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 Bhatia, 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 Sciences (V.M., W.W., 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*).

Authors declared no funding for this work.

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

Radiology: Artificial Intelligence 2022; 4(2):e210114 • **https://doi.org/10.1148/ryai.210114** • **Content code:** AI

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.

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, NLP = natural language processing, TI-RADS = Thyroid Imaging Reporting and Data System

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 protocoling, 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 Protocoling

Protocoling 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 protocoling studies, totaling 50 hours per week across the department, which is the equivalent of the workload for one full-time equivalent radiologist. Protocoling 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 protocoling 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 protocoling 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

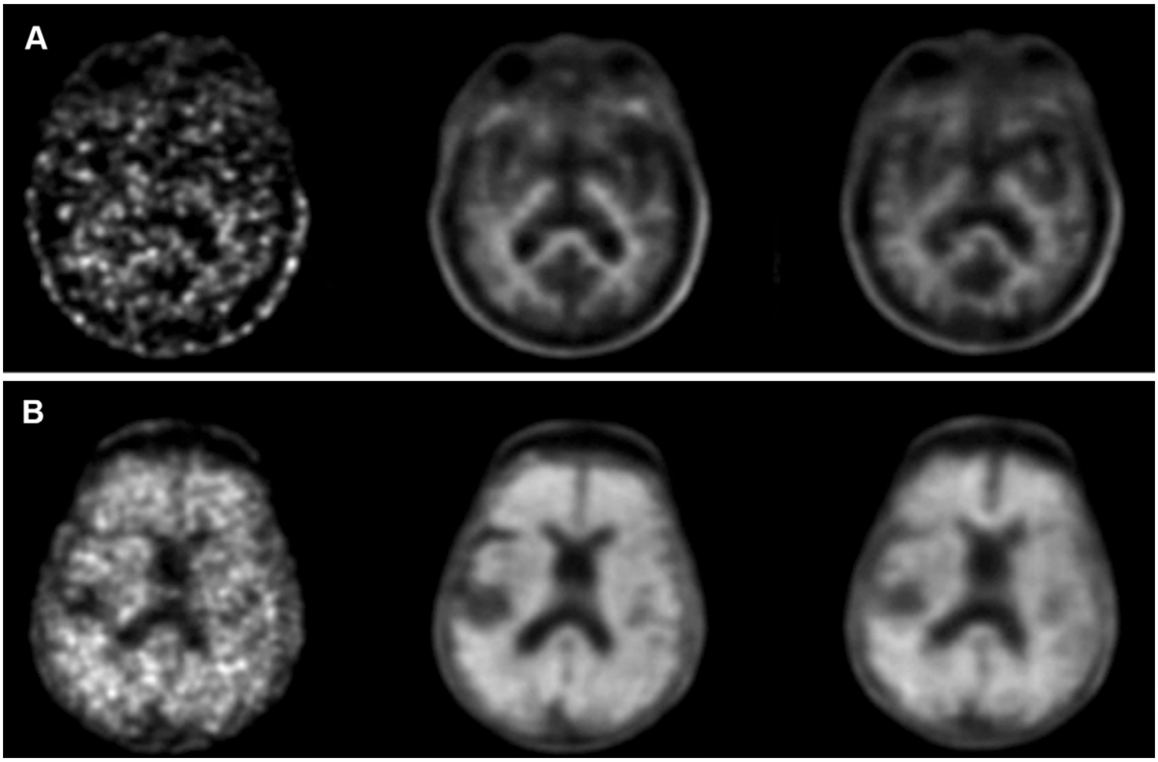

Figure 1: (A, B) Two examples of low-dose PET (left), ground truth standard-dose PET (middle), and low-dose PET with generative adversarial network-synthesized images (right). (Adapted, with permission, from reference 32.)

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 (25,29,30). 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

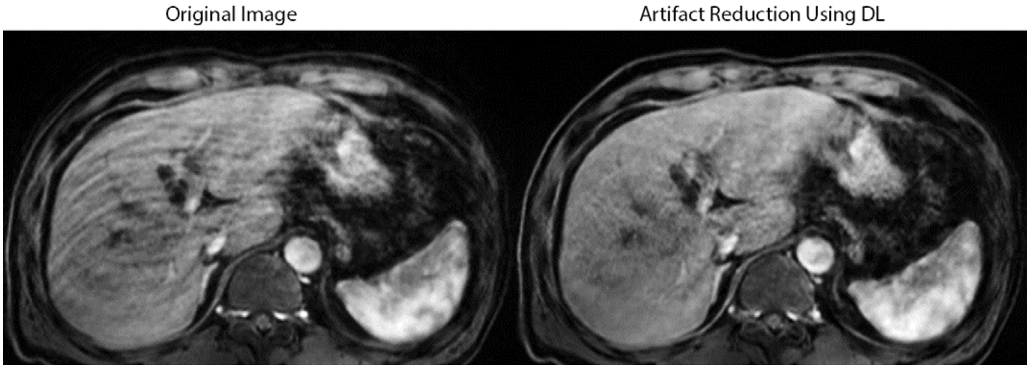

Figure 2: MRI with image aliasing, specifically respiratory artifact and blurring suppression **(A)** before and **(B)** after artifact reduction. DL = deep learning. (Adapted, with permission, from reference 44.)

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).

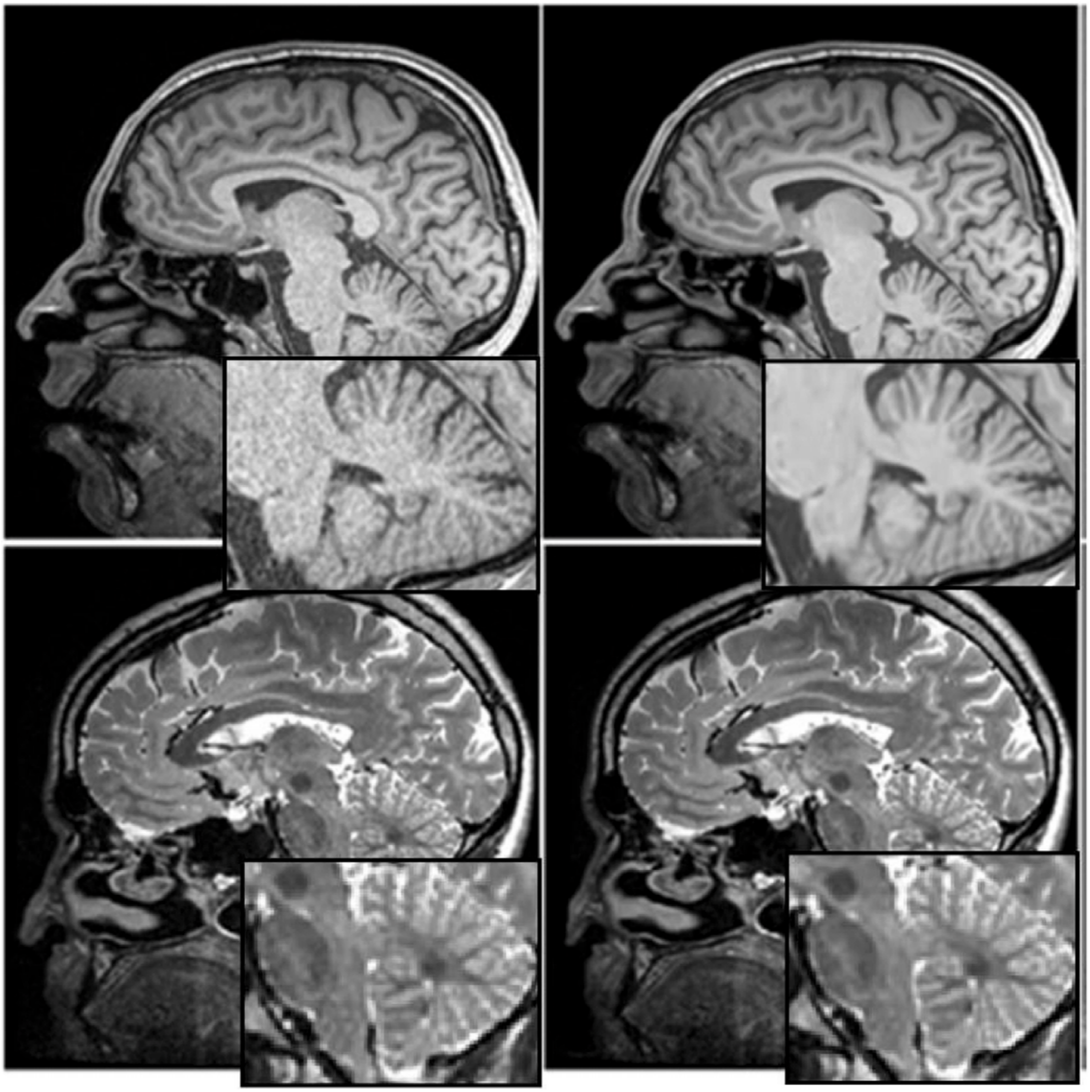

Figure 3: Noise suppression of (top) T1- and (bottom) T2-weighted images. Original images (left) and processed images (right). (Adapted, with permission, from reference 50.)

