



## OPEN ACCESS

## EDITED BY

Zhan Zhang,  
Pace University, United States

## REVIEWED BY

Zhe He,  
Florida State University, United States  
Claus Bossen,  
Aarhus University, Denmark

## \*CORRESPONDENCE

Mara Ulloa  
mara.ulloa@northwestern.edu

## SPECIALTY SECTION

This article was submitted to  
Human-Media Interaction,  
a section of the journal  
Frontiers in Computer Science

RECEIVED 15 September 2022

ACCEPTED 18 November 2022

PUBLISHED 01 December 2022

## CITATION

Ulloa M, Rothrock B, Ahmad FS and  
Jacobs M (2022) Invisible clinical labor  
driving the successful integration of AI  
in healthcare.  
*Front. Comput. Sci.* 4:1045704.  
doi: 10.3389/fcomp.2022.1045704

## COPYRIGHT

© 2022 Ulloa, Rothrock, Ahmad and  
Jacobs. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Invisible clinical labor driving the successful integration of AI in healthcare

Mara Ulloa<sup>1\*</sup>, Blaine Rothrock<sup>1</sup>, Faraz S. Ahmad<sup>2,3</sup> and  
Maia Jacobs<sup>1,3</sup>

<sup>1</sup>Department of Computer Science, Northwestern University, Evanston, IL, United States,

<sup>2</sup>Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States, <sup>3</sup>Department of Preventive Medicine, Northwestern University, Evanston, IL, United States

Artificial Intelligence and Machine Learning (AI/ML) tools are changing the landscape of healthcare decision-making. Vast amounts of data can lead to efficient triage and diagnosis of patients with the assistance of ML methodologies. However, more research has focused on the technological challenges of developing AI, rather than the system integration. As a result, clinical teams' role in developing and deploying these tools has been overlooked. We look to three case studies from our research to describe the often invisible work that clinical teams do in driving the successful integration of clinical AI tools. Namely, clinical teams support data labeling, identifying algorithmic errors and accounting for workflow exceptions, translating algorithmic output to clinical next steps in care, and developing team awareness of how the tool is used once deployed. We call for detailed and extensive documentation strategies (of clinical labor, workflows, and team structures) to ensure this labor is valued and to promote sharing of sociotechnical implementation strategies.

## KEYWORDS

artificial intelligence, healthcare, sociotechnical systems, decision support systems, human-AI collaboration

## 1. Introduction

Since the widespread adoption of electronic health records (EHRs), prospects of utilizing data to improve patient care have been in the spotlight of medical and informatics research. Artificial intelligence (AI) may improve healthcare decision-making by considering far more data than any individual in making recommendations for common medical decisions (Wang and Summers, 2012). Recent advancements in AI used to make predictions from large clinical datasets have generated a barrage of applications, though many attempts fail during implementation (Coiera, 2019; Emanuel and Wachter, 2019; Yang et al., 2019; Matthiesen et al., 2021).

Much of the work on the design and development of medical-AI tools favor algorithmic performance (Beede et al., 2020; Johnson et al., 2021). A less discussed aspect of these tools is the required labor of the end-users throughout development and deployment (Anderson and Aydin, 1997; Pratt et al., 2004).

For example Sepsis Watch, a successful AI system, reported personnel time as the largest resource for implementation (Sendak et al., 2020). Therefore, understanding how new technologies impact people's work is a critical step in fostering technology usefulness, acceptance, and adoption (Kushniruk, 2002; Khairat et al., 2018). However, recent reviews of clinical AI tools have exposed barriers to adoption in real-world settings (Middleton et al., 2016; Tonekaboni et al., 2019). Consistently, these reviews show that poor adoption is not due to inadequate algorithmic performance but rather issues related to how the tools are designed and implemented into existing workflows (Yang et al., 2019). Pratt et al. highlighted a similar paradigm at the peak of the transition to EMRs, showing that a focus on Computer-Supported Cooperative Work (CSCW) research is necessary for successful implementation: "Often, the causes of system failures cannot be explained in purely technical terms. Rather, the complex network of relationships among people in an organization strongly affects the success of a technology" (Pratt et al., 2004).

