Overview of Noninterpretive Artificial Intelligence Models for Safety, Quality, Workflow, and Education Applications in Radiology Practice

Yasasvi Tadavarthi, MD • Valeria Makeeva, MD • William Wagstaff, MD • Henry Zhan, MD • Anna Podlasek, MD • Neil Bhattia, BS • Marta Heilbrun, MD • Elizabeth Krupinski, PhD • Nabile Safdar, MD • Imon Banerjee, PhD • Judy Gichoya, MD • Hari Trivedi, MD

From the Department of Medicine, Medical College of Georgia, Augusta, Ga (Y.T.); Department of Radiology and Imaging Science (V.M., WW, H.Z., M.H., E.K., N.S., J.G., H.T.), School of Medicine (N.B.), and Department of Biomedical Informatics (I.B.), Emory University, 1364 E Clifton Rd NE, Atlanta, GA 30322; and Southend University Hospital NHS Foundation Trust, Westcliff-on-Sea, UK (A.P.). Received April 28, 2021; revision requested June 24; revision received December 17; accepted January 11, 2022. Address correspondence to H.T. (e-mail: hari.trivedi@emory.edu).

Conflicts of interest are listed at the end of this article.

Authors declared no funding for this work.

Artificial intelligence has become a ubiquitous term in radiology over the past several years, and much attention has been given to applications that aid radiologists in the detection of abnormalities and diagnosis of diseases. However, there are many potential applications related to radiologic image quality, safety, and workflow improvements that present equal, if not greater, value propositions to radiology practices, insurance companies, and hospital systems. This review focuses on six major categories for artificial intelligence applications: study selection and protocoling, image acquisition, worklist prioritization, study reporting, business applications, and resident education. All of these categories can substantially affect different aspects of radiology practices and workflows. Each of these categories has different value propositions in terms of whether they could be used to increase efficiency, improve patient safety, increase revenue, or save costs. Each application is covered in depth in the context of both current and future areas of work.

© RSNA, 2022

The radiology community has had a leading role in exploring medical applications of artificial intelligence (AI), and one of the primary drivers for this is the desire for increased accuracy and efficiency in clinical care. Radiologist responsibilities extend beyond image interpretation. AI tools have the potential to improve essential tasks in the imaging value chain, from image acquisition to generating and disseminating radiology reports (1). These applications are crucial in current medical environments with increasing workloads, increasing scan complexity, and the need to decrease costs and reduce errors (2–4). AI applications related to radiologic quality, safety, and workflow improvements can be grouped by their influence on various steps in the typical radiology workflow, as follows in their approximate order of occurrence: study selection and protocoling; image acquisition; worklist prioritization; study reporting, business applications, and resident education. This qualitative review is a discussion of current research and commercial models regarding these applications within the entire imaging chain.

Methods

Studies published from 1980 through 2019 were retrieved nonsystematically from academic search engines including PubMed, ScienceDirect, and Google Scholar by using search terms related to each application of interest. Public legal documents were also accessed including the Medicare Physician Fee Schedule and Other Revisions to Part B, Quality Payment Program requirements, and Shared Savings Program requirements.

Public news sources, such as Becker’s Hospital Review, Healthcare Finance, Optum, and Healthcare IT News, and vendor lists from meetings of the Radiological Society of North America and the Society for Imaging Informatics in Medicine were used to find any commercial efforts in each space. All searches were performed by the authors, all of whom are attending radiologists or trainees with a research interest in radiology AI.

Study Selection and Protocoling

Automated Study Vetting and Clinical Decision Support

Inappropriate imaging studies are inefficient because they expend health care resources, increase payer costs, increase patient risk, and delay care (5,6). Inappropriate imaging orders may represent up to 10% of ordered examinations, and not all are caught before the examination is performed (6–10). Imaging ordering errors have multifactorial causes but can include a lack of knowledge of appropriate imaging types, over-ordering by providers because of constrained resources, erroneous clicks in the computerized physician order entry system, and unnecessary duplicate examinations if a similar study was already performed (eg, chest radiography performed immediately after chest CT).

To address concerns regarding inappropriate imaging, the Protecting Access to Medicare Act of 2014 requires the use of an appropriate use criteria system for any advanced diagnostic imaging service. Many automated clinical decision support systems have been developed to meet these
Abbreviations

AI = artificial intelligence, BI-RADS = Breast Imaging Reporting and Data System, BT-RADS = Brain Tumor Reporting and Data System, EMR = electronic medical record, LI-RADS = Liver Imaging Reporting and Data System, MRI = magnetic resonance imaging, NLP = natural language processing, TI-RADS = Thyroid Imaging Reporting and Data System, BT-RADS = Brain Tumor Reporting and Data System, EMR = electronic medical record, LI-RADS = Liver Imaging Reporting and Data System, MRI = magnetic resonance imaging, NLP = natural language processing

Summary

Many noninterpretive artificial intelligence applications with the potential to improve multiple aspects of radiology practice, including workflow, efficiency, image acquisition, reporting, billing, and education, are either currently available or in development.

Essentials

- Artificial intelligence (AI) models to improve workflow efficiency and safety include automated clinical decision support, study protocolling, examination scheduling, and worklist prioritization.
- Models to improve image acquisition focus on patient positioning, multimodal image registration, dose reduction, noise reduction, and artifact reduction.
- Models to improve reporting include automatic finding categorization using classification systems (eg, Breast Imaging Reporting and Data System, Liver Imaging Reporting and Data System), provider notification of incidental findings, and closing the loop on patient follow-up.
- Business applications include automated billing and coding, obtaining preauthorization, and optimization of performance on quality measures to increase reimbursement.
- Use of AI in resident education is somewhat controversial, but AI can be used to help flag high-risk cases for faster review by an attending physician, customize teaching files based on residents’ needs, and help improve resident reporting.

Keywords

Use of AI in Education, Application Domain, Supervised Learning, Safety

requirements, including by vendors that license the American College of Radiology ACR Select database (11). Implementation of clinical decision support systems in the hospital setting has resulted in decreased inappropriate imaging and advanced imaging overall (12,13). For example, Yan et al (14) reported that the yield of CT angiography in detecting pulmonary embolism doubled after implementation of a clinical decision support system. Doyle et al (15) reported an overall 6% decrease in imaging with the use of a clinical decision support system in a randomized clinical trial of 3500 health care providers. Existing systems, however, are not without substantial limitations: They are largely based on a branching decision tree structure that can be exploited to arrive at the desired examination type. A more advanced system that relies on natural language processing (NLP) of free-text input and integration of electronic medical record (EMR) data could decrease the so-called click fatigue associated with current systems by allowing more flexible input. However, our research did not reveal any advanced NLP-based system currently in existence or development.

Study Protocolling

Protocolling is the process of selecting the appropriate sequences for an MRI or CT examination to ensure that the desired anatomy and abnormalities are adequately captured; it is typically performed by the radiologist because of their domain expertise. This is a time-consuming process, however. At our institution, approximately 1–2 hours per day in each division is spent protocolling studies, totaling 50 hours per week across the department, which is the equivalent of the workload for one full-time equivalent radiologist. Protocolling is time-consuming for many reasons, including the frequent presence of dozens of protocol options, the need to look up information from the EMR, and the lack of intelligent aids within the protocol workflow.

In recent years, NLP has shown good results for automating study protocols. For example, Lee (16) automated the selection of routine versus tumor or infection protocols for musculoskeletal MRI, and Trivedi et al (17) distinguished between musculoskeletal studies with and without gadolinium contrast enhancement. Both models achieved overall accuracies of greater than 90%. Brown and Marotta (18) automated three tasks for brain MRI (protocol selection, need for intravenous contrast agent, and examination prioritization) and achieved overall accuracies between 83% and 88%. More recent work focused on a model that functioned beyond a single anatomic region or imaging modality, achieving a precision of 76%–82% when tested on 18 000 diverse CT and MRI examinations (19). Overall, we found that models with more advanced deep learning approaches had higher performance than those with traditional machine learning techniques.

A limitation of current protocolling model performance is the input data to which the model has access. Just as a radiologist may access EMR data to correctly protocol an examination, AI models also need access to these additional data to maximize their performance. This is challenging, however, because these data are stored in various locations within the EMR and often within free-text clinical notes, the interpretation of which is a difficult machine learning challenge. Approaches such as long short-term memory networks and bidirectional encoder representations of transformers have been used to automatically extract information from the EMR and could be leveraged to provide more data to a protocolling model (20–23). In the meantime, human in-the-loop verification of automatically selected protocols is likely necessary to ensure patient safety and optimal imaging.

Image Acquisition

Successful interpretation of medical imaging requires proper image acquisition. Radiation dose, imaging dimensions, patient positioning and motion, implanted hardware, and sensor variability affect image quality for interpretation. Machine learning techniques in this domain have been shown to reduce radiation exposure, decrease scan times, reduce rates of false-positive findings, and reduce unnecessary repeat imaging while maintaining image quality (24).

Dose Reduction

As the use of CT and PET increases worldwide, radiation exposure to patients undergoing frequent examinations is a concern. Radiology departments often must balance radiation dose and image quality against the practices of “as low as reasonably achievable” to avoid unnecessary radiation exposure (25). The
conventional method to reduce CT radiation dose is to decrease tube current but this increases noise and reduces diagnostic confidence (26). However, machine learning techniques for image reconstruction have recently demonstrated impressive results that provide higher-quality images than traditional techniques while maintaining lower radiation doses (27,28). These denoising algorithms are discussed in further detail in the Image Reconstruction section below.

In PET imaging, radiotracer dose reduction has been targeted with models that reconstruct low-dose examinations to appear similar to full-dose examinations by using noise-reduction algorithms. One commercial company has been able to use only one 200th of the standard tracer dose and a reduced scan time of up to 75% while achieving image quality comparable to the industry standard by using encoder-decoder residual deep learning networks (29,30). Generative adversarial networks have been used to reconstruct PET images acquired with 1%–25% of the standard radiotracer dose with quality similar to that of normal-dose PET images (31,32) (Fig 1).

MRI does not produce ionizing radiation, but researchers have explored machine learning techniques to reduce gadolinium-based intravenous contrast agent dosage (33). Gong et al (33) used machine learning to achieve a 10-fold reduction in gadolinium-based contrast agent administration with no significant reduction in image quality or contrast information.