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

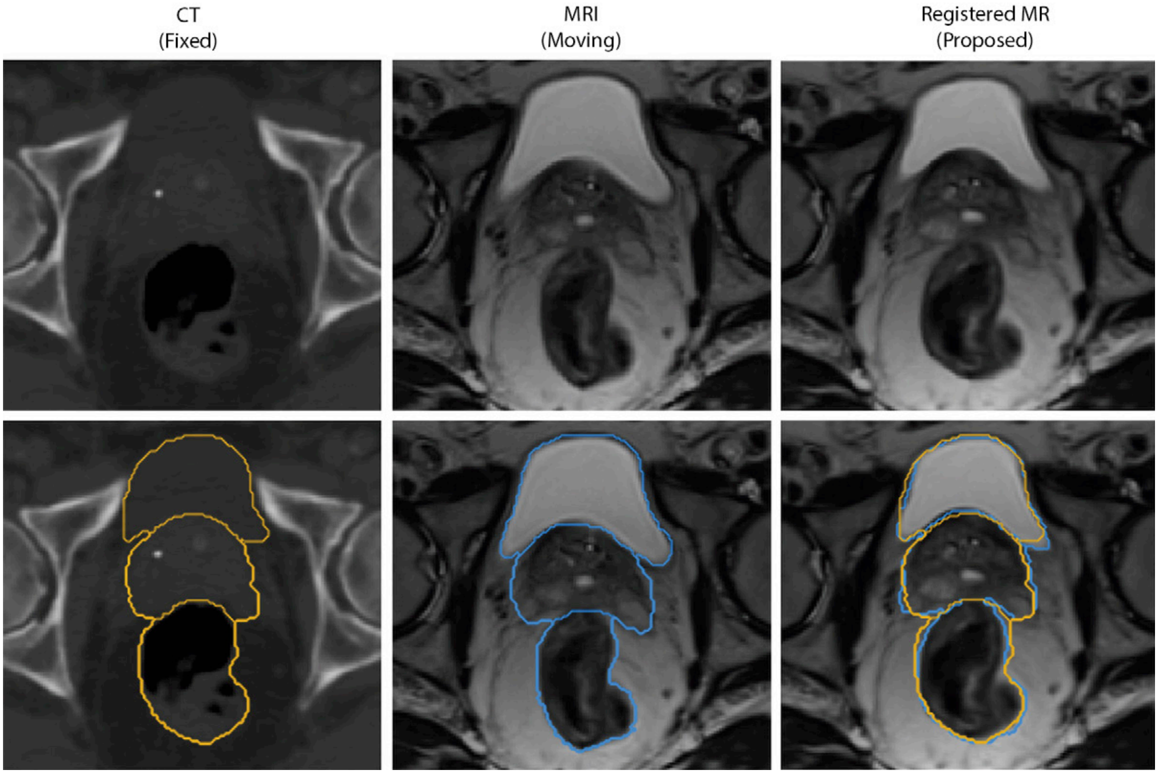

Figure 4: Sample image registration between CT and MRI scans shows original CT image with the manual contour in yellow (left), MRI scan with manual contour in blue (middle), and colocalized section and contour carried from the CT image to the MRI scan with a good overlap between contours (right). (Adapted, with permission, from reference 54.)

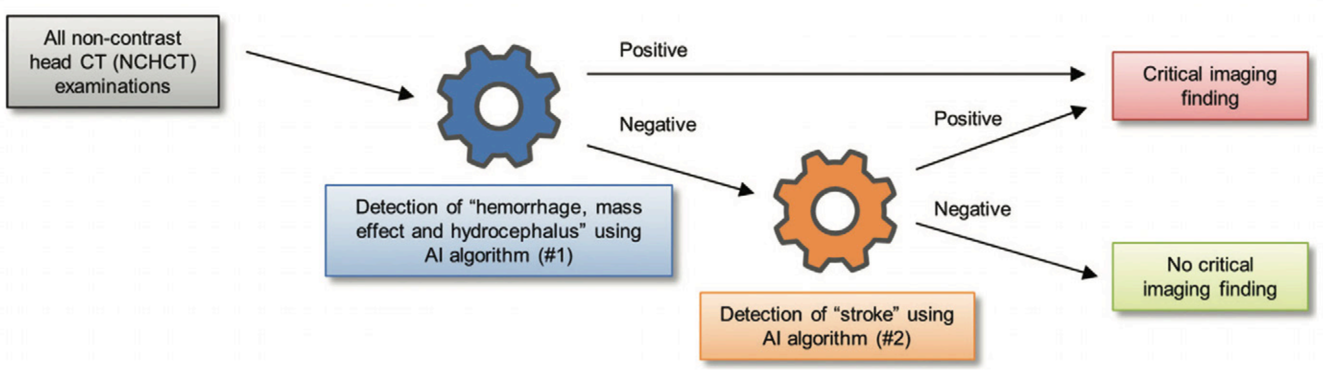

Figure 5: Analytic algorithm of noncontrast head CT examinations for urgent findings. AI = artificial intelligence. (Adapted, with permission, from reference 62.)

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])

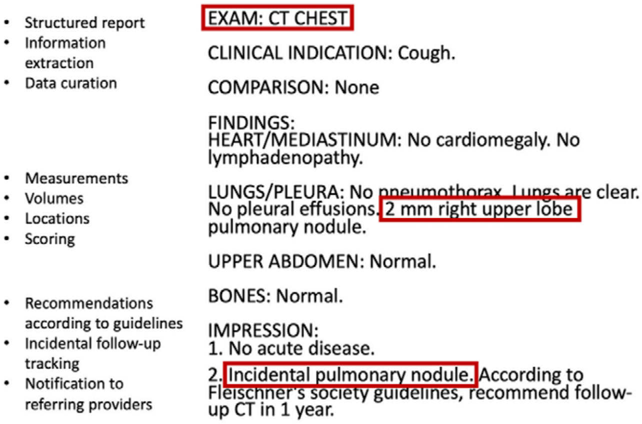

Figure 6: Sample of potential automaton for detection of an incidental pulmonary nodule in the report and appropriate follow-up recommendation generation. Exam = examination. Red boxes = portions of report model would use to generated follow-up recommendation.

(84), and primary brain malignancies (Brain Tumor Reporting and Data System [BT-RADS]) (85). Each of these scoring systems relies on imaging characteristics and change over time to guide diagnosis or follow-up management. Many AI algorithms have been developed to automate the tasks associated with these scoring systems, including lesion measurement, image segmentation, and comparison with prior images. Some systems measure lesions that must first be identified by the radiologist (86–88), whereas others detect candidate lesions and their characteristics and predict the likelihood of future cancer (89). For example, algorithms have been developed to derive BI-RADS scores and breast densities or to highlight lesions that are suspected for cancer directly from breast MRI, US, or mammography. These algorithms have achieved areas under the curve of greater than 0.9 (90–92). For liver lesions, models have been created to identify lesions at multisequence imaging and perform sequence coregistration to help measurement and interpretation (93,94) or to derive the LI-RADS score directly from the images, with accuracies ranging from 57% to 85% (95). In the BT-RADS, NLP algorithms have been able to derive BT-RADS classification scores directly from the MRI report, achieving F1 scores of up to 0.98 (96).

Machine learning algorithms have been incorporated into the data curation process used to update recommendations within the classification systems, as in the case of TI-RADS (97). A model trained with thyroid US lesions and their respective TI-RADS scores was able to improve the specificity of thyroid biopsy from 47% to 65% (ie, decreased biopsy of nonmalignant nodules) while maintaining sensitivity (98).

