

## Editorial

---

### Who will be the Third Umpire: AI or Radiologists?

It is now accepted that the reports of the death of radiologists, imminently replaceable by artificial intelligence (AI)—the canaries in the coalmine of the AI revolution, were grossly exaggerated.<sup>1</sup> Futurists still caution that the question is not ‘if’ but ‘when’ AI will replace radiologists. It’s difficult predicting the future, particularly of a technology which improves exponentially per Moore’s Law. Nevertheless, an analysis of the present is important in understanding how the radiologist–AI symbiosis may unfold.

#### *Can AI think like radiologists?*

The Turing Test,<sup>2</sup> proposed by the father of AI, Alan Turing, determines whether computers have reached human intelligence. The computer and human are asked a series of questions, such as ‘what do you think of Picasso?’ If human observers, blinded to the source of the answer, can’t determine whether the response was from a computer or a human, then the computer has passed the test. It is a pragmatic measure of computer intelligence, in so far as it demands from it no abstract reasoning. With the emergence of Chat GPT, AI has arguably passed the Turing test. For example, submissions to journals now often include an attestation that generative AI was not used. This means that peer reviewers cannot distinguish real from artificial intelligence.

Imaging interpretation has two components: detection and inference. Detection, for example, is spotting a lung nodule, whereas inference is determining whether that nodule is malignant or benign. The field has evolved in a manner that makes it propitious for AI to pass a radiology-specific Turing test, for multiple reasons.

Radiologists convert visual information into text, the radiology report, which has been standardized, ostensibly to reduce variability and, by creating a universal vernacular, to reduce clinician uncertainty. Whatever the gains of standardization, the end result is that it is no longer possible to tell one radiologist from another—all radiologists sound similar. For AI, the lack of individualism means that the training data is more uniform and easier to master. Now there is only one master chef they must learn from.

Radiology is becoming increasingly algorithmic, partly due to evidence-based medicine, and partly from efforts, misguided to some degree, to quell inter-observer variability. Though actuated by science, the plethora of rules tries vainly to reduce uncertainty. If all radiologists report 5 mm lung nodules in the same way, the net accuracy does not improve. Instead, all radiologists, like Tolstoy’s happy families, are wrong in the same way.

There is a rise of quantification. Numbers prognosticate and diagnose, particularly in cardiac imaging, where numbers, such as ejection fraction, strain, and valve regurgitation fraction, guide management. Since biopsies are eschewed, numbers have taken their place. Hypertrophic cardiomyopathy, for instance, is defined by maximal left ventricular wall thickness and prognosticated by the total amount of myocardial fibrosis.

The net effect of these trends is that the space for judgment has shrunk. Importantly, the case for a highly trained specialist has diminished. One does not need 5 years of medical school and another few of postgraduate training to measure the ejection fraction on cardiac MRI. Paradoxically, in the era of chest X-ray, before the march of CT and MRI, when imaging was far more imprecise, the radiologist’s clinical judgment was far more valuable. Thus, the role of radiologists has changed from inferring to detecting, and measuring. Quantification, which is labour-intensive, not necessarily cognitively-challenging, is ripe for AI takeover.

*Is AI here yet?*

The short answer is, no, which begs the question: why not? AI has been trained in data generated by us, a net average of our strengths and weaknesses, our homunculus. It has faithfully inherited our biases, some of which deepen societal inequities. Even when AI does something clever, like tell a person's race from their chest X-ray,<sup>3</sup> it's unclear how it does that, through genuine clairvoyance or by spotting hidden variables opaque to radiologists.

AI development has taken a reductionist approach. Let's consider AI for chest X-rays. Algorithms are trained for specific findings in the chest X-ray, such as mediastinal widening, cardiomegaly, pleural effusion and pneumothorax. Each trained algorithm, a one trick pony, must be approved by the Food and Drug Administration (FDA). The process is not only costly but betrays how radiologists actually interpret chest X-rays, rarely atomistic, and mostly ensemble. We diagnose conditions such as cardiac failure, pneumonia and chronic obstructive pulmonary disease (COPD). We have not achieved artificial general intelligence, one that takes an ensemble approach, even for chest X-rays.

*More information*

One strength of AI is that it extracts more diagnostic and prognostic information from the image than the human eye. It may show early signs of cancer, heart disease and dementia that radiologists simply cannot see, and so instead of being asymptotic to human performance, it extends it. This prowess, stuff of science fiction, poses two problems. First, how do we know that AI is correct? With no external frame of reference, the ground truth, AI becomes its own ground truth, and a tautological fallacy emerges. The verification problem is exacerbated by the fact that how algorithms arrive at their answer is mysterious, unbeknownst even to their developers, the so-called black box problem. The second issue is that in detecting sub-clinical disease, AI detects clinically irrelevant disease, unleashing an epidemic of overdiagnosis, treating which could cause net harm.