**Image Reconstruction**

Image reconstruction is fundamental to medical imaging to create high-quality diagnostic images while managing cost, reconstruction time, and risk to the patient (34,35). The details of image reconstruction are beyond the scope of this review, but there have been extensive research efforts to use machine learning techniques to improve image reconstruction in CT, MRI, and PET. Examples of targets for improvement include noise reduction, artifact suppression, motion compensation, faster image acquisition, and multimodal image registration. These goals are often codependent and closely related, and it is therefore possible to reduce both radiation dose and contrast agent dose with the use of successful image reconstruction techniques.

Image quality is often a trade-off between radiation dose in CT and scan times for MRI. Filtered back projection (36,37), iterative reconstruction (38,39), and newer model-based iterative reconstruction techniques function by filtering raw sensor data or by considering noise statistics, optics, physics, and scanner parameters (38). However, all of these techniques are specific to the vendor and can have substantial overhead costs because of their long computational time (27).

Early machine learning–based CT reconstruction techniques caused over-smoothing, resulting in so-called waxy images (26). Since then, several subtypes of convolutional neural networks have been developed to denoise CT and MR images without loss of technical detail (25,40). One method combines deep learning techniques with standard filtered back projection principles to produce high-quality images with low noise, even with a 20-fold reduction in CT input data (41). Another vendor-agnostic CT solution achieved higher spatial resolution than filtered back projection and model-based iterative reconstruction for processing low-dose CT and has been granted U.S. Food and Drug Administration clearance (ClariCT.AI; ClariPi). A different company

---

*Figure 1: (A, B) Two examples of low-dose PET (left), ground truth standard-dose PET (middle), and low-dose PET with generative adversarial network-synthesized images (right). (Adapted, with permission, from reference 32.)*
has commercialized a deep learning–based CT reconstruction product that provides quality similar to that of model-based iterative reconstruction but with a three- to fourfold reduction in reconstruction time (42,43).

In MRI, longer acquisition times can produce higher image quality, but they also increase the risk of motion artifacts (44). As a result, several machine learning approaches have targeted MRI noise reduction and artifact suppression (44) (Fig 2). Most of these applications are in the research phase, although a few vendor-agnostic denoising products have been approved by the U.S. Food and Drug Administration. These products reduce MRI acquisition times by 30%–40% (45,46).

### Image Quality Control

Poor image quality can be particularly challenging in MRI because of suboptimal scan parameters, artifacts, or inappropriate coverage (47). Repeat MRI sequences are required in up to 20% of examinations, at a cost to hospitals of up to $115 000 per scanner annually (24). Various methods have been proposed to automatically assess image quality prospectively or retrospectively.

Prospective image quality control can benefit scan protocols with high acquisition times, such as brain MRI (24) or real-time T2-weighted liver MRI (48). In these cases, models have shown value in assessing for nondiagnostic scan quality during acquisition so technologists can adjust scan parameters during the examination rather than after its completion (24,48). Retrospective image quality control explores techniques to mitigate metal artifact, respiratory motion, and banding artifact at MRI. Multiple groups have developed models that target noise and artifact suppression (44,49,50) (Fig 3).

One company has developed algorithms for image quality issues in radiography, US, and conventional angiography (ContextVision). They offer products to reduce over- or underexposure and metal artifact in radiography, suppress noise to improve contrast and tissue differentiation at US, and reduce noise and motion artifact for improved visibility of stents and catheter tips in coronary artery angiography.

### Image Registration

Image registration refers to linking the same anatomic region together within an examination or across examinations, and it is a frequent and repetitive task for radiologists during study interpretation. Several permutations of this mathematical problem exist because several variables can be considered, including modality, region of interest, temporality, dimensionality, and elasticity of tissues (51).

Several techniques for automatic image registration have been explored. Section-to-volume registration is a common implementation in which a two-dimensional image section is registered to an existing three-dimensional volume. The primary example of this type of application is registration of two-dimensional transrectal US with an existing three-dimensional MRI for targeted prostate biopsy (52). Cross-modality registration is also performed between three-dimensional volumes (eg, registration of a preoperative CT or MRI to an intraoperative CT for targeted thermal ablation of liver lesions [53] or registration of prostate lesions across CT and MRI [54]; Fig 4). Haskins et al (52) published a comprehensive list of image registration applications.

### Patient Positioning

Radiation dose exposure to different organs depends on patient positioning within the CT gantry, and an inexperienced technologist may inadvertently over- or underexpose the region of interest because of miscalculations of patient size on the basis of the localizer radiograph (55,56). An offset of as little as 20 mm can result in significant changes in effective organ dose (55,56). Advances in patient positioning include a three-dimensional depth-sensing camera that recognizes the anatomic landmarks and models that automatically calculate the patient’s center, which is used to optimize the patient bed position for dose and image quality. This implementation is commercially available by one vendor and has been shown to be more accurate and less variable than manual positioning by technologists (55,57,58).

In mammography, poor positioning can result in missed breast cancers or technical recalls (59). Strict adherence to positioning and technique optimizes breast coverage and diagnostic quality while minimizing radiation (59,60). Models to automatically evaluate image quality at the time of acquisition to ensure compliance with the Mammography Quality Standards Act and Program (61) could reduce technical recalls, and one such solution is registered with the U.S. Food and Drug Administration (Mia IQ: Kheiron Medical Technologies).
findings) and detected five new intracranial hemorrhages with a reduction in reporting time for these cases from 8.5 hours to 19 minutes. Multiple similar models exist for detection of intracranial hemorrhage (65,66) and emergency findings at abdominal CT (67) and chest CT angiography (68,69).

Typically, AI is used to detect positive findings that require emergency intervention (eg, pulmonary embolism, hemorrhage, and pneumoperitoneum), but this narrowed focus addresses only part of the problem in a resource-limited setting such as the emergency department. Prolonged turnaround times for examinations with negative findings also equate to prolonged turnaround times for the emergency department, in which staff may be awaiting a negative result to discharge a patient (70,71). Negative results may also be necessary for taking appropriate steps in patient care, for example, clearing a noncontrast head CT for hemorrhage before a patient can undergo thrombolysis for acute stroke. In this scenario, rapid confirmation of the absence of a finding is crucial for patient care (72). As of the writing of this review, there is no U.S. Food and Drug Administration–approved model for detection of examinations with definitively negative findings.)

Figure 3: Noise suppression of (top) T1- and (bottom) T2-weighted images. Original images (left) and processed images (right). (Adapted, with permission, from reference 50.)

Worklist Prioritization

Radiologist worklists are typically populated by examinations on the basis of preset criteria, such as body part, modality, patient location, and priority. However, nonemergency examinations are often mistakenly ordered as emergency examination in an effort to expedite imaging, thereby preventing the radiologist from differentiating between routine and emergency studies and potentially delaying the interpretation of truly emergency cases.

Many AI algorithms have been developed across multiple body regions to prioritize examinations with emergent findings (62) (Fig 5). These models must be adequately sensitive and specific to identify emergency findings while avoiding excessive false-positive results. Annarumma et al (63) tested such a system to simulate a triage system for retrospective adult chest radiographs, resulting in a theoretical reduction in reporting delay for critical studies from 11.2 to 2.7 days. Arbabshirani et al (64) prospectively implemented a prioritization system for detection of intracranial hemorrhage at head CT, which flagged 94 of 347 routine cases (60 true-positive findings, 34 false-positive findings) and detected five new intracranial hemorrhages with a reduction in reporting time for these cases from 8.5 hours to 19 minutes. Multiple similar models exist for detection of intracranial hemorrhage (65,66) and emergency findings at abdominal CT (67) and chest CT angiography (68,69).

Typically, AI is used to detect positive findings that require emergency intervention (eg, pulmonary embolism, hemorrhage, and pneumoperitoneum), but this narrowed focus addresses only part of the problem in a resource-limited setting such as the emergency department. Prolonged turnaround times for examinations with negative findings also equate to prolonged turnaround times for the emergency department, in which staff may be awaiting a negative result to discharge a patient (70,71). Negative results may also be necessary for taking appropriate steps in patient care, for example, clearing a noncontrast head CT for hemorrhage before a patient can undergo thrombolysis for acute stroke. In this scenario, rapid confirmation of the absence of a finding is crucial for patient care (72). As of the writing of this review, there is no U.S. Food and Drug Administration–approved model for detection of examinations with definitively negative findings.)
results; however, such models have the potential to substantially affect patient care and throughput.

**Reporting**

**Structured Reporting**

Integration of AI applications into radiology reporting has the potential to increase the clarity, accuracy, and quality of reporting and decrease report variability in some situations (73). For example, models have been created to improve patient care by automatically populating recommendations for follow-up of incidental findings (74–77). NLP models have also been developed as smart assistants. For example, Do et al (78) developed a tool that detected when the radiologist was reporting a fracture and displayed additional information regarding pertinent classifications, associated injuries, and further clinical recommendations. Whereas multiple frameworks have been developed to convert unstructured findings in reports into structured templates to improve legibility (79–81), we were unable to find any recent system that has been systematically tested for performance or implemented clinically.

**Classification Systems**

Several classification systems have been developed for frequently encountered lesions, including thyroid (Thyroid Imaging Reporting and Data System [TI-RADS]) (82), breast (Breast Imaging Reporting and Data System [BI-RADS]) (83), liver (Liver Imaging Reporting and Data System [LI-RADS])...
Radiology: Artificial Intelligence

Volume 4: Number 2—2022

Abstract

Communication of critical diagnoses is mandated by the Joint Commission as a part of National Patient Safety Goal 2, “Improving the Effectiveness of Communication Among Caregivers” (99). In practice, implementing this trail of communication is inefficient and can disrupt workflow, contributing to burnout among radiologists (100). Communication failure is also one of the leading causes of malpractice lawsuits (101). Hiring reading room coordinators or medical students to help with communication increases work satisfaction among radiologists; however, hiring personnel is costly. Therefore, AI has been a topic of interest in automating provider notification (62,102–104). A notable implementation of this technology was described by Do et al (105), who used AI in outpatient oncologic CT images to detect actionable incidental findings such as pulmonary embolism, gastrointestinal obstruction, hydrenephrosis, and pneumothorax, resulting in a median 1-hour decrease in notification time to referring physicians and a 37% improvement in radiologist interpretation time.

