

# AI model transferability in healthcare: a sociotechnical perspective

Batia Mishan Wiesenfeld, Yin Aphinyanaphongs and Oded Nov

 Check for updates

To deliver value in healthcare, artificial intelligence and machine learning models must be integrated not only into technology platforms but also into local human and organizational ecosystems and workflows. To realize the promised benefits of applying these models at scale, a roadmap of the challenges and potential solutions to sociotechnical transferability is needed.

Predictive model transferability, traditionally defined as “the ability to produce accurate predictions among patients drawn from a different but plausibly related population”<sup>1</sup>, is receiving increasing attention as healthcare organizations attempt to implement artificial intelligence (AI)-based prediction tools<sup>2–4</sup>. Although some machine learning (ML)-based models fail when subjected to retrospective validation across institutions and patient populations<sup>5</sup>, technical improvements (e.g., foundation models) show promise for addressing this model efficacy problem. To address the engineering challenges, a technical subfield labelled MLOps has emerged, promising to address technical transferability by injecting needed discipline into the development, integration, deployment, monitoring, iteration and governance of ML models<sup>6,7</sup>. These developing solutions open the door to deploying models developed for localized applications in new contexts, thereby realizing AI’s promise of scalability.

The focus of MLOps on technical transferability may be obscuring a larger set of obstacles to sociotechnical transferability: organizational, social and individual challenges of deploying models at scale across contexts, whether institutions, teams or individual roles<sup>8</sup>. The challenge may be particularly acute in healthcare, where electronic health record systems have not standardized workflows and practices in the way that business process technology implementations standardized core processes in other industries. Variation in sociotechnical systems influences what we term model effectiveness, which is how well the model works in practice. Challenges to model effectiveness often arise when models are transferred across institutions and implementation settings. Effectiveness has received far less attention than model efficacy challenges.

The “efficacy–effectiveness gap”<sup>9</sup> refers to the fact that efficacy of drugs in clinical trials is often not replicated in real-world settings owing to differences between institutions and practices. Based on three years of multi-method research (ethnographic, interview and survey) on fully and partially implemented diagnostic and prognostic prediction models for clinical practice in a multi-hospital healthcare system, we see a similar efficacy–effectiveness gap emerging in ML-based prediction model transfer. To develop generalizable insights,

we studied models implemented across departments (e.g., radiology, medicine, paediatrics), roles (e.g., physicians, nurses) and conditions (e.g., COVID-19 adverse events, sepsis, clinical deterioration, screening mammograms for breast cancer). We unpack model transferability and review the sociotechnical challenges that transferability introduces at the intersection of institutions, providers, care teams and roles. We then offer guidance on ways to address these challenges.

## Challenges

Beyond differences between training and deployment populations, transferring ML-based models to a new institution introduces challenges to effective use at multiple levels.

### **Institutional challenges across units, organizations or systems.**

When models are transferred across institutions, a host of structural, cultural and incentive factors reshape their performance, acceptance and use.

### **Culture of innovation and/or risk.**

For most healthcare staff, ML models are a relatively new technology, so organizational cultures that are less innovation-focused may discourage model adoption<sup>10</sup>. Less well understood are the structural barriers to innovation strategies and cultures. Union contracts or legal and regulatory regimes may limit responsibility for using models to certain providers, such as physicians<sup>11</sup>. Institutions whose providers are 100% clinical or have high levels of alert fatigue may have lower bandwidth for experimenting with new technologies, and the availability of professional training for clinical staff on AI models may vary by institution<sup>11</sup>.

### **Institutional model owners.**

Research demonstrates that model trust and use in single, local applications depends upon users’ involvement in model development and on the social capital of clinical champions who legitimize the model<sup>10,12</sup>. However, social capital is often local and does not transfer across institutions, and it degrades rapidly when clinical champions leave an organization, endangering trust in models. Dedicated predictive analytics units can develop, test, win support for and implement models, but such units require sufficient scale, making them impractical for small hospitals. Large hospital networks may be siloed, reducing the likelihood that innovations developed centrally will be adopted throughout the system.

### **Costs vs. benefits.**

Leadership buy-in for ML-based models’ adoption depends on the degree to which models align with the organization’s strategy, business model and care delivery pathways<sup>11</sup>. Healthcare institutions are under pressure to maintain high-quality healthcare while lowering costs. Reliable evidence of cost savings from model use is often equivocal, unavailable or not generalizable. Regulation and reimbursement may also vary across healthcare systems. For example, in the United Kingdom, breast sonograms are read by two radiologists,

while in the United States they are commonly read by one. Therefore, a model may generate cost savings by substituting for a human 'second opinion' in the United Kingdom but not in the United States.

**Use cases.** As models are transferred across institutions, their use may change. For example, a tool that identifies low-risk patients from 3D mammograms may be used by radiologists screening for breast cancer in one institution, but another institution may use it for triage or to order the queue of images radiologists review. Commercial tools approved by the US Food and Drug Administration may be used differently from 'home-grown' tools because of their cost and reimbursement implications. The roles, responsibilities, vulnerabilities and requirements of these models will differ across such use cases.

