BOOK REVIEW

Can We Trust AI?

Mark K. Huntington, MD, PhD[a,b]

AUTHOR AFFILIATIONS:

[a] Center for Family Medicine, Sioux Falls, SD

[b] Sanford School of Medicine, University of South Dakota, Sioux Falls, SD

HOW TO CITE: Huntington MK. Can We Trust AI?. *Fam Med.* 2024;56(X):1-2. doi: 10.22454/FamMed.2024.627355

PUBLISHED: 14 October 2024

Book Title: Can We Trust AI?

Author: Rama Chellappa, PhD

Publication Details: Johns Hopkins University Press, 2022, 197 pp., $16.95, paperback

The rise of ChatGPT and other large language models (LLMs) just over a year ago thrust artificial intelligence (AI) into the spotlight. No longer the purview of a few experts and slightly nerdy aficionados, AI dominates headlines and conversations. What is the role of AI? What is its future? What ethics surround its use? To address these questions, a basic understanding of AI is needed, and Dr Rama Chellappa's book is just the thing to provide it.

An eminently readable, concise summary of AI, this book will provide a solid introduction for the layperson—including physicians—for this field. Though it is a bit dated, an inevitable result for a printed text in this rapidly developing field (it was published before ChatGPT's revolution), it lays a solid foundation upon which to build one's knowledge.

The author's description of the state of AI is still apt today: "more at the intellectual level of a curious toddler than a silver-haired genius planning world domination" (p. 2). For those who've raised toddlers, the potential for destruction is not necessarily reassuring. The introductory chapter presents differences among North America, Europe, and Asia (specifically, China) in the adoption and regulation of AI. It reviews beneficial applications such as autonomous vehicles as well as those more ominous such as predictive policing and assigning social credit scores. The technology that makes AI feasible, specifically the use of graphic processing units rather than central processing units to increase computing power, is explained in a way that is comprehensible, without getting too deep in the weeds. The ethics of AI are introduced by citing the author and thinker Isaac Asimov's "Three Laws of Robotics":[1]

- A robot (ie, AI) may not injure a human being or, through inaction, allow a human being to come to harm.
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. (p. 18)

Following the introductory chapter, a deeper dive into several areas of application is made. The first area is the field of medicine and the potential for using AI to predict patient outcomes, interpret clinical tests such as electroencephalograms, and aid in diagnosis. An opportunity for AI to demonstrate its capabilities came unexpectedly with COVID-19. "This pandemic was a big test for AI and medicine. . . . But I don't think we passed the test" (p. 56). The algorithms employed often were based on mislabeled data, unverified data, or simply too few points of data. This outcome provided a poignant reminder of the GIGO acronym from computing's early days: garbage in, garbage out.

Facial recognition, with its image analysis and interpretation capabilities, is another area of obvious relevance to medicine (eg, reading X-rays, pathology slides, and endoscopic screening). Accuracy of AI has varied between 68% and 96%, depending on the study. Bias adversely affects accuracy; how this might manifest itself in medical image analysis remains to be seen.

Autonomous vehicles are a final example of active areas of AI development. The application has clear relevance to improving health by reducing motor vehicle accidents and improving supply chains, or to worsening it through applications such as autonomous warfare.

In discussing the future of AI, the author suggests a variety of promising directions. Regarding medicine, these include improved disaster relief, more effective home health care, and personalized medicine.

The author identifies four potential areas of concern as AI emerges: bias, data domain shifts (altered performance on datasets that differ from the specific ones used to train the AI), privacy concerns, and vulnerability to cyberattacks. While this list is a good place to start, other significant areas of potential concern exist beyond these four.

Dr Chellappa's book is a useful orientation to the general concept of AI and examples of its applications. The book is a good starting point. For those already possessing minor familiarity with AI, this book may be too basic. Similarly, for those interested in orientation to hot topics such as LLMs and generative AI, other sources—likely online, continuously updated ones—will be far more current and complete.

REFERENCES

1. Asimov I. Runaround. *Astounding Science-Fiction.* 1942;29:94-103. https://archive.org/details/Astounding_v29n01_1942-03_dtsg0318/page/n93/mode/2up