Healthcare

VANTAGE POINT®

A Risk Management Resource for Hospitals and Health Systems | 2020 Issue 1

Artificial Intelligence: Examining Five Key Sources of Liability

Artificial intelligence (AI) – a term encompassing such concepts as machine learning, pattern recognition, natural language processing, robotics and neural networks designed to replicate human thought processes – has become a driving force within the healthcare information technology field. (See "Artificial Intelligence Defined," page 2.) With its potential to enhance diagnostics and treatment, streamline administrative and operational processes, and engage patients in innovative preventive care programs, AI is poised to revolutionize medical practice and healthcare management. (See "Clinical Applications of Artificial Intelligence Tools," page 4.)

However, as with any revolutionary development, the advent of sophisticated medical AI tools has produced some degree of uncertainty and anxiety. Clinicians wonder whether this potentially disruptive technology will serve to augment their training and judgment, or instead render their hard-won knowledge and skills superfluous. Administrators note the lack of legal and regulatory parameters governing AI, a situation that presents compliance, ethical and liability questions.

In a 2015 survey, 86 percent of healthcare and life sciences respondents reported that they were using some form of cognitive technology, with the remaining 14 percent noting that they planned to do so by 2020. (Healthcare industry findings are found on page 42 of the study.) The COVID-19 crisis has helped jump-start clinical acceptance of this advanced technology, as existing AI systems are being retooled to assist providers in predicting the course of the illness in individual patients and identifying who is likely to require intensive care.

In this issue…

Until concrete guidance is available from government agencies and professional associations, healthcare providers and organizations must take the initiative to familiarize themselves with the uses, benefits, limitations and hazards of AI tools, and to create a framework to evaluate their safety and effectiveness. This edition of *Vantage Point®* examines five major AI-related sources of liability that are of potential concern in clinical care settings:

1. Data inaccessibility
2. Data breach
3. Data or outcome bias
4. Black-box reasoning
5. Automation bias

The issue concludes with a brief discussion of the professional liability implications of clinical AI and a convenient checklist of measures (see page 7) designed to address these five areas of risk and minimize liability exposure.

1. Data Inaccessibility

To achieve predictive accuracy, AI algorithms (i.e., sequences of problem-solving instructions) must be continuously "fed" large amounts of reliable data. For this reason, the long-term success of clinical AI tools depends upon the ability to convert diverse types of information from many sources into an integrated, structured database. Unfortunately, while healthcare organizations are inundated with data from information sources scattered across the continuum of care – e.g., electronic healthcare records, laboratory and imaging findings, patient texts and SMS messages, physician notes, billing and claims files – the systems that produce and store these data are not necessarily engineered for easy computer interface.

Compiling and arranging data into useable form often requires a sizable investment in improving information technology infrastructure and training staff. In addition, organizations must be prepared to address technical and administrative barriers to effective data sharing among key players, including healthcare providers and facilities, pharmaceutical companies, diagnostic services and insurers.

The long-term **success of clinical AI tools** depends upon the **ability to convert** diverse types of **information** from many sources **into an integrated, structured database**.

Artificial Intelligence Defined

Artificial intelligence (AI) is the science of creating tools that process masses of data, recognize subtle patterns, and perform other tasks with minimal supervision and intervention, using technology designed to mimic human cognitive functioning. In a clinical context, AI's analytical and predictive powers can help providers expedite and enhance the diagnostic process, avoid unnecessary lab tests, select more precise and efficacious treatments, detect looming crises, encourage preventive self-care and reduce hospital readmissions, among other benefits.

Two Common AI Methodologies

Machine learning:	Deep learning:
A type of algorithm that allows software applications to more accurately predict clinical outcomes by reviewing data in an iterative manner and reaching a conclusion (known as an *output*) through statistical analysis. For example, so-called "smart" healthcare records may be useful in identifying population health risk factors, predicting illness and modeling disease progression.	A subset of machine learning involving AI systems that perform human-like tasks, such as recognizing speech, analyzing images and rendering diagnoses. For example, computer-assisted MRI workstations may potentially improve the speed and accuracy of cancer tumor diagnosis.

Source: Ross, D. and Surgenor, V. "Artificial Intelligence and Healthcare: FAQs." *Business Law Today*, a publication of the American Bar Association. Posted February 8, 2019.

2. Data Breach

Clinical AI applications utilize a centralized database of protected health information (PHI), which must be guarded against data breaches and other types of unauthorized disclosures. Safe implementation of AI thus requires effective security strategies, as well as clear policies that define who has access to patient data and for what purposes, what types of data can be stored centrally, and how PHI is safeguarded and rendered anonymous.