Automatic Notification to Provider of Incidental and Emergent Findings

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, hydronephrosis, 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).

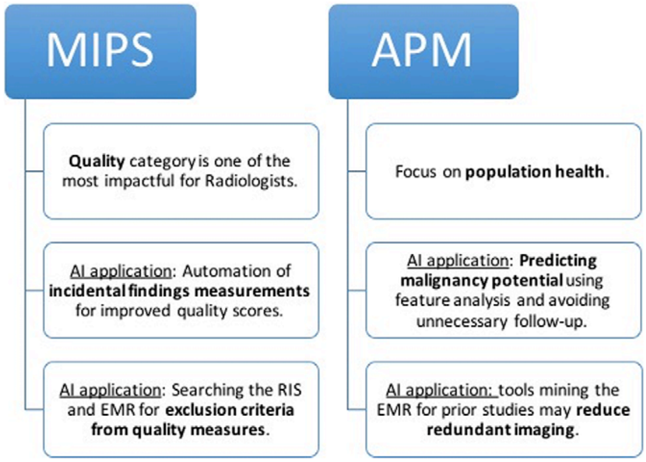

Figure 7: A comparison of the Merit-based Incentive Payment System (MIPS) and the alternative payment model (APM) pathways and possible artificial intelligence (AI) applications under each model. EMR = electronic medical record, RIS = radiology information system.

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 (eg, 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.

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 noninterpretive 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, noninterpretive 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 protocoling 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 multivariant 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 protocoling, 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 HL-7 Shield (Emory University is applying for a software patent for an AI approach to findings notification, this author listed as contributor). **I.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 Kheiron Medical for institution; owner of Lightbox AI.

References

1. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. Insights Imaging 2019;10(1):44.
2. Boland GWL, Guimaraes AS, Mueller PR. The radiologist's conundrum: benefits and costs of increasing CT capacity and utilization. Eur Radiol 2009;19(1):9–11; discussion 12.
3. Bhargavan M, Kaye AH, Forman HP, Sunshine JH. Workload of radiologists in United States in 2006-2007 and trends since 1991-1992. Radiology 2009;252(2):458–467.
4. Forsberg D, Rosipko B, Sunshine JL. Radiologists' variation of time to read across different procedure types. J Digit Imaging 2017;30(1):86–94.
5. Bernardy M, Ullrich CG, Rawson JV, et al. Strategies for managing imaging utilization. J Am Coll Radiol 2009;6(12):844–850.
6. Lehnert BE, Bree RL. Analysis of appropriateness of outpatient CT and MRI referred from primary care clinics at an academic medical center: how critical is the need for improved decision support? J Am Coll Radiol 2010;7(3):192–197. [Published correction appears in J Am Coll Radiol. 2010 Jun;7(6):466.]
7. Bairstow PJ, Persaud J, Mendelson R, Nguyen L. Reducing inappropriate diagnostic practice through education and decision support. Int J Qual Health Care 2010;22(3):194–200.
8. Blackmore CC, Mecklenburg RS, Kaplan GS. Effectiveness of clinical decision support in controlling inappropriate imaging. J Am Coll Radiol 2011;8(1):19–25.
9. Hendee WR, Becker GJ, Borgstede JP, et al. Addressing overutilization in medical imaging. Radiology 2010;257(1):240–245.
10. Dunnick NR, Applegate KE, Arenson RL. The inappropriate use of imaging studies: a report of the 2004 Intersociety Conference. J Am Coll Radiol 2005;2(5):401–406.
11. American College of Radiology. ACR Select. http://nationaldecisionsupport.com/acrselect/. Accessed August 30, 2021.
12. Stern RG. "Can we all get along?" Cooperative strategies to reduce imaging overuse. Am J Med 2013;126(8):657–658.
13. Ip IK, Schneider L, Seltzer S, et al. Impact of provider-led, technology-enabled radiology management program on imaging. Am J Med 2013;126(8):687–692.
14. Yan Z, Ip IK, Raja AS, Gupta A, Kosowsky JM, Khorasani R. Yield of CT Pulmonary Angiography in the Emergency Department When Providers Override Evidence-based Clinical Decision Support. Radiology 2017;282(3):717–725.
15. Doyle J, Abraham S, Feeney L, Reimer S, Finkelstein A. Clinical decision support for high-cost imaging: A randomized clinical trial. PLoS One 2019;14(3):e0213373.
16. Lee YH. Efficiency Improvement in a Busy Radiology Practice: Determination of Musculoskeletal Magnetic Resonance Imaging Protocol Using Deep-Learning Convolutional Neural Networks. J Digit Imaging 2018;31(5):604–610.
17. Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH. Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM watson's natural language processing algorithm. J Digit Imaging 2018;31(2):245–251.
18. Brown AD, Marotta TR. A Natural Language Processing-based Model to Automate MRI Brain Protocol Selection and Prioritization. Acad Radiol 2017;24(2):160–166.
19. Kalra A, Chakraborty A, Fine B, Reicher J. Machine learning for automation of radiology protocols for quality and efficiency improvement. J Am Coll Radiol 2020;17(9):1149–1158.
20. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med 2021;4(1):86.
21. Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One 2019;14(7):e0218942.
22. Rajendran S, Topaloglu U. Extracting Smoking Status from Electronic Health Records Using NLP and Deep Learning. AMIA Jt Summits Transl Sci Proc 2020;2020:507–516.
23. Li Y, Rao S, Solares JRA, et al. BEHRT: transformer for electronic health records. Sci Rep 2020;10(1):7155.
24. Sreekumari A, Shanbhag D, Yeo D, et al. A Deep Learning-Based Approach to Reduce Rescan and Recall Rates in Clinical MRI Examinations. AJNR Am J Neuroradiol 2019;40(2):217–223.
25. Xu J, Gong E, Pauly J, Zaharchuk G. 200x Low-dose PET Reconstruction using Deep Learning. arXiv 1712.04119 [preprint] https://arxiv.org/abs/1712.04119. Posted December 12, 2017. Accessed August 30, 2021.
26. Richardson ML, Garwood ER, Lee Y, et al. Noninterpretive uses of artificial intelligence in radiology. Acad Radiol 2021;28(9):1225–1235.
27. Chen H, Zhang Y, Kalra MK, et al. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. IEEE Trans Med Imaging 2017;36(12):2524–2535.
28. Shan H, Padole A, Homayounieh F, et al. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. Nat Mach Intell 2019;1(6):269–276.
29. Katsari K, Penna D, Arena V, et al. Artificial intelligence for reduced dose 18F-FDG PET examinations: a real-world deployment through a standardized framework and business case assessment. EJNMMI Phys 2021;8(1):25.
30. Le V, Frye S, Botkin C, et al. Effect of PET Scan with Count Reduction Using AI-Based Processing Techniques on Image Quality. J Nucl Med 2020;61(Suppl 1):3095. https://jnm.snmjournals.org/content/61/supplement_1/3095.
31. Wang Y, Yu B, Wang L, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage 2018;174:550–562.
32. Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Med Phys 2019;46(8):3555–3564.
33. Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging 2018;48(2):330–340.
34. Zhang HM, Dong B. A review on deep learning in medical image reconstruction. J Oper Res Soc China 2020;8(2):311–340.
35. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature 2018;555(7697):487–492.
36. Lauritsch G, Haerer WH. Theoretical framework for filtered back projection in tomosynthesis. In: Hanson KM, ed. Proceedings of SPIE: medical imaging 1998—image processing. Vol 3338. Bellingham, Wash: International Society for Optics and Photonics, 1998; 1127–1137.
37. Singh S, Kalra MK, Hsieh J, et al. Abdominal CT: comparison of adaptive statistical iterative and filtered back projection reconstruction techniques. Radiology 2010;257(2):373–383.
38. Klink T, Obmann V, Heverhagen J, Stork A, Adam G, Begemann P. Reducing CT radiation dose with iterative reconstruction algorithms: the influence of scan and reconstruction parameters on image quality and CTDIvol. Eur J Radiol 2014;83(9):1645–1654.
39. Schofield R, King L, Tayal U, et al. Image reconstruction: Part 1 - understanding filtered back projection, noise and image acquisition. J Cardiovasc Comput Tomogr 2020;14(3):219–225.
40. Bermudez C, Plassard AJ, Davis LT, Newton AT, Resnick SM, Landman BA. Learning implicit brain MRI manifolds with deep learning. In: Angelini ED, Landman BA, eds. Proceedings of SPIE: medical imaging 2018—image processing. Vol 10574. Bellingham, Wash: International Society for Optics and Photonics, 2018; 105741L.
41. Jin KH, McCann MT, Froustey E, Unser M. Deep Convolutional Neural Network for Inverse Problems in Imaging. IEEE Trans Image Process 2017;26(9):4509–4522.
42. Bash S. Enhancing Neuroimaging with Artificial Intelligence. Appl Radiol 2020;49(1):20–21. https://appliedradiology.com/communities/Artificial-Intelligence/enhancing-neuroimaging-with-artificial-intelligence.
43. Loria K. 2020 Vision — A Look at the Latest Trends and Best Practices in Dose Safety and Image Acquisition. Radiology Today Magazine 2020;21(2):10–12. https://www.radiologytoday.net/archive/rt0220p10.shtml.
44. Tamada D. Review: Noise and artifact reduction for MRI using deep learning. arXiv 2002.12889 [preprint] https://arxiv.org/abs/2002.12889. Posted February 28, 2020. Accessed August 30, 2021.
45. Tanenbaum L, Gibbs W, Johnson B, et al. Machine learning-based iterative image reconstruction algorithm allows significant reduction in brain MRI scan times. In: Proceedings of American Society of Neuroradiology 56th Annual Meeting, 2018. https://www.asnr.org/wp-content/uploads/Proceedings/ASNR2018_Proceedings.pdf.

