AAAI-HUMAN.AI Symposium Nov 4-6, 2021: Preprint.

# Technical Feasibility, Financial Viability, and Clinician Acceptance: On the many challenges to AI in Clinical Practice

**Nur Yildirim,[1] John Zimmerman, [1] Sarah Preum [2]**
[1] Human-Computer Interaction Institute, Carnegie Mellon University
[2] Computer Science Department, Dartmouth College
yildirim@cmu.edu, johnz@cs.cmu.edu, sarah.masud.preum@dartmouth.edu

## Abstract

Artificial intelligence (AI) applications in healthcare offer the promise of improved decision making for clinicians, and better healthcare outcomes for patients. While technical AI advances in healthcare showcase impressive performances in lab settings, they seem to fail when moving to clinical practice. In this position paper, we reflect on our experiences of designing for AI acceptance and discuss three interrelated challenges to AI in clinical practice: technical feasibility, financial viability, and clinician acceptance. We discuss each challenge and their implications for future research in clinical AI. We encourage the research community to take on these lenses in collaboratively tackling the challenges of moving AI systems into real-world healthcare applications.

## Introduction

Over the last decade, there has been lots of excitement about what AI might do for healthcare. AI offers the promise of improved cancer diagnosis, faster discovery of new drugs, and even personalization of patients' healthcare experiences. The transition to electronic health records has produced a wealth of data ripe for mining. Interestingly, AI systems that work great in computer labs largely fail when they move to clinical practice, and the number one reason they fail is the lack of adoption by clinicians (Yang, Zimmerman, and Steinfeld 2015).

Our team has been investigating how AI might function more effectively in the enterprise: How AI systems might help professionals both make better decisions and also feel like they are becoming better at their jobs. We have done work in education, business, and healthcare. Currently, we are working on a project to identify how AI might improve ICU care. Can it automate mundane tasks? Can it discover low-value care? Can it detect deviations between standards of care and actual practice?

For this position paper, we reflect on what we have learned about AI acceptance in the workplace across many domains, and we tailor our insights to aspects most relevant to clinical practice. In our experience, AI only flourishes when it is technically feasible, financially viable, and acceptable or even desired by end-users. Research on clinical AI has largely focused on some aspects of technical feasibility. Researchers have made many, stunning technical advances that confirm this is a great space for innovation. Little to no work has investigated how AI systems might pay for themselves within the complex landscape of healthcare reimbursement, and little work has explored when, where, and in what form AI inferences might be viewed as valuable by clinicians. Below we touch on each of these three areas that we feel must be addressed for AI to thrive in the clinic.



## Technical Feasibility

While there have been great technical advances around AI in healthcare, much of the work is not clinically relevant (Seneviratne, Shah, and Chu 2020); research has largely reproduced human decision making, and researchers have tended to focus on more difficult problems than searching for low-hanging fruit. When we use the term "clinically relevant" and apply it to AI innovation, we are talking about research that offers empirical evidence that clinicians want the AI output researchers are developing. Drawing from our work on the ICU, researchers have created systems that predict medication (Suresh et al. 2017), predict a patient will need a ventilator (Suresh et al. 2017), predict if a patient will die (Song et al. 2018) or be discharged (Zhang et al. 2020), and predict the onset of conditions like sepsis (Nemati et al. 2018), tachycardia (Liu et al. 2021), or hypotension (Yoon et al. 2020). Researchers detail the importance of this information to a patient's health, but they do not provide evidence that clinicians currently find the use of their own expertise to make these predictions challenging, that clinicians make a high rate of errors, or that clinicians have expressed an explicit desire to know this prediction.

Technology follows a familiar adoption process. First, there is a need for technical capabilities. Once capabilities exist, there is a need to map these capabilities to situations that might benefit from this capability. Finally, there is a need for evidence that applying the new capability actually created a desirable benefit. Almost all AI innovation in healthcare is working on new capabilities. That is an important first step. But now, we need more work showing the capabilities actually map to authentic needs. When we have a body of work showing AI can address real needs, the next step will be deployment studies showing the new technology has real-world impact, that it improves health outcomes and lowers the cost of delivering care.

One driver of the disconnect between AI capability and clinical need is the fact that most AI innovations in health focus on reproducing the work of human decision-makers. These are often approaching problems from a machine learning perspective because the diagnosis (human decision) and selected treatment (human decision) are well documented. The labeled data exists for training a system. Most healthcare decisions like these are "textbook", they are obvious to a clinical expert. And an AI system trained on this data will work the best for textbook cases. This is not where clinicians need help, unless the intent is to automate clinicians out of existence. Clinicians need the most help with the unusual cases, cases with high clinician uncertainty (Yang et al. 2016). Clinicians also need help with the social aspects of their work, with getting a team to work better together in order to benefit from the collective intelligence.