We see a clear need to discuss the often-invisible labor required to create and deploy medical-AI (Anderson and Aydin, 1997; Pratt et al., 2004). The concept of invisible labor was first introduced as undervalued and often unpaid work that serve critical roles in society (Daniels, 1987). The term was later extended to CSCW to encompass the redistribution of work with technical solutions (Suchman, 1995; Star and Strauss, 1999). More recently, 'invisible labor' represents crowd-workers who support many tasks related to web automation and AI (Gray and Suri, 2019). To avoid undervaluing the labor required for successful implementations, we argue that we must understand the complex relationship between people in healthcare organizations and AI tools.

In this perspective, we examine three case studies, outlined in Figure 1, two from an author's previous work and one based on a recently-deployed system. These studies identify the labor performed by clinical teams to create and use AI technology (Jacobs et al., 2021a,b; Cheema et al., 2022). Using these findings and related literature, we describe four examples of how labor is needed in the creation of medical-AI that requires clinical expertise (summarized in Figure 2). We do not consider this list exhaustive. Rather, we present these to motivate more documentation and sharing of the effort and workflow strategies used in the design and implementation of medical-AI.

## 2. Examples of labor behind clinical AI developments

We describe four types of labor involving clinical staff, occurring at four distinct phases of the development process. We refer to the first phase of development as **preparing data**, as it involves preparing data for algorithmic use. Once data is prepared, **training algorithms** on the data

