and action-making tool that can be deployed anywhere with a maintainable computer infrastructure.

### Humanity and humility

Human interaction is a fundamental philosophy of current medical practice. The character and conduct of the clinician influence how patients interact with their diagnosis and can even improve outcomes. A *Lancet* review reported that a consistent finding for improving outcomes was a warm, friendly and reassuring clinician.<sup>32</sup> This ‘doctor effect’ has been reinforced in the primary care setting whereby, typically, human skills of empathy and reassurance are shown to improve patient outlook and promote change in behaviour.<sup>33</sup> However, if patients cannot discern between machines and humans, can this effect still be achieved?

Narrow AI can act as an adjunct to these interactions through providing information, recommendations or results to a clinician that maintains the human relationship with the patient. However, AGI would automate the whole process, including the patient-facing interaction. Early work has shown promise in the use of chatbots as a conversational narrative, but wider (patient-led) validation and factor identification is still required to fully understand what aspect of human interaction contributes to the emotional care of patients and how this affects outcomes.<sup>34,35</sup>

Furthermore, doctors are bound by a series of duties that are proclaimed during their initiation into the role of becoming a physician, a process herein described as ‘doctorhood’. To doctors, a medical career is a transformational experience that involves empathetic engagement, intimacy and detachment, centred around the formation of human relationships. With over a third of patient symptoms existing without medical diagnosis, doctors play a role in their patients’ lives in the absence of curative options through listening, managing dignity and comfort, acknowledgement and appreciation. This facet of the patient–doctor relationship has been shown to have a small but significant effect on improving

patient outcomes.<sup>36</sup> It has been termed as ‘emotional care’ and ‘cognitive care’, and refers to the process of exchanging emotions and information between patients and doctors.<sup>32</sup> However, for patients with medically unexplained symptoms, AI provides a potentially invaluable source through learning algorithms that mine patient information to see correlations that clinicians may not be able to, so providing subsets of this population with a possible treatable condition or valuable diagnosis.<sup>37</sup> The cancer detection tool ‘C the Signs’ has aided thousands of GPs in England to consider and diagnose cancer earlier in patients that present with seemingly disparate symptoms.<sup>38,39</sup>

Although world-leading experts believe AGI may be achieved in the next 10 years, there still exists a large deficit in the literature pertaining to patient attitudes on the use of AI in medical encounters. It is not known in what capacity a patient would be willing to interact with AGI, or even the cultural or demographic permeability of such an approach. Often, patients may attend appointments with a medical problem that is not their primary concern and are reliant on the doctor to elucidate the true reason for their presence. The Academy of Medical Royal Colleges have expressed their concerns regarding the ‘gaming’ of coding that may prioritise diagnosis over patient care.<sup>40</sup> However, the technology for detection of mental health disorders, early brain degenerative disorders and subtle micro-expressions is in development and is likely to improve in the near future; this may mitigate these issues and expand the holistic diagnostic ability of these tools but this is yet to be proven.<sup>41–43</sup>

