

AI AS MEDICAL DEVICES

BUILDING TRUST

Text by Dr Scott Wong

Artificial intelligence (AI) technology has captured the headlines and our collective imaginations. Reading through news and academic articles, it would seem that AI is an impending and inevitable silver bullet that will transform healthcare. However, many clinicians may find it difficult to fluently define AI or its capabilities, or describe its uses in the clinical setting. This poses a challenge for doctors to identify the AI *koyok* (slang for snake oil) and allow us to protect ourselves and our patients from potential harms. Furthermore, AI can be directly used in medical and clinical care and are officially classified as AI medical devices (AI-MDs). A look at the US Food and Drug Administration (FDA) shows 1,015 such AI-MDs from 2020 to 2025 that have achieved FDA regulatory clearance,¹ including AI-MDs that can analyse ECGs or diagnose cancer from CT scans. These AI-MDs carry a greater risk of harm to patients compared to AI tools used for administrative, research and policy development purposes, and hence have a higher regulatory requirement before clinical adoption. I hope to provide a common-sense understanding of what AI and AI-MDs

are, and use my own experience developing an AI tool to detect positive and negative Antigen Rapid Test (ART) results from photos to review how clinicians can implement and deploy AI in a safe and ethical manner.

How do we define AI?

AI is an umbrella term for several loosely related technologies, allowing computers to perform tasks we normally associate with human intelligence and cognition – think of recognising complex patterns, making predictions, understanding language and creating new art and music. A definition is provided by the Ministry of Health's first Artificial Intelligence in Healthcare Guidelines (AIHGle), where AI is a set of general-purpose technologies allowing machines to (i) model and optimise, (ii) automate, (iii) forecast and (iv) classify/detect a required result. The AIHGle 2.0 guidelines released on 13 March 2026 moved toward focusing on the subsets of AI called machine learning (ML) and deep learning, which are involved in clinical practice and clinical operations which can directly influence clinical care. Given that these definitions lack

specific explanations of any particular AI technologies, clinicians and patients may focus on AI technologies that are currently in vogue such as ChatGPT and self-driving cars, while paying less attention to AI tools already being used, such as for facial recognition in patient gantries or autocomplete in a search bar.

Three common AI models explained

There are three common approaches to AI technology: ML, neural networks and large language models (LLMs). ML is the most popular form of AI that can make future predictions based on past data. A model is created largely using statistical methods (eg, linear regression, support vector machines, random forests) to identify patterns and classifications in large amounts of data. These models then predict outcomes based on new data, with actual outcomes fed back to the model so that it can further correct itself. Compared to computer programs that need manual updates to improve the execution of specific tasks, ML models can self-improve without explicit programming and can become more

accurate (or less accurate in some cases) with large quantities of data. In theory, people can take a supervisory role to focus on getting more accurate and clean data for training, and decide on accurate and meaningful results for the model to retrain with. A common use case would be determining the length of stay of hospitalised patients, where ML models can analyse inpatient data (including presenting complaint and diagnosis, patient demographics, vital signs, medications and investigations), to predict the estimated length of stay and resource utilisation.

Neural networks are a subset of ML that are loosely inspired by how our brains work. Basic computer units called nodes simulate a neuron; every node takes an input, processes it with an ML algorithm, and then fires the output to the next neuron if it meets a threshold. By connecting many nodes side by side and stacking them hundreds or thousands of layers deep, we create a computational network that seems to simulate how a brain learns and makes decisions, similar to how we may recognise patterns and solve problems. Image recognition, such as analysing chest X-rays or classifying negative and positive ART test kits, benefits greatly from this method.

LLMs, such as generative pre-trained transformers (GPTs), are neural networks designed to understand and generate human language. Training focuses mostly on text scraped from the Internet, with the models learning patterns and structures of language. In the medical context, LLMs may be further trained on research papers, guidelines and even referral replies to produce novel hypotheses and provide research papers, or summarise medical notes to provide a somewhat coherent blue letter response. Two key innovations allow easy LLM interaction with everyday spoken language, where LLMs can understand our context and reply in an eerily human-like manner. The first is by converting text into smaller “tokens” which are assigned numerical IDs. For example, the sentence “A chest X-ray showed a consolidation” can be split into tokens “A”, “chest”, “X”, “ray”, “showed” and “consolidation”. The LLM assigns relationships between each token based on the probability they are used together in sequence (eg, “chest X-ray” rather than “chest ray X”, and often followed by “consolidation”), and the context these tokens may tend to appear

in (eg, desaturation with chest X-ray, and consolidation with pneumonia). For context, the GPT-5 model was trained on almost 20 trillion tokens. The second innovation for LLMs is the concept of “attention”, where the model weighs the importance of different tokens when processing sentences, similar to how we focus on key terms when listening to our patient’s presenting complaint. LLMs then generate novel and coherent responses to our questioning by analysing the tokens and relationships of the question and outputting the words with the highest probability of appearing.

