Volume Title: The Economics of Artificial Intelligence: Health Care Challenges

Volume Authors/Editors: Ajay Agrawal, Joshua Gans, Avi Goldfarb, and Catherine Tucker, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-83311-8 (cloth); 978-0-226-83312-5 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges

Conference Date: September 22–23, 2022

Publication Date: March 2024

Chapter Title: Comment on Chapters 1 and 4: Health AI, System Performance, and Physicians in the Loop

Chapter Author(s): W. Nicholson Price II

Chapter URL: https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges/comment-chapters-1-and-4-health-ai-system-performance-and-physicians-loop

Chapter pages in book: p. 147 – 150

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018b. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Press.
Bernheim, B. Douglas, Stefano DellaVigna, and David Laibson. 2019. *Handbook of Behavioral Economics-Foundations and Applications 2*. Amsterdam: North-Holland.
Bundorf, M. Kate, Maria Polyakova, and Ming Tai-Seale. 2019. "How Do Humans Interact with Algorithms? Experimental Evidence from Health Insurance." NBER Working Paper No. 25976. Cambridge, MA: National Bureau of Economic Research.
Morgenstern, Oskar, and John Von Neumann. 1953. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
Resnicow, Ken, Delwyn Catley, Kathy Goggin, Sarah Hawley, and Geoffrey C. Williams. 2022. "Shared Decision Making in Health Care: Theoretical Perspectives for Why It Works and for Whom." *Medical Decision Making* 42 (6): 755–64.
Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley & Sons.
Stacey, Dawn, France Légaré, Krystina Lewis, Michael J. Barry, Carol L. Bennett, Karen B. Eden, Margaret Holmes-Rovner, Hilary Llewellyn-Thomas, Anne Lyddiatt, Richard Thomson, and Lyndal Trevena. 2017. "Decision Aids for People Facing Health Treatment or Screening Decisions." *Cochrane Database of Systematic Reviews* 4 (4): CD001431.

Comment on Chapters 1 and 4 W. Nicholson Price II

Health AI, System Performance, and Physicians in the Loop

Accounts of artificial intelligence (AI) in medicine must grapple, in one way or another, with the interaction between AI systems and the humans involved in delivering healthcare. Humans are, of course, involved throughout the process of developing, deploying, and evaluating AI systems, but a particular role stands out: the human in the loop of an algorithmic decision. In medicine, when an algorithm is involved in a decision, a typical view of the system envisions a human healthcare professional mediating that algorithm—deciding whether and how to implement or react to any recommendation, prediction, or other algorithmic output. This person is the human in the loop, and their role is often central, complicated, and contested.

Principal contributions to this volume recognize the key role of humans in the loop. Dranove and Garthwaite (2023) focus on the value chain of healthcare and the physician's role within it, recapitulating the notion of the

W. Nicholson Price II is a professor at University of Michigan Law School and a senior fellow at the University of Copenhagen Centre for Advanced Studies in Biomedical Innovation Law.

For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges/health-ai-system-performance-and-physicians-loop.

physician as captain of the ship—a key human-in-the-loop conception—and recognizing that AI's contribution, and who will capture its value in the healthcare delivery setting, depends in some part on its role in substituting for that human role versus complementing it. Lakkarju and Farronato (2022), in ongoing work presented at the conference but not included in this volume, elucidate the complexities of physician–algorithm interactions in terms of system performance. They demonstrate that when physicians using AI systems are provided with explanations of the system's recommendations, their performance may be better or worse than the system alone, depending on how accurate and complete those explanations are. Physicians provided with small amounts of accurate information about the model's recommendations performed worse than the model alone; those who received substantial accurate information did better, but those who received substantial information that was somewhat accurate did worse. The effects are complex and nonintuitive. And even the regulatory regime described by Stern (2023) turns on the role of the human in the loop, whether that human is the intended user of an FDA-regulated system (and thus intended but not required to use the system as labeled) or the adequately informed human user of clinical decision support software (and thus rendering the system outside FDA's regulatory authority) (US FDA 2022).

Conceptions of AI systems in health typically envision the human in the loop as a well-trained, well-resourced physician with adequate resources and adequate time. But variations in who the human in the loop is, what they can do, the context in which they do it, and what they are supposed to be doing in the first place may substantially change how the whole system functions and the economic implications of that function. Consider two perturbations: what the human in the loop is supposed to be doing, and the capabilities of the human who occupies that role.

First, system designers, patients, health systems, and physicians may have very different visions of what the human in the loop is supposed to be doing, whether implicitly or explicitly, and these visions may be in substantial tension (Crootof, Kaminski, and Price 2023). The most obvious role for a physician in the loop of an algorithmic system is to increase accuracy—indeed, this conception underlies Lakkarju and Farronato's evaluation of the success of different AI explanatory models (2023). But a conception of physicians as only improving the system's *overall* performance may align poorly with the expectations of patients, who would likely prefer that physicians prioritize their own individual outcomes, regardless of that prioritization's impact on overall systemic accuracy or efficiency. Or patients and patient advocates may prefer that physicians not defer to AI systems because they consider machine decision making to be deleterious to patient dignity, and want a human role in protecting that dignity (Crootof Kaminski, and Price 2023). On the other hand, whatever the impact on efficiency or accuracy, physicians themselves may prefer to remain in control—and avoid deference—because the role of physician as the knowledgeable captain of the healthcare ship is essential to physician job security and prestige. These implicit or explicit roles may easily conflict, and unless system designers and regulators construct the physician–system interactions quite carefully, those conflicts can easily go unnoticed and unresolved.

