How to Smoothen AI Implementation in Healthcare

Three domain-focused prescriptions.

by Adrian Yeow and Foong Pin Sym

In the field of precision medicine, where physicians aim to tailor medical treatments specific to individual patients, Artificial Intelligence (AI) tools are being used to augment complex medical decision-making. For example, in pharmacogenomics, the branch of medicine that studies how genetics affects medical treatment, the unique genetic profile of patients is used to determine whether they have genes that are clinically relevant in certain drug metabolisms.[1] To apply such therapy requires an understanding of the science that determines the epigenetic profile, as well as the data modelling that determines the adverse drug reaction data. Such convergence of advanced data modelling and medicine is a key feature of recent Health Information Technology (HIT) endeavours.

However, as training in genetics and data modelling among physicians is uneven, the development, testing, integration, and implementation of medical AI tools create many challenges. Such HIT challenges are also emerging in other sectors of healthcare. In this article, we describe these challenges and draw on current research, as well as our collective experience in developing AI-enhanced decision aid tools for HIT, to offer pointers on how to manage the challenges of implementation.

AI IN HEALTHCARE

Healthcare organisations have experimented with embedding AI tools in various diagnostics, administrative, and therapeutics tasks, e.g., medical imaging, clinical diagnosis, clinical skills benchmarking, and pharmacogenomics.[2] Healthcare AI tools are not homogeneous; rather, they encompass a broad range of technologies including biometrics, cognitive robotics, robotic process automation (RPA), machine learning (ML), natural language processing (NLP), and speech recognition. These technologies differ in terms of the specific technical platforms used, programming technologies, and their ability to learn.

While the underlying technologies for healthcare AI tools may be different, and how they support healthcare tasks can vary based on how they are integrated with the clinical work, the core of each AI tool is similar in that they are made up of two interrelated components–the AI model and data. For example, in medical imaging, the data would involve functional magnetic resonance imaging (fMRI) data. The AI model is an aggregation of this source data for identifying abnormal findings. In cognitive robotics or RPA, the model would be the sequence of tasks based on relationships between specific data inputs and specified output task data. Using this conceptualisation of the healthcare AI core structure, we discuss its development and implications for AI implementation.

How the AI model is developed

First, the AI team, comprising clinicians and data scientists, must define the application of the AI model. This involves determining the scope of the clinical tasks, analysis, and decisions.

Second, after defining the application scope, the AI team would need to simultaneously develop the AI model (a form of algorithmic classification) and acquire the necessary data for the AI model. The AI model can be based on human classification and domain expertise–this involves determining the potential predictors or data that relate to the clinical problem and 'ground truth labelling', i.e., assigning human-sourced labels to the algorithmic outcomes that reflect the correct outputs.[3] For example, the AI model could include image recognition and classification models using ML for detecting breast cancer, brain tumours, or diabetic retinopathy. Ground truth labelling would involve labelling outputs based on the diagnostic decisions of professional radiologists. Alternatively, an AI model could include natural language text mining and use classification models to detect signs and symptoms of sepsis, make predictions of Intensive Care Unit transfers, or forecast the likelihood of hospital readmissions. The ground truth labelling here would be labelling outputs based on diagnostic decisions of clinicians. As part of developing the AI model, the AI team usually uses part of the data as training data, and another part of the data as validation data to test for predictive accuracy of the AI model.

Finally, the AI team must evaluate the AI model's output. Part of this evaluation includes removing inaccuracies, such as those arising from spurious correlations and statistical biases. This may involve a highly iterative process of significant data cleaning and model refinement. One key aspect of this process is how the AI team measures and evaluates the quality of the AI model. Most AI teams rely on output-based accuracy metrics that measure how well an AI model's predicted outputs match the human classified outputs (ground truth labels) within the testing or validation data.

AI model development ends when the AI team can develop a model where the output-based accuracy metric is robust and above the industry standard. The industry standard is often based on agreed-upon measurements–in some cases, it is based on human experts' performance or aggregate measures espoused in the industry's analyses.

CHALLENGES IN IMPLEMENTING HEALTHCARE AI

Despite the considerable potential of AI, most healthcare AI tools are still in the development and proof-of-concept stages. Notwithstanding their high predictive accuracy, some models fail when they are used with new data; some others fail because they may not be easy to implement. These difficulties mean that there are still very few successful implementations of healthcare AI.[4] As such, healthcare AI research institutions, governments, and industry groups have released frameworks and best practice guidelines to assist healthcare organisations in the development and implementation of healthcare AI technologies.[5,6]

In Singapore, the Ministry of Health (MOH), in coordination with Integrated Health Information Systems (IHiS), a national healthtech agency, has developed best practice recommendations for the development and integration of healthcare IT systems that use AI. Their recommendations suggest that healthcare organisations need to invest in resources to test the data, validate the model with both retrospective and representative data, and ensure that the AI model works according to the ground truths. These are good practices, but we believe there are other concerns that must be addressed beyond model accuracy.

