

From Code to Bedside: Implementing Artificial Intelligence Using Quality Improvement Methods



Margaret Smith, MBA, Amelia Sattler, MD, Grace Hong, and Steven Lin, MD

Stanford Healthcare AI Applied Research Team, Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.

Despite increasing interest in how artificial intelligence (AI) can augment and improve healthcare delivery, the development of new AI models continues to outpace adoption in existing healthcare processes. Integration is difficult because current approaches separate the development of AI models from the complex healthcare environments in which they are intended to function, resulting in models developed without a clear and compelling use case and not tested or scalable in a clinical setting. We propose that current approaches and traditional research methods do not support successful AI implementation in healthcare and outline a repeatable mixed-methods approach, along with several examples, that facilitates uptake of AI technologies into human-driven healthcare processes. Unlike traditional research, these methods do not seek to control for variation, but rather understand it to learn how a technology will function in practice coupled with user-centered design techniques. This approach, leveraging design thinking and quality improvement methods, aims to increase the adoption of AI in healthcare and prompt further study to understand which methods are most successful for AI implementations.

KEY WORDS: artificial intelligence; quality improvement; design thinking; implementation science.

J Gen Intern Med
DOI: 10.1007/s11606-020-06394-w
© Society of General Internal Medicine 2021

INTRODUCTION

Artificial intelligence (AI) is a transformational force that has the potential to solve many important problems impacting patients, providers, and health systems. From more accurate risk prediction and decision support, to digital health coaching,¹ the potential is far-reaching and significant. This is fueled by the growing availability of healthcare data, projected to grow at a compounding annual rate of 36%² and reinforced by the 1988 assertion by Dr. David Eddy that the complexity of modern medicine exceeds the capacity of the unaided expert mind.³

Prior presentations None.

Received April 22, 2020

Accepted December 3, 2020

With all the hype, the challenges of AI adoption in healthcare are coming into focus. Regulation, data quality, privacy, and interoperability⁴ are some of the most cited challenges. Yet, when senior leaders are asked what factors most often derail AI implementations, the primary issue cited is the challenge of integrating with existing human processes.⁵ Integration is difficult because current approaches separate the development of AI solutions from the complex healthcare environments in which they are intended to function.

In this article, we outline a repeatable mixed-methods approach that facilitates uptake of AI technologies into human-driven healthcare processes (Table 1). This approach draws on methods and techniques from quality improvement (QI) and design thinking.^{6,7} Unlike traditional research, these methods do not seek to control for variation, but rather to understand and contend with it to learn how a technology will function in practice. Research methods, by comparison, aim to isolate and prove an intervention's ability to influence a controlled system, independent of where and how it is implemented. For AI technology, proving efficacy is wholly dependent on the data in a given system and how it is implemented. The approach outlined in this article organizes work into two main categories: work to be done prior to AI model development and work to be done concurrently with model development.

PART I: PRIOR TO AI MODEL DEVELOPMENT

Define the Problem

Before an AI model is considered or developed, a key first step is to articulate the problem and why it is important to patients, providers, and health systems. QI methodologies all start with the identification and expression of a measurable problem. Some examples are “longer than medically necessary length of stay” or “high rates of infection.” Aligning the project around a problem ensures that technology and workflows are developed to address a genuine need. At this stage, you may have an AI solution already in mind, but it is critical to first clearly define the problem you are trying to solve. The problem should be linked to an important outcome related to the construct your organization uses to prioritize and measure performance; for example, the quadruple aim,⁸ the quintuple aim,⁹ or the value framework.¹⁰ Alignment helps secure adequate leadership support and resources.