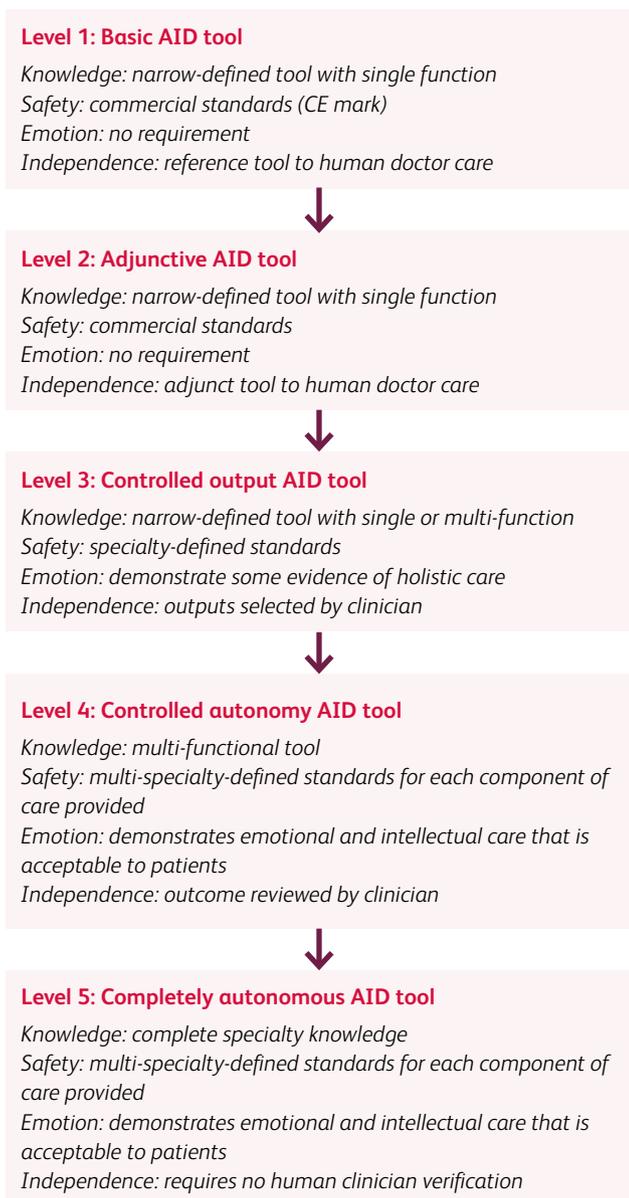
### A test of doctorhood

With the potential imminent arrival of AGI and its foreseeable use in medicine, it is beneficial to classify different forms of AI within medicine and create a framework from which leading institutions can define specific criteria for AGI. While other groups have looked at ways to classify specific applications of AI in medicine (such as for radiology, data learning algorithms and clinical decision-making), there is no unifying or central framework from which regulatory or specialist groups can develop their policies.<sup>13,44,45</sup> When considering long-term goals of AI within medicine, this could lead to significant heterogeneity in standards across specialties and locations. In deriving a normative framework, one could contemplate the topics discussed thus far and, specifically, the characteristics that tools require in order to achieve AGI in the healthcare field.

In the progression from augmented intelligence to AGI (worthy of *doctorhood*), this article has explored how a system needs to achieve minimum levels of competency in knowledge, safety and emotional care. Additional supplementary skills can be determined by specialty bodies; these may include practical ability for surgery, administrative abilities for interns/junior doctors or higher empathetic/conversational thresholds for palliative care and psychiatry. However, this proposal does not discriminate between ‘good’ and ‘bad’ forms of doctorhood, which will only be reflective of the quality of the input of the system.

This Automation of Doctors and Machines (ADAM) framework is set out in Fig 1. As the level progresses, so do the minimum requirements in the core competencies.

- > Knowledge: sufficient levels of information and decision making to carry out the task.
- > Safety: a permissible level of accuracy and independence in function.



**Fig 1. Automation of Doctors and Machines framework for five levels of artificial intelligence doctor tools.** AID = artificial intelligence doctor.

- > Emotion: provision of holistic and emotional care for the patient.
- > Independence: indistinguishable from a human doctor, or patients are aware that it the tool is not human and this is permissible.

Thus, the simplest form of AI in medicine are the machine learning tools that can be applied to sets of data to aid but not determine care (level 1 artificial intelligence doctor (AID) tool). These will typically demonstrate sufficient levels of knowledge, with no requirement for emotion. They will have commercial levels of safety but necessitate a doctor to ensure they are medically safe for application. Examples include reference tools, such as risk stratification scores that produce severity scores (eg for common presentations like pancreatitis or appendicitis). These are widely used already, and could be further 'automated' through the automatic generation of scores for patients at presentation.

When the tool develops an independent function outside of the human doctor's care, it becomes a level 2 AID tool. This is where the majority of currently AI applications in medicine are aiming. They demonstrate a clear and narrow function that is both accurate and safe, usually a diagnostic tool (with sufficient knowledge and safety), but require human instrumentation in order to prove effective. A triaging service is an illustration of this stage, as it provides a complete and narrow medical expertise service independent of the clinician, but that ultimately will be reviewed by a clinician. This application is already being explored in triage heavy services such as emergency medicine and radiology. Models have been successfully designed to triage patients with common presentations (eg abdominal pain), likelihood for hospitalisation, for resource (personnel) allocation in remote environments and severity of (radiology) diagnosis.<sup>46-49</sup> The cancer diagnostic tool (C the Signs) discussed earlier would fall into this category.