> AI model development ends when the AI team can develop a model where the output-based accuracy metric is robust and above the industry standard.

A recent review study on implementing ML products for healthcare delivery highlighted that it is also challenging for AI products to move from *in silico* settings (where the AI model is tested on retrospective data) to actual care settings (where the AI model is evaluated in different 'live' settings).[7] This could arise for multiple reasons, such as intrinsic differences in the data, the interactions between the AI model and local conditions, and other aspects of the 'live' context. The researchers behind the study also argue that the clinical integration step–where the AI model is linked to the clinical work–may be the most difficult step to execute in the entire model.

Drawing from current research, and our experiences in developing and implementing AI tools for various healthcare settings in Singapore, we highlight the following obstacles that healthcare organisations need to be aware of when embarking on this process of AI implementation.

Transparency of AI model

While the recommendations and reviews inform us of the need to ensure that appropriate data testing is done at different stages of development and track the ground truths during implementation, the reality is that the transparency of AI models remains a major obstacle hindering the implementation of those recommendations.

> A key obstacle noted in current research and from our experience deploying AI tools is the significant data-related challenges present when contextualising the tool in view of local conditions.

Specifically, it is often not clear how an AI model's ground truth labels are established in development. In a study conducted in a US hospital system,[8] the medical diagnostic AI evaluation teams were unable to access the source of ground truth labels in the AI model for some of the tools being evaluated. In other cases, the evaluation teams realised that there were significant discrepancies between the AI model's ground truths and the ones used by their local experts. By digging deeper, they found that for certain AI models, the ground truths were labelled using only current images, which were limited or narrowly defined training data, instead of the typical practice of comparing current with prior images or using messier and nuanced data. Finally, for a specific set of AI models, the teams realised that the ground truths were hard to establish in practice as they were either determined by costly professional standards or there were no agreed-upon standards for the ground truths.

Context of AI model

Another key obstacle noted in current research and from our experience deploying AI tools is the significant data-related challenges present when contextualising the tool in view of local conditions. First, the process of integrating the AI model into actual operations (or what we call 'production environment') is not trivial. It requires the coordinated efforts of the AI team, health IT infrastructure team, and clinicians to test and validate that the AI model can work in the environment. Second, significant effort may be required to ensure the AI technology is compatible with the existing IT systems, and it is able to retrieve and transform the required data. For example, the data may be stored in different parts of the IT infrastructure. The coordination costs, development, and testing efforts required are not trivial and often hard to enact, given differences among stakeholders' organisational objectives.

Apart from the data work required, the AI tool needs to be integrated with existing clinical user tasks and overall workflows. A workflow refers to a set of interlinked routine and novel tasks performed by clinicians and supporting staff as part of care delivery. This may require deliberate changes in tasks and even the workflow. For example, as part of the AI-enabled protocol, there may be a need to check the AI diagnostic scores and new procedures may be required, such as what the clinician should do when the scores are above a specified threshold. It may also require work on designing how the AI outputs are presented as part of the existing digital and physical work environment. A research study on the implementation of an AI-enabled readmission prediction model within a hospital system found that significant barriers emerge during the integration phase.[9] In that study, the researchers found that variations in the readmission risk assessment workflow across different stakeholders (e.g., case manager, pharmacists, physicians and nurses, or social workers) led to different concerns about how the AI model should be integrated.

Supporting AI tool 'explainability'

As mentioned, a defining feature of using AI to create predictive models is that the AI model itself is inscrutable. The functions used to create the models are uninterpretable, or several different algorithms are applied in such a way that they cannot be broken into its parts. This is known as the 'blackbox' of AI. Thus, an AI-enhanced decision aid may have its internal logic hidden from the user. When applied to high-stakes medical decision-making, this opacity challenges both the patient and the clinician.

For example, in an AI project for a Singapore hospital, the team of one of the authors had built a highly robust NLP-based model for sepsis prediction. However, one key validation issue was the NLP variables that were critical to the model's high level of predictive accuracy. It was challenging to explain clearly and fully how these variables drawn from clinicians' patient notes could predict whether a patient will suffer from sepsis. Furthermore, these NLP-derived variables are partly dependent on the clinician's documentation in that hospital. Without further external validation, it was unclear how well this model would perform with other hospitals' clinician notes.