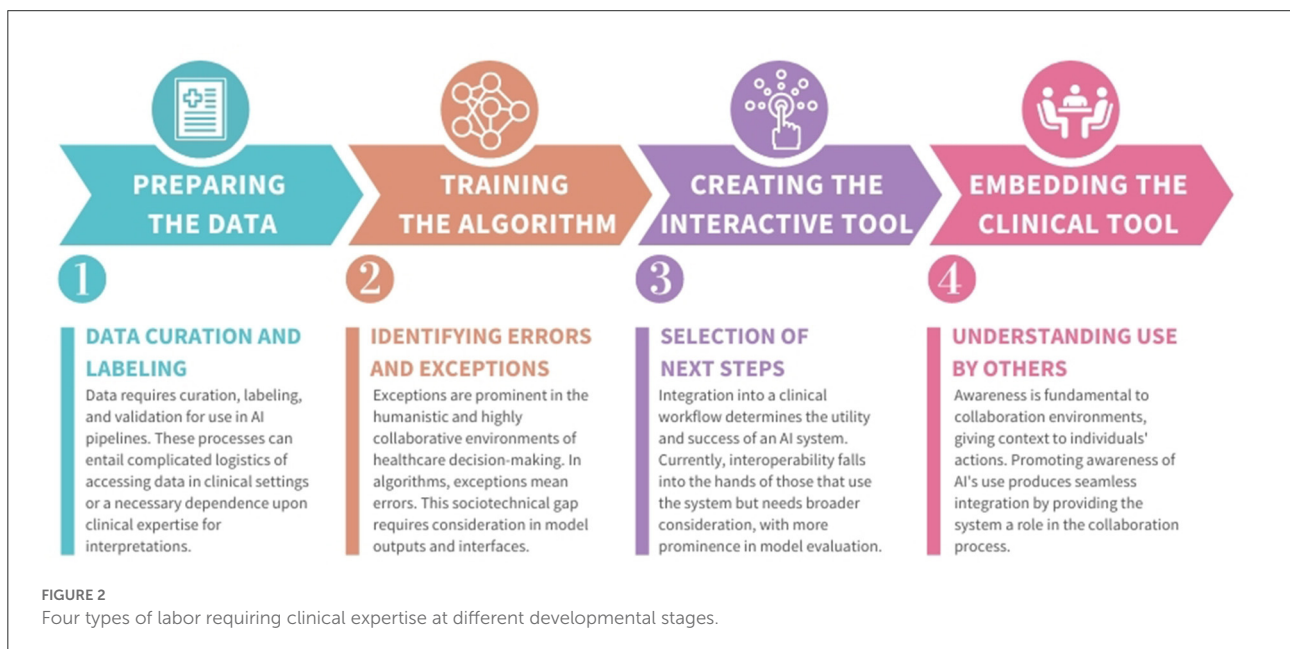
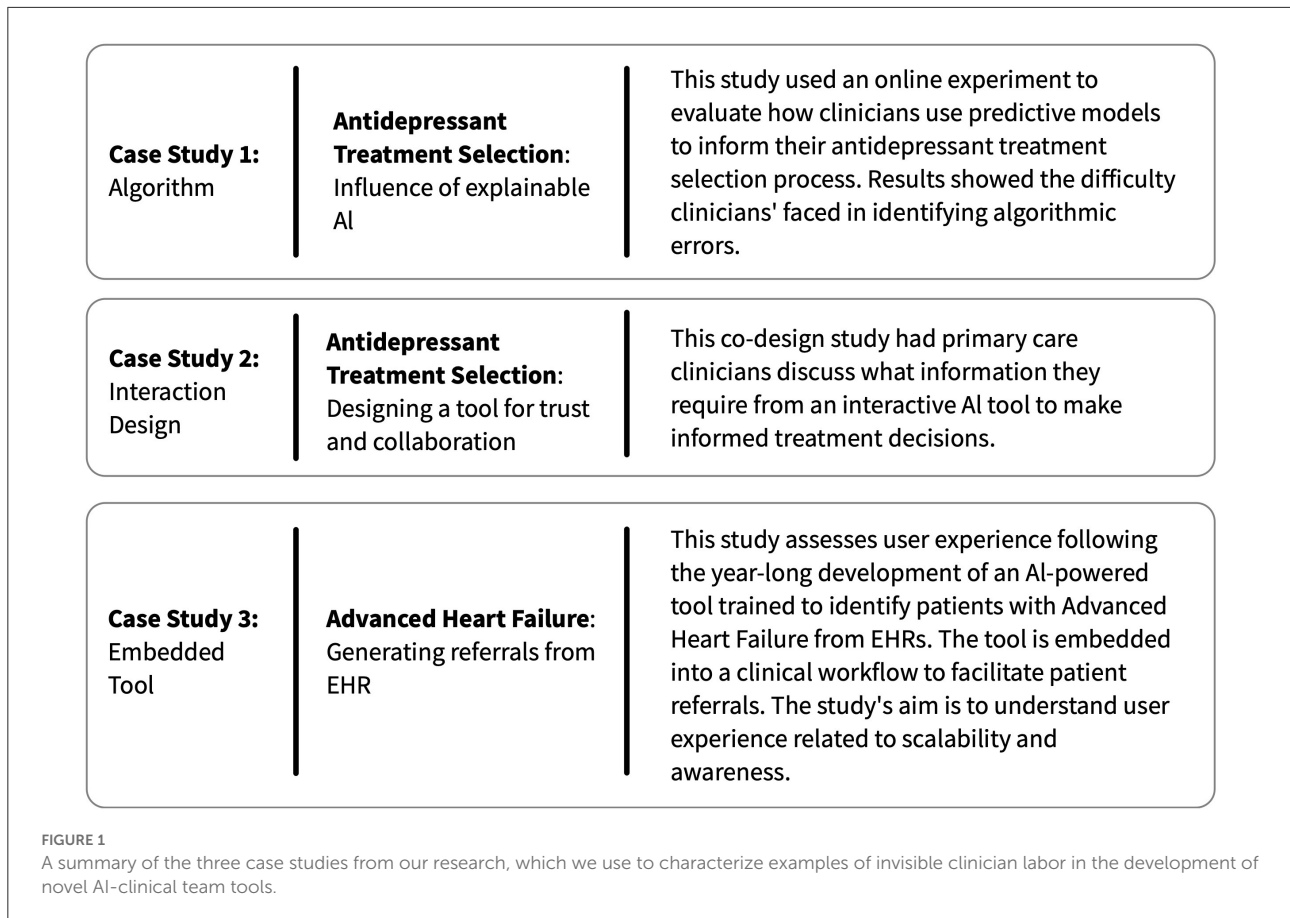
takes place to make predictions. Achieving an acceptable prediction accuracy can take substantial time and fine-tuning. The third phase requires **creating an interactive tool**, which can present algorithmic output to user(s). Finally, once deployed into a clinical setting, **embedding of the clinical tool** occurs, as it now exists within a broader clinical context. Other development phases can describe this process (Nascimento et al., 2019). We include these four phases to illustrate how labor requirements can change over time.