Progress in imaging has led to more information.<sup>4</sup> We see more, know more, but the added knowledge has not translated linearly to better patient outcomes. AI may accelerate that trend leading to too much information. The resultant information flood will need a non-machine to discard information, so that an optimal course is found for the individual patient.

*Automation*

AI is an automating technology, meaning that it does tasks that formerly required human labour. In that sense, it is like tractors, which replaced scythe-wielding farmers, but not farmers. History shows that automation created more jobs than it destroyed.<sup>5</sup> Automation in medicine has not negated human participation. Let's consider the automated external defibrillator (AED), which detects ventricular fibrillation, but still needs permission to deliver a shock. This is sensible because there is a small, though not insignificant, risk of a false positive with catastrophic consequences. The first responder checks for the one piece of information that the AED does not have: the presence of a pulse.

*The third umpire*

At this point, it is worth visiting an esoteric but pertinent concept: Gödel's Incompleteness Theorem,<sup>6</sup> formulated by the Austrian mathematician, Kurt Gödel, which is as follows:

1. In any consistent formal system  $F$  within which a certain amount of arithmetic can be carried out, there are statements of the language of  $F$  which can neither be proved nor disproved in  $F$ .
2. Such a formal system cannot prove that the system itself is consistent (assuming it is indeed consistent).

The limitation was stated more poetically by Rudyard Kipling<sup>7</sup> who once lamented 'what do they know of England, who only England know?' Kipling, frustrated by the ignorance of fellow Englishmen, appreciated that to truly appreciate England one needed to see the country from outside. Gödel's Incompleteness arises because systems can't see themselves through the looking glass, looking in. This is a feature, not a bug of AI.

The major task of the man-machine collaboration is, therefore, to bridge Gödel's

incompleteness. How it does that is the imminent question. One answer comes from an unexpected quarter, likely familiar to readers—cricket. Despite the progress in technology, decisions such as leg before wicket are still made by the on-field umpires. The third umpire, who uses all available technology, is called upon rarely. The third umpire can also be wrong. But if the judgment of the third umpire and on-field umpires are concordant, one can be nearly certain that it is correct. Importantly, the third umpire has the last say when asked.

In the radiologist–AI symbiosis the question is whether the third umpire will be AI or the radiologist. If radiologists are to be the third umpires, their role must evolve from the inference-detection paradigm. They must become overseers of AI, the integrators of information. They will have to see what AI can't. Their training must evolve to enable this symbiosis. Future radiologists may consider merging with other information extractors such as pathologists, creating a new discipline: Information Specialists,<sup>8</sup> with AI being to them what Dr Watson was to Sherlock Holmes—a thoughtful assistant, always respected, often overruled.

#### REFERENCES

- 1 Langlotz CP. Will artificial intelligence replace radiologists? *Radiol Artif Intell* 2019;**1**:e190058.
- 2 Turing AM. Computing machinery and intelligence. *Mind* 1950;**59**:433–60.
- 3 Gichoya JW, Banerjee I, Bhimoreddy AR, Burns JL, Celi LA, Chen LC, *et al*. AI recognition of patient race in medical imaging: A modelling study. *Lancet Digit Health* 2022;**4**:e406–e414.
- 4 Jha S. Automation and radiology: Part 1. *Acad Radiol* 2020;**27**:147–9.
- 5 Acemoglu D, Restrepo P. Artificial intelligence, automation and work. NBER (Working Paper) 2018;**24196**:1–41.
- 6 Raatikainen P. Gödel's incompleteness theorems. In: Zalta EN (ed). *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition). Available at <https://plato.stanford.edu/archives/spr2022/entries/goedel-incompleteness> (accessed 6 Jun 2024).
- 7 The English Flag. Available at [www.kiplingsociety.co.uk/poem/poems\\_englishflag.htm](http://www.kiplingsociety.co.uk/poem/poems_englishflag.htm) (accessed 6 Jun 2024)
- 8 Jha S, Topol EJ. Adapting to artificial intelligence: Radiologists and pathologists as information specialists. *JAMA* 2016;**316**:2353–4.

SAURABH JHA  
*Department of Radiology*  
*University of Pennsylvania*  
*Philadelphia, USA*  
*saurabh.jha@penncmedicine.upenn.edu*

[To cite: Jha S. Who will be the Third Umpire: AI or Radiologists? *Natl Med J India* 2024;**37**: 121–3. DOI: 10.25259/NMJI\_1018\_2024]

*The National Medical Journal of India* is covered in  
**Medline, Current Contents: Clinical Medicine, Science  
Citation Index, Science Citation Index Expanded,  
SCOPUS, Biosis Previews and SciSearch.**