Second, even assuming alignment of roles and incentives (and here let us assume a role focused on purely on accuracy), the human who is actually in the loop of an algorithmic system may be substantially different from the human assumed by system designers, users, or regulators. In particular, expectations may often be significantly higher than reality. Not all healthcare providers will be able, adequately trained, or well resourced enough to catch errors in the system or to ensure that it works as intended, especially in settings that differ significantly from those in which the algorithm was developed (Price 2020).

Consider regulators. In October 2022, FDA issued a final guidance laying out how it evaluates whether software is clinical decision support software (US FDA 2022). Such software, which is intended merely to inform physicians and to give them adequate information to evaluate the software's recommendations, has been congressionally defined as not a medical device and therefore outside FDA's regulatory authority (21 U.S.C. § 360j[o][1][E]). And so FDA carefully considered what software needs to do to fall within this exception: It cannot be time-critical, because humans would tend to rely on the software in a crunch; it cannot provide only one recommendation, because humans would tend to defer; and it must provide a large amount of information to support its recommendations, because humans need that information to evaluate that recommendation (US FDA 2022). In this vision, an adequately enabled human in the loop ensures that the system will not dominate the care decision process, which therefore removes the need for a heavy regulatory hand. But of course, all this relies on a healthcare professional who has the time, training, and inclination to evaluate and review recommendations even when the decision is *not* time-critical, rather than just picking the top off the list of ranked recommendations and going along with that recommendation. That vision of a healthcare professional in the loop may not accurately reflect the human filling the role; healthcare professionals are always pressed for time and already burned out on computer-related tasks (Downing, Bates, and Longhurst 2018).

Overly prescriptive or demanding assumptions about humans in the loop are likely to result in systemic underperformance. If system designers, implementers, and regulators assume a time-rich expert and bake that assumption into AI design, the absence of that expert could easily cause system failure—though whether that failure is obvious, catastrophic, or insidiously opaque will depend on the particulars of the system and its failure mode (Choi 2019). But an expert human in the loop need not be assumed. Sometimes, indeed, AI systems may need to be designed to be almost totally agnostic as to who is the human in the loop, what role they might perform, and what training or resources they may have available.

Taking it one step further, the value of AI systems in healthcare settings may in fact be greatest in situations where human experts are *least* available. A key potential role of AI systems is to democratize expertise and to make formerly specialist capabilities available more widely (Price 2019). Some systems do this explicitly—the IDx-DR diabetic retinopathy diagnosis system aims to provide broad access to a diagnostic tool formerly only available with the assistance of an ophthalmologist, and other systems in development aim for similar breadth (Grzybowski et al., 2019). Other systems may achieve such broad reach only by use far beyond their FDA-cleared label or other intended use. But they may nonetheless bring healthcare capacity to many otherwise underserved individuals, whether domestically or internationally, and this result should be celebrated even if it is harder to evaluate or reimburse.

Humans in the loop are key to considerations of AI systems in health as in other fields. What those humans are supposed to do, the resources they have available, and even who stands in the room in the first place all underlie how well the system works and who ultimately benefits.

References

Choi, Bryan. 2019. "Crashworthy Code." *Washington Law Review* 94: 39–117.

Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson Price II. 2023. "Humans in the Loop." *Vanderbilt Law Review* 76 (2): 429–510.

Downing, N. Lance, David W. Bates, and Christopher A. Longhurst. 2018. "Physician Burnout in the Electronic Health Record Era: Are We Ignoring the Real Cause?" *Annals of Internal Medicine* 169 (1): 50–52.

Dranove, David, and Craig Garthwaite. 2023. "Artificial Intelligence, the Evolution of the Healthcare Value Chain, and the Future of the Physician." This volume.

Grzybowski, Andrzej, Piotr Brona, Gilbert Lim, Paisan Ruamviboonsuk, Gavin S. W. Tan, Michael Abramoff, and Daniel S. W. Ting. 2020. "Artificial Intelligence for Diabetic Retinopathy Screening: A Review." *Eye* 34: 451–60.

Lakkarju, Himabindu, and Chiara Farronato. 2022. "When Algorithms Explain Themselves: AI Adoption and Accuracy of Experts' Decisions." Presented at NBER Conference on Economics of Artificial Intelligence, Toronto, September 22–23, 2022.

Price, W. Nicholson, II. 2019. "Artificial Intelligence in the Medical System: Four Roles for Potential Transformation." *Yale Journal of Health Policy, Law, and Ethics* 18 (3): 122–32.

Price, W. Nicholson, II. 2020. "Medical AI and Contextual Bias." *Harvard Journal of Law & Technology* 33 (1): 66–116.

Stern, Ariel Dora. 2023. "The Regulation of Medical AI: Policy Approaches, Data, and Innovation Incentives." This volume.

US FDA (Food and Drug Administration). 2022. "Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff." https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software.