### 2.1. Data labeling with clinical expertise

Machine learning (ML) and AI pipelines require data first to be curated, labeled and validated. Data curation is the process of collecting data about a particular problem, which can be difficult and generally produces datasets siloed to single health systems (Wiens et al., 2019; Wu et al., 2021). Data labeling entails classifying each data sample, which is then used to train models to identify those classifications from new samples. While methods exist for labeling automation (Wang et al., 2018), manual intervention is regularly required (Rajkomar et al., 2019). Data validation compares data labels or model outputs against ground truth methods (e.g., -expert analysis) which can introduce measurement bias in interpretations (Suresh and Guttag, 2021). Highly specialized classification problems, like medical diagnosis, require experts at each of these steps, which is expensive and laborious.

In case study 3, an AI-powered tool was embedded within the EHR to develop regular reports of patients at risk of advanced heart failure (Cheema et al., 2022). Labeling of datasets comprised a delicate balance between clinical expertise and the availability of human labor. Determining the appropriate heart failure stage for a patient is a complex process with no clear-cut lines between stages and requires data across multiple sources. As one participant shared: "Chart review takes time. It takes time to go through all the clinical notes. We're looking at [multiple electronic medical records like] what's in Epic, what's in [external systems like] Care Everywhere, making sure that we've got a thorough understanding of what's going on with that patient." The team developed an approach in which the algorithm was tested on a holdout set of patients labeled by a cardiology fellow using criteria from clinical consensus statements. Once the model and workflow went live, the nurse clinician documented the label for each patient. These labels were then used to evaluate model performance prospectively.

Thus far, we have not seen standardized methodologies for continually incorporating clinical expertise to data collection and labeling, resulting in many research efforts relying on few, stale datasets (Sendak et al., 2019). Ultimately,



an AI tools' aim is to ease the workload for clinicians, but developing these tools requires devoting effort to data curation, labeling, and validation. AI development

has an uphill labor commitment before it produces relief in clinical processes, which must be realized and planned accordingly.

## 2.2. Identifying algorithmic errors and exceptions

AI provides technological assistance by analyzing datasets larger than any individual could interpret to identify patterns for decision-making. However, these algorithms will never be perfect due to the unpredictability of dynamic variables like human behavior and environmental factors. For an algorithm to make an accurate prediction, it must see many comparable examples. With the numerous variables involved in medicine, both erroneous predictions and exceptional cases are probable.

Yu and Kohane provide a hypothetical example where a technician accidentally leaves ECG leads on a patient before a chest X-Ray (Yu and Kohane, 2019). A clinician examining the X-Ray would immediately know the problem and order a new image or ignore the anomaly. By contrast, an AI algorithm encountering such an example could produce an utterly invalid recommendation. We found a real example of this in case study 3, where the Nurse Coordinator (NC) dismisses a referral due to COVID-19 and extracorporeal membrane oxygenation (ECMO) treatment results: “The computer says stage D, well yeah, I can see why it would of say that, the patient has been hospitalized five times. However, the patient has been hospitalized because of COVID, or they were on ECMO, but you know what? They’re recovered.”

In the clinical context, when errors like these are a matter of patient safety, a common approach is to present a model’s output to a clinical expert who will make a final determination. However, in case study 1, we evaluated mental health providers’ ability to identify errors in predictive algorithms. Results showed that identifying algorithmic errors can be a complex process, even for healthcare providers with years of experience. In this study, presenting clinicians with incorrect antidepressant treatment recommendations did significantly change the accuracy of clinicians’ treatment decisions, where accuracy was defined as “concordance with expert psychopharmacologist consensus” (Jacobs et al., 2021b).

