2024, Volume 56, Issue 7, 462-463, e-ISSN 1938-3800

Family Medicine

BOOK REVIEW

Guardrails: Guiding Human Decisions in the Age of AI

Franklin Berkey, DO

AUTHOR AFFILIATION:

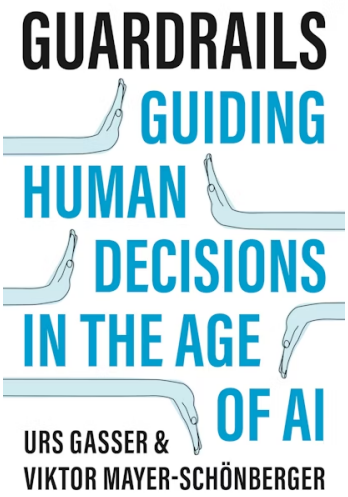
Family and Community Medicine, Penn State College of Medicine, State College, PA

CORRESPONDING AUTHOR:

Franklin Berkey, Family and Community Medicine, Penn State College of Medicine, State College, PA, fberkey@pennstatehealth.psu.edu

HOW TO CITE: Berkey F. Guardrails: Guiding Human Decisions in the Age of AI. *Fam Med.* 2024;56(7):462-463.
doi: 10.22454/FamMed.2024.383539

PUBLISHED: 21 May 2024



Book Title: Guardrails: Guiding Human Decisions in the Age of AI

Authors: Urs Gasser, Viktor Mayer-Schönberger

Publication Details: Princeton University Press, 2024, 226 pages, $27.95 hardcover

We make a staggering number of daily decisions, mostly dominated by the mundane, inconsequential choices of everyday life. In addition to these decisions of daily living, physicians are tasked with potentially life-altering clinical decisions that make decision fatigue an unfortunate component of medical care. Studies have demonstrated that surgeons schedule fewer patients for the operating room when nearing the end of their shift,[1] and office-based medical physicians order fewer tests as their day progresses.[2]

Enter artificial intelligence (AI), the remedy to eliminate physician fatigue, bias, and innate cognitive limitations in decision-making. Hailed as a profession-changing answer in the late 1950s, the integration of AI into clinical medicine has been slowed by unforeseen stumbling blocks.[3] These limitations are not unique to medicine and, in fact, are ubiquitous in nearly all professions. In *Guardrails: Guiding Human Decisions in the Age of AI*, authors Urs Gasser and Viktor Mayer-Schönberger argue that the key to integrating AI is not in the governance of data, but rather in the governance of decisions. Gasser, a professor of public policy, governance, and innovative technology at the Technical University of Munich, and his coauthor Mayer-Schönberger, professor of Internet governance and regulation at the University of Oxford, use historical events and everyday references to illustrate the consequences of the absolute substitution of AI for human decision-making.

While the reader may know of the stories, the behind-the-scenes decision-making pathway of those involved are what illustrate the authors' points. In a scenario pitting human decisions against that of a computer, the authors reference a 2002 airline collision in the skies over Germany in which the airline crews received conflicting information from the air traffic controller and the plane's automatic collision warning system. In making critical decisions, the pilots needed to select between good and bad information, although they were uncertain which was which. While the crew's ultimate decision to follow an air traffic controller's recommendation over the computer-provided direction proved fatal, Gasser and Mayer-Schönberger argue that blindly accepting AI-generated decisions is a deeply flawed practice.

Similar to the pilots' dilemma, discerning correct information from falsities is paramount in making clinical decisions. However, identifying misinformation has proven difficult even for machines. Internet giants Google and Facebook use complicated algorithms to identify misinformation—how could they not when they receive millions to billions of take-down requests each year? However, the machines have continual difficulty understanding shifting societal norms and social context, forcing Facebook to supplement these machines with 15,000 employed content moderators.

Bias is a widely cited innate human characteristic leading to poor decision-making; yet, when identifying misinformation, computers struggle with similar bias. Gasses and Mayer-Schönberger relay the story of a Black couple with six-digit salary jobs and very good credit scores who were denied a mortgage. The authors cite a recent study of mortgage approval algorithms that found lenders were 80% more likely to deny Black applicants than White applicants with similar financials. The reason for this, the duo explains, is rather simple: bias in, bias out. Because data-driven AI is trained with previous, real-life,

and inherently biased decisions, human biases are embedded into the artificial algorithms. Subsequently, as newer data based on a greater proportion of AI-made decisions is added, AI begins to learn from its own decisions, thus magnifying biases and limiting innovation.

While the authors do not provide the ultimate answer, they foresee solutions that will be more societal than technical, portraying a flexible framework that seeks compromise. As in the examples of the ill-fated flight, social media's regulation of misinformation, and the mortgage industry's screening algorithms, the ultimate pathway is a system of guardrails that prioritize social concern through human oversight.

Guardrails provides a broad overview of a dense academic topic, but the embedded stories provide a foundation for understanding. While Grasser and Mayer-Schönberger include a few medical references, the issues related to critical decision making, misinformation, and bias are easily applicable to medicine. Improving individual decision-making while ensuring both human agency and progress, they contend, is paramount—and this argument should leave the family physician reassured: Our jobs are secure.

REFERENCES

1. Persson E, Barrafrem K, Meunier A, Tinghög G. The effect of decision fatigue on surgeons' clinical decision making. *Health Econ.* 2019;28(10):1994-1203.
2. Trinh P, Hoover DR, Sonnenberg FA. Time-of-day changes in physician clinical decision making: A retrospective study. *PLoS One.* 2021;16(9):257500.
3. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine. *N Engl J Med.* 2023;388(13):1201-1208.