46. Wang L, Bash S, Dupont S, et al. Deep Learning Enables Accurate Quantitative Volumetric Brain MRI with 2x Faster Scan Times. In: Proceedings of American Society of Neuroradiology 58th Annual Meeting, 2020. https://www.asnr.org/wp-content/uploads/2021/07/ASNR20-Proceedings.pdf.
47. Lakhani P, Prater AB, Hutson RK, et al. Machine learning in radiology: applications beyond image interpretation. J Am Coll Radiol 2018;15(2):350–359.
48. Esses SJ, Lu X, Zhao T, et al. Automated image quality evaluation of T2 -weighted liver MRI utilizing deep learning architecture. J Magn Reson Imaging 2018;47(3):723–728.
49. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. Jpn J Radiol 2019;37(1):73–80.
50. Manjon JV, Coupe P. MRI denoising using Deep Learning and Non-local averaging. arXiv 1911.04798 [preprint] https://arxiv.org/abs/1911.04798. Posted November 12, 2019. Accessed August 30, 2021.
51. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 2019;29(2):102–127.
52. Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. Mach Vis Appl 2020;31(1-2):8.
53. Wei D, Ahmad S, Huo J, et al. SLIR: Synthesis, localization, inpainting, and registration for image-guided thermal ablation of liver tumors. Med Image Anal 2020;65:101763.
54. Cao X, Yang J, Wang L, Xue Z, Wang Q, Shen D. Deep Learning based Inter-Modality Image Registration Supervised by Intra-Modality Similarity. Mach Learn Med Imaging 2018;11046:55–63.
55. Eberhard M, Alkadhi H. Machine learning and deep neural networks: applications in patient and scan preparation, contrast medium, and radiation dose optimization. J Thorac Imaging 2020;35(Suppl 1):S17–S20.
56. Saltybaeva N, Alkadhi H. Vertical off-centering affects organ dose in chest CT: Evidence from Monte Carlo simulations in anthropomorphic phantoms. Med Phys 2017;44(11):5697–5704.
57. Saltybaeva N, Schmidt B, Wimmer A, Flohr T, Alkadhi H. Precise and Automatic Patient Positioning in Computed Tomography: Avatar Modeling of the Patient Surface Using a 3-Dimensional Camera. Invest Radiol 2018;53(11):641–646.
58. Booij R, Budde RPJ, Dijkshoorn ML, van Straten M. Accuracy of automated patient positioning in CT using a 3D camera for body contour detection. Eur Radiol 2019;29(4):2079–2088.
59. Taplin SH, Rutter CM, Finder C, Mandelson MT, Houn F, White E. Screening mammography: clinical image quality and the risk of interval breast cancer. AJR Am J Roentgenol 2002;178(4):797–803.
60. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. RadioGraphics 2003;23(4):881–895.
61. U.S. Food and Drug Administration. Mammography Quality Standards Act and Program. https://www.fda.gov/radiation-emitting-products/mammography-quality-standards-act-and-program. Accessed August 13, 2021.
62. Prevedello LM, Erdal BS, Ryu JL, et al. Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 2017;285(3):923–931.
63. Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks. Radiology 2019;291(1):196–202. [Published correction appears in Radiology. 2019 Apr;291(1):272.]
64. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digit Med 2018;1(1):9.
65. Wismüller A, Stockmaster L. A prospective randomized clinical trial for measuring radiology study reporting time on Artificial Intelligence-based detection of intracranial hemorrhage in emergent care head CT. In: Krol A, Gimi BS, eds. Proceedings of SPIE: medical imaging 2020—biomedical applications in molecular, structural, and functional imaging. Vol 11317. Bellingham, Wash: International Society for Optics and Photonics, 2020; 113170M.
66. Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. Neuroradiology 2020;62(3):335–340.
67. Winkel DJ, Heye T, Weikert TJ, Boll DT, Stieltjes B. Evaluation of an AI-Based Detection Software for Acute Findings in Abdominal Computed Tomography Scans: Toward an Automated Work List Prioritization of Routine CT Examinations. Invest Radiol 2019;54(1):55–59.
68. Huang SC, Kothari T, Banerjee I, et al. PENet-a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. NPJ Digit Med 2020;3(1):61. [Published correction appears in NPJ Digit Med 2020;3:102.]
69. Banerjee I, Sofela M, Yang J, et al. Development and performance of the pulmonary embolism result forecast model (PERFORM) for computed tomography clinical decision support. JAMA Netw Open 2019;2(8):e198719.
70. Pines JM, Griffey RT. What we have learned from a decade of ED crowding research. Acad Emerg Med 2015;22(8):985–987.
71. Jha S. Value of triage by artificial intelligence. Acad Radiol 2020;27(1):153–155.
72. Potter CA, Vagal AS, Goyal M, Nunez DB, Leslie-Mazwi TM, Lev MH. CT for treatment selection in acute ischemic stroke: A code stroke primer. RadioGraphics 2019;39(6):1717–1738.
73. Goldberg-Stein S, Chernyak V. Adding value in radiology reporting. J Am Coll Radiol 2019;16(9 Pt B):1292–1298.
74. Cai T, Giannopoulos AA, Yu S, et al. Natural language processing technologies in radiology research and clinical applications. RadioGraphics 2016;36(1):176–191.
75. Lou R, Lalevic D, Chambers C, Zafar HM, Cook TS. Automated Detection of Radiology Reports that Require Follow-up Imaging Using Natural Language Processing Feature Engineering and Machine Learning Classification. J Digit Imaging 2020;33(1):131–136.
76. Dutta S, Long WJ, Brown DFM, Reisner AT. Automated detection using natural language processing of radiologists recommendations for additional imaging of incidental findings. Ann Emerg Med 2013;62(2):162–169.
77. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: A systematic review. Radiology 2016;279(2):329–343.
78. Do BH, Wu AS, Maley J, Biswal S. Automatic retrieval of bone fracture knowledge using natural language processing. J Digit Imaging 2013;26(4):709–713.
79. Taira RK, Soderland SG, Jakobovits RM. Automatic structuring of radiology free-text reports. RadioGraphics 2001;21(1):237–245.
80. Zimmerman SL, Kim W, Boonn WW. Informatics in radiology: automated structured reporting of imaging findings using the AIM standard and XML. RadioGraphics 2011;31(3):881–887.
81. Jungmann F, Arnhold G, Kämpgen B, et al. A hybrid reporting platform for extended radlex coding combining structured reporting templates and natural language processing. J Digit Imaging 2020;33(4):1026–1033.
82. Delfim RLC, Veiga LCGD, Vidal APA, Lopes FPPL, Vaisman M, Teixeira PFDS. Likelihood of malignancy in thyroid nodules according to a proposed Thyroid Imaging Reporting and Data System (TI-RADS) classification merging suspicious and benign ultrasound features. Arch Endocrinol Metab 2017;61(3):211–221.
83. Sickles EA, D'Orsi CJ, Bassett LW, et al. American College of Radiology. Breast imaging reporting and data system (BI-RADS) atlas. 5th ed. Reston, Va: American College of Radiology, 2013.
84. Mitchell DG, Bruix J, Sherman M, Sirlin CB. LI-RADS (Liver Imaging Reporting and Data System): summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. Hepatology 2015;61(3):1056–1065.
85. Zhang JY, Weinberg BD, Hu R, et al. Quantitative Improvement in Brain Tumor MRI Through Structured Reporting (BT-RADS). Acad Radiol 2020;27(6):780–784.
86. Sedghi Gamechi Z, Bons LR, Giordano M, et al. Automated 3D segmentation and diameter measurement of the thoracic aorta on non-contrast enhanced CT. Eur Radiol 2019;29(9):4613–4623.
87. Barnard R, Tan J, Roller B, et al. Machine Learning for Automatic Paraspinous Muscle Area and Attenuation Measures on Low-Dose Chest CT Scans. Acad Radiol 2019;26(12):1686–1694.
88. Ironside N, Chen CJ, Mutasa S, et al. Fully automated segmentation algorithm for hematoma volumetric analysis in spontaneous intracerebral hemorrhage. Stroke 2019;50(12):3416–3423.
89. Saha A, Grimm LJ, Ghate SV, et al. Machine learning-based prediction of future breast cancer using algorithmically measured background parenchymal enhancement on high-risk screening MRI. J Magn Reson Imaging 2019;50(2):456–464.
90. Reig B, Heacock L, Geras KJ, Moy L. Machine learning in breast MRI. J Magn Reson Imaging 2020;52(4):998–1018.
91. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. Med Phys 2018;45(1):314–321.
92. Zhang E, Seiler S, Chen M, Lu W, Gu X. BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis. Phys Med Biol 2020;65(12):125005.
93. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. Voxelmorph: A learning framework for deformable medical image registration. IEEE Trans Med Imaging 2019;38(8):1788–1800.
94. Das A, Rajendra Acharya U, Panda SS, Sabut S. Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. Cogn Syst Res 2019;54:165–175.