Thus, final decision-making requires interpreting the algorithmic output, understanding the reasoning behind the output, and then deciding whether to agree or disagree with the algorithm’s determination. Therefore, practicing methods of building trust and transparency must be prioritized to avoid increasing errors in healthcare decision-making (Gretton, 2018). Yet, identifying errors may be more difficult for healthcare providers than making decisions independently, without influence from algorithm output. Similar studies have shown people’s difficulty in identifying algorithmic errors (Green and Chen, 2019; De-Arteaga et al., 2020). Little recognition has been given to the challenging task placed on clinicians to overcome the technical limitations of medical-AI, and we see a need for more work considering ways to reduce this burden while optimizing for patient safety.

## 2.3. Translating algorithmic output to patient care decisions

Often, clinical AI output does not directly translate into healthcare decisions. For example, an AI tool that groups medical similarities for pathological image analysis sporadically returns results lacking medical insight (Cai et al., 2019). In our case studies, clinicians described the complex work of translating predictions to appropriate next steps, as well as the desire for more support in this translation work.

In case study 2, primary care providers (PCPs) participated in co-design activities to design an interactive tool that used ML to support antidepressant treatment selection decisions. Clinicians discussed the desire for interoperability between the proposed tool, which identified patients at risk of dropping out of treatment, and current patient follow-up protocols. Clinicians demonstrated differing abilities to connect predictions to actionable next steps (Jacobs et al., 2021a). For example, when the tool identified a patient at high dropout risk, some clinicians were unsure how to respond appropriately. In contrast, other clinicians stated that this should initiate the involvement of a broader care team who could follow up with the patient frequently. The obscure link between predictions and clinical next steps signals a need for this translation work to be formalized, documented, and shared.

In case study 3, a successful manual workflow has been created. First, a nurse reviews the model’s recommendations, a report which lists heart failure patients who may have progressed to advanced stages. The nurse decides if a patient on the report needs to be seen by the heart failure clinic and initiates a manual process to contact the patient. This involves looking for the patient’s current provider, contacting the provider directly to notify them that the patient with heart failure is high risk and may benefit from a specialist evaluation, and then following up with the patient to schedule an appointment if the provider places the referral. The nurses shared that while a more automated procedure was considered, it could not be implemented due to legal and compliance issues. Such roadblocks are common in the regulated and high-stakes environment of medicine and may attribute to adoption issues due to limitations of seamless integration, regardless of technical feasibility.

These case studies show a need for medical-AI tools to integrate into existing clinical workflows seamlessly. Few examples exist outlining this translation, but some practices are highlighting the importance of workflow integration (Sendak et al., 2020). While clinical team members may be the best at translating model output to care decisions, especially in light of potential algorithmic errors, we identify an opportunity to share these strategies and decisions across team members and clinical settings to establish best practices.

## 2.4. Developing awareness of AI use within a clinical setting

How a system is used in a more extensive sociotechnical process is key to its adoption and overall evaluation (Pratt et al., 2004). This includes understanding how model outputs are intended to support team goals and how individuals interpret results for clinical decisions. Ehsan et al. coins this sense of awareness Social Transparency, where all system users can gain insight from decisions influenced by the model, for example, with a tool that captures contextual detail of a decision-making process (Ehsan et al., 2021).

The importance of social transparency and awareness of a tools' use within a broader team came up in multiple case studies. In case study 2, PCPs stated that understanding if and how their colleagues use new technology is a notable factor in determining whether they would trust and use a tool. Several clinicians shared that they are more comfortable adopting new tools if they hear their colleagues say it is useful. In case study 3, social transparency was implemented through the NC, which is described as "key to how [the clinic has] been able to impact and reach the patients." The NC directs other clinicians on sorting criteria and priorities of the model output, which is informed by weekly meetings with physician leads.