**Regulatory explainability/interpretability demands.** Different legal jurisdictions place different demands on the explainability of models. This raises challenges in transferring ML-based black-box models to institutions in locations in which explainability demands are stricter.

**Knowledge sharing.** Electronic health record vendors hold conferences in which providers and data scientists share their experiences with locally developed models (e.g., Epic's XGM and UGM). The reporting is not standardized but rather delivered as free-form anecdotes about hospital practices, a highly inefficient communication pathway not addressing cross-institution generalizability of structures, workflows and practices.

**Healthcare teams, composition and design.** Zooming in from institutions to the healthcare teams, model transferability involves additional challenges. In most instances, a team of providers working together is responsible for patient care. One of the poorly understood values of predictive models is that they serve as a possible mechanism for team coordination and information-sharing across roles<sup>13</sup>. The fact that models relate to multiple interdependent team roles in different ways can hinder transferability, particularly when the teams adopting the model are configured differently from the original teams. For example, we developed a Covid Adverse Event model and assigned its 'ownership' to a team of clinical alert nurses, who regularly round on patient units and monitor at-risk patients. The clinical alert team facilitates coordination between bedside nurses, attending physicians, respiratory therapists and other roles. The team used the model to prioritize patients, help identify those who could benefit from early transfer to the intensive care unit and work with nurses to develop proactive care strategies. Transferring this model into a setting without a clinical alert team requires remapping workflows in the recipient teams and assigning the responsibilities associated with interpreting, monitoring and acting upon the model to providers in other roles.

**Team outcome focus.** When a model developed for individual decision-makers is transferred to the team level, the outcomes the model predicts may change. Many models in healthcare predict interventions (e.g., resuscitation in the case of sepsis) rather than downstream health outcomes (see refs.<sup>14,15</sup>). If a model predicting intervention is transferred to an organization that has different intervention practices, it may fail to be validated. In such cases, it is essential to determine whether the problem is the model or workflows of the recipient team. In an extreme example, transferring a model customized for an institution with low rates of sepsis, reflecting effective workflows for

recognition and intervention, to another institution that does not have these workflows may not realize the same outcome.

**Individual providers.** ML models offer advice that individual decision-makers act upon, making it essential to consider differences in individual providers and their model use.

**Expertise and specialization.** Providers in the same occupation vary in expertise and specialization. These factors shape how they think. For example, an inexperienced resident may have a less nuanced mental model of a disease. Prediction tools may help such clinicians create a mental model by drawing attention to the most important features and helping them translate these features into outcome probabilities. An experienced provider may benefit from models that identify cases deviating from the norm to reduce their superstitious learning<sup>16</sup> from experience, and enable them to update prior routines. Also, experts may be annoyed by higher frequency of model advice, while novices may appreciate it. Provider specialization may influence the value of model advice. For example, a breast radiologist may gain less value from model advice on breast sonogram images than a general radiologist who reads a variety of different scans and has a less differentiated mental model of any one scan type.

**Advice taking.** Different occupations use the same advice to address different needs. For example, while physicians are trained to make predictions, nurses are generally trained to respond to patient conditions. Nurses may have to shift that orientation to understand and leverage predictive advice. Unit leads may be concerned about collective outcomes, such as average length of stay, and clinical alert teams seek patients to monitor more closely. Consequently, transferred models may be less useful in occupational contexts that differ from the ones for which they were developed.

## Moving forward

Overcoming the model transferability challenges requires fully integrating the consideration of contextual differences into MLOps at all stages of model design, implementation and use. Although technology can be efficiently standardized, enabling interoperability, the complexity and diversity of sociotechnical systems requires a more modular and flexible approach. This may be achieved by expanding and standardizing the content of what is transferred, thus enabling translation and flexibility in the model transfer process. A metric of success would be providing decision-makers and users with the necessary information to infer whether a deployed model would be successful in their local environment. Most model deployment descriptions do not provide sufficient detail about model scope and limitations, implementation plans, workflow integration, roles and responsibilities, and the environment to enable potential users to effectively assess transferability. Below we discuss what could be required in terms of content and process to improve model implementation.

**Content.** To make models more modular, a number of components must be transferred that enable new users to reconcile differences across healthcare institutions, teams and individual providers. First, users must understand a model's scope and limitations. A 'model facts' label designed to facilitate model transfer across locations, use cases and contexts through greater transparency should accompany transferred models. Model facts should ideally include information about the model's intended goal or health outcome, its output, target

population, time of prediction, input data source and type, training data location and time period, and model type, as well as important implementation information such as application domain, directions and warnings (e.g., refs. <sup>17–19</sup>). In addition, fully developed and standardized implementation plans should be transferred. At a minimum, these should include training materials, notification pathways and systems for measurement and analysis of performance. The roles and responsibilities of the providers associated with the model's use and outputs must be communicated to practitioners in a generalizable way, referring to work functions rather than institution-specific positions. For example, when transferring a model developed in a care setting in which clinical alert nurse teams actively prioritize at-risk patients to a setting that does not have clinical alert teams, a new workflow must be designed to handle alerts created by the model and new model 'owners' designated. In addition, details of the environment are crucial. For example, the sepsis rate of the hospital where a model was deployed should be reported. A sepsis prediction model successfully deployed at an institution with a starting high rate of sepsis may not transfer to an institution whose sepsis rates are already low. Conversely, holding other factors constant, low prior rates of sepsis at an institution originating a model may hold the promise of transferring sociotechnical best practices along with the model if the originating institution's implementation plan is followed. Likewise, clarifying the specific actions to be taken by team members in reaction to model predictions and recommendations increases the likelihood that desirable model outcomes will be replicated in new contexts.