95. Banerjee I, Choi HH, Desser T, Rubin DL. A Scalable Machine Learning Approach for Inferring Probabilistic US-LI-RADS Categorization. AMIA Annu Symp Proc 2018;2018:215–224.
96. Lee SJ, Weinberg BD, Gore A, Banerjee I. A Scalable Natural Language Processing for Inferring BT-RADS Categorization from Unstructured Brain Magnetic Resonance Reports. J Digit Imaging 2020;33(6):1393–1400.
97. Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. Artif Intell Med 2016;66:29–39.
98. Wildman-Tobriner B, Buda M, Hoang JK, et al. Using Artificial Intelligence to Revise ACR TI-RADS Risk Stratification of Thyroid Nodules: Diagnostic Accuracy and Utility. Radiology 2019;292(1):112–119.
99. The Joint Commission. National Patient Safety Goals Effective July 2020 for the Hospital Program. https://www.jointcommission.org/-/media/tjc/documents/standards/national-patient-safety-goals/2022/npsg_chapter_hap_jan2022.pdf. Accessed August 31, 2021.
100. Chetlen AL, Chan TL, Ballard DH, et al. Addressing burnout in radiologists. Acad Radiol 2019;26(4):526–533.
101. Cannavale A, Santoni M, Mancarella P, Passariello R, Arbarello P. Malpractice in radiology: what should you worry about? Radiol Res Pract 2013;2013:219259.
102. Lakhani P, Langlotz CP. Automated detection of radiology reports that document non-routine communication of critical or significant results. J Digit Imaging 2010;23(6):647–657.
103. Lakhani P, Kim W, Langlotz CP. Automated detection of critical results in radiology reports. J Digit Imaging 2012;25(1):30–36.
104. Meng X, Ganoe CH, Sieberg RT, Cheung YY, Hassanpour S. Assisting radiologists with reporting urgent findings to referring physicians: A machine learning approach to identify cases for prompt communication. J Biomed Inform 2019;93:103169.
105. Do HM, Spear LG, Nikpanah M, et al. Augmented radiologist workflow improves report value and saves time: A potential model for implementation of artificial intelligence. Acad Radiol 2020;27(1):96–105.
106. Cochon LR, Kapoor N, Carrodeguas E, et al. Variation in Follow-up Imaging Recommendations in Radiology Reports: Patient, Modality, and Radiologist Predictors. Radiology 2019;291(3):700–707.
107. Trivedi G, Dadashzadeh ER, Handzel RM, Chapman WW, Visweswaran S, Hochheiser H. Interactive NLP in clinical care: identifying incidental findings in radiology reports. Appl Clin Inform 2019;10(4):655–669.
108. Kang SK, Garry K, Chung R, et al. Natural language processing for identification of incidental pulmonary nodules in radiology reports. J Am Coll Radiol 2019;16(11):1587–1594.
109. Carrodeguas E, Lacson R, Swanson W, Khorasani R. Use of Machine Learning to Identify Follow-Up Recommendations in Radiology Reports. J Am Coll Radiol 2019;16(3):336–343.
110. Pham AD, Névéol A, Lavergne T, et al. Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. BMC Bioinformatics 2014;15(1):266.
111. Oliveira L, Tellis R, Qian Y, Trovato K, Mankovich G. Follow-up Recommendation Detection on Radiology Reports with Incidental Pulmonary Nodules. Stud Health Technol Inform 2015;216:1028.
112. Hammer MM, Kapoor N, Desai SP, et al. Adoption of a Closed-Loop Communication Tool to Establish and Execute a Collaborative Follow-Up Plan for Incidental Pulmonary Nodules. AJR Am J Roentgenol 2019;1–5(5):1–5.
113. Sanborn BJ. Change Healthcare analysis shows $262 billion in medical claims initially denied, meaning billions in administrative costs. Healthcare Finance News. https://www.healthcarefinancenews.com/news/change-healthcare-analysis-shows-262-million-medical-claims-initially-denied-meaning-billions. Accessed February 24, 2022.
114. Combatting denials using machine intelligence: How it works and why now is the time for it. https://www.beckershospitalreview.com/finance/combatting-denials-using-machine-intelligence-how-it-works-and-why-now-is-the-time-for-it.html. Accessed February 24, 2022.
115. Jiwani A, Himmelstein D, Woolhandler S, Kahn JG. Billing and insurance-related administrative costs in United States' health care: synthesis of micro-costing evidence. BMC Health Serv Res 2014;14(1):556.
116. Institute of Medicine Roundtable on Evidence-Based Medicine. Excess administrative costs. In: Yong PL, Saunders RS, Olsen L, eds. The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary. Washington, DC: The National Academies Press, 2010. https://www.ncbi.nlm.nih.gov/books/NBK53942/.
117. Chan L, Beers K, Yau AA, et al. Natural language processing of electronic health records is superior to billing codes to identify symptom burden in hemodialysis patients. Kidney Int 2020;97(2):383–392.
118. Jagannathan V, Mullett CJ, Arbogast JG, et al. Assessment of commercial NLP engines for medication information extraction from dictated clinical notes. Int J Med Inform 2009;78(4):284–291.
119. Duszak R, Nossal M, Schofield L, Picus D. Physician documentation deficiencies in abdominal ultrasound reports: frequency, characteristics, and financial impact. J Am Coll Radiol 2012;9(6):403–408.
120. Denck J, Landschütz W, Nairz K, Heverhagen JT, Maier A, Rothgang E. Automated Billing Code Retrieval from MRI Scanner Log Data. J Digit Imaging 2019;32(6):1103–1111.
121. Syed AB, Zoga AC. Artificial intelligence in radiology: current technology and future directions. Semin Musculoskelet Radiol 2018;22(5):540–545.
122. Khanna R, Yen T. Computerized physician order entry: promise, perils, and experience. Neurohospitalist 2014;4(1):26–33.
123. Artificial Intelligence Adoption and Investments Growing Rapidly Among Health Industry Leaders. Business Wire. https://www.businesswire.com/news/home/20191008005194/en/Artificial-Intelligence-Adoption-and-Investments-Growing-Rapidly-Among-Health-Industry-Leaders. Accessed February 24, 2022.
124. At RadNet, AI-fueled prior authorization tech shows promise. Healthcare IT News. https://www.healthcareitnews.com/news/radnet-ai-fueled-prior-authorization-tech-99-accurate. Accessed February 24, 2022.
125. Allen B. Valuing the professional work of diagnostic radiologic services. J Am Coll Radiol 2007;4(2):106–114.
126. Golding LP, Nicola GN. A business case for artificial intelligence tools: the currency of improved quality and reduced cost. J Am Coll Radiol 2019;16(9 Pt B):1357–1361.
127. Centers for Medicare & Medicaid Services. Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY 2019; Medicare Shared Savings Program Requirements; Quality Payment Program; Medicaid Promoting Interoperability Program; Quality Payment Program-Extreme and Uncontrollable Circumstance Policy for the 2019 MIPS Payment Year; Provisions From the Medicare Shared Savings Program-Accountable Care Organizations-Pathways to Success; and Expanding the Use of Telehealth Services for the Treatment of Opioid Use Disorder Under the Substance Use-Disorder Prevention That Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act. https://www.federalregister.gov/documents/2018/11/23/2018-24170/medicare-program-revisions-to-payment-policies-under-the-physician-fee-schedule-and-other-revisions. Published November 23, 2018. Accessed February 24, 2022.
128. Tajmir SH, Alkasab TK. Toward augmented radiologists: changes in radiology education in the era of machine learning and artificial intelligence. Acad Radiol 2018;25(6):747–750.
129. Rubin DL, Kahn CE Jr. Common data elements in radiology. Radiology 2017;283(3):837–844.
130. Kamel PI, Nagy PG. Patient-Centered Radiology with FHIR: an Introduction to the Use of FHIR to Offer Radiology a Clinically Integrated Platform. J Digit Imaging 2018;31(3):327–333.
131. AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. Ann Saudi Med 2019;39(6):373–381.
132. Kurasawa H, Hayashi K, Fujino A, et al. Machine-Learning-Based Prediction of a Missed Scheduled Clinical Appointment by Patients With Diabetes. J Diabetes Sci Technol 2016;10(3):730–736.
133. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. NPJ Digit Med 2019;2(1):26.
134. Levine WC, Dunn PF. Optimizing operating room scheduling. Anesthesiol Clin 2015;33(4):697–711.
135. Luo L, Zhou Y, Han BT, Li J. An optimization model to determine appointment scheduling window for an outpatient clinic with patient no-shows. Health Care Manage Sci 2019;22(1):68–84.
136. Mieloszyk RJ, Rosenbaum JI, Hall CS, Hippe DS, Gunn ML, Bhargava P. Environmental Factors Predictive of No-Show Visits in Radiology: Observations of Three Million Outpatient Imaging Visits Over 16 Years. J Am Coll Radiol 2019;16(4 Pt B):554–559.
137. Chong LR, Tsai KT, Lee LL, Foo SG, Chang PC. Artificial Intelligence Predictive Analytics in the Management of Outpatient MRI Appointment No-Shows. AJR Am J Roentgenol 2020;215(5):1155–1162.