Hence, establishing awareness is essential for the adoption of new tools and for helping the team understand how the technology operates in a collaborative environment. Promoting awareness of the development of AI-tools in medicine can prevent clinicians from being isolated in deciphering the validation of the algorithm. An absence of awareness hinders team collaboration and, by extension, prevents the delivery of optimum patient care. Our case studies demonstrate a desire for greater awareness of how medical-AI tools are used and how such awareness may be created through the additional efforts of nurses or other teammates.

## 3. Discussion

This paper characterizes invisible labor requirements placed on clinical teams to make medical-AI tools function in the realworld. The types of labor we identified in our work include: 1) data curation, labeling, and validation prior to clinical data being used to train an algorithm, 2) identifying errors made by the algorithm, 3) translating AI output to clinically meaningful next steps, and 4) developing team awareness of how a tool is being used once deployed. Characterizing the labor placed on members of clinical care teams, such as nurses, nurse coordinators, nurse clinicians, nurse practitioners, physicians, physician assistants, and technicians, is a vital component of laying the foundation of usable and ethical AI for medicine.

From a usability standpoint, CSCW research provides decades-long grounding in the complexities of sociotechnical

workflows and how technology enables broader collaborative systems (Heath and Luff, 1996; Symon et al., 1996). Medicine is at the forefront of merging human and technological expertise, although new solutions must integrate within well-established, highly collaborative workflows. Evaluating how a particular tool fits within a clinical workflow, and the effects of its operation, are necessary to claim a tool's utility for implementation. The ultimate goal for all is continued improvement in patient care. If a tool's implementation requires a high burden or impacts existing productivity due to workflow turbulence, it will not gain adoption in the clinic.

In reflecting on the labor costs of implementing medical-AI tools, we observe a clear need for more, and easier, **documentation of labor, workflows, and team structures**. One useful area for future work is in creating platforms that encourage multidisciplinary teams to capture detailed design and implementation decisions about clinical-AI tools. Comprehensive documentation can inform scaling lessons learned to similar technologies or contexts. Whether this documentation is of successful or failed implementations, its existence can steer future undertakings in the appropriate direction. In the words of case study 3's participant: patterns or similar aspects could reveal themselves, preventing the need to start over.

We have seen documenting implementation successes and challenges used to support rapid progress, such as in COVID clinics (Gilbert et al., 2020). Similarly, this approach can scale the development and availability of clinical AI by reducing replicated efforts when similar systems must be embedded into medical settings. Further, building on methods of transparency and explainability, providing user-friendly and consistent ways to document decisions and strategies throughout the development pipeline can help develop a responsible AI clinical practice.

The practice of ethical and responsible AI in medicine is rapidly growing (Gretton, 2018; Wiens et al., 2019; McCradden et al., 2022). Documenting the workflows used at each phase of design and development is critical in identifying invisible labor. Transparency is required to ensure that those who assume these workloads are not undervalued, as can often occur with invisible labor (Gray and Suri, 2019). Frequently, invisible labor can lead to inequities, and ignoring human behavior devalues those who make the technology work. This culminates in being detrimental to both system implementation and society as a whole.

## 4. Conclusion

The often-invisible clinical work throughout the development process of clinical-AI tools is a vital factor in the technology's success in the real world. Yet, thus far, most research has centered around the technical challenges rather than the significant effort required by clinical teams to make

these tools work. By reviewing three case studies from our research and related literature, we describe several types of clinical labor used to support AI development and deployments in healthcare settings. We call for greater attention and documentation of these workloads and how clinical-AI tools are changing healthcare workflows. Recognition of this labor as part of the development process is necessary for successful, ethical deployments of medical-AI. Further, by standardizing the documentation and sharing of sociotechnical implementation strategies (e.g., workflow changes, team structures), we may help to accelerate the successful creation of medical-AI tools with a real-world impact on patient care.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

MJ, MU, and BR performed data analysis on case studies presented and iterated over earlier versions of the manuscript. All authors contributed to conception and design of the study, as well as manuscript revision, read, and approved the submitted version.