Table 1 A Method for Implementing Artificial Intelligence in Healthcare

Part I: Activities preceding AI model development		
Task or concept	Key activities	Relevance for AI
Identify and articulate the problem	<ul style="list-style-type: none"> • Define the problem and why it is important to patients, providers, or health systems • Ensure a clear linkage with organizational priorities and constructs 	<ul style="list-style-type: none"> • Clarifies the potential use case for AI capabilities from the beginning • Ensures organizational resourcing and support throughout development and implementation
Form the team	<ul style="list-style-type: none"> • Identify sponsors, project manager, process owners, end users, and relevant subject matter experts • Determine accountability vs. responsibility for project success 	<ul style="list-style-type: none"> • For the most effective design and implementation, the team should represent the technical and operational perspectives from the beginning – including sponsorship and project team composition • Accountability (i.e., “yes-no” authority and veto power) should reside within operations
Analyze current system and determine key features for success	<ul style="list-style-type: none"> • Observe and map out the current processes, identify issues, and analyze and prioritize root causes • Articulate key features for success from the list of prioritized root causes 	<ul style="list-style-type: none"> • Assists in determining when and how an AI technology might be helpful in solving the problem • Brings into focus the possible use cases for AI for various end users • Broadens the possibilities for workflow design and increases the likelihood of a successful AI deployment
Assess the utility of AI	<ul style="list-style-type: none"> • Determine if AI could address any of the key features for success derived from the last phase • Consider the following costs: <ul style="list-style-type: none"> ○ Cost to build the model ○ Potential for model bias ○ Ongoing model maintenance needs 	<ul style="list-style-type: none"> • Identifies when AI is needed vs. situations where a non-AI solution is superior • Encourages stewardship of technical resources • Mitigates downstream surprises related to cost and feasibility
Part II: Activities concurrent with AI model development		
Task or concept	Key activities	Relevancies for AI
Ideate on key features for success and develop clinical integration workflows	<ul style="list-style-type: none"> • Conduct future state process mapping sessions leveraging design thinking methods such as applying lenses and extremes to stimulate creativity • Develop prototypes, and conduct body storming sessions using real scenarios in a simulation 	<ul style="list-style-type: none"> • Ensures clinical integration workflow design and AI delivery design that balance feasibility, acceptability, efficiency, and effectiveness • Design simulations allow for rapid testing of clinical integration workflows and AI model acceptability
Plan, do, study, adjust (PDSA)	<ul style="list-style-type: none"> • Organize and run iterative live clinical tests of the full intervention (model and workflows) • After sufficient iteration, learning and success, the team can plan more widespread implementation and evaluation 	<ul style="list-style-type: none"> • Allows the technical and operational teams to observe, assess and adjust the model and workflow in usual clinical practice where variation is rife • Use of methods and frameworks, such as the unified theory of acceptance and use of technology (UTAUT), enables an iterative data-driven to design of the intervention (model and workflow)

Form the Team

There are five main roles needed on an AI implementation project team: sponsors (technical and clinical), project manager, process owners, end users, and subject matter experts (technical and clinical) (Table 2). Many technically successful AI models are never implemented because process owners and end users are not involved until after the model is developed, preventing the operational and clinical teams from informing the design. For AI projects, it is important to note that the technical sponsors and subject matter experts are responsible for completing a key piece of the overall project, but they should not be accountable for the project. Accountability must lie with the clinical or operational team because they are ultimately accountable for the outcome, and therefore have “yes” or “no” authority. Assigning responsibility versus accountability is often overlooked, leading to confusion when key decisions need to be made.

Analyze Current System and Determine Key Features for Success

To ensure that model and workflow design are rooted in reality, the team should perform a current system analysis. This includes process mapping, issue identification, root cause analysis, and prioritization (Table 2). The team should first

observe and map the current processes related to the problem articulated in step 1. Visualizing the flow of current processes helps the team identify inefficiencies, variability, and bottlenecks.

With a visual of the current process, the team can identify barriers and build consensus about the issues causing the problem; brainstorm issues related to current technology and tools, organization, physical environment, tasks, and professional or social norms; and prioritize these issues for impact and frequency.