TBM

COMMENTARY/POSITION PAPER

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

Elizabeth Chen,[1] Gila Neta,[2] Megan C. Roberts[3]

[1]Department of Health Behavior, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[2]Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD 20850, USA

[3]Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Correspondence to: E. Chen, lizcchen@unc.edu

Cite this as: *TBM* 2021;11:1115–1121
doi: 10.1093/tbm/ibaa079

Keywords

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

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.

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.

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 meetings with providers may be an effective implementation 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, simulation modeling, intervention mapping, and concept mapping, among others [8–10].

In addition to implementation frameworks, outcomes, and strategies, a broad variety of study designs can be used to study implementation, including effectiveness-implementation hybrid designs (which includes effectiveness and implementation research aims and data collection); mixed methods (integrating qualitative and quantitative methods); factorial designs (e.g., sequential multiple assignment randomized implementation trial); two-level nested randomized designs; cluster randomized control 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 economically 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 systems 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 individuals 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 diverse 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 commentary, the authors rely on the HCD process as defined by IDEO, a leading global design company, which has successfully used HCD to create groundbreaking products like Palm pilots and Oral-B toothbrushes [19].

IDEO's HCD process for problem-solving consists 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 engaged in HCD) first aim is to build empathy toward and draw inspiration from individual users (e.g., patients, 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 barriers (i.e., "pain points" in HCD) they have experienced 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 implementation 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 recombination 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 project [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 delivery of evidence-based practices [5]. We might consider HCD when developing a new intervention. Both fields acknowledge the importance of multiple stakeholder perspectives, iterative study cycles to optimize outcomes of interest, and consideration of the end users to improve implementation in real-world settings. Based on these complementary 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

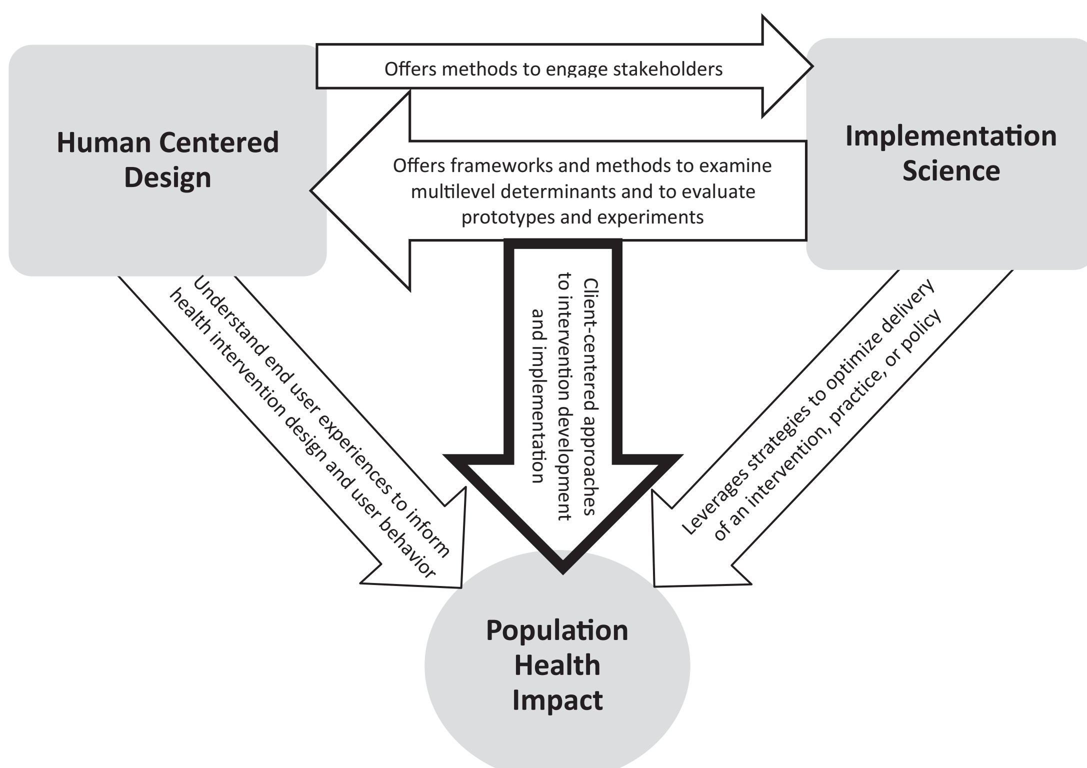

Fig 1 | How human-centered design and implementation science can lead to public health impact.

Table 1 | Implementation strategies and aligned design thinking methods from IDEO.org's *The Field Guide to Human-Centered Design*

Implementation strategy [6]	Implementation strategy description [6]	Human-centered design method [13]	Real Talk case study example
Involve patients/ consumers and family	Engage or include patients/ consumers and families in the implementation effort	Card sort Co-creation session Conversation starters Guided tour Resource flow	Card sort: Teenagers sorted cards that presented options related to where they were most comfortable talking about sex and/or relationships, who they were most comfortable talking with, and what topics interested them the most. Results of these card sorts informed the content and mode of delivery for the intervention.
Conduct cyclical small tests of change	Implement changes in a cyclical fashion using small tests of change before taking changes systemwide. Tests of change benefit from systematic measurement, and results of the tests of change are studied for insights on how to do better. This process continues serially over time, and refinement is added with each cycle.	Rapid prototyping Integrate and iterate Define success Measure and evaluation	Rapid prototyping: Team members used InVision to design multiple versions of a feature, like the story submission experience, and tested the viability of these options (e.g., form vs. chatbot) with potential users. Data from these tests were used to make decisions.
Intervene with patients/consumers to enhance update/ adherence	Develop strategies with patients to encourage and problem solve around adherence	Co-creation session Live prototyping	Co-creation: Team members hosted co-creation sessions with potential users where they were given paper with blank iPhone screens and were asked to (a) draw their ideal app to learn about sex education and (b) pitch their app to the other potential users for feedback. Team members listened to the pitches and used similar language for the iTunes App Store description.