Patient Follow-up

Radiologist reporting and recommendations for incidental findings is variable (106), and patient chain management can be challenging in large, complex health systems, sometimes resulting in lack of follow-up care. Many groups have used NLP to identify incidental follow-up findings in the radiology report to reduce the variability of recommendations or the number of patients for whom follow-up recommendations are not suggested or are not followed (107–111). Implementation of such systems into the live clinical environment remains rare; however, Hammer et al (112) implemented a closed-loop system for follow-up of incidental pulmonary nodules, resulting in a significantly higher rate of appropriate follow-up by primary care physicians ($P < .001$). A sample report from such a system is shown in Figure 6.

Business Applications

Billing and Coding

AI applications in business analytics present an opportunity to create value and shape radiology practice. A major area of focus has been billing and coding because of the combined potential effect of increased revenue and decreased errors.

It has been estimated that health care organizations lose between 3% and 5% of net revenue annually because of insurance claim denials (113,114). In 2010, the National Academy of Medicine synthesized one of the most extensive datasets of U.S. administrative costs related to billing and insurance, estimating that billing-related costs account for 13% of physician care spending and 8.5% of hospital care spending (115,116). More than 100 variables contribute to claim denial by insurance companies, and although this number is too vast to assess manually for each report, NLP can automatically ensure that reports are billed and coded appropriately (117,118).

Research that uses NLP has shown that incomplete documentation is common for many examinations. For example, documentation deficiencies have been identified in 9.3%–20.2% of abdominal US reports, representing a 2.5%–5.5% loss in professional reimbursement (119). AI can assist by creating predictive classification models for automated procedure coding. A study investigating the coding of MRI examinations demonstrated that the AI system achieved the same performance as manual coding by a technologist and did not require any human intervention (120). Therefore, automated coding techniques may...
optimize reimbursement, improve workflow efficiency, and assess rejected claims to help reduce future denials (121).

**Preauthorization**

Lack of clinical documentation from referrals often leads to delays in authorization of procedures and imaging. Whereas computerized physician order entry was created as a tool to decrease errors in ordering and to help with preauthorization, it has had variable success depending on the use case and method of implementation (122). Even with computerized physician order entry, many referrals must be manually reviewed and are subject to time-consuming telephone calls to insurance companies. Examples of missing information include incomplete patient demographics, outdated or inactive insurance information, and incomplete clinical documentation. According to a survey of 500 health industry leaders in the United States, automation of preauthorization was seen as the AI application with the most potential (123).

A substantial amount of these relevant data resides in the radiology information system and EMR, which may contain data pertinent to preauthorization such as patient orders, insurance, and clinical history that may be amenable to query by using NLP techniques. Prior authorization software enables health care organizations to identify authorization requirements at the time of scheduling by mining the radiology information system and EMR, therefore reducing manual administrative burden and patient scheduling delays (124).

**Value-based Payment Models**

Data-driven quality improvement lies at the intersection of new value-based payment models and AI. The Quality Payment Program arose as part of the Medicare Access and CHIP Reauthorization Act of 2015 and represented the shift to value-based care by enumerating a series of value-based paradigms for physician reimbursement (125). To understand AI applications within the Quality Payment Program, it is important to understand how reimbursement processes differ between the two major Quality Payment Program pathways—the Merit-based Incentive Payment System and the alternative payment model (Fig 7).

The Merit-based Incentive Payment System involves a 100-point score related to quality, cost, interoperability, and improvement and results in positive, negative, or neutral adjustments to reimbursements based on physician performance. For radiologists, the quality category is the most important, and approximately 85% of radiologist Merit-based Incentive Payment System scores were directly affected by the quality category in 2019 (126,127). Many quality metrics center on reducing unnecessary imaging and ensuring appropriate documentation and follow-up. AI-based tools may be used to optimize performance on quality measures such as carotid artery stenosis measurements or appropriate follow-up for incidentally discovered lesions (126). Similarly, AI could be used to develop tools to automatically measure and track lesion progression, place information into reports, or even search the radiology information system and EMR to evaluate inclusion or exclusion criteria for certain patients (126).

The alternative payment model pathway has a greater focus on population health compared with the Merit-based Incentive Payment System, such that tools that improve the health of the entire population are specifically incentivized. AI applications that reduce cost while maintaining or improving quality are especially relevant to alternative payment model pathways and encourage team-based accountability within a health care organization. In 2019, up to 15% of the final alternative payment model scores were related to cost (127). Within this context, AI that is focused on reducing unnecessary procedures and imaging is especially valuable (e.g., models that predict the malignancy potential of a lesion to decrease unnecessary follow-up scans or a tool that mines the EMR for prior studies to reduce redundant imaging) (126). In the future, primary drivers of AI applications in radiology business analytics, such as applications in quality improvement, will likely continue to correlate with the regulatory landscape and payer reimbursement patterns.

**Resident Education**

There are many potential use cases for AI in radiology education. As AI tools become ubiquitous in the daily workflow for radiologists, care must be taken to ensure that radiology trainees learn adequate interpretation skills and do not rely on AI software to locate abnormal findings or assign diagnoses. Beyond these potential risks, however, there are many opportunities to improve resident education by using AI tools.

Tajmir and Alkasab (128) list various potential applications of AI in radiology education, including selection of trainee cases, improved supervision of residents by attending physicians, analysis of report differences between trainees of various levels, and facilitation of lifelong learning. For example, AI algorithms could identify cases that have educational value based on parameters such as common diseases; rare, interesting, or unique findings; complexity; and acuity. These cases could be automatically incorporated into a trainee’s worklist or into a teaching file for dedicated teaching sessions. Conceivably, such a process could be tailored to specific residents, thereby creating individualized learning opportunities.

---

**Figure 7:** A comparison of the Merit-based Incentive Payment System (MIPS) and the alternative payment model (APM) pathways and possible artificial intelligence (AI) applications under each model. EMR = electronic medical record, RIS = radiology information system.
Receiving feedback from supervising attending physicians is an integral part of clinical education; however, a balance must be struck between complete trainee autonomy and overbearing supervision. AI could help by silently alerting a supervising radiologist when a junior resident opens a complex or high-acuity case (128). This workflow would allow the resident an opportunity to independently review a case while ensuring that an attending physician is also aware of the case, thereby maintaining patient safety and simultaneously allowing for the effective educational growth of residents.

There is also an opportunity for NLP-based applications to affect resident education. NLP and AI algorithms may be used to compare reporting differences between trainees and nontrainees of various levels (128). Although this is a potentially sensitive area, a theoretical use case would demonstrate to junior residents how their reporting differs from that of more senior trainees. The AI system could then provide suggestions for changes that could be made by the junior resident. Care must be taken in implementation, however, so trainees do not feel unnecessarily “watched over” during interpretation.

AI applications could also facilitate lifelong education by incorporating new data and recent updates in imaging guidelines into a radiologist’s reporting (128), for example, the newest guidelines for incidental pulmonary nodule follow-up. Such an application could benefit both trainees and attending physicians alike.

Despite these potential benefits, AI must be used judiciously in resident training to avoid interfering with development of the resident’s skills. Residents must be educated in the appropriate use and interpretation of AI results because understanding how AI models are developed will better equip them to identify and appropriately manage model errors.

Areas of Future Work

A limitation of most machine learning applications for non-interpretive use cases is the relative lack of exploration of clinical effect and generalizability. Most research models described herein were developed and validated at a single institution. There is a vast technical, resource, time, and cost gap between developing a well-performing model on the basis of retrospective data and implementing the model in a live clinical setting at multiple disparate sites. Unlike imaging-based AI models that work on standardized Digital Imaging and Communications in Medicine imaging, non-interpretive models rely on heterogeneous data from multiple sources that are complex and varied across institutions. In our own institution, more than 80 interconnected software products are used in the radiology department and accessing data from these software products and integrating models into them is complex, requiring the agreement of multiple stakeholders. Those who are interested in trying publicly available research models at their own institution must be prepared to devote the time and personnel for implementation, even if the software is available free of charge. Companies developing products in this space should understand the potential complexity of implementation, which may be unique for every customer.

Ordering, imaging, and billing patterns are also diverse across institutions and patient populations. To ensure models are generalizable, they must be developed and tested by using data from multiple sites. For example, brain MRI protocols likely differ across institutions. A protocoting model must have access to these varied data for training and testing; however, these data must be harmonized to a common schema to be combined. This increases the complexity, time, and cost of model development. The ongoing adoption of standardized lexicons and communications standards such as common data elements (129) and Fast Healthcare Interoperability Resources (130) could help mitigate these issues by reducing variations in the input data structure, thereby allowing easier collection of multisite data.

There are also some underexplored areas in the radiology value chain that could benefit from machine learning applications. Missed appointments, particularly for MRI examinations, represent substantial lost revenue for radiology departments. Several studies have described the use of machine learning to predict no-shows for hospital and outpatient visits (131–133) and outpatient appointment and surgery scheduling (134,135). However, this work has not yet been extended to the radiology domain. The largest study in this area used a multivariate model to show the effect of median income and commute distance on missed or canceled appointments, but it did not use more advanced modeling or any EMR data (136). Another study used an XGBoost model only on structured data from the hospital radiology information system and appointment system and achieved an area under the receiver operating characteristic curve of 0.746; however, the model did not include more diagnostic information from the EMR (137). NLP and machine learning-based techniques could be used to process structured and unstructured data from the EMR to potentially achieve improved performance. Intelligent hanging protocols could be trained to automatically extract series information and display examinations according to the preferences of a radiologist, saving time during interpretation. Intelligent worklist optimization to ensure that radiologists read examinations for which they have the most experience or efficiency could improve diagnostic quality and turnaround times. Additionally, chatbots that interface with patients to answer questions or explain report findings could improve health literacy and patient confidence. These are just a few of the many potential areas of exploration in the development of radiology AI models.