Each current state analysis exercise should be done as a group, including technical team members, making each aspect of the analysis visual using sticky notes and flip charts, or an online collaboration platform.¹⁸

After analyzing the current state, the team is almost ready to dive into model and workflow design. To bridge the current system analysis findings with the intervention design phase, the project team should develop a list of key features for success derived from the prioritized issues. Key features for success are the conditions, structures, and processes that must exist for the problem to be solved. For example, if a high impact issue is that “the current clinical deterioration detection methods are irregular because they rely heavily on provider intuition,” the key feature for success might be “clinical teams need a mechanism that

Table 2 Definition of Terms and Team Member Roles

Term	Definition
Process mapping	Exercise creating a graphical representation of the steps in a process, drawn to better understand processes and identify opportunities to improve ¹¹
Root cause analysis	A collective term that describes a wide range of approaches, tools, and techniques used to uncover causes of problems in a process ¹¹
Prioritization decision matrix	Exercise that evaluates and prioritizes a list of issues or ideas, using predetermined weighted criteria ¹¹
Bottleneck	Anything that limits a system from achieving higher performance or throughput ¹¹
Applying lenses	A design exercise looking and evaluating the design in question through a lens that captures a key design principle ¹²
Bodystorming	Technique of physically experiencing and testing a process to derive new ideas ¹³
Design simulation	Creating the physical or logical representations of a scenario to enable testing of design specifications with potential users before the formal process or product is launched ¹⁴
Prototyping	The iterative development of artifacts – digital, physical, or experiential – intended to elicit qualitative or quantitative feedback ¹⁵
Team members	Role
Clinical or operational sponsor	An individual (director or executive) with overall accountability for all aspects of the project; typically responsible for initiating, ensuring, approving, and establishing a series of key aspects in relation to the project
Technical sponsor	An individual (often a manager or executive) with responsibility for the technical aspects of the project
Project manager	An individual with delegated responsibility for the successful initiation, planning, design, execution, monitoring, controlling, and closure of a project
Process owner(s)	Individuals who have the responsibility and authority for ensuring that a process meets specified requirements and achieves objectives
End user(s)	Individuals who participate in the current system and will likely act on the end product of design (often report directly to process owners)
Clinical subject matter expert(s)	Individuals with a high level of expertise in performing a specialized clinical job, task, or skill relevant for a given project
Technical subject matter expert(s)	Individuals with a high level of expertise in performing a specialized technical job, task, or skill relevant for a given project

predicts deterioration based on clinically relevant data.” This intentionally broad description of a key feature for success guides the design of the model task and workflow for implementation.

Assess whether AI Is Needed

A common pitfall with AI projects is prioritizing the technology over the problem it is intended to solve. To

ensure AI solutions are designed to serve the current system and not vice versa, the project team needs to determine if a clear AI task has emerged. Here are a few additional considerations:

1. The cost to build the model including the human effort and expertise needed to acquire the necessary data, ensure data quality, and establish the infrastructure and data pipelines.
2. The potential for model bias influenced by the ability to procure a population-representative dataset and the extent to which inherent biases can be mitigated.¹⁹
3. The ongoing model maintenance and support needs to detect and allay decline in model performance or systematic data shifts.⁹

The aforementioned should be weighed against the available capacity in the current system and the potential benefits to patients, providers, and health systems.

**PART II: CONCURRENT WITH AI MODEL DEVELOPMENT
Ideate on Key Features and Possible Clinical Integration Workflows**

Design thinking and QI methods are key to designing models and workflows that balance feasibility, acceptability, efficiency, and effectiveness. Some of the most helpful tools for AI implementation are future state process mapping, applying lenses, prototyping, bodystorming, and design simulation (Table 2).

Start by having each of the end users create an ideal future state process map, including when the model output is received. Then, have each person share their ideal process and begin to combine the “best of” each of the individual process maps into one representation, removing redundancies. After consolidation, the team should evaluate and edit the consolidated process map through lenses such as patient safety, feasibility, and acceptability. Finally, the team should review the consolidated process map through the lens of the key features for success developed in step 3. For example, perhaps the team has outlined a process where the AI model output triggers an auto-consult for another specialty. From a patient safety and feasibility perspective, this workflow might seem reasonable, but through the acceptability lens, the team may realize they need to explore alternative workflows.