## Funding

FA was supported by grants the National Institutes of Health/National Heart, Lung, and Blood Institute

## References

- Anderson, J. G., and Aydin, C. E. (1997). Evaluating the impact of health care information systems. *Int. J. Technol. Assess Health Care* 13, 380–393. doi: 10.1017/S0266462300010436
- Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., et al. (2020). “A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI), 1–12.
- Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G. S., Stumpe, M. C., and others (2019). “Human-centered tools for coping with imperfect algorithms during medical decision-making,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–14.
- Cheema, B., Mutharasan, R. K., Sharma, A., Jacobs, M., Powers, K., Lehrer, S., et al. (2022). Augmented intelligence to identify patients with advanced heart failure in an integrated health system. *JACC Adv.* 1, 1–11. doi: 10.1016/j.jacadv.2022.100123
- Coiera, E. (2019). The last mile: where artificial intelligence meets reality. *J. Med. Internet Res.* 21, e16323. doi: 10.2196/16323
- Daniels, A. K. (1987). Invisible work. *Soc. Probl.* 34, 403–415. doi: 10.2307/800538
- De-Arteaga, M., Fogliato, R., and Chouldechova, A. (2020). “A case for humans-in-the-loop: decisions in the presence of erroneous algorithmic scores,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI: ACM), 1–12.
- Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., and Weisz, J. D. (2021). “Expanding explainability: towards social transparency in ai systems,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama: ACM), 1–19.
- Emanuel, E. J., and Wachter, R. M. (2019). Artificial intelligence in health care: will the value match the hype? *JAMA* 321, 2281–2282. doi: 10.1001/jama.2019.4914
- Gilbert, A. W., Billany, J. C. T., Adam, R., Martin, L., Tobin, R., Bagdai, S., et al. (2020). Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. *BMJ Open Quality* 9, e000985. doi: 10.1136/bmjopen-2020-000985
- Gray, M. L., and Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*. Boston, MA: Houghton Mifflin Harcourt; OCLC: 1052904468.
- Green, B., and Chen, Y. (2019). “Disparate Interactions: an algorithm-in-the-loop analysis of fairness in risk assessments,” in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA: ACM), 90–99.

(K23HL155970) and the American Heart Association (AHA number 856917).

## Conflict of interest

Author FA has received consulting fees from Teladoc Livongo and Pfizer unrelated to this manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Author disclaimer

The statements presented in this work are solely the responsibility of the author(s) and do not necessarily represent the official views of the National Institutes of Health or the American Heart Association.