Conclusion

Radiology AI software has become increasingly popular over the past several years. Whereas the majority of research and commercial software focuses on diagnostic or interpretive applications, there are large areas of potential improvement in upstream workflow, including protocoting, acquisition, reconstruction, and worklist management, and downstream applications such as reporting, follow-up, and billing and coding. In aggregate, these solutions could have a similar or even larger effect than most diagnostic AI software because of their applicability to a large number of cases and at multiple points in the radiology workflow.
Disclosures of conflicts of interest: Y.T. No relevant relationships. V.M. No relevant relationships. W.W. No relevant relationships. H.Z. No relevant relationships. A.P. Consulting fee or honorarium from QMC. N.B. No relevant relationships. M.H. No relevant relationships. E.K. Publications Ethics Committee for RSNA journals. N.S. Advisor to a resident who received an RSNA Research & Education Foundation grant. The project is entitled HL7: Shield: A Versatile HL7 Listener Software for Automated Follow-up Tracking; pending patent for HL7: Shield (Emory University is applying for a software patent for an AI approach to findings notification, this author listed as contributor). J.B. No relevant relationships. J.G. Future of Work Grant NIH NIBIB MIDRC grant (not related to this work); associate editor and trainee editorial board lead for Radiology: Artificial Intelligence. H.T. Consultant for BioData Consortium and Sirona Medical; grant from Kleiner Medical for institution; owner of Lightbox AI.

References


62. Tadavarthi et al.

63. doi: 10.21037/jim-2001-001.


Complementary approaches to problem solving in healthcare and public health: implementation science and human-centered design

Elizabeth Chen, Gila Neta, Megan C. Roberts

Keywords
Implementation science, Evidence-based strategies, Human-centered design, Design thinking, User-centered design, Translation, Health interventions, Health policy

INTRODUCTION
The timely, effective adoption and implementation of evidence-based practices, interventions, tools, programs, and policies (hereafter referenced as evidence-based practices) is important to improve health care delivery and patient outcomes. Implementation science (IS), user-centered design (UCD), and human-centered design (HCD) are three research approaches that focus on translating research evidence into the real world. In their recent article titled “A glossary of user-centered design strategies for implementation experts,” Dopp et al. [1] established a precedent for combining IS and UCD approaches and offered a new glossary of UCD strategies IS experts could use. In this commentary, we build upon this work by combining IS and HCD approaches and offering a how-to guide for IS experts to operationalize implementation strategies using HCD methods. Combining IS and HCD approaches is novel to health care research and practice, and we believe that these complementary approaches can be applied together to optimize the integration of evidence-based practices within clinical and public health settings.

IMPLEMENTATION SCIENCE
IS explores methods to effectively translate evidence-based care, interventions, and policies into practice to improve health [2]. By accounting for context and multilevel determinants, researchers and practitioners may better address implementation challenges for evidence-based practices and maximize their potential benefits on population health. The field leverages dozens of frameworks, theories, and conceptual models [3] to inform IS and uses a variety of measures and study designs [4] to understand implementation processes and develop and test implementation strategies [5]. More specifically, IS theories and frameworks can help (a) identify factors that may influence implementation processes or outcomes, (b) provide guidance for conceptualizing an implementation challenge and inform study hypotheses, including how to overcome barriers to implementation, and (c) select and tailor implementation strategies to address delivery gaps.

Implications
Practice: Human-centered design (HCD) methods can be used to consistently operationalize implementation strategies.

Policy: HCD and implementation science (IS), when used together, can provide an avenue for developing stakeholder engaged policy interventions and implementation strategies.

Research: Integrating HCD and IS is a novel approach and future research should be aimed at understanding which HCD strategies are most effective for operationalizing implementation strategies and how IS can be used to inform and evaluate HCD research.

Implementation strategies promote the integration of evidence-based practices into public health and health care settings. Powell et al. [6] identified 73 implementation strategies in their Expert Recommendations for Implementing Change study, of which many involve stakeholder engagement, such as conducting educational meetings, clinical reminders, and conducting local needs assessments to improve implementation outcomes, such as acceptability, adoption, appropriateness, costs, feasibility, fidelity, penetration, and sustainability [7]. These strategies can be selected to address specific multilevel barriers to implementation and improve implementation outcomes, which, in turn, strengthens the health impact of evidence-based practices [7]. For example, if a needs assessment uncovers low provider awareness of an evidence-based practice to improve...
asthma inhaler adherence, then educational meet-
ings with providers may be an effective implementa-
tion strategy for increasing adoption of this practice.
Methods for selecting and refining implementation
strategies for a given context are continuing to be
developed. Some recommended approaches for
selecting strategies include conjoint analysis, simu-
lation modeling, intervention mapping, and concept
mapping, among others [8–10].

In addition to implementation frameworks, out-
comes, and strategies, a broad variety of study de-
signs can be used to study implementation, including
effectiveness-implementation hybrid designs (which
includes effectiveness and implementation re-
search aims and data collection); mixed methods
(integrating qualitative and quantitative methods);
factorial designs (e.g., sequential multiple assign-
ment randomized implementation trial); two-level
nested randomized designs; cluster randomized con-
tral trials; crossover designs; and simulation models
among others [5,11,12]. Taken together, the field has
utilized a set of research methods to rigorously study
and evaluate the implementation of evidence-based
practices into public health and clinical settings.

HUMAN-CENTERED DESIGN
HCD is a repeatable, creative approach to problem-
solving that brings together what is desirable to
humans with what is technologically feasible and eco-
nomically viable [13]. Dopp et al. [1] offer a glossary
of USD strategies for IS experts, which focuses on
the individual for which a solution is designed (e.g.,
patient or practitioner), whereas HCD focuses on the
individual, those who are around them, and the sys-
tems in which the individual is a part. Dopp et al. [1]
offered this when comparing HCD to UCD:

The closely related approach of human-centered design
more explicitly seeks to integrate an innovation into
human activities and systems by considering indi-
guals beyond primary users (including those who
interact indirectly with the innovation, such as clinic
leaders who oversee implementation, as well as those
who are unintentionally affected by it, such as family
members of patients) in the design process.

Given the multiple levels of influence (e.g., patient,
provider, clinic, organization, and system) that can
impact successful implementation, IS experts could
benefit from combining a multilevel, HCD approach
to operationalizing implementation strategies.

Over the past 30 years, HCD has evolved from di-
verse disciplines, including computer science, visual
design, and architecture, and has been primarily
embraced in the private sector [14,15]. However,
the public sector has started to embrace HCD [16].
Recently, public health researchers have started to
apply HCD approaches and methodologies to
community-based participatory research projects as
a way to better understand the experiences of end
users (i.e., intended beneficiaries) and to codevelop
health interventions with them [17,18]. For this com-
mentary, the authors rely on the HCD process as de-
signed by IDEO, a leading global design company,
which has successfully used HCD to create ground-
breaking products like Palm pilots and Oral-B tooth-
brushes [19].

IDEO’s HCD process for problem-solving con-
sists of three distinct phases: the inspiration phase,
the ideation phase, and the implementation phase
[13,19]. After identifying a particular problem for
which a solution is desired, designers’ (i.e., those en-
gaged in HCD) first aim is to build empathy toward
and draw inspiration from individual users (e.g., pa-
tients, patients’ families, clinicians, and staff) through
in-depth conversations and experiences in Phase 1
of HCD [18]. The purpose of this first phase is not
to arrive at a solution; instead, the goals are to more
completely understand the intended users, the bar-
riers (i.e., “pain points” in HCD) they have experi-
enced given the problem, and the solutions (i.e.,
“workarounds”) they have found [13]. Second, in the
ideation phase, designers generate numerous ideas
for how to solve the problem, informed by the users’
thoughts, feelings, and experiences. Third, in the im-
plementation phase, designers quickly prototype (i.e.,
test) the different ideas with users to solicit immediate
feedback. This is achieved through designing short
experiments with low-fidelity prototypes. Low-fidelity
prototypes are simple versions of a solution, often
paper based, that are quickly produced to test broad
concepts [13]. Prototyping allows for the recombin-
ation and refinement of these concepts into a solution
that is desirable, feasible, and viable for a specific set
of users. These short iteration cycles help to secure
buy-in by repeatedly engaging collaborators, which
also allows for a smoother, broader implementation
of the product or service at the conclusion of the pro-
ject [13,18]. To make HCD more accessible to the
general public, IDEO’s nonprofit arm, IDEO.org,
published The Field Guide to Human-Centered Design
in 2015 [13]. This field guide includes HCD mindsets,
methods, case studies, and resources.

COMBINING HCD AND IS APPROACHES
We typically consider IS when there is an evidence-
based practice with proven efficacy that has not
yet been effectively implemented in health care
or community settings. Through IS, researchers
can develop and test strategies to improve care de-
livery of evidence-based practices [5]. We might
consider HCD when developing a new interven-
tion. Both fields acknowledge the importance of
multiple stakeholder perspectives, iterative study
cycles to optimize outcomes of interest, and consid-
eration of the end users to improve implementation
in real-world settings. Based on these complemen-
tary strengths, we believe that IS and HCD can be
combined to provide “client-centered” approaches for implementing health care and public health practices, and we offer two ways to conceptualize how to integrate the two approaches.

First, we could view HCD as a process that occurs toward the beginning of the translational research pathway (i.e., discovery), and IS on the distal end of the pathway. Indeed, the final phase of HCD includes an implementation phase, so there are opportunities to integrate these two fields in the effort to develop patient-, provider-, and system-centered implementation strategies across the research continuum. IS frameworks, measures, and study designs could play a key role in strengthening the rigor of HCD research projects in the implementation phase.

Second, we could view HCD as a practical method for operationalizing implementation strategies. As previously outlined, IS leverages strategies to optimize the delivery of interventions and stakeholder engagement is paramount. HCD offers IS a set of methods (i.e., activities) to engage with intended beneficiaries [13,18,19]. Therefore, HCD may provide a new approach for selecting, optimizing, and operationalizing implementation strategies.

HCD methods may be particularly useful for operationalizing implementation strategies [6] within four of the nine broader implementation strategy categories identified by Waltz et al. [20]: use evaluative strategies, adapt and tailor to context, develop stakeholder interrelationships, and engage consumers. Publications have provided guidance on how to select, tailor, and specify the 73 implementations strategies [6,9], but there is still little guidance for how to execute specific implementation strategies; that is—how do researchers actually apply these implementation strategies in the field? For example, if researchers want to employ “involving patients/consumers and family members” as an implementation strategy in their research, how do they operationalize this implementation strategy? Operationalizing implementation strategies through the use of low-cost, accessible HCD methods could help researchers and practitioners assess which implementation strategies are most acceptable and feasible, as well as how these strategies should be executed. Using HCD methods to operationalize implementation strategies will also provide implementation scientists a shared language with those who practice HCD and vice versa. Fig. 1 below summarizes the interrelationship between HCD and IS and illustrates how combining these approaches can impact population health. In order to further illustrate how HCD can enhance IS and how IS can enhance HCD, we present the case study below.