Next, sketching prototypes and conducting iterative design simulations are effective methods to ascertain more detailed specifications for model delivery and accompanying workflows. Like the future state process mapping exercise, have each member of the team sketch what they see in their mind’s eye as the output from the model in question. The sketches will be highly variable. Shuffle the sketches, then design and run simulations using the prototypes and observe the results. For

the highest yield, simulations should be built around real scenarios and early model prototypes. It is vitally important that the technical team is present for these simulations to gather model customization requirements.

Plan, Do, Study, Adjust

To move from theory to application, the team needs to conduct live clinical testing. This is where traditional research methods fail to meet the needs of AI implementation projects, and an approach that studies variation rather than controlling for it is far better suited to help the team understand how the technology and workflow function in practice. Rapid, structured ideation can be broken into four parts, plan, do, study and adjust, or PDSA.²⁰

Plan:

- Define the questions the test will answer. For example, are you trying to determine model accuracy, acceptability, workflow feasibility, impact on clinical outcomes, or a combination of measures?
- Based on the pilot aims, determine pilot design, activities, data collection tools, and timeline. Leverage evidence-based frameworks such as the unified theory of acceptance and use of technology (UTAUT).²¹
- Acquire the necessary approvals, including compliance, privacy, risk management, institutional review board (IRB), and any other relevant groups.
- Recruit and train pilot participants and leadership on the AI model, workflow, timeline, and mechanisms for gathering feedback.

Do and study:

- Carry out the experiment, monitor progress, and address any unexpected workflow or technical issues through regularly scheduled meetings with the project team and pilot participants.
- Observe the process to understand firsthand what is working well and what is not.
- Review quantitative data trends and assess statistical significance.
- Conduct inductive or deductive coding of qualitative data to derive useful insights.

Adjust (occurs iteratively with do and study):

- Maintain a high degree of organization and ensure clear communication to impacted stakeholders as adjustments are made.
- Schedule structured adjustment sessions where quantitative data and coded qualitative findings are reviewed and deliberated.
- Anchoring on the previously developed key features for success, decide on changes to the workflow, and model as a team.

To illustrate the PDSA process, let us review an example:

A multidisciplinary team is convened to increase the number of documented end-of-life advance care planning discussions in the inpatient acute care setting (Table 3). The team identifies a clear AI task through the current state analysis and decides to implement a homegrown and validated mortality predictive model to assist with patient selection based on prognosis. As an initial pilot, the team decides to test the acceptability of sharing the predictive model output with physicians before morning bedside rounds in the hopes that this new information will prompt end-of-life advance care planning conversations with their patients.

The data pipelines are in place, and the team holds a brief training session with the 6 physicians on service. After one week, the team learns that the model was acted on 5% of the time, much lower than expected. Qualitative analysis reveals that providers struggle to make time to have these difficult conversations with their patients. In addition, although providers agree with the model's prognosis predictions, they believe it is too early to engage with patients in these conversations on the first day of admission as the model often suggests.

The team is encouraged to hear that providers agree clinically with the model output and settles on two adjustments for the next PDSA cycle. First, the team decides to run the model on patients that are past the first 24 h of their hospitalization. To address the physicians' expressed time constraints, extended team members, such as nursing, physical therapy, and social work, suggest that they also receive the model output and share the responsibility for the workflow along with the physicians. A revised pilot workflow and evaluation criteria are developed, and the team starts a new PDSA cycle.

Conduct as many PDSA cycles as necessary to refine the workflow and model. If a current system analysis was completed prior to the PDSA experimentation, the team will likely reach the desired result quickly, and can then determine if widespread implementation is warranted.

CONCLUSION

Current approaches and traditional research methods do not support successful AI implementation in healthcare. Different than new drugs or medical devices, the efficacy of AI technology is wholly dependent on the data in a given system and how the tool is integrated into a relevant environment. Research methods aim to isolate and prove an intervention's ability to influence a controlled system independent of where and how it is implemented. QI methods seek to understand the sources of variation in a current system and facilitate model and workflow design that withstand normal variation. Research answers the question, "does it work?," while QI answers the question, "how does it work?"