In the ICU, patients on a ventilator receive input from the Interventionist, the ICU nurse, and the Respiratory Therapists (RT). The RTs will perform breathing tests on ventilated patients in the early morning. When the Interventionist arrives, they use the results of this test to decide if a patient should be extubated. But RTs and Interventionists do not always agree on who should get a breathing test, leading to situations where the doctor wants the results but the test has not been conducted. An AI system that tries to predict expert disagreement could raise this issue ahead of time, allowing the experts to make a decision before the window for decision making has closed. Situations like expert disagreement are only indirectly captured in EHRs, but they show real moments where clinicians would benefit from a machine prediction.

In our experience, AI researchers working in healthcare are most interested in working on difficult challenges. This helps researchers publish, as they can offer clear evidence that they have advanced the state of the art for AI and machine learning. However, by doing this, they most often overlook the low-hanging fruit, situations where a little bit of well-known AI might actually help accelerate or enhance clinical practice. In our ICU work, we noticed that clinicians must frequently input orders for new medication. This is not a difficult task, but it is a tedious task. As they type, they see a list of possible medications and doses they might be looking for, shown as an alphabetical list. Sometimes this helps. In examining this mundane, tedious task, we noticed that a list of medications ordered by frequency as opposed to alphabetically significantly reduced the task completion time by more than 50%. This is not a "sexy" innovation nor even a use of AI. But it does help to illustrate how the mundane labor of interaction with IT systems is largely ignored by data scientists and AI researchers.

## Financial Viability

In our work on AI innovation in the enterprise, we have observed that the biggest barrier for getting a new AI capability off of the whiteboard and onto a product roadmap is a strong business case. Software product managers want to know that the value of the innovation will be much higher than the development costs and the operational costs for the innovation. This can be challenging due to the near-monopoly held by the tiny number of EHR vendors and from the software development culture that has shifted to lean-agile, with a focus on making an MVP – the minimum viable product. Technical AI healthcare research never addresses how the advance will make money, how it will pay for itself. The work does not detail the changes needed to current data pipelines. It does not talk about the increased amount of computing required. It does not specify how this will save time, allowing clinicians to treat more people in the same amount of time, and it does not detail how clinicians might be able to charge more, because the quality of the decision-making should get better. The flow of money in healthcare is complicated, from patient to insurance company to the many intersecting clinicians delivering care. Unlike with consumer goods, a better product (healthcare decision) does not directly relate to increased demand or higher price. While this challenge may seem out of scope for technical AI innovation, it still constitutes a significant barrier to AI adoption compared to other industries.

EHR are an expensive problem. Many hospitals will have 10 or more different systems that all seem designed to be inoperable with one another (Reisman 2017; Glaser 2020). Clinicians and healthcare providers are not software companies, yet they must constantly make large IT investments to make systems run and to get them to trade data with one another (He et al. 2019). Adding on an AI system in this environment is expensive in ways that are not the case in other industries. The development cost never really ends, as each time an EHR vendor offers a major upgrade, almost all of the additional code and enhancements a healthcare provider previously developed must be rewritten.

An additional financial barrier comes from the current culture of software development. With the rapid growth of agile development, software development has become much more risk averse, with development teams searching for clear evidence of value. Increasingly, teams are working toward defining and rapidly deploying an MVP that can produce clear evidence of its beneficial impact in weeks. This is fine for retail companies that might want to try out new personalization approaches, where they can deploy and run A/B studies within a few weeks to collect evidence of the positive impact for their innovation. This seems to be out of reach for almost any AI healthcare system that focuses on high-risk clinician decision-making. Healthcare is unprepared for A/B testing, and we are not suggesting this would be a good thing. Our point is that new employees coming into software development bring a mindset that is in conflict with the pace AI in healthcare will need.

## Clinician Acceptance and Desire

Unfortunately, very little research documents the details of clinician decision-making and identifies the time and place an AI inference might be experienced as most valuable or explores which forms of AI output are most useful (Cai et al. 2019; Yang et al. 2016). It does seem like increased funding for the human-AI interaction aspects of AI in healthcare is one way this might be addressed. The lack of human-centered approach in healthcare AI innovation is evident. The current standard, the assumptions made by AI researchers working on healthcare systems often indicates a lack of understanding of clinical practice workflows (Topol 2019). Many if not most research systems are built with the assumption that clinicians recognize that they need help with a decision, and that in their "free time" the clinicians will walk up and use a separate IT system to get advice on what they should do (Yang et al. 2016). The reality is that clinicians have no free time and they are unsure when a smart system might help. Their main experience with clinical decision support systems mostly involves continuous, irrelevant alerts that distract them from their work, and that provide a negative orientation to spending time on the computer (Rajkomar, Dean, and Kohane 2019).