**CASE STUDY: THE REAL TALK APP**

This case study reports the development of a new mobile app where the first author and her team used HCD as the approach for intervention development and implementation. We will report the activities completed by the team in the development and implementation of the Real Talk app and note (a) where HCD methods offered ways to operationalize IS strategies and (b) where IS could have enhanced this HCD project in identifying determinants of implementation and offering ways to evaluate both effectiveness and implementation outcomes.

**About the Real Talk app**

In 2017, the first author and the two other cofounders of the technology nonprofit MyHealthEd, Inc., applied HCD to build and launch the first version of their Real Talk app for teenage users aged 13–15 [24]. To date, the app has more than 15,000 users in all 50 states and in more than 125 countries. The purpose of the app is to build a community for teens around taboo health issues, such as sexual health and mental health, and let users know that they are not alone. In the app, users can browse, share, and react to stories on a variety of topics, as well as connect with high-quality online resources from organizations like amaze.org and TeensHealth.

**How HCD can enhance IS**

While the MyHealthEd, Inc., team did not apply an explicit IS framework through their design work, they did apply several implementation strategies, including: (a) involve patients/consumers and family; (b) conduct cyclical small tests of change; and (c) intervene with patients/consumers to enhance uptake/adherence (Table 1). The team applied these implementation strategies by using the Inspiration, Ideation, and Implementation methods from IDEO.org’s field guide [13] as described below. The examples below illustrate how HCD methods could be used to operationalize IS strategies.

**Involving patients/consumers and family**

In order to operationalize “involving patients/consumers and family” as an implementation strategy, the MyHealthEd, Inc., team involved teenagers aged 13–15 (intended users) early in the HCD process. IDEO.org’s field guide [13] offers a number of specific HCD methods (i.e., activities) to involve end users that include activities like Card Sorts, Conversation Starters, a Guided Tour, or a Resource Flow. The team used the field guide’s Card Sort method to answer questions regarding where teenagers felt most comfortable talking about sex and/or relationships. In order to do this, the team created cards with the following options: school, home, bus, church, friend’s house, and other. Then, the team asked the teenagers to rank the cards in terms of comfort level. After meeting with teenagers across the country and completing the same activity, the team quickly realized that teenagers did not want to talk about sex and/or relationships in school, so they...
moved away from thinking that they might implement their intervention in schools. This, along with other insights gained through the formative research process, led to a direct-to-consumer approach via a native smartphone app rather than a school-based approach.

*Conduct cyclical small tests of change*

As part of the ideation phase, the team then conducted dozens of small cyclical tests of change to get feedback from teenagers and other stakeholders (e.g., parents, health teachers, school administrators, and faith leaders) on different features and design elements in the app. This HCD phase directly relates to the implementation strategy for conducting cyclical small tests of change but adds a more specific methodology. The Real Talk app’s user interface and user experience designers used the software InVision to create clickable prototypes of the different app screens. Then, the team used the IDEO.org’s field guide [13] Rapid Prototyping method to share the InVision prototypes with intended users, collect reactions and data, and make adjustments. For example, the team heard from intended users that they would prefer to interact with sexual health content via stories rather than facts or statistics. Teens also wanted the ability to share their own stories through the app, so the MyHealthEd, Inc., team rapidly tested different versions of the story submission experience. One major test compared a form-based study submission experience (e.g., users submit their entire story by typing it into a box) with a chatbot experience (e.g., users respond to prompts from a chatbot to share their stories piece by piece). After testing these two options, the team found that the majority of their intended users preferred the more interactive chatbot because it was as easy and familiar as text messaging a friend. This resulted in building the interactive story submission feature rather than the form-based feature.

*Intervene with patients/consumers to enhance uptake/adherence*

Prior to implementation and dissemination, the MyHealthEd, Inc., team was also very intentional about engaging with teenagers to develop strategies together to increase uptake (i.e., app downloads) and adherence (app usage). Through using the Co-Creation Session method from IDEO.org’s field guide [13], the team convened a group of teenagers to design alongside them by empowering them to jointly create and brand the solution. Specifically, the team worked with teenagers to name the health app. Teenagers came up with the name “Real Talk” because it captured the raw or “cringey” nature of the stories submitted by other teenage users, but it did not overtly signal that the app covers sexual health education topics. Teenagers wanted a resource like this to be discreet and this insight informed the app logo (two generic white chat bubbles without signals to sexual health content). Lastly, the team held multiple Co-Creation Sessions for teenagers..
to design and pitch their own sexual health apps. Both the drawings and language that the teenagers used to pitch their app concepts to other teenagers shaped the language and images shown for the app as it is advertised in the iTunes App Store. The description reads:

Real Talk is a community for teens packed with real stories about cringey moments. Browse through stories, search for topics that matter most to you, and use emojis to share your reactions. You can also share your own story directly in the app - it’s as easy as texting with a friend. Join thousands of teens who already use and love Real Talk. With totally relatable stories, you won’t feel as alone as you go through the struggles of growing up [24].

Additional language offered by teenagers in Co-Creation sessions was used in other marketing and outreach materials. Applying this HCD Co-Creation Session method led to an app description that was more teen-friendly than what the MyHealthEd, Inc., team initially envisioned before collaborating with the teens. The use of this specific HCD method provided a protocol to inform the language used to attract new users of the app. In the first year of launching the app, Real Talk was downloaded more than 10,000 times by teenagers across the globe.

### How IS can enhance HCD

While the MyHealthEd, Inc., team considered implementation from the start, they did not employ an IS framework or study design. As mentioned earlier, IS can enhance HCD by identifying multilevel determinants of implementation and offering more rigorous evaluation of an evidence-based practice. IS frameworks that focus on multilevel determinants of implementation can provide structure to studying the implementation of an evidence-based practice. HCD largely considers determinants for implementation on the individual level in the case of Real Talk app from the perspective of teens and their families. However, when considering the implementation of an evidence-based practice, it is essential to consider the multilevel determinants that impact an individual’s use of that practice. Investigating multilevel determinants iteratively throughout the development and evaluation of the Real Talk app could help hone in on the appropriate implementation strategies, as well as provide a more holistic view of the effectiveness of the practice. For example, exploration of multilevel determinants for implementing Real Talk outside of...
patients and families could prevent disconnects between the patient and their providers, clinics, retail stores, and pharmacists, who also play a role in their sexual health. As individuals act within systems, it is important to study, act upon, and evaluate across multiple levels rather than within a vacuum on the individual level. A number of IS frameworks do well in systematically providing a multilevel perspective on implementation determinants and processes.

Next, IS frameworks, measures, and study designs could provide structure for evaluating the effectiveness and implementation of practices, particularly throughout the rapid prototyping and cyclical experiments. In addition to assessing determinants of implementation, as mentioned above, IS frameworks are available to provide structure to the evaluation of the implementation of evidence-based practices and newly developed innovative solutions to improve health [7,23]. Implementation outcomes have been outlined by the field and include measures such as acceptability, appropriateness, feasibility, and costs, among others [7]. Assessing these implementation outcomes, as well as effectiveness outcomes, is essential for understanding the total impact of Real Talk on adolescent sexual health outcomes. Hybrid effectiveness-implementation designs allow for more rigorous testing and documentation of rapid prototyping cycles by incorporating the exploration of not only effectiveness outcomes but also implementation outcomes. For example, teenagers could have been randomized to view one of three sets of marketing materials each with different content. Then, the team could assess the implementation outcomes (e.g., acceptability, appropriateness, and download rates) and the effectiveness outcomes (e.g., sexual health knowledge) of the teenagers and determine which of these three sets of marketing materials leads to the strongest outcomes. These data on implementation outcomes are key for optimizing, scaling-up, and implementing the intervention in different settings (i.e., scale out) if found to be effective and, if not effective, may point to reasons why the intervention failed to have the intended impact on health.

CONCLUSION

Overall, HCD offers specific methods that can readily operationalize implementation strategies to improve the translation of health innovations into practice. Using HCD to execute implementation strategies provides a set of tools for implementation researchers to develop and test implementation strategies associated with health interventions. Additionally, IS offers specific approaches to identifying and analyzing multilevel systems and barriers to implementation, as well as rigorous study designs that will enhance HCD research by providing guidance for how to document and evaluate the iterative, cyclical experiments [22,23]. By combining the processes and tools from HCD and IS, we believe that health care and public health researchers can develop a common language to improve implementation outcomes and health outcomes for patients and communities.

Funding: Megan Roberts is funded through the National Center for Advancing Translational Sciences, National Institutes of Health through Grant KL2TR002490.

Compliance with Ethical Standards

Conflicts of Interest: E.C. was a cofounder and employee of MyHealthEd, Inc., the technology 501 (c)3 nonprofit that created the Real Talk app mentioned in this manuscript, from August 2016 through June 2019. G.N. and M.C.R. declare that they have no conflicts of interest.

Authors’ Contributions: E.C. and M.C.R. conceptualized this study; E.C., M.C.R., and G.N. analyzed the data; and E.C., M.C.R. and G.N. co-authored this manuscript.

Ethical Approval: This article does not contain any studies with human participants performed by the authors. This article does not contain any studies with animals performed by any of the authors.

Informed Consent: This study does not involve human participants and informed consent was, therefore, not required.

References

Artificial Intelligence and the Implementation Challenge

James Shaw1,2, PhD; Frank Rudzicz3, PhD; Trevor Jamieson1,4, MD, MBI; Avi Goldfarb5, PhD

1 Women's College Hospital, Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada
2 Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada
3 International Centre for Surgical Safety, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada
4 St Michael's Hospital, Toronto, ON, Canada
5 Rotman School of Management, University of Toronto, Toronto, ON, Canada

Corresponding Author:
James Shaw, PhD
Women's College Hospital
Institute for Health System Solutions and Virtual Care
76 Grenville Street
Toronto, ON, M5G2A2
Canada
Phone: 1 4163236400
Email: jay.shaw@wchospital.ca

Abstract

Background: Applications of artificial intelligence (AI) in health care have garnered much attention in recent years, but the implementation issues posed by AI have not been substantially addressed.