At present, HIPAA privacy regulations apply to statutorily defined PHI used to train AI systems. Among other provisions, these regulations require that confidential information be "de-identified" before being used in an AI context or, alternatively, that it be input into AI systems only with the consent of the patient. It is yet to be seen whether HIPAA regulations will be amended to permit the feeding of "raw" data into AI programs. For now, organizations should rigorously monitor data use and transfer in light of HIPAA rules.

What is clear is that the advent of data-driven AI systems creates significant privacy concerns, as a centralized PHI database seems a natural target for cyber security threats. In a 2018 study, only 35 percent of 500 surveyed patients expressed confidence that data utilized for AI purposes are stored securely. As many AI tools require active patient involvement, privacy questions could potentially derail the widespread adoption and effective functioning of some forms of machine learning technology.*

The key issue of data security and confidentiality in relation to AI is evolving, with many questions still unanswered. For more information about privacy expectations as they apply to providers, see the broad policy statement issued by the American Medical Association about the potential effects of "augmented intelligence" on healthcare delivery. The statement notes that, as AI systems can easily identify even "anonymized" data, traditional privacy expectations "are simply no longer attainable."

3. Data or Outcome Bias

To function effectively with all patient groups, AI platforms and projects must have access to data that are accurate, up-to-date and inclusive. The potential for bias may arise when the data used to "train" AI algorithms fail to represent the entirety of a population, or when algorithms are tainted by racial, gender, socioeconomic and/or age-related biases. Data bias also may develop due to faulty utilization, such as applying an AI tool to an unintended patient population, or failing to update data to reflect changes in disease patterns.

Data-related disparities in AI tools can significantly affect care provided to underrepresented or vulnerable patient groups, potentially resulting in missed or failed diagnoses, as well as other clinical oversights. Whereas human practitioners who have a relationship with and a psychological understanding of individual patients can often identify and compensate for inaccuracies or omissions in their presentations, AI systems operate in a relatively rigid manner, processing the information they are given without additional input or insight. Excessive reliance on these tools may produce the following types of lapses, among others, which in turn could potentially affect patient safety and outcomes: **

- **Over-focus on a predetermined end-goal,** as when a robotic surgery device bypasses critical clinical findings not directly related to the operation and initiates a procedure despite contraindications.
- **"Reward hacking,"** as when an AI predictive model connected to an automated medication administration system gives a dose of heparin just before a patient's activated partial thromboplastin time (aPTT) is measured. While this action controls clotting in the short term, it does little to achieve long-range stability or control.
- **Unsafe exploration,** as when an AI-driven diagnostic algorithm does not undergo regular human review and testing as it absorbs data and develops its capacities, eventually leading to faulty results.
- **Unsafe failure mode,** as when an AI decision-support system neglects to inform the user that it lacks sufficient information to render a reliable recommendation or otherwise cannot function as designed, resulting in an inaccurate output.

1 To enhance AI-related data security and allay patients' doubts, some healthcare alliances are embracing blockchain technology, which stores data in a decentralized manner on servers, laptops and other computing devices within an interconnected network. While this innovative technique can help prevent major data breaches, it should be noted that it is expensive to implement, consumes a great deal of energy and currently lacks adequate regulatory protocols.

2 For more information about these and other AI-related hazards, see "The Dangers of AI in the Healthcare Industry [Report]." *Thomas Insights*, May 7, 2019.

Clinical Applications of Artificial Intelligence Tools

Clinical diagnosis:	Treatment and care:	Patient and community support:
Clinical decision-support software that identifies diseases faster and with greater accuracy, using a combination of historical medical data and patient records.	**Virtual bedside voice assistants** that monitor doctor-patient interactions, suggest treatment approaches, and alert caregivers to patient requests or impending emergencies. (See "Privacy Guidelines for Virtual Voice Assistants" on page 5.)	**Interactive kiosks** used to register patients and refer them to appropriate providers.
Automated testing tools, such as blood glucose monitors that analyze data generated from sensors attached to a patient's body and interface with provider records.	**Wearable devices** that collect data and automatically connect to electronic healthcare record systems in real time.	**Population health analytic applications** that reduce hospitalizations and treatment costs by detecting gaps in healthcare delivery and encouraging providers to address them.
Specialized algorithms used for radiological image analysis.	**Robotics and telehealth systems** that virtually connect providers to patients.	**Programs that utilize medical and environmental factors** to forecast patient behavior, calculate disease probabilities, and advise both providers and patients.
Computerized vision and other machine learning technologies that analyze bodily fluids and tissues, in order to detect potential pathology.	**"Smart" electronic healthcare records** that generate and extract data in real time, and that enhance physician orders using predictive technology.	**Health profiles based upon genetics and blood markers** that help patients understand and manage their specific risk factors.