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 would 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

1. Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. *Transl Behav Med.* 2019;9(6):1057–1064.
2. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. *BMC Psychol.* 2015;3(1) :1–12. doi:10.1186/S40359-015-0089-9
3. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. *Am J Prev Med.* 2012;43(3):337–350.
4. Lewis C, Mettert K, Dorset C, Weiner B. Measures and outcomes in implementation science. In: Chambers D, Vinson C, Norton W, eds. *Advancing the Sciences of Implementation Across the Cancer Continuum.* New York, NY: Oxford University Press; 2018:57–77.
5. Brown CH, Curran G, Palinkas LA, et al. An overview of research and evaluation designs for dissemination and implementation. *Annu Rev Public Health.* 2017;38(1):1–22.
6. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci.* 2015;10:21.
7. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health.* 2011;38(2):65–76.
8. Powell BJ, Beidas RS, Lewis CC, et al. Methods to improve the selection and tailoring of implementation strategies. *J Behav Health Serv Res.* 2017;44(2):177–194.
9. Leeman J, Birken SA, Powell BJ, Rohweder C, Shea CM. Beyond "implementation strategies": Classifying the full range of strategies used in implementation science and practice. *Implement Sci.* 2017;12(1):125.
10. Fernandez ME, ten Hoor GA, van Lieshout S, et al. Implementation mapping: using intervention mapping to develop implementation strategies. *Front Public Health.* 2019;7(Jun):158.
11. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care.* 2012;50(3):217–226.
12. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. *Adm Policy Ment Health.* 2011;38(1):44–53.
13. IDEO.org. 2015. The field guide to human-centered design. Available at http://www.designkit.org/resources/1. Accessibility verified November 19, 2019.

14. McKim RH. *Experiences in Visual Thinking*. 2nd ed. Monterey, CA: Cengage Learning; 1980.
15. Rowe PG. *Design Thinking*. Cambridge, MA: The MIT Press; 1987.
16. Nanos, J. 2016. Designers bring private sector ideas to public policy. Available at https://www.bostonglobe.com/business/2016/10/28/designers-bring-private-sector-ideas-public-policy/oQS7HSvcSmEhUpN9gJ6QqM/story.html. Accessibility verified October 28, 2019.
17. Kia-Keating M, Santacrose DE, Liu SR, Adams J. Using community-based participatory research and human-centered design to address violence-related health disparities among Latino/a youth. *Fam Community Health.* 2017;40(2):160–169.
18. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. *Health Promot Pract.* 2020;21(1):37–48.
19. Brown T, Wyatt J. 2010. Stanford social innovation review design thinking for social innovation. Available at https://ojs.unbc.ca/index.php/design/article/viewFile/1272/1089. Accessibility verified November 19, 2019.
20. Waltz TJ, Powell BJ, Matthieu MM, et al. Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: Results from the Expert Recommendations for Implementing Change (ERIC) study. *Implement Sci.* 2015;10:109.
21. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci.* 2015;10:53.
22. Moen RD, Norman CL. 2010. Circling back. Quality progress. Available at https://search-proquest-com.libproxy.lib.unc.edu/docview/816914443?pq-origsite=360link. Accessibility verified August 7, 2020.
23. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. *BMJ Qual Saf.* 2014;23(4):290–298.
24. MyHealthEd, Inc. 2017. Real Talk: Stories by teens. Available at https://apps.apple.com/us/app/real-talk-stories-by-teens/id1274703738. Accessibility verified November 19, 2019.

Viewpoint

Artificial Intelligence and the Implementation Challenge

James Shaw[1,2], PhD; Frank Rudzicz[3], PhD; Trevor Jamieson[1,4], MD, MBI; Avi Goldfarb[5], 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.

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

[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 contributed to writing and revising 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

1. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. J Am Med Assoc 2016 Dec 13;316(22):2353-2354. [doi: 10.1001/jama.2016.17438] [Medline: 27898975]
2. Naylor CD. On the prospects for a (deep) learning health care system. J Am Med Assoc 2018 Sep 18;320(11):1099-1100. [doi: 10.1001/jama.2018.11103] [Medline: 30178068]
3. Thiébaut R, Thiessard F, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics. Artificial intelligence in public health and epidemiology. Yearb Med Inform 2018 Aug;27(1):207-210. [doi: 10.1055/s-0038-1667082] [Medline: 30157525]
4. Harwich E, Laycock K. Wilton Park. London: National Health Service; 2018. Thinking On Its Own: AI in The NHS URL: https://www.wiltonpark.org.uk/wp-content/uploads/Thinking-on-its-own-AI-in-the-NHS.pdf
5. Ajay A, Gans J, Goldfarb A. The Economics of Artificial Intelligence: An Agenda. Washington DC: National Bureau of Economic Research; 2019.
6. Nilsen P, Ståhl C, Roback K, Cairney P. Never the twain shall meet?-a comparison of implementation science and policy implementation research. Implement Sci 2013;8(1):63. [doi: 10.1186/1748-5908-8-63]
7. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]

8. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Dec 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
9. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond 'implementation': digital health innovation and service design. NPJ Digit Med 2018;1(1):48. [doi: 10.1038/s41746-018-0059-8]
10. May C. Agency and implementation: understanding the embedding of healthcare innovations in practice. Soc Sci Med 2013;78:26-33. [doi: 10.1016/j.socscimed.2012.11.021published] [Medline: 23246396]
11. Agrawal A, Gans J, Goldfarb A. Prediction Machines: The Simple Economics of Artificial Intelligence. Brighton, Massachusetts: Harvard Business Review Press; 2018.
12. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014 Jul;33(7):1139-1147. [doi: 10.1377/hlthaff.2014.0048] [Medline: 25006139]
13. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005 Jan;39(1):98-106. [doi: 10.1111/j.1365-2929.2004.01972.x] [Medline: 15612906]
14. Victor-Chmil J. Critical thinking versus clinical reasoning versus clinical judgment: differential diagnosis. Nurse Educ 2013;38(1):34-36. [doi: 10.1097/NNE.0b013e318276dfbe] [Medline: 23222632]
15. Esteva A, Kuprel B, Novoa R, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
16. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
17. Rajpurkar P, Polamreddi V, Balakrishnan A. ResearchGate.: Cornell University Library; 2017. Malaria likelihood prediction by effectively surveying households using deep reinforcement learning URL: https://www.researchgate.net/publication/321324506_Malaria_Likelihood_Prediction_By_Effectively_Surveying_Households_Using_Deep_Reinforcement_Learning
18. Chinaei H, Currie LC, Danks A, Lin H, Mehta T, Rudzicz F. Identifying and avoiding confusion in dialogue with people with alzheimer's disease. Comput Linguist 2017 Jun;43(2):377-406. [doi: 10.1162/COLI_a_00290]
19. Liu N, Logan B, Liu N, Xu Z, Tang J, Wang Y. Deep reinforcement learning for dynamic treatment regimes on medical registry data. Healthc Inform 2017 Aug;2017:380-385 [FREE Full text] [doi: 10.1109/ICHI.2017.45] [Medline: 29556119]
20. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 2016 Mar;3(3):243-250. [doi: 10.1016/S2215-0366(15)00471-X] [Medline: 26803397]
21. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. Acad Emerg Med 2008 Feb;15(2):159-170 [FREE Full text] [doi: 10.1111/j.1553-2712.2007.00032.x] [Medline: 18275446]
22. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. IEEE Access 2017;5:8869-8879. [doi: 10.1109/ACCESS.2017.2694446]
23. Miller S, Nimri R, Atlas E, Grunberg EA, Phillip M. Automatic learning algorithm for the MD-logic artificial pancreas system. Diabetes Technol Ther 2011 Oct;13(10):983-990. [doi: 10.1089/dia.2010.0216] [Medline: 21774690]
24. Casey D, Beyan O, Murphy K, Felzmann H. Robot-Assisted Care for Elderly With Dementia: Is There a Potential for Genuine End-User Empowerment? In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts. 2015 Presented at: HRI'15 Extended Abstracts; March 2-5, 2015; Portland, OR p. 247-248.
25. Chaudoir SR, Dugan AG, Barr CH. Measuring factors affecting implementation of health innovations: a systematic review of structural, organizational, provider, patient, and innovation level measures. Implement Sci 2013 Feb 17;8:22 [FREE Full text] [doi: 10.1186/1748-5908-8-22] [Medline: 23414420]
26. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
27. Légaré F, O'Connor AM, Graham ID, Saucier D, Côté L, Blais J, et al. Primary health care professionals' views on barriers and facilitators to the implementation of the Ottawa decision support framework in practice. Patient Educ Couns 2006 Nov;63(3):380-390. [doi: 10.1016/j.pec.2006.04.011] [Medline: 17010555]
28. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc 2017 Mar 1;24(2):423-431. [doi: 10.1093/jamia/ocw105] [Medline: 27516495]
29. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
30. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med 2018;15(11):e1002689 [FREE Full text] [doi: 10.1371/journal.pmed.1002689] [Medline: 30399149]
31. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