Current approaches compartmentalize technical develop-

Table 3 Example Projects and Key Learnings

Project description	Key problem identified	Quadruple aim(s) addressed	Current project status (as of June 2020)	Key learning(s)
Implementing a predictive model for clinical deterioration in the inpatient acute care setting	Unanticipated clinical deterioration resulting in unexpected escalations to the intensive care units and mortality	Better care	In implementation; conducting multiple PDSA cycles	Designing simulations built around real scenarios and early model prototypes accelerates learning the key features for successful workflow integration and model output delivery features. Involving the technical team from the beginning, along with the clinical team, allows for more rapid development and ideation.
Implementing a predictive mortality model to enable patient selection for end-of-life advance care planning discussions	Lower than desired incidence of end-of-life care planning conversations resulting in distress for patients and providers	Better care; workforce wellness	In implementation; conducting multiple PDSA cycles	Involving clinical, operational, and technical team members from the beginning leads to surprising and more feasible clinical integration workflows. Selecting a small focus group of front-line staff to engage in multiple PDSA cycles allows rapid multidisciplinary learning, and adjustments to model output delivery and workflows.
Designing an outpatient AI-enabled patient risk stratification and prediction tool	Ambulatory sensitive admissions and existing “hot spotting” tools that do not consider social determinants and clinician impressions found in unstructured medical data	Lower cost; better health	Assessing the utility of an AI solution; ideating on key features for success	Identifying a clinically relevant problem and conducting a current state analysis ensures that, before building the model, a clear AI task has been characterized and key drivers for operational success inform design early on. This also mitigates wasted time and resources.
Designing an AI-enabled at-home monitoring platform to close the data gap and enhance ability to engage patients between clinic visits	Low incidence of necessary behavior changes in patients with chronic diseases	Lower cost; better health	Analyzing the current system; determining key features for success	Involving the clinical, operational, and technical team members in an in-depth current state analysis leads to more innovative and surprising key features for clinical and operational success.
Designing an AI algorithm that can detect depression and anxiety based on audio and visual cues	Current depression screening methods require significant cognitive and clerical burden, limiting the ability to screen and identify patients with depression and anxiety	Better care, better health	Model in development; ideating on key features for success	Aligning with existing organizational priorities and quality improvement efforts helps secure leadership support and resourcing. Leveraging early-stage clinical integration workflows to determine the approach for curating training set data can be helpful.
Assessing the feasibility and acceptability of an AI-powered tool to assist primary care providers with diagnosing skin conditions	Suboptimal referrals, delays in care, and errors in diagnoses due to a shortage of dermatologists, and the burden of diagnosis placed on primary care providers	Lower cost; better care	Ideating on key features for success and clinical integration workflow for the pilot study	Including technical, clinical, and operational stakeholders, and leveraging frameworks such as PDSA and UTAUT ²¹ enables the team to learn together iteratively <i>how</i> an AI model works versus <i>does</i> an AI model work.
Designing, and assessing the acceptability of an AI-enabled model for pre-visit planning and intra-visit care management	Scarcity of integrated tools to help patients efficiently and effectively organize and manage their personal care plans between visits	Better care	Analyzing the current system; ideating on key features for success	Conducting extensive user experience and current system analyses from the patient and provider perspective ensures that technology features and data infrastructure are designed with workflow integration in mind from the beginning.
Assessing the acceptability of AI-enabled documentation of various aspects of the clinic visit note	Physician burnout due to burden of documentation	Workforce wellness	Project completed	Conducting extensive user experience and current state analyses of existing human-driven documentation practices provides important insights into key features for success and factors that influence trust and acceptability of proposed AI technologies.

ment, workflow design, and clinical integration, and as a result, many AI models are created without a compelling use case co-developed with patients, providers, or health systems, and are not tested or scalable in a clinical context. This is illustrated by the dearth of translational or implementation literature on AI in contrast with the breadth of research on technically successful (but not clinically integrated) models. The approach we outlined in this article leverages design thinking and QI methodologies with the aims of increasing adoption and prompting a further study of AI in healthcare in order to better understand which implementation methods are most successful.