Objective: In this paper, we have focused on machine learning (ML) as a form of AI and have provided a framework for thinking about use cases of ML in health care. We have structured our discussion of challenges in the implementation of ML in comparison with other technologies using the framework of Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies (NASSS).

Methods: After providing an overview of AI technology, we describe use cases of ML as falling into the categories of decision support and automation. We suggest these use cases apply to clinical, operational, and epidemiological tasks and that the primary function of ML in health care in the near term will be decision support. We then outline unique implementation issues posed by ML initiatives in the categories addressed by the NASSS framework, specifically including meaningful decision support, explainability, privacy, consent, algorithmic bias, security, scalability, the role of corporations, and the changing nature of health care work.

Results: Ultimately, we suggest that the future of ML in health care remains positive but uncertain, as support from patients, the public, and a wide range of health care stakeholders is necessary to enable its meaningful implementation.

Conclusions: If the implementation science community is to facilitate the adoption of ML in ways that stand to generate widespread benefits, the issues raised in this paper will require substantial attention in the coming years.

(J Med Internet Res 2019;21(7):e13659) doi: 10.2196/13659

KEYWORDS
artificial intelligence; machine learning; implementation science; ethics

Introduction

Artificial intelligence (AI) has become a topic of central importance to the ways in which health care will change in the coming decades, with recent commentaries addressing potential transformations in clinical care [1,2], public health [3], and health system planning [4]. AI is a general purpose technology (GPT), which means it represents a core set of capabilities that can be leveraged to perform a wide variety of tasks in different contexts of application [5]. Understanding the core capabilities of AI as a GPT, and the ways in which it stands to be incorporated into health care processes, is essential for the implementation research community to contribute to promoting a positive place for AI in the future of health care. We believe that AI has the potential to substantially reconfigure health care, with implications that reach beyond enhancing the efficiency and effectiveness of care delivery. Due to this potential, we
suggest that implementation science researchers and practitioners make a commitment to more fully consider the wider range of issues that relate to its implementation, which include health system, social, and economic implications of the deployment of AI in health care settings.

We suggest that the most appropriate language for discussions of AI in health care is actually to discuss machine learning (ML), which is the specific subfield of AI that is currently making the most impact across industries. We then focus on 2 questions about the deployment of ML in health care. First, how should ML be understood in terms of its actual use cases in health care? This question addresses the nature of ML as an implementation object [6,7] in health-related contexts. We present a basic framework for thinking about use cases of ML in terms of decision support versus automation and elaborate clinical, operational, and epidemiological categories of these use cases.

Second, what are the unique challenges posed by ML that may require consideration during an implementation initiative? As opposed to focusing on strategies for the adoption of digital technologies in general, which has been addressed extensively in other literature [8-10], we focus on what we understand to be the most important risks arising from the implementation of ML in health care. Our discussion of the risks associated with implementing ML in health care is guided by the work of Greenhalgh et al in the framework for theorizing and evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability (NASSS) of health and care technologies [8].

The NASSS framework is based on the premise that when considering influences on whether and how a technology is successfully taken up and used, it is important to keep in mind that “it is not individual factors that make or break a technology implementation effort but the dynamic interaction between them” [8]. The NASSS framework outlines a range of considerations that are relevant to understanding how a technology might be adopted across an entire region or health system, ranging from a focus on the particular health condition in the clinical scenario to the wider political, regulatory, and sociocultural system in which it is to be embedded. In our paper, we examine ML as a GPT that has the potential to apply across clinical conditions and focus our analysis on elements of the NASSS framework: the technology, its value propositions, and the adopters, organizations, and systems into which it might be introduced. We emphasize the evolutionary nature of ML as a GPT and explicitly acknowledge that it will continue to develop and change over the coming years, which is also an important feature of the NASSS framework. We conclude by advocating for further research on the risks posed by ML from an implementation science perspective.

AI has been described in many ways. Using the framing in Agrawal et al, we emphasize that recent advances in AI can be best understood as “prediction technology” [11]. Quite simply, prediction is defined for this purpose as “taking information you have, often called ‘data’, and using it to generate information you don’t have” (PM, p. 24). This newly generated information estimates the true information that is missing, leading to the potential for people and technology to take actions that may have otherwise been based on less accurate information.

Predicting illness episodes that might be experienced in the future is an obvious application of AI in this sense, but prediction as we have defined it has many other uses as well. Examples include an automatic translator predicting the phrases of Spanish that correspond to a particular set of phrases in English or a chat bot predicting the most appropriate cluster of words in response to a given query. These examples might not represent the very intuitive understanding of prediction that we have become used to in everyday usage or the way we tend to think of prediction of health-related events and outcomes in health care. However, they represent the prediction of information that we do not have based on information we do have and point toward the potentially widespread applications of AI as a GPT.

The phrase “predictive analytics” is very intuitive with regard to defining AI as a prediction technology, using advanced computer algorithms to predict health-related events from existing data in ways that exceed the ability of individual researchers applying individual analyses [12]. However, AI opens new opportunities for prediction beyond the familiar predictive analytics for hospital admissions, length of stay, and patient survival rates. As a process of filling in missing information, better and cheaper prediction is already being used in new areas, from transcribing audio to enhancing security to informing diagnoses.

At its core, current applications of AI bring statistical modeling, computer code, and advanced computing power to bear on large amounts of representative data. In his recent commentary on the potential of deep learning (a form of AI) to transform health care, Hinton gave the example of deciding whether a patient has a particular disease and explained that a common approach would be to use a simple logistic regression (using data to predict a binary outcome: the patient has the disease or does not). However, he suggested that if there are extremely high numbers of potential influences or predictors of whether the person has the disease, many of which may interact with one another, the prediction challenge becomes much more complex. This is especially the case where we have imperfect knowledge of the causes and correlates of a particular disease. This example also pertains only to binary queries specifically about whether a patient has a single disease, which is different from the typical reasoning processes involved in differential diagnosis among clinicians, where multiple confounding, interdependent outcomes must be considered [13,14].

Specific applications of AI can fall under distinct categories, with AI serving as an umbrella concept, covering more specific frameworks. In this paper, we are primarily concerned with the subdomain of AI referred to as ML in which statistical models are automatically (or semiautomatically) induced from data according to some criterion (eg, best expected discriminative power or maximum likelihood given to training data). This means that complex statistical models capable of executing advanced predictions are generated in part by using data to train the model to achieve a particular goal.
Often, ML involves *supervised* methods that categorize data points, for example, as images of skin cancer or otherwise given datasets in which all data points (or at least a substantial subset) are associated with a label, ordinal, or category that is meant to be predicted or inferred [15]. This process requires datasets that have the appropriate labels indicating what the data means; in the example of images of skin cancer, each data point would be labeled according to its representation of a mass as *malignant* or *benign* or some variation thereof. Given these labels and the statistical models they help to train, ML can be very effective at determining the category in which any newly available individual data point belongs, thereby being useful in the effort to, for example, identify malignant cancers based on particular images [16].

Much of the power of modern ML also derives from *unsupervised* pattern recognition, in which hidden (or *latent*) aspects of the data are automatically identified by the algorithms and exploited according to the aforementioned criteria. Unsupervised ML can often identify patterns in the data that humans do not even think to look for. Often, these hidden aspects are nonlinear combinations of many parts of the input.

ML can also improve its ability to *take actions* according to these induced hidden patterns and particular functions of cost and reward in a process called *reinforcement learning*. For example, ML can dynamically adapt survey questions to more quickly identify possible diseases [17], dynamically avoid potential communication breakdowns during speech conversation in the assessment of dementia [18], and even recommend treatments directly when using structured institutional data [19]. As so much health care information can be represented digitally, the potential of ML to improve health care practices is profound.

## Methods

### Use Cases of Machine Learning in Health Care

In the remainder of our paper we refer primarily to ML as opposed to AI, focusing our analysis on the concrete possibilities of ML in health care. We can think about use cases of ML in health care in 2 broad ways. The first is through *decision support*, wherein ML algorithms are used to provide some form of input into human decision making. An example is where an algorithm is used to provide more accurate predictions of the outcome of a particular procedure given a particular clinical presentation. This helps to inform a human decision about whether a given procedure is the best course of action. The second is through *automation*, wherein algorithms are used not only to predict an output but also to take action to achieve a particular outcome. An example is the automatic transcribing of a clinical note when dictated into a computer program, resulting in a complete note being added to a patient’s record (technically referred to as Automated Speech Recognition).

These 2 broadly defined categories of use cases can be thought of as applying to various types of tasks in health care, and we suggest it is instructive to consider 3 types of tasks as most relevant for the implementation of ML for health: clinical, operational, and epidemiological.

Clinical tasks refer to health-related assessment, intervention, and evaluation, generally performed by qualified health care providers, for example, determining a differential diagnosis. Operational tasks are those related to activities that are ancillary to clinical tasks but necessary or valuable in the delivery of services, such as generating, storing, and retrieving medical records. Finally, epidemiological tasks are those related to more accurately identifying the health needs and outcomes of a set of people within a given population. An example is the development of a warning system for disease outbreak. As epidemiological use cases of ML are related to enhancing the ability of humans to make decisions in the other categories described here (clinical or operational), there are no examples of pure automation for epidemiological tasks that contain an output other than informing a human decision. Hypothetical examples of both decision support and automation are given under each of these categories in Table 1.

This table presents a basic framework for thinking about use cases of ML in health care as falling into 2 primary categories: decision support and automation. These use cases apply in categories of clinical, operational, and epidemiological tasks. As no examples of pure automation exist for epidemiological tasks, no example is presented in that cell.