To develop further, a tool must demonstrate some element of emotional and cognitive intelligence to provide care specific for the patient. These tools are still 'limited' as they have a precise output but are independent in this end. For example, a tool that can diagnose and prescribe accurately while demonstrating consideration for the patient's life through personalised selection (eg dosing regimen) would fulfil this criterion for level 3. In contrast to polypharmacy rationalisation tools, items in this category would place greater focus on patient preference. For example, they may aid a patient diagnosed with depression pick the best form of therapy or medication after consideration of their priorities, character traits, employment/availability, aims and likelihood of engagement.

The distinction between these tools and level 4 is the 'personable' factor that studies attribute to improved health results. While subject to generational influence, this criterion would involve a complete doctor-patient interaction conducted by AI. Due to medical ethics of transparency and honesty, a Turing-like test is not permissible as patients should be informed of the specifics of their service provider. However, at level 4, the use of AI is totally acceptable to the patient. With regard to safety, a similar system is adopted to specialty trainees, whereby a senior doctor verifies and monitors the output and practice of this stage. This had the added benefit of enabling parallel streaming of patients in a conventionally safe manner.

Finally, level 5 is achieved when the AGI functions as a totally independent and autonomous practitioner with sufficient knowledge and ability to complete the patient-doctor interaction without validation. The pre-validation of these tools will happen in their development. The potential for level 5 AID tools is enormous, and it is difficult to predict whether these will be single-specialty based or demonstrate cross-specialty autonomy. This level also presents its own complexities with uncertainty regarding the ultimate responsibility and liability for the care provided.<sup>50-54</sup>

The benefit of a graded system, such as the ADAM framework, is the enhanced ability of unified monitoring, evaluation and learning. Its stepwise progression lends itself to a layered approach that would allow external regulatory and specialty bodies to define complexity, safety and intervention in a graduated manner.

### Implementation the ADAM framework in practice

The ADAM framework is analogous to the classification of driving automation proposed by the Society of Automotive Engineers (SAE; Fig 2).<sup>55</sup> It provides a simple tiered framework with clear designations of the characteristics and distinction of human versus automated involvement at each level. This framework is unique in that it directly

	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?	You <b>are</b> driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You <b>are not</b> driving when these automated driving features are engaged – even if you are seated in “the driver’s seat”		
	You <b>must constantly supervise</b> these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
What do these features do?	These are driver support features			These are automated driving features		
	These features are limited to providing warnings and momentary assistance	These features provide steering <b>OR</b> brake/acceleration support to the driver	These features provide steering <b>AND</b> brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met	This feature can drive the vehicle under all conditions	
	<ul style="list-style-type: none"> <li>• automatic emergency braking</li> <li>• blind spot warning</li> <li>• lane departure warning</li> </ul>	<ul style="list-style-type: none"> <li>• lane centering <b>OR</b></li> <li>• adaptive cruise control</li> </ul>	<ul style="list-style-type: none"> <li>• lane centering <b>AND</b></li> <li>• adaptive cruise control at the same time</li> </ul>	<ul style="list-style-type: none"> <li>• traffic jam chauffeur</li> </ul>	<ul style="list-style-type: none"> <li>• local driverless taxi</li> <li>• pedals/steering wheel may or may not be installed</li> </ul>	<ul style="list-style-type: none"> <li>• same as level 4, but feature can drive everywhere in all conditions</li> </ul>
Example Features						

Fig 2. SAE J3016™ levels of driving automation. Reproduced with permission from SAE International.

calls for emotive aspect of patient care and doctorhood to be a part of the expansion of AI, and provides a way to quickly and easily compare multiple new technologies. It would act as a centralised framework from which specialist bodies (eg royal colleges) and regulators could classify new and emerging technologies. In some developed form, it may act as a way to assess automated tools in order to make them suitable for medical environments.

In implementing AI in medicine, it will be important to determine the level of standards and regulation for different tools. With so many groups working on innovations, it will soon become challenging to set individual standards for each form of technology. Specialist bodies and regulators will be able to define their own requirements for the four domains at each level of the ADAM framework. In turn, this will provide a simple way to implement broad rules that hospitals and innovators can work towards in order to gain some form of accreditation from the standard-setting groups.