How about AI-MDs?

As defined by the Health Sciences Authority (HSA), AI-MDs are AI solutions meant for the investigation, detection, diagnosis, monitoring, treatment or management of any medical condition, disease, anatomy or physiological process. Compared with AI meant for clinical, administrative, research and in policy development, AI-MDs typically have a direct or indirect impact on patient safety and outcomes and are therefore regulated as medical devices.

Safety and efficacy – what matters for clinicians

Clinicians play a pivotal role in ensuring the efficacy and safety of AI-MDs throughout their development and deployment. The critical considerations of data hygiene and cybersecurity have already been covered in the November 2024 *SMA News* issue, and I will elaborate here on seven key guiding principles in the MOH AIHGLE 2.0 guidance, as well as how clinicians can apply them in practice. These seven principles are:

1. **Safety to patients/Patient-centricity** – Safeguards in the design, development and implementation of AI-MDs allow us to uphold the ethical principle of non-maleficence, and ensure that patients’ interests, including their safety and well-being, are protected.
2. **Fairness** – AI should not be a dividing force, resulting in discriminatory or unjust clinical outcomes for patients across different demographic groups.
3. **Transparency** – End-users of AI-MDs (eg, medical practitioners, patients) should be adequately informed that they are interacting

with an AI-MD, upholding patient and physician autonomy.

4. **Explainability** – Decisions and/or recommendations from an AI-MD should be explainable and reproducible. End-users should be consulted during development or adoption of the AI-MD to ensure that the level of explainability meets their expectations. Examples include knowing the data sets, testing protocols, and algorithmic model uses.
5. **Robustness** – AI-MDs should perform consistently when deployed in different circumstances.
6. **Security and data protection** – AI-MDs should be secure-by-design, to maintain confidentiality, integrity and availability to the patients we care for.
7. **AI alignment to human values or goals**

The crucial task of the clinician is in identifying the unmet clinical needs amid current clinical practices. We are well placed to see the clinical problem as it affects our patients, and to be responsible for the outcomes that can be achievable by AI-MDs. At the beginning of the Delta wave of the COVID-19 pandemic in 2021, I was asked to help patients understand how to do self-testing with an ART kit and also to track the results of positive or negative ART results based on their phone photos. While I was able to rapidly produce a working progressive web application that explained the steps for self-ART testing, the challenge remained how to accurately track thousands of ART results. I was the only person manually classifying thousands of ART images a day, and it was clear that to have the capability of detecting emerging COVID-19 hotspots in real time and verifying ART results before patients are released from quarantine, I needed a way to detect thousands of ART test kit results from photos with around 95% accuracy. A clear need statement defining the problem, population and outcome is key to defining the intended use of an ART image recognition AI-MD. This helped to define the technical ground work for the engineers and allowed clinicians to understand the responsibilities associated with ensuring that patients who were found to be newly positive via self-ART testing could be given appropriate medical advice and quarantine if needed.

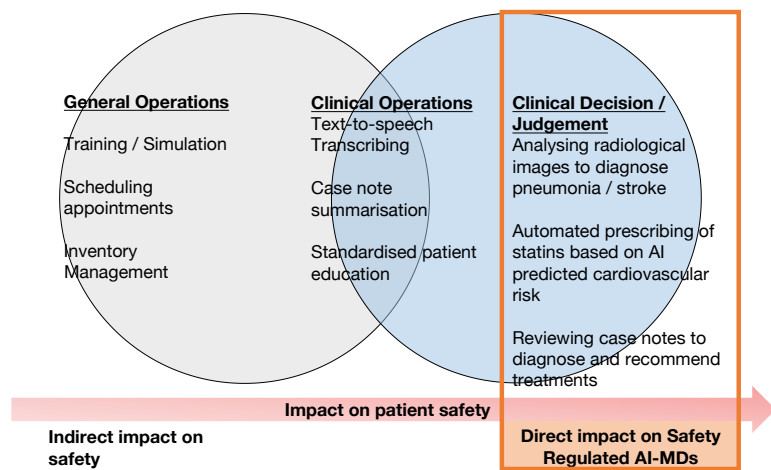


Figure 1: Examples of AI in healthcare and their impact on patient safety. AI-MDs that have the most direct impact on clinical decision-making and judgement are regulated by the Health Sciences Authority