Source: Thomas, M. "Ultra-modern Medicine: Examples of Machine Learning in Healthcare." *Built In*, post updated February 24, 2020.

With its potential to **enhance diagnostics and treatment, streamline administrative** and **operational processes,** and **engage patients** in innovative preventive care programs, **AI** is poised to **revolutionize medical practice** and healthcare management.

4. Black-box Reasoning

"Black-box reasoning" refers to the breakdown in control and communication that may occur when providers lack full understanding of how an AI algorithm reaches a diagnostic or therapeutic conclusion. In some cases, the clinical decision support (CDS) software program or code reveals the system's logic. However, decision-making criteria are less discernible in more complex tools that utilize neural networks or deep learning systems, which continue to develop over time through ongoing absorption of training data and analysis of outcomes.

The lower the degree of human input in an AI system, the more difficult it is for healthcare professionals not only to explain the clinical rationale for treatment recommendations, but also to gauge the tool's safety and effectiveness. In addition to weakening patient trust in both physicians and AI applications, black-box scenarios could potentially lead to lapses in patient care due to flaws or biases in algorithms, as well as a weakening of accountability for medical decisions made.

The Food and Drug Administration has begun the process of defining the types of CDS software it will regulate, as part of the agency's effort to create a risk-based oversight system for these computerized tools. Until this effort materializes, organizations may wish to consult the voluntary industry guidelines published in 2017 by the Clinical Decision Support Coalition, an industry watchdog group that advocates for patient safety and the continued central importance of clinicians in AI-enabled decision-making. (Scroll down to page 12 of the guidelines.)

Privacy Guidelines for Virtual Voice Assistants

Alexa, Siri and other virtual voice assistants are assuming an ever-wider range of healthcare-related tasks. Patients are increasingly making use of these robotic helpers, which incorporate artificial intelligence (AI) technology, to connect to their healthcare information record, obtain on-demand medical advice, receive postoperative instructions and check medication regimens. Voice assistants are also becoming virtual members of the medical team, serving such practical functions as monitoring interactions between providers or clinical staff and patients, suggesting treatment options and alerting clinicians to impending medical emergencies.

Because of their versatility and convenience, virtual voice assistants have become popular in healthcare settings. However, the expanding use of this form of AI does create privacy concerns. Amazon has addressed this issue directly, offering software that allows the company's Alexa system to securely transmit sensitive patient data.* A variety of providers and healthcare organizations – from hospitals to pharmacy benefit managers to insurance companies – are expected to adopt the new generation of HIPAA-compliant technology.

At this point, though, not all virtual voice assistants satisfy HIPAA privacy regulations, and organizations must take care when using any such system to prevent unwanted disclosure of identifiable health information. The following usage guidelines can help ensure that Alexa, Siri and related products serve only as useful tools and not as electronic eavesdroppers:

- **Engage the device's mute feature when it is not in use,** in order to prevent continual recording of voices.
- **Prohibit access to sensitive patient data**, limiting device use to medication profiles, appointment reminders, care schedules and similar routine functions.
- **Automatically erase old recording history from the system's privacy dashboards**, or selectively delete individual queries and requests.
- **Periodically change and strengthen passwords**, following established cyber security guidelines.
- **Restrict third-party use of the device** by adjusting settings to limit access.

* See Ross, C. "Amazon Alexa Is Now HIPAA-compliant. Tech Giant Says Health Data Can Now Be Accessed Securely." *STAT*, posted April 4, 2019.

5. Automation Bias

Automation bias – a situation in which users favor the suggestions of digital decision-support systems over their own experience, training and professional judgment – presents a credible risk in the healthcare sector, where medical situations may involve subtle variables that cannot be incorporated easily into a computer algorithm. Over-trust in automation can erode clinical decision-making skills and encourage complacency, potentially resulting in providers unthinkingly accepting questionable AI-related care recommendations.

In addition, the tendency to rely excessively upon AI tools could potentially affect key risk management processes. As human input decreases, errors are less apt to be reported through customary reporting channels. Moreover, extreme automation also makes it more difficult to detect sources of error through traditional quality and performance assessment activities, such as record audits and peer review.

Over-trust in automation can erode clinical decision-making skills and **encourage complacency**, potentially resulting in **providers** unthinkingly **accepting questionable** AI-related **care recommendations**.

Professional Liability Implications of Artificial Intelligence

AI may fundamentally alter the patient-physician relationship, as "self-learning" algorithms not only support, but also potentially compete with, human clinicians' ability to diagnose and treat diseases. As traditional notions of provider responsibilities change, so perhaps will the concept of negligence. In the future, it is possible that documented use of AI-based diagnostic or treatment output could help reduce liability exposure, especially when the tool