The considerations most pertinent to the implementation of ML will depend on the particular use case being proposed in a given implementation initiative, and the categories outlined in Table 1 provide a framework for understanding those use cases. The NASSS framework and other work in implementation science for digital health technologies emphasize the importance of attending to the particular value proposition that a new technology offers for health care stakeholders [8,9]. The value proposition of digital technology might be different for different stakeholder groups, and implementation frameworks direct attention to the implications of newly introduced technologies for patients, health care providers, managers, health policymakers, and others [8,25,26]. The clinical, operational, and epidemiological task types presented in Table 1 will correspond to different value propositions for different stakeholder groups, meaning that specific applications of ML might preferentially benefit one group over another, for example, identifying a scheduling process to maximize efficiency in operating costs might preferentially benefit managers over health care providers inconvenienced by a new system. Understanding how value propositions differ for the various stakeholders implicated in a given implementation of ML is an essential consideration for successful adoption and use.
Table 1. Examples of use cases in each category of application.

<table>
<thead>
<tr>
<th>Type of use case</th>
<th>Clinical</th>
<th>Operational</th>
<th>Epidemiological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision support</td>
<td>Producing a more accurate prediction of the likely outcome of a particular intervention [20]</td>
<td>Identifying potential staff scheduling changes related to forecasted emergency room volumes [21]</td>
<td>Warning systems for disease outbreak [22]</td>
</tr>
<tr>
<td>Automation</td>
<td>Automatically altering insulin treatment in response to monitored glucose-insulin dynamics [23]</td>
<td>Use of robotics for operational tasks in dementia care, such as meal delivery [24]</td>
<td>N/A[^d]</td>
</tr>
</tbody>
</table>

[^a]: Tasks related to the assessment, intervention, and evaluation of health-related issues and procedures, generally performed by qualified health care providers.

[^b]: Tasks related to activities that are ancillary to clinical tasks but necessary or valuable in the delivery of services.

[^c]: Tasks related to more accurately identifying the health needs and outcomes of people within a given population.

[^d]: Not applicable.

The potential value propositions of an ML technology offering decision support versus one offering automation are very different and bring along different sorts of implementation issues. The implementation of decision support systems in health care that do not include applications of ML have been well studied and the difficulties include perceived challenges to autonomy, lack of time, and dissatisfaction with user interfaces [27,28]. Implementation initiatives involving decision support applications of ML will need to consider this past work to develop implementation strategies that more effectively address known challenges.

Implementation initiatives involving automation are likely to face some similar and some different challenges. For example, stakeholder views on the introduction of automated robotics into a variety of health care settings found a widespread lack of interest and understanding and fear of the ways in which work would be disrupted and distributed [29]. Although automation has existed in health care for decades through technologies such as heart rate monitors, the question of how acceptable stakeholders will perceive new forms of automation to be remains an important issue. This point raises the overarching issue of the extent of automation that is possible through applications of ML, linked to speculation about whether ML will mostly augment or actually replace health care providers’ work [1,30].

**Augmentation and Replacement of Health Care Work**

We agree with a growing chorus of health care providers and researchers who suggest that ML will primarily serve to augment as opposed to replace the work of humans in the provision of health care in the near term [31], despite applications of automation in health care. This is because the role of ML in the current generation of capabilities functions at the level of the task, and not at the level of an entire job. Agrawal et al explained that “the actual implementation of AI is through the development of tools. The unit of AI tool design is not ‘the job’ or ‘the occupation’ or the ‘the strategy’, but rather ‘the task’.” (p. 125). Therefore, for a health care provider to be entirely replaced, every single task performed by that provider would need to be automated by an ML tool or handed off to a different human.

The complete automation of the full range of human tasks involved in providing clinical care is not yet possible; activities such as making treatment decisions based on a differential diagnosis that integrates data from laboratory investigation, visual observation, and patient history are still too complex for automation. In emphasizing this point, we are suggesting that although much of the hype about AI (and specifically ML) in health care has focused on its potential role in automating processes of health service delivery, it is more likely that near-term applications of ML will fall under the category of decision support.

Further comments about prediction tasks and decision tasks will help to clarify this point. As stated earlier, ML applications fundamentally perform some form of prediction. The specific instance of prediction that the application is performing may be thought of as the prediction task, which may be paired with a complementary decision task. The decision task is where the newly generated information is used to select a particular action in a given context. In applications of ML that function as decision support, the decision task is performed by a human. As ML diffuses, an important new challenge for health care providers is to make choices using the predictions that arise from decision support applications of ML, involving new forms of input to clinical thought processes related to risks, benefits, and previously unrecognized influences on health. The examples of decision support in Table 1 involve generating better information to inform human decision making.

In applications of ML that function as automation, both the prediction task and the decision task are accomplished by machines. A clear example is self-driving cars. The sensors surrounding the car enable predictions of the best direction in which the car should travel. However, it is the selection from a predetermined set of actions and execution of one action over another that makes self-driving cars an example of automation as distinct from one of decision support. ML is not yet sophisticated enough to complete these selection and execution functions for many health care tasks, across both clinical and operational levels.

As prediction tasks become more amenable to being performed by ML, decision tasks become more valuable [5,32]. This is because predictions are improved, meaning that decisions can be made with greater confidence and impact. The enhanced value of these decisions represents the potential value of ML as a decision support tool and illustrates the potential breadth of value propositions that could arise from this technology with...
a wide range of implications for the implementation process. However, for decision support to be valuable in health care, the outputs of algorithms must have a clear entryway into the human decision-making processes that pervade health service delivery. This points us toward one of a series of important issues raised from an implementation science perspective on the introduction of ML in health care settings, which we turn to next.

Results

Unique Considerations for Implementation Science

We have described use cases (and attendant value propositions) of ML in health care as more likely relating to decision support and less likely to automation, which begins to illustrate the implementation object of focus in ML initiatives [6,7]. In many cases of decision support, the implementation object is actually not all that different from the statistical tools that are already used as part of common practice, such as risk prediction. In cases of automation, there are similarly many examples of technologies that have already been successfully implemented in health care settings (such as automatic transcription mentioned earlier). However, ML as a GPT raises a number of issues that run across use cases and might be anticipated as unique in comparison with implementation projects for other digital technologies.

Best practices of implementation for digital innovations [8,9,33] will be fundamental to the adoption of ML in health care. Here, we discuss considerations that might appear in implementation projects involving ML that may be less likely to appear in implementation projects involving other digital technologies and yet stand to have a potentially strong influence on the success of such projects. We organize this section based on distinct levels of consideration that are presented in the NASSS framework that we have not yet addressed [8,26]: health care providers, patients and the public, health care organizations, and health policy and systems. Although we consider the primary considerations of health technology vendors working on the development of ML application in health care to be outside the scope of this paper, we acknowledge this is a gap in the literature that requires attention.

Health Care Providers

Health care providers are those responsible for doing the actual work of health care delivery and are being increasingly expected to adopt and use new technologies in health care environments. We suggest that the core considerations or risks of the implementation of ML for health care providers will fall into the categories of meaningful decision support and explainability.

Meaningful Decision Support

For ML to function as decision support in a way that is valuable to health care stakeholders, the outputs of algorithms must have a meaningful entryway into decision making. From an operational or epidemiological perspective, isolated analyses of risk prediction may help to inform resource allocation and subsequent analysis decisions fairly simply. However, from a clinical perspective, algorithms that perform isolated risk prediction may be less useful. Clinical decision making is a complex process involving the integration of a variety of data sources, incorporating both tacit and explicit modes of intelligence [34-36]. To inform this decision-making process more intuitively, attention is increasingly being devoted to communication tools such as data visualization [37]. The nature and value of these communication tools are central to the implementation process, helping to determine whether and how algorithmic outputs are incorporated in everyday routine practices. This point primarily relates to the decision support use case across clinical, operational, and epidemiological tasks.

Explainability

There is a growing concern in the AI community related to the explainability of the results achieved by ML algorithms, wherein the ways in which algorithms enhance the performance of prediction can often not be understood [38]. As a result of the processes described earlier in this paper, the ways in which data are being used to train algorithms cannot be traced out in sequential, logical detail. Hence, the actual ways in which models achieve their results are in some instances not knowable even to the computer scientists who create them. Evidence-based medicine rests on a foundation of the highest standards of explainability: medical decision making aspires to incorporate a sound understanding of the mechanisms by which diseases and their treatments function and the particular treatments that have demonstrated the greatest benefits under particular experimental circumstances (in addition to patient needs and values [35,39,40]). The lack of understanding of those mechanisms and circumstances poses challenges to the acceptability of ML to health care stakeholders. Although the issue of explainability relates clearly to decision support uses cases of ML as explained here, the issue may apply even more profoundly to automation-focused use cases as they gain prominence in health care.

Patients and the Public

The issues of public trust and public input into the governance of ML initiatives in health care have been widely discussed as the popularity of AI has grown, with advocates suggesting that future developments of AI ought to be explicitly supporting a broader public interest. We suggest that 2 pairs of issues frame the risks of ML related to patients and the public. The first pair is privacy and consent and the second is representative data and algorithmic bias.

Privacy and Consent

The training of ML models requires large amounts of data, which means that applications of ML in health will likely rely on health-related data from patients and the public. As governments and other actors internationally become interested in developing applications of ML, health-related data are increasingly made available to private entities with the capability of producing AI applications that are relevant to peoples’ health [41-43]. Currently, data from wearable devices such as smart watches and mobile apps are not widely covered by health information legislation [44], and many health-related apps have unclear consenting processes related to the flow of data generated through their use [45]. Furthermore, data that are de-identified may be reidentifiable when linked with other datasets [46]. These considerations create major risks for
initiatives that seek to make health data available for use in the development of ML applications, potentially leading to substantial resistance from health care providers such as that seen in primary care in Denmark in recent years [42]. This will be particularly important for population and public health use cases that require data from very large segments of the population. The meaning of consent and strategies to maintain patient privacy are central considerations to ML implementation initiatives. The related issues of privacy and consent pertain especially to clinical and epidemiological use cases of ML in both decision support and automation categories, as data from patients or the public are essential to train algorithms in these areas (whereas operational use cases may only rely on other forms of data, such as clinical scheduling histories).