Because of these challenges in the explainability of the AI tool in justifying the diagnosis, clinicians are put in a position where they take on the agency for the choice made by the AI model, without being able to grasp the conditions under which such a decision was made. This creates an undesirable situation where the authority of clinicians does not arise from examinable knowledge, but from their role as an AI tool operator. Patients are also placed in an untenable position because they are being asked to trust a decision that they cannot query, and one made by a clinician who may not have the expert skills to offer an explanation.

PRESCRIPTIONS FOR IMPLEMENTING HEALTHCARE AI

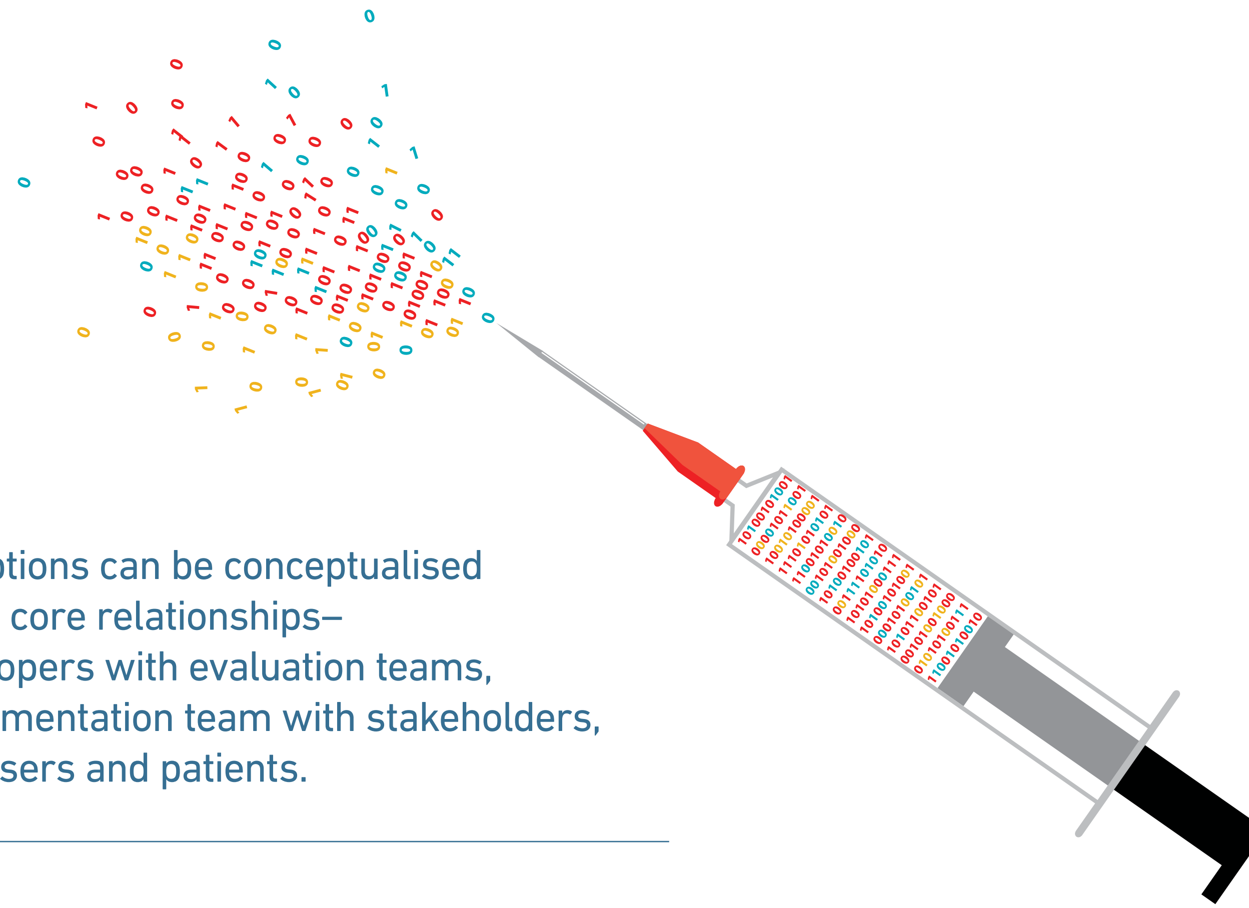
Having reviewed some of the key challenges of implementing AI in healthcare, we now explicate three prescriptions for healthcare organisations to consider as they start to implement AI tools for healthcare processes. These prescriptions can be conceptualised as three core relationships–AI developers with evaluation teams, AI implementation team with stakeholders, and AI users and patients.