In practical terms, this would enable quick comparison of the function of multiple technologies and aid healthcare organisations to differentiate between their capabilities. It is highly likely that further work on this framework would be needed, particularly to subcategorise technologies within each level. This is because, at current progression, each level will be achieved in a time-dependent manner, with level 5 occurring furthest in the future. As such, time urgency may necessitate further requirements within each domain at the most imminent level (at this point, level 2). However, the hope is that the introduction of uniform considerations for new technologies will help to align the thoughts of the groups that are attempting to rationalise this ever expanding and complicated field.

In making the framework simpler in its first iteration, it aims to also aid the clarity and understanding for patients who will be able to concisely see what form of technology they are engaging with.

### Conclusion

The development of AI brings with it an exciting era of modern medicine. In order to fully enhance, expand and regulate this field, the ADAM framework provides a tool to classify its use in medicine. In being able to categorise forms of medical AI, this allows clinicians, patients and regulators to delineate different forms of AI, and a foundation is created from which governing bodies can set and standardise levels of care. ■

### References

- 1 McCarthy J. Ascribing mental qualities to machines. In: Ringle M (ed). *Philosophical Perspectives in Artificial Intelligence*. Humanities Press, 1979.
- 2 Maloof M. *Artificial intelligence: an introduction*. Georgetown University, 2017.
- 3 Searle JR. Minds, brains, and programs. *Behav Brain Sci* 1980;3:417–24.
- 4 Hildt E. Artificial intelligence: does consciousness matter? *Front Psychol* 2019;10:1535.
- 5 Turing AM. Computing machinery and intelligence. *Mind*. 1950;56:433–60.
- 6 Fast Company, Worzniak S. *Worzniak: Could a computer make a cup of coffee?* YouTube, 2010. [www.youtube.com/watch?v=MowergQR5Y](http://www.youtube.com/watch?v=MowergQR5Y).
- 7 Goertzel B. What counts as a conscious thinking machine? *New Sci* 2012.