32. Agrawal A, Gans J, Goldfarb A. WSB Wiki. Toronto: University of Toronto; 2018. The Labor Market Impact of Artificial Intelligence URL: https://wiki.bus.wisc.edu/download/attachments/45908081/AGG_2018_10_15.pdf?version=1&modificationDate=1549310002037&api=v2
33. Shaw J, Shaw S, Wherton J, Hughes G, Greenhalgh T. Studying scale-up and spread as social practice: theoretical introduction and empirical case study. J Med Internet Res 2017;19(7):e244 [FREE Full text] [doi: 10.2196/jmir.7482] [Medline: 28687532]
34. Greenhalgh T, Wieringa S. Is it time to drop the 'knowledge translation' metaphor? A critical literature review. J R Soc Med 2011 Dec;104(12):501-509 [FREE Full text] [doi: 10.1258/jrsm.2011.110285] [Medline: 22179293]
35. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? Br Med J 2014 Jun 13;348:g3725 [FREE Full text] [doi: 10.1136/bmj.g3725] [Medline: 24927763]
36. Patel VL, Kaufman DR, Arocha JF. Emerging paradigms of cognition in medical decision-making. J Biomed Inform 2002 Feb;35(1):52-75 [FREE Full text] [doi: 10.1016/S1532-0464(02)00009-6] [Medline: 12415726]
37. Rossi RA, Ahmed NK. The Network Data Repository with Interactive Graph Analytics and Visualization. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015 Presented at: AAAI'15; January 25-30, 2015; Austin Texas p. 4292-4293.
38. Samek W, Wiegand T, Müller KR. arXiv. 2017. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models URL: https://arxiv.org/pdf/1708.08296.pdf
39. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. J Am Med Assoc 1992 Nov 4;268(17):2420-2425. [doi: 10.1001/jama.1992.03490170092032] [Medline: 1404801]
40. Tracy CS, Dantas GC, Upshur RE. Evidence-based medicine in primary care: qualitative study of family physicians. BMC Fam Pract 2003 May 9;4:6 [FREE Full text] [doi: 10.1186/1471-2296-4-6] [Medline: 12740025]
41. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-367 [FREE Full text] [doi: 10.1007/s12553-017-0179-1] [Medline: 29308344]
42. Wadmann S, Hoeyer K. Dangers of the digital fit: rethinking seamlessness and social sustainability in data-intensive healthcare. Big Data Soc 2018 Jan 11;5:205395171775296. [doi: 10.1177/2053951717752964]
43. Vezyridis P, Timmons S. Understanding the care.data conundrum: new information flows for economic growth. Big Data Soc 2017 Jan;4(1):205395171668849. [doi: 10.1177/2053951716688490]
44. Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. J Am Med Assoc 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]
45. Grundy Q, Held FP, Bero LA. Tracing the potential flow of consumer data: a network analysis of prominent health and fitness apps. J Med Internet Res 2017;19(6):e233 [FREE Full text] [doi: 10.2196/jmir.7347] [Medline: 28659254]
46. Culnane C, Rubinstein B, Teague V. arXiv. 2017. Health data in an open world URL: https://arxiv.org/ftp/arxiv/papers/1712/1712.05627.pdf
47. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc 2016 Dec;3(2):205395171667967. [doi: 10.1177/2053951716679679]
48. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
49. Higginbottom K. Forbes. 2018. The pros and cons of algorithms in recruitment URL: https://www.forbes.com/sites/karenhigginbottom/2018/10/19/the-pros-and-cons-of-algorithms-in-recruitment/#79f3658b7340
50. Tsymbal A. ResearchGate.: Computer Science Department, Trinity College Dublin; 2004. The problem of concept drift: definitions and related work URL: https://www.researchgate.net/publication/228723141_The_Problem_of_Concept_Drift_Definitions_and_Related_Work
51. Žliobaitė I, Pechenizkiy M, Gama J. An overview of concept drift applications. In: Big Data Analysis: New Algorithms for a New Society. Switzerland: Springer International Publishing; 2016:91-114.
52. Martin G, Martin P, Hankin C, Darzi A, Kinross J. Cybersecurity and healthcare: how safe are we? Br Med J 2017;358:j3179. [doi: 10.1136/bmj.j3179] [Medline: 28684400]
53. Clarke R, Youngstein T. Cyberattack on Britain's national health service - a wake-up call for modern medicine. N Engl J Med 2017 Aug 3;377(5):409-411. [doi: 10.1056/NEJMp1706754] [Medline: 28591519]
54. Sittig DF, Singh H. A socio-technical approach to preventing, mitigating, and recovering from ransomware attacks. Appl Clin Inform 2016;7(2):624-632 [FREE Full text] [doi: 10.4338/ACI-2016-04-SOA-0064] [Medline: 27437066]
55. Youssef AE. A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments. Int J Ambient Syst Appl 2014;2(2):1-11. [doi: 10.5121/ijasa.2014.2201]
56. Jogova M, Shaw J, Jamieson T. The regulatory challenge of mobile health: lessons for Canada. Healthc Policy 2019;14(3):19-28. [doi: 10.12927/hcpol.2019.25795] [Medline: 31017863]
57. Torbica A, Cappellaro G. Uptake and diffusion of medical technology innovation in Europe: what role for funding and procurement policies? J Med Market 2010 Jan;10(1):61-69. [doi: 10.1057/jmm.2009.48]
58. Allen B, Wade E, Dickinson H. Bridging the divide - commercial procurement and supply chain management: are there lessons for health care commissioning in England? J Public Procure 2009 Mar;9(1):79-108. [doi: 10.1108/JOPP-09-01-2009-B003]
59. Perrow C. Normal Accidents: Living With High Risk Technologies. Princeton, NJ: Princeton University Press; 2011.

60. Vogus TJ, Welbourne TM. Structuring for high reliability: HR practices and mindful processes in reliability-seeking organizations. J Organ Behav 2003 Nov;24(7):877-903. [doi: 10.1002/job.221]
61. Zuboff S. Big other: surveillance capitalism and the prospects of an information civilization. J Inf Technol 2015;30(1):75-89. [doi: 10.1057/jit.2015.5]
62. CB Insights. 2018. AI In Healthcare Heatmap: From Diagnostics To Drug Discovery, Deals Heats Up URL: https://www.cbinsights.com/research/artificial-intelligence-healthcare-heatmap-expert-intelligence/
63. Scott WR. Institutions and Organizations: Ideas, Interests, and Identities. New York: Sage Publications; 2013.
64. North DC. Institutions. J Econ Perspect 1991;5(1):97-112. [doi: 10.1257/jep.5.1.97]
65. Battilana J, Leca B, Boxenbaum E. How actors change institutions: towards a theory of institutional entrepreneurship. Acad Manag Ann 2009 Jan;3(1):65-107. [doi: 10.1080/19416520903053598]

Abbreviations

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

Edited by G Eysenbach; submitted 07.02.19; peer-reviewed by KL Ong, S Chen, S Zheng; comments to author 27.04.19; revised version received 16.05.19; accepted 31.05.19; published 10.07.19

Please cite as:
Shaw J, Rudzicz F, Jamieson T, Goldfarb A
Artificial Intelligence and the Implementation Challenge
J Med Internet Res 2019;21(7):e13659
URL: https://www.jmir.org/2019/7/e13659/
doi: 10.2196/13659
PMID: 31293245