1. AI developers with evaluation teams

The first set of prescriptions focuses on recontextualising the AI model within each hospital or healthcare organisation setting. As discussed, one of the biggest challenges faced by healthcare organisations is in evaluating a new AI tool for its internal use. While there are some existing guidelines provided by MOH and industry groups for recommendations concerning the understanding of how the AI tool was developed, the data used to train the model, and how to validate its predictive outputs, we argue that these should be explicitly codified as part of the AI evaluation team's work.

As such, the first prescription is to set up a cross-functional AI evaluation team, comprising clinical innovators, data scientists, and medical informatics representatives. The scope and responsibilities of this AI evaluation team is to understand and validate the AI model's performance, accuracy, and reliability. Given that there is currently no standardised model to measure the above, the team's responsibility is to understand and validate the AI model for the adopting organisation's local conditions.

The team's first task is to review the AI model's reported measures. The review would include collecting data on the AI model's reported output-based accuracy metrics, the sources used to establish the ground truth, the members who assisted in the ground truth labelling, and the data used in training and validating the model. This review should enable the team to answer questions concerning the AI model's core assumptions, variables, relationships, and the data that it was based on. In certain cases, the AI evaluation team would require information on the different ML models used or core AI technologies utilised in the AI model. While these may not be fully interpretable, it provides the team some information on the method through which the data was used to predict the outcomes.



Prescriptions can be conceptualised as three core relationships–
AI developers with evaluation teams,
AI implementation team with stakeholders,
and AI users and patients.

The team's next task would be to verify the AI model's performance with the organisation's local data. At the same time, the team should work with its local clinical experts to cross-check this version of ground truth labels for the phenomenon predicted by the AI model. After examining the AI model's performance using local data and cross validating the model's output with the local experts' version of ground truth labels, the team would be able to ascertain if the AI model can perform accurately and reliably within the current organisation.

By doing the above, the AI evaluation team would have a clearer sense as to how well the AI model works in the local context in terms of the difference, if any, between the AI model's ground truth labelling and the local expertise's ground truth, the ability of the AI model to work with local data, and its performance in local conditions versus its reported performance.

2. AI implementation team with stakeholders

The second set of prescriptions focuses on the integration of the AI tool with target departments' workflow and tasks. This integration challenge is a multi-dimensional problem as it encompasses the AI tool's integration with existing technical infrastructure, data, as well as operational and clinical tasks and workflow.[10] Given that such an integration encompasses the clinical, technical, and operational domains, the AI implementation team must be carefully set up and managed by the senior members of the healthcare organisation. As such, the AI implementation team structure would follow other established enterprise system project structures. This may include a steering committee, an AI implementation working committee, and various AI implementation project teams.

The steering committee would usually be chaired by senior clinical and/or executive leaders, and should include senior clinicians, technical experts, operational executives, as well as legal and ethical experts. It would provide the leadership, oversight, and direction for the AI implementation for the healthcare organisation. As part of its leadership role, it can help to secure resources for the implementation team, deliberate and approve budgets and plans, and get buy-in from the different stakeholders across all levels of the organisation.

The working committee would be led by the AI leads, and comprise target clinical department heads, senior clinical users, as well as the heads of the healthcare IT systems, medical informatics, technical infrastructure, clinic operations, and legal/ethics departments. It would focus on deliberating, designing, and overseeing the technical, clinical, and operational integration of the AI tool; developing appropriate process outcomes and goals to be achieved by the AI tool; as well as considering how the integration would address or mitigate privacy, ethics, and safety issues related to the implementation of AI tools. Specifically, we would expect this committee to focus on AI tool design such as the design of clinical systems to reflect specific AI inputs, the data indicators, and predictors and their thresholds. These in turn would lead to the AI tool's impact on a) changes in roles and responsibilities of clinical or non-clinical users, b) changes to coordination of tasks and handovers, and c) changes to intermediate process outputs and patient outcomes.

Finally, the approved AI-enabled workflows, protocols, and tasks would be implemented by respective AI implementation project teams. These teams would not just be responsible for the actual deployment of the AI tool, they would also be responsible for evaluating and monitoring the process metrics to validate the efficacy and effectiveness of the AI tool. The implementation team should therefore take note of user issues, data drifts, unexpected outcomes, and data risk, and bring this up to the working committee. We should expect multiple iterations and adjustments for each AI tool implementation, and these would require close coordination between the working committee and the implementation teams.