- 8 Rajkumar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–58.
- 9 Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *npj Digit Med* 2020;3:126.
- 10 Milea D, Najjar RP, Zhuho J *et al*. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med* 2020;382:1687–95.
- 11 Department for Business, energy and Industry Strategy. *The Grand Challenge missions*. GOV.UK, 2021. [www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/missions#artificial-intelligence-and-data](http://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/missions#artificial-intelligence-and-data)
- 12 Tran B, Vu G, Ha G *et al*. Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. *J Clin Med* 2019;8:360.
- 13 Jiang F, Jiang Y, Zhi H *et al*. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–43.
- 14 Esteva A, Kuprel B, Novoa RA *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
- 15 Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci* 2017;5:457–69.
- 16 Calvert JS, Price DA, Chettipally UK *et al*. A computational approach to early sepsis detection. *Comput Biol Med* 2016;74:69–73.
- 17 Galeon D. For the first time, a robot passed a medical licensing exam. *Futurism* 2017. <https://futurism.com/first-time-robot-passed-medical-licensing-exam> [Accessed 24 March 2019].
- 18 Blease C, Kaptchuk TJ, Bernstein MH *et al*. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. *J Med Internet Res* 2019;21:e12802.
- 19 Intel. *AI adoption in healthcare*. Intel, 2018. <https://newsroom.intel.com/wp-content/uploads/sites/11/2018/07/healthcare-iot-infographic.pdf>. exam [Accessed 26 March 2019].
- 20 Weber C. Engineering bias in AI. *IEEE Pulse* 2019;10:15–7.
- 21 Keskinbora KH. Medical ethics considerations on artificial intelligence. *J Clin Neurosci* 2019;64:277–82.
- 22 Challen R, Denny J, Pitt M *et al*. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28:231–7.
- 23 Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. *J Am Med Informatics Assoc* 2003;10:478–83.
- 24 Hill AB. Reflections on controlled trial. *Ann Rheum Dis* 1966;25:107–13.
- 25 Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 2018;363:k4245.
- 26 Shung DL, Au B, Taylor RA *et al*. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. *Gastroenterology* 2020;158:160–7.
- 27 Loftus TJ, Tighe PJ, Filiberto AC *et al*. Artificial intelligence and surgical decision-making. *JAMA Surg* 2020;155:148–58.
- 28 Xu F, Sepúlveda MJ, Jiang Z *et al*. Artificial intelligence treatment decision support for complex breast cancer among oncologists with varying expertise. *JCO Clin Cancer Informatics* 2019;3:1–15.
- 29 Lamanna C, Byrne L. Should artificial intelligence augment medical decision making? the case for an autonomy algorithm. *AMA J Ethics* 2018;20:E902–10.
- 30 Rid A, Wendler D. Treatment decision making for incapacitated patients: is development and use of a patient preference predictor feasible? *J Med Philos* 2014;39:130–52.
- 31 Sommers BD, Beard CJ, D'Amico AV *et al*. Predictors of patient preferences and treatment choices for localized prostate cancer. *Cancer* 2008;113:2058–67.
- 32 Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet* 2001;357:757–62.
- 33 Moreau A, Boussageon R, Girier P, Figon S. [The 'doctor' effect in primary care]. *Presse Med* 2006;35:967–73.
- 34 Greene A, Greene CC, Greene C. Artificial intelligence, chatbots, and the future of medicine. *Lancet Oncol* 2019;20:481–2.
- 35 Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth uHealth* 2018;6:e12106.
- 36 Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. *PLoS One* 2014;9:e94207.
- 37 Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* 2020;11:3923.
- 38 Bakshi B. *Written evidence submitted by C the Signs (DEL0047): Impact of COVID-19 on suspected cancer referrals and early diagnosis of cancer*. UK Parliament, 2020. <https://committees.parliament.uk/writtenevidence/2672/pdf>
- 39 NHS England. *C the Signs – How artificial intelligence (AI) is supporting referrals*. NHS, 2018. [www.england.nhs.uk/publication/c-the-signs-how-artificial-intelligence-ai-is-supporting-referrals](http://www.england.nhs.uk/publication/c-the-signs-how-artificial-intelligence-ai-is-supporting-referrals) [Accessed 29 November 2020].
- 40 Academy of Medical Royal Colleges. *Artificial intelligence in healthcare*. AOMRC, 2019. [www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial\\_intelligence\\_in\\_healthcare\\_0119.pdf](http://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf) [Accessed 26 March 2019].
- 41 IBM Research. *IBM 5 in 5: With AI, our words will be a window into our mental health*. IBM, 2017. [www.ibm.com/blogs/research/2017/01/ibm-5-in-5-our-words-will-be-the-windows-to-our-mental-health](http://www.ibm.com/blogs/research/2017/01/ibm-5-in-5-our-words-will-be-the-windows-to-our-mental-health)
- 42 Luxton DD (ed). *Artificial intelligence in behavioral and mental health care*. Elsevier Academic Press, 2016.
- 43 Khoury N, Attal F, Amirat Y, Oukhellou L, Mohammed S. Data-driven based approach to aid parkinson's disease diagnosis. *Sensors* 2019;19:242.
- 44 Tang A, Tam R, Cadrin-Chênevert A *et al*. Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J* 2018;69:120–35.
- 45 Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. *Artif Intell Med* 2013;57:9–19.
- 46 Farahmand S, Shabestari O, Pakrah M *et al*. Artificial intelligence-based triage for patients with acute abdominal pain in emergency department; a diagnostic accuracy study. *Adv J Emerg Med* 2017;1:e5.
- 47 Raita Y, Goto T, Faridi MK *et al*. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019;23:64.
- 48 Kim D, You S, So S *et al*. A data-driven artificial intelligence model for remote triage in the prehospital environment. *PLoS One* 2018;13:e0206006.
- 49 Weisberg EM, Chu LC, Fishman EK. The first use of artificial intelligence (AI) in the ER: triage not diagnosis. *Emerg Radiol* 2020;27:361–6.
- 50 Crigger E, Khoury C. Making policy on augmented intelligence in health care. *AMA J Ethics* 2019;21:E188–91.
- 51 Anderson M, Anderson SL. How should AI be developed, validated, and implemented in patient care? *AMA J Ethics* 2019;21:E125–30.
- 52 Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? *AMA J Ethics* 2019;21:E160–6.
- 53 Keskinbora KH. Medical ethics considerations on artificial intelligence. *J Clin Neurosci* 2019;64:277–82.
- 54 Ho CWL, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clin Radiol* 2019;74:329–37.
- 55 Shuttleworth J. *J3016 automated-driving graphic*. Society of Automotive Engineers International, 2019. [www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic](http://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic) [Accessed 30 November 2020].

**Address for correspondence: Dr Fawz Kazzazi, Mason Institute for Medicine, Life Sciences and Law, University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, UK.**  
**Email: fawzkazzazi@gmail.com**  
**Twitter: @drkazzazi**