**Representative Data and Algorithmic Bias**

Algorithms are only as good as the data used to train them. In cases where training data are partial or incomplete or only reflect a subset of a given population, the resulting model will only be relevant to the population of people represented in the dataset [47]. This raises the question about data provenance [30,48] and represents a set of issues related to the biases that are built into algorithms used to inform decision making. One high profile example was the hiring bias exhibited when algorithms were used to make hiring decisions at Amazon, resulting in only men being advanced to subsequent stages of hiring [49]. This is notable in part because the algorithm performed extremely well based on the available data, simply extending the bias that already existed in hiring practices at the company. When applied to health care of public health, data provenance and potential bias in training data represent important issues that are likely to be of major concern for the stakeholders involved in the implementation of an ML initiative. Public health has health equity as a primary goal, and representativeness in terms of which populations can be addressed by an ML initiative will be a central consideration.

A further challenge with the nature of the data on which algorithms are trained relates to concept drift, a phenomenon where data on which an algorithm is trained change over time (or become out of date), which changes the performance of the algorithm as new data are acquired [50]. The possibility of concept drift means that those overseeing the performance of ML-based technologies in health care must identify strategies to determine how well the algorithm deals with new data and whether concept drift is occurring. Applications to support this effort are emerging in the literature [51].

The issues addressed here apply most clearly to ML applications that use patient data to inform clinical and epidemiological use cases that enhance clinical care and health system planning. And although the use of public data will likely be the most contentious issue in this domain, the challenges of representativeness and bias apply to all ML use cases across decision support and automation domains.

**Health Care Organizations**

Health care and public health systems are composed of independent organizations that need to develop and execute strategies within the limits of the resources available to them. Organizations have been the driving force behind the adoption of many innovations in health care and have a collection of considerations that are unique from the broader systems of which they are a part. We suggest that the issue of security and computational resources become particularly important for organizations as they adopt ML initiatives in health care and public health.

**Security**

As data are collated and stored for training ML models, the risk and potential severity of security breaches grows. The global attack of health care organizations using WannaCry ransomware in May 2017 shows the vulnerabilities of even well protected health data to malicious interests. This particular attack is estimated to have affected 200,000 systems in over 150 countries, indicating the potential scope of security problems as the value of data grows [52]. Strategies to prevent such security breaches on Web accessible health data are now being proposed in the literature [53,54], and the high profile of security issues makes this a particularly important issue as ML applications develop in health care and public health. The issue of security transcends any particular use case of ML and includes any applications or analysis that relies on big data more generally.

**Computational Resources**

Advanced applications of ML require substantial computing power, with some predictive analyses and training models requiring up to several weeks to run. The more extensive the computing support, the more efficient ML applications will become, raising the question of the cost and availability of such advanced computing power for health care organizations. Health care is publicly funded in many countries around the world, and public support to secure the resources to fund the necessary computing power may not be present. Cloud-based analytics present an opportunity and a challenge for health-related organizations in relation to the issue of computational resources.

Cloud-based data analysis means that organizations would not need to own computational resources directly [55] but also introduces the potential challenges of data safety. These issues are relevant to the training phase of a newly developed algorithm, but of course, less computing power is required to simply apply algorithms that have been generated and trained elsewhere. How data are stored and processed is thus also an important consideration in ML implementation initiatives. The issue of computational resources also applies more generally than any given ML use case, related to the development and functioning of many kinds of AI algorithms.

**Health Policy and Systems**

The challenges associated with ML initiatives at the level of health policy and systems are extensive. These include broad legislative frameworks related to emerging health-related technologies more generally [56] and to the innovation procurement systems that vary across health system settings [57,58]. The policy issues presented by ML in health care are beginning to garner more attention [42,43], but here we present one issue that we have not seen addressed in health care or public health literature: the challenge of scalability.
Scalability and Normal Accidents

A major challenge that extends beyond any single implementation of ML, and therefore requires a system-wide view, relates to the scalability of ML. Scalability in this sense refers to the unanticipated effects of the appearance of multiple ML technologies that will inevitably interact with one another by some means. As applications of ML proliferate across health care and public health, eventually some algorithmic outputs will confront others. The effects of this interaction are impossible to predict in advance, in part because the particular technologies that will interact are unclear and likely not yet implemented in the course of usual care.

Health care represents what Charles Perrow referred to as a complex system or a system in which processes are tightly linked to one another and interact in unintended ways in the effort to achieve the goals of the system [59]. This acknowledgement has led to the high reliability movement in health care and other industries [60], intending to implement management strategies that could mitigate against the risk of disasters arising from such immense complexity. Perrow’s work was titled Normal Accidents: Living with High Risk Technologies, suggesting that in systems characterized by complexity and the use of advanced technologies, accidents are bound to happen [59]. This basic point about the seeming inevitability of accidents in the context of complex systems and new technologies underscores the significance of the scalability challenge of ML in health care. We suggest that implementation scientists will need to consider the unintended consequences of the implementation and scale of ML in health care, creating even more complexity and greater opportunity for risks to the safety of patients, health care providers, and the general public. ML safety will likely need to become a dedicated focus of patient safety research internationally. This point about scalability frames the broader challenge for implementation scientists who are committed to a system-wide perspective on health innovations and relates not only to each type of use case identified in our framework but also to the interactions between them as well.

Discussion

Intersecting Issues in the Future of Health Care

In our brief Discussion section, we outline 2 overarching issues that we consider to frame the challenges facing health care systems that are hoping to adopt ML in the coming years. The discussion here is informed by the explicit recognition in the NASSS framework that both the technology and context in which innovations are being introduced shift and change over time. Greenhalgh et al suggest that although the levels of the framework can be distinguished analytically, “at an empirical level they are inextricably interlinked and dynamically evolving, often against a rapidly shifting policy context or continued evolution of the technology” (p. 14). Our assessment of the 2 issues we address here is intended to represent the connections between the changes that will be required as the policy context and technology evolve concurrently. The first is the issue of the role of corporations in health-related applications of ML, and the second is the issue of the role of ML in the evolving nature of health care.

The Role of Corporations

As the innovations enabled by ML have taken on a more powerful role in driving global economies, corporations have strategically sought to acquire larger amounts of more diverse data to boost their capacity to develop ML algorithms [61]. The shifting focus of many large corporations to the collection and manipulation of data characterizes what Zuboff refers to as surveillance capitalism, a relatively recent phenomenon in the global economy that relies on data for innovation and corporate success. The more that large corporations enter the health care industry with the power to collect, store, and use data, the more intertwined health care will become with the corporate realities of these large, multinational companies [62].

As large corporations acquire more data and develop more sophisticated forms of ML that transcend any individual geographical region, the implications for domestic health care policy are at risk of being overlooked. Although recent efforts to create regional protections around data collection and use have appeared to make an impact, such as the General Data Protection Regulation in Europe, health care policy is well behind. In cases where health-related data are already being stored in a country other than where the user is living, what are the regulations on how those data can be used? Where users voluntarily engage with technologies that collect their data for explicit health-related use by a corporation outside of their political jurisdiction, what legislative frameworks apply to protect patients and the public? These issues represent the important challenge of making health policy matter when conventional political boundaries are less able to contain the potential of large corporations to develop and use their technological capabilities.

The Changing Nature of Artificial Intelligence–Enabled Health Care

AI applications represent a potential impetus for major change in the institutions that constitute health care. In this sense, the term institution refers not just to the organizations in which health care providers work but to a complex collection of cognitive, cultural, regulative, and moral influences that shape the way that health care workers see their work and their lives [63]. The social sciences have worked to provide clear definitions of institutions through decades of research and theory [63-65]. Scott explained that institutions are combinations of 3 pillars: norms of the way things are usually done around here (cultural-cognitive influences), laws and regulations (regulative influences), and assumed moral codes (normative influences) [63]. Health care represents a confluence of institutions understood in this sense, many of which are naturally oriented toward maintaining some version of the status quo. Particularly for members of institutions who maintain power over resources, such as the medical profession, embracing institutional change is a point of resistance and difficulty.

We suggest that ML will confront the realities of entrenched institutions through issues such as meaningful decision support and explainability described earlier. These 2 issues represent
the authority of health care providers over the decisions that come to define health care as a multi-institutional field, both in terms of their rightful positions within the system and the fabric of decision making that has always defined health care processes. These issues point toward an important challenge that we suggest implementation scientists must grapple with: the changing nature of health care work. In Prediction Machines, the authors explain that as AI technology develops, “the value of substitutes to prediction machines, namely human prediction, will decline. However, the value of complements, such as the human skills associated with data collection, judgment, and actions, will become more valuable.” (p. 81). As the implementation science community considers how to encourage the adoption of ML technologies, it will also need to consider how such technologies stand to change the ways in which health care planning, decision making, and delivery are understood and the evolving role of human health care providers within that context.

The challenges described here refer to unique considerations of ML that pose novel challenges to implementation beyond the work of promoting the routine use of technologies among health care providers. We suggest that the hype and high stakes of ML make these issues more prominent in the mindsets of health care stakeholders and therefore more likely to impact upon an ML implementation project. The implementation science community will need to establish strategies to address these issues as ML becomes more prominent, each of which requires ongoing work to be adequately addressed.

Conclusions

In this paper, we have provided an overview of ML for implementation scientists informed by the NASSS framework, outlining the use cases of ML as falling into the categories of decision support and automation. We suggest these use cases apply to clinical, operational, and epidemiological tasks and that the primary ways in which ML will enter into health care in the near term will be through decision support. We then outlined unique implementation issues posed by ML initiatives from 4 perspectives, those of health care providers, patients and the public, health care organizations, and health policy and systems.

Ultimately, we suggest that the future of ML in health care remains positive but uncertain, as support from patients, the public, and a wide range of health care stakeholders is necessary to enable its meaningful implementation. However, as applications of ML become more sophisticated and investment in communications strategies such as data visualization grows, ML is likely to become more user-friendly and more effective. If the implementation science community is to facilitate the adoption of ML in ways that stand to benefit all, the issues raised in this paper will require substantial attention in the coming years.

Acknowledgments

This research is supported by an Associated Medical Services Phoenix Fellowship to the corresponding author.

Authors’ Contributions

JS led the writing of the manuscript. JS, TJ, AG, and FR contributed to the conceptualization, design, and approach for the manuscript. JS, TJ, AG, and FR contributed to analysis and interpretation of the argument made in the manuscript. All authors provided the final approval of the manuscript. All authors agree to be accountable for the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

AI: artificial intelligence
GPT: general purpose technology
ML: machine learning
NASSS: Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability

©James Shaw, Frank Rudzicz, Trevor Jamieson, Avi Goldfarb. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 10.07.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.