To support business model decision-making, evidence of the model's financial implications should accompany evidence of its health-care outcomes, including the costs of implementation, long-term cost savings and any revenue-generating opportunities.

**Process.** Model transfer requires a sociotechnical model localization process to enable buy-in, which requires flexibility. Designing the workflows associated with model implementation should involve parallel iteration between the design of the tool's output (e.g., when and what alerts are triggered given a certain prediction) and what a provider needs to do given a certain alert. Model localization is ideally led by, or heavily involves, clinical champions who invest their social capital in the implementation and build trust. Lining up strategic champions and financial champions may also be valuable.

To support flexibility, personalization should be done for the intended user by the organization medical informatics team in consultation with clinical leadership. It should take into account role-specific information, including the desirable level of explainability and interpretability. Customization should be done by the user, based on their preferences for the type, timing, location and level of detail of the information presented to them. Where dedicated analytics or informatics units exist, customization may also be possible for teams and institutions.

Finally, there is an important role for communities of practice to empower and support the emergence of local champions through

intra- and inter-organizational social and information networks. Community norms can standardize the sharing of explicit knowledge but should also facilitate personal outreach for sharing tacit information, making sociotechnical model transfer more common, and likely more successful.

**Batia Mishan Wiesenfeld** <sup>1</sup>✉, **Yin Aphinyanaphongs** <sup>2</sup> and **Oded Nov** <sup>2,3</sup>

<sup>1</sup>Department of Management and Organizations, NYU Stern School of Business, New York, NY, USA. <sup>2</sup>Department of Population Health, NYU Grossman School of Medicine and Langone Health, New York, NY, USA. <sup>3</sup>Department of Technology Management, NYU Tandon School of Engineering, Brooklyn, NY, USA.

✉ e-mail: [bmw1@stern.nyu.edu](mailto:bmw1@stern.nyu.edu)

Published online: 17 October 2022

## References

1. Justice, A. C., Covinsky, K. E. & Berlin, J. A. *Ann. Intern. Med.* **130**, 515–524 (1999).
2. Song, X. et al. *Nat. Commun.* **11**, 5668 (2020).
3. Pajouheshnia, R., Van Smeden, M., Peelen, L. & Groenwold, R. J. *Clin. Epidemiol.* **105**, 136–141 (2019).
4. Prosperi, M., the Dataiku Team. *Nat. Mach. Intell.* **2**, 369–375 (2020).
5. Panch, T., Mattie, H. & Celi, L. A. *NPJ Digit. Med.* **2**, 77 (2019).
6. Treveil, M. et al. *Introducing MLOps: How to Scale Machine Learning in the Enterprise* (O'Reilly Media, 2020).
7. Tamburri, D. A. in *Proceedings of the 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASCO 2020)* 17–23 (IEEE, 2020)).
8. Jacobs, M. et al. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* 659 (IEEE, 2021).
9. Nordon, C. et al. *Value Health* **19**, 75–81 (2016).
10. Strohm, L. G. D. *The Augmented Radiologist—Challenges and Opportunities for Widescale Implementation of AI-based Applications in Dutch Radiology Departments*. Master's thesis, Utrecht Univ. (2019).
11. Petersson, L. et al. *BMC Health Serv. Res.* **22**, 850 (2022).
12. Winter, P. & Carusi, A. *Sci. Technol. Stud.* <https://doi.org/10.23987/sts.102198> (2022).
13. Bechky, B. *Am. J. Sociol.* **109**, 720–752 (2003).
14. Agniel, D., Kohane, I. S. & Weber, G. M. *J. Br. Med. J.* **361**, k1479 (2018).
15. Beaulieu-Jones, B. K. et al. *NPJ Digit. Med.* **4**, 62 (2021).
16. Levitt, B. & March, J. *Annu. Rev. Sociol.* **14**, 319–338 (1988).
17. Mitchell, M. et al. in *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* '19)* 220–229 (Association for Computing Machinery, 2019).
18. Sendak, M. P., Gao, M., Brajer, N. & Balu, S. *NPJ Digit. Med.* **3**, 41 (2020).
19. Stoyanovich, J. & Howe, B. *IEEE Data Eng. Bull.* **42**, 13–23 (2019).

## Acknowledgements

The authors receive financial support from the US National Science Foundation (awards no. 1928614 and 2129076) for the submitted work. The funding source had no further role in this study.

## Author contributions

B.M.W.: conceptualization, writing—original draft; Y.A.: writing—review and editing; O.N.: conceptualization, writing—review and editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Peer review information** *Nature Machine Intelligence* thanks Parisa Rashidi, Peter Winter, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.