- Gretton, C. (2018). "Trust and transparency in machine learning-based clinical decision support," in *Human and Machine Learning* (Springer), 279–292.
- Heath, C., and Luff, P. (1996). "Documents and professional practice: "bad" organisational reasons for "good" clinical records," in *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (Boston, MA: ACM), 354–363.
- Jacobs, M., He, J., F., Pradier, M., Lam, B., Ahn, A. C., et al. (2021a). "Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama: ACM), 1–14.
- Jacobs, M., Pradier, M. F., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., and Gajos, K. Z. (2021b). How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Transl. Psychiatry* 11, 108. doi: 10.1038/s41398-021-01224-x
- Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., et al. (2021). Precision medicine, AI, and the future of personalized health care. *Clin. Transl. Sci.* 14, 86–93. doi: 10.1111/cts.12884
- Khairat, S., Marc, D., Crosby, W., and Al Sanousi, A. (2018). Reasons for physicians not adopting clinical decision support systems: critical analysis. *JMIR Med. Inform.* 6, e24. doi: 10.2196/medinform.8912
- Kushniruk, A. (2002). Evaluation in the design of health information systems: application of approaches emerging from usability engineering. *Comput. Biol. Med.* 32, 141–149. doi: 10.1016/S0010-4825(02)00011-2
- Matthiesen, S., Diederichsen, S. Z., Hansen, M. K. H., Villumsen, C., Lassen, M. C. H., Jacobsen, P. K., et al. (2021). Clinician preimplementation perspectives of a decision-support tool for the prediction of cardiac arrhythmia based on machine learning: near-live feasibility and qualitative study. *JMIR Hum. Factors* 8, e26964. doi: 10.2196/26964
- McCadden, M. D., Anderson, J. A., A., Stephenson, E., Drysdale, E., Erdman, L., et al. (2022). A research ethics framework for the clinical translation of healthcare machine learning. *Am. J. Bioethics* 22, 8–22. doi: 10.1080/15265161.2021.2013977
- Middleton, B., Sittig, D. F., and Wright, A. (2016). Clinical decision support: a 25 year retrospective and a 25 year vision. *Yearb Med. Inform.* 25, S103–S116. doi: 10.15265/IYS-2016-s034
- Nascimento, E. d. S., Ahmed, I., Oliveira, E., Palheta, M. P., Steinmacher, I., et al. (2019). "Understanding development process of machine learning systems: challenges and solutions," in *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (Porto de Galinhas: ACM; IEEE), 1–6.
- Pratt, W., Reddy, M. C., McDonald, D. W., Tarczy-Hornoch, P., and Gennari, J. H. (2004). Incorporating ideas from computer-supported cooperative work. *J. Biomed. Inform.* 37, 128–137. doi: 10.1016/j.jbi.2004.04.001
- Rajkomar, A., Dean, J., and Kohane, I. (2019). Machine learning in medicine. *N. Eng. J. Med.* 380, 1347–1358. doi: 10.1056/NEJMr1814259
- Sendak, M., Gao, M., Nichols, M., Lin, A., and Balu, S. (2019). Machine learning in health care: a critical appraisal of challenges and opportunities. *EGEMs* 7. 1. doi: 10.5334/egems.287
- Sendak, M. P., Ratliff, W., Sarro, D., Alderton, E., Futoma, J., Gao, M., et al. (2020). Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med. Inform.* 8, e15182. doi: 10.2196/15182
- Star, S. L., and Strauss, A. (1999). Layers of silence, arenas of voice: the ecology of visible and invisible work. *Comput. Support. Cooperat. Work* 8, 9–30. doi: 10.1023/A:1008651105359
- Suchman, L. (1995). Making work visible. *Commun. ACM* 38, 56–64. doi: 10.1145/223248.223263
- Suresh, H., and Guttig, J. (2021). "A framework for understanding sources of harm throughout the machine learning life cycle," in *Equity and Access in Algorithms, Mechanisms, and Optimization* (New York, NY) 1–9.
- Symon, G., Long, K., and Ellis, J. (1996). The coordination of work activities: cooperation and conflict in a hospital context. *Comput. Support. Cooperat. Work* 5, 1–31. doi: 10.1007/BF00141934
- Tonekaboni, S., Joshi, S., McCadden, M. D., and Goldenberg, A. (2019). "What clinicians want: contextualizing explainable machine learning for clinical end use," in *Machine Learning for Healthcare Conference*, 359–380. Available online at: <https://proceedings.mlr.press/v106/tonekaboni19a.html>
- Wang, J., Li, S., Song, W., Qin, H., Zhang, B., and Hao, A. (2018). "Learning from weakly-labeled clinical data for automatic thyroid nodule classification in ultrasound images," in *2018 25th IEEE International Conference on Image Processing (ICIP)* (Athens: IEEE), 3114–3118.
- Wang, S., and Summers, R. M. (2012). Machine learning and radiology. *Med. Image Anal.* 16, 933–951. doi: 10.1016/j.media.2012.02.005
- Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., et al. (2019). Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med.* 25, 1337–1340. doi: 10.1038/s41591-019-0548-6
- Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., and Zou, J. (2021). How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat. Med.* 27, 582–584. doi: 10.1038/s41591-021-01312-x
- Yang, Q., Steinfeld, A., and Zimmerman, J. (2019). "Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow: ACM), 1–11.
- Yu, K.-H., and Kohane, I. S. (2019). Framing the challenges of artificial intelligence in medicine. *BMJ Quality Safety* 28, 238–241. doi: 10.1136/bmjqs-2018-008551