I place patient-centricity as the core guiding value as it enables clinicians to provide transparency, explainability and fairness in its execution. I initially thought that it would surely be easy to create an accurate image classifier that can distinguish one red line from two on an ART kit. After all, I used the positive and negative control swabs of ARTs, to create thousands of positive and negative ART images for training. However, pictures from patients told a different story, with some pictures showing ART kits placed on the instruction sheet containing a drawing of a positive ART test. Some patients even took selfies with their ID and a clock showing the current date and time. Migrant workers also had cameras with lower definition in poor dormitory lighting, with around 30% to 40% of them being illiterate in their native tongue. No AI-MD solution can work if the training photos contains too much data that is **not** the ART test kit. I had to sit down with patients young and old, foreign or local, to co-develop clear instructions and ensure that most people are able to take the best image. The best accuracy happened when patients were simply asked to crop the ART kit, with the cropping tool being an upright rectangle, like cropping a shopping receipt. This subconsciously made patients place ART kits in an upright position, and minimised the image file size while ensuring that only the ART kit was in the image.

Throughout the AI-MD deployment, transparency and explainability was important between not just patients

but also with the parties involved in its deployment. This should not be limited to just the AI portion or its technicalities but should be extended to the entire solution and what it means for the patients and clinicians. For example, I had to liaise closely with patients and providers on the ground to determine how a positive ART test kit could be verified and acted upon. This was even more important when things went wrong, with false positives potentially meaning a work site closure, or the system crashing when overloaded with the activation of contingency screening measures. In the dynamic nature of deploying AI-MDs, regular post-deployment monitoring and constant engagement should be taken to ensure its continued safety and efficacy. If the above principles are not followed, we may end up with poor validation and performance, and end up relying on a poorly performing AI-MD such as the Epic Sepsis Model.²

Ethics and AI-MDs – the era of emerging artificial general intelligence

In medical ethics, there are four core tenets: beneficence, non-maleficence, autonomy and justice. They serve as overriding duties for all medical practice, and the ethical considerations always precede the adoption of AI regardless of the promise that it can hold. While the promise of artificial general intelligence that can exceed the performance of medical doctors is present, we must be more mindful of the current reality that AI

as a technology is backed by large technological corporations with their own vested interests and marketing.

Clinicians should adopt continuous learning and seek to update our understanding of AI and its clinical uses. By first learning to understand, we can then fulfil our responsibility in ensuring that AI-MDs actually support beneficial treatments and not cause harm, either inadvertently or by design. Medical confidentiality is also a key consideration, especially when the training of AI-MDs involves extensive use of the patient's healthcare and social data. We should be able to explain the purpose of the AI-MD before its use and let the patient know if their data is being used for training. The process of obtaining informed consent is key, and patients should be allowed to voluntarily withdraw consent from being treated with any AI-MD or have their data removed from any training.

However, the only certainty is that AI will rapidly change, along with innovations in processing power and its generative capabilities that seem otherworldly. It may be that modern medical practice starts to resemble a constantly evolving tango between technological innovation and our humanistic values. Maintaining open communications through professional advocacy, transparency and consultation both within and beyond the medical fraternity will be key to maintaining our duty of care towards our patients. ♦

References

1. Artificial Intelligence-Enabled Medical Devices. In: US Food and Drug Administration. Available at: <https://bit.ly/4cEFxNg>. Accessed 13 March 2026.
2. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med* 2021; 181(8):1065-70.

Dr Wong is an avid clinician-innovator and a Singapore-Stanford Biodesign fellow focusing on digital medical devices. He focused on regulatory and policy work previously with the Ministry of Health (MOH) and the MOH Office for Healthcare Transformation, and hopes to help patients with neuro- and cardiovascular diseases on the ground and with AI-MDs. Aside from work, he loves lifting weights and drinking good coffee.