The lack of a human-centered approach to AI innovation means that many innovation avenues are under-investigated. For example, little work has been done to apply techniques such as business process mining to healthcare. This would help reveal what the actual standard of care is, and clinicians could use this type of insight on their own behaviors to better understand and identify areas for improvement. As we mentioned previously, predicting events like expert disagreement empowers human decision-makers to reflect and consider before they have committed to a path of action. Healthcare practice has many goals; for example, rounding often mixes goals of patient care and health worker training. More human-learning focused AI systems might capture the dialog and provide feedback to an attending physician on the quality of their rounding – feedback such as waiting longer after asking a question and monitoring for implicit bias in who they ask questions of and who they compliment. Systems could also help with orchestration, the work of effectively coordinating work across the many experts. For example, in the ICU, a system might recommend an order for visiting patients during rounding based on the estimated time needed, the importance of early decision making for a patient (will they likely be extubated), and the physical layout of the rooms. There are many types of insights around human performance and processes that have been largely ignored by current technically focused research.

## Conclusion

In this position paper, we elaborated on the interrelated challenges of feasibility, viability, and acceptance for moving clinical AI into the real world. These challenges will require thinking about not only the AI capabilities that are ripe for application, but also the business of healthcare, and the needs and desires of the frontline healthcare workers. We envision a future where researchers from AI, HCI, healthcare, design, and business research communities work together to take on these challenges. HCI and design researchers can focus on how technical advances in clinical AI might match to current needs and workflows of clinicians. AI and business researchers can work on low hanging fruit – worker needs and desires design research reveals that are likely to be solved with well known AI capabilities and existing healthcare data and infrastructure. We invite clinical AI researchers to advance collaborative research practices, effectively bridging the gap between research communities.


## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. (2007501) and work supported by the National Institutes of Health (R35HL144804). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.


## References


Cai, C. J.; Winter, S.; Steiner, D.; Wilcox, L.; and Terry, M. 2019. " Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-computer Interaction* 3(CSCW): 1–24.

Glaser, J. 2020. Its time for a new kind of electronic health record. *Harvard Business Review. June* 12.

He, J.; Baxter, S. L.; Xu, J.; Xu, J.; Zhou, X.; and Zhang, K. 2019. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine* 25(1): 30–36.

Liu, X.; Liu, T.; Zhang, Z.; Kuo, P.-C.; Xu, H.; Yang, Z.; Lan, K.; Li, P.; Ouyang, Z.; Ng, Y. L.; et al. 2021. TOP-Net Prediction Model Using Bidirectional Long Short-term Memory and Medical-Grade Wearable Multisensor System for Tachycardia Onset: Algorithm Development Study. *JMIR Medical Informatics* 9(4): e18803.

Nemati, S.; Holder, A.; Razmi, F.; Stanley, M. D.; Clifford, G. D.; and Buchman, T. G. 2018. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical care medicine* 46(4): 547.

Rajkomar, A.; Dean, J.; and Kohane, I. 2019. Machine learning in medicine. *New England Journal of Medicine* 380(14): 1347–1358.

Reisman, M. 2017. EHRs: the challenge of making electronic data usable and interoperable. *Pharmacy and Therapeutics* 42(9): 572.

Seneviratne, M. G.; Shah, N. H.; and Chu, L. 2020. Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations* 6(2).

Song, H.; Rajan, D.; Thiagarajan, J. J.; and Spanias, A. 2018. Attend and diagnose: Clinical time series analysis using attention models. In *Thirty-second AAAI conference on artificial intelligence*.

Suresh, H.; Hunt, N.; Johnson, A.; Celi, L. A.; Szolovits, P.; and Ghassemi, M. 2017. Clinical intervention prediction and understanding using deep networks. *arXiv preprint arXiv:1705.08498* .

Topol, E. J. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine* 25(1): 44–56.

Yang, Q.; Zimmerman, J.; and Steinfeld, A. 2015. Review of Medical Decision Support Tools: Emerging Opportunity for Interaction Design. *IASDR 2015 Interplay Proceedings* .

Yang, Q.; Zimmerman, J.; Steinfeld, A.; Carey, L.; and Antaki, J. F. 2016. Investigating the heart pump implant decision process: opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4477–4488.

Yoon, J. H.; Jeanselme, V.; Dubrawski, A.; Hravnak, M.; Pinsky, M. R.; and Clermont, G. 2020. Prediction of hypotension events with physiologic vital sign signatures in the intensive care unit. *Critical Care* 24(1): 1–9.

Zhang, D.; Thadajarassiri, J.; Sen, C.; and Rundensteiner, E. 2020. Time-Aware Transformer-based Network for Clinical Notes Series Prediction. In *Machine Learning for Healthcare Conference*, 566–588. PMLR.