3. AI users and patients

The last set of prescriptions focuses on the AI users–clinicians, and the patients affected by such AI-enabled healthcare processes. One possible approach to resolve the issue of the AI tool's explainability is to create interpretable explanations for the prediction using explainable models. This approach assumes that uninterpretable AI models may have interpretable statistical correlates that perform similarly. Explaining the model's prediction using the non-AI models may be more trust-generating than offering no explanation. If the model remains stubbornly opaque, another strategy is to enhance its interpretability by allowing the user/clinician to query the conditions under which the model was constructed.

> An explainable AI model allows the clinician to query the data used in its training, and how well or badly it performed when the population changed.

Clinicians already use these strategies today in evidence-based medicine. They are often already aware of what assumptions were made in preparing the drug trial or how the drugs were applied only to certain sub-populations. Along the same vein, an explainable AI model allows the clinician to query the data used in its training, and how well or badly it performed when the population changed. In the same way, patient-facing decision aids for AI-enhanced tools may benefit from permitting patients to play with the input parameters to explore the response of the tool. While decidedly less scientific than statistical knowledge, the ability to 'get a feel' for the model promotes trust in the decision, and the physician who is acting as the agent responsible for wielding the AI tool.

The key takeaway here is that developing interfaces that support interpretability of an AI tool can be of benefit to the end-users–clinicians and their patients. We recommend creating interfaces that help users query the factors, the assumptions of the model, and the way the predictions change as the key factors vary. These will serve to increase trust and confidence in the shared medical decision derived from AI tools.

CONCLUSION

Even as AI tools in HIT continue to advance in exciting and incredible ways, healthcare organisations are paradoxically finding it harder to leverage and implement these newer, cutting-edge AI tools. We propose three domain-focused prescriptions, which can be used as practical springboards, as healthcare organisations embark on their implementation journey. We believe that by carefully following these prescriptions, healthcare organisations can successfully navigate known AI implementation pitfalls and challenges and be able to repeatedly implement AI tools in an effective manner.

Dr Adrian Yeow
is Associate Professor at the Singapore University of Social Sciences

Dr Foong Pin Sym
is Research Fellow and Head of Design (TeleHealth Core) at the Saw Swee Hock School of Public Health, National University of Singapore

References

[1] Ramón Cacabelos, Natalia Cacabelos, and Juan C. Carril, "The Role of Pharmacogenomics in Adverse Drug Reactions", Expert Review of Clinical Pharmacology, 12(5), 407-442, 2019.

[2] Thomas H. Davenport and Dwight McNeill, "Analytics in Healthcare and the Life Sciences: Strategies, Implementation Methods, and Best Practices", Pearson Education, 2013.

[3] Sarah Lebovitz, Natalia Levina, and Hila Lifshitz-Assaf, "Is AI Ground Truth really 'True'? The Dangers of Training and Evaluating AI Tools based on Experts' Know-what", Management Information Systems Quarterly, 45(3), 1501-1525, 2021.

[4] Mark P. Sendak, Joshua D'Arcy, Sehj Kashyap, et al., "A Path for Translation of Machine Learning Products into Healthcare Delivery", EMJ Innovations, 10, 19-172, 2020.

[5] Federico Girosi, Sean Mann, and Vishnupriya Kareddy, "Artificial Intelligence in Clinical Care", RAND Corporation, 2021.

[6] Sehj Kashyap, Keith E Morse, Birju Patel, et al., "A Survey of Extant Organizational and Computational Setups for Deploying Predictive Models in Health Systems", Journal of the American Medical Informatics Association, 28(11), 2445-2450, 2021.

[7] Mark P. Sendak, Joshua D'Arcy, Sehj Kashyap, et al., "A Path for Translation of Machine Learning Products into Healthcare Delivery", EMJ Innovations, 10, 19-172, 2020.

[8] Sarah Lebovitz, Natalia Levina, and Hila Lifshitz-Assaf, "Is AI Ground Truth really 'True'? The Dangers of Training and Evaluating AI Tools based on Experts' Know-what", Management Information Systems Quarterly, 45(3), 1501-1525, 2021.

[9] Sally L. Baxter, Jeremy S. Bass, and Amy M. Sitapati, "Barriers to Implementing an Artificial Intelligence Model for Unplanned Readmissions", ACI Open, 4(02), e108-e113, 2020.

[10] Sandeep Reddy, Sonia Allan, Simon Coghlan, et al., "A Governance Model for the Application of AI in Health Care", Journal of the American Medical Informatics Association, 27(3), 491-497, 2020.