Corresponding Author: Steven Lin, MD; Stanford Healthcare AI Applied Research Team, Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA (e-mail: stevenlin@stanford.edu).

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

REFERENCES

1. **Lin SY, Mahoney MR, Sinsky CA.** Ten ways artificial intelligence will transform primary care. *J Gen Intern Med.* 2019;34:1626-1630.
2. **Reinsel D, Gantz J, Rydning J.** The Digitization of the World – From Edge to Core. International Data Corporation: Analyze the Future. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. Published November 2018. Accessed April 9, 2020.
3. **Eddy DM, Billings J.** The quality of medical evidence: implications for quality of care. *Health Aff (Millwood).* 1988;7(1):19-32.
4. **Kelly CJ, Karthikesalingam A, Suleyman M** et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195.
5. **Davenport TH, Ronanki R.** Artificial Intelligence for the Real World. *Harv Bus Rev.* 2018;96(1):108-116.
6. **Roberts JP, Fisher TR, Trowbridge MJ,** et al. A design thinking framework for healthcare management and innovation. *Healthc (Amst).* 2016;4(1):11-4.
7. **Ferreira FK, Song EH, Gomes H,** et al. New mindset in scientific method in the health field: Design Thinking. *Clinics (Sao Paulo).* 2015;70(12):770-772.
8. **Bodenheimer T, Sinsky CA.** From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med.* 2014;12(6):573-576.
9. **Matheny M, Thadaney SI,** et al. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. *NAM Special Publication.* Washington, DC: National Academy of Medicine. 2019.
10. **Porter ME.** What Is Value in Health Care? *N Engl J Med.* 2010;363:2477-2481.
11. **Quality Glossary.** American Society for Quality. Website. <https://asq.org/quality-resources/quality-glossary>. Updated January 2020. Accessed April 9, 2020.
12. **Scott B.** Designing with Lenses. UX Booth. Website. <https://www.uxbooth.com/articles/designing-with-lenses/>. Published April 2010. Accessed April 9, 2020.
13. **Oulasvirta A, Kurvinen E, Kankainen T.** Understanding contexts by being there: case studies in bodystorming. *Pers Ubiquit Comput.* 2003;7:125-134.
14. **Lamé G, Dixon-Woods M.** Using clinical simulation to study how to improve quality and safety in healthcare. *BMJ STEL.* 2020;6:87-94.
15. **Dam RF, Teo YS.** Design Thinking: Get Started with Prototyping. Interaction Design Foundation. Website. <https://www.interaction-design.org/literature/article/design-thinking-get-started-with-prototyping>. Published March 2020. Accessed April 9, 2020.
16. **Schibi O, Lee C.** Project sponsorship: senior management's role in the successful outcome of projects. Project Management Institute. <https://www.pmi.org/learning/library/importance-of-project-sponsorship-9946>. Published October 2015. Accessed April 9, 2020.
17. **Munns AK, Bjeirmi BF.** The role of project management in achieving project success. *Int J Proj Manag.* 1996;14(2):81-87.
18. **20 Best Online Tools for Design Thinking.** SessionLab. Website. <https://www.sessionlab.com/blog/design-thinking-online-tools/>. Published March 19, 2019. Accessed April 9, 2020.
19. **Char DS, Shah NH, Magnus D.** Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N Engl J Med.* 2018;378(11):981-983.
20. **Speroff T, O'Connor GT.** Study Designs for PDSA Quality Improvement Research. *Qual Manag Health Ca.* 2004;13(1):17-32.
21. **Venkatesh V, Morris MG, Davis GB,** et al. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly.* 2003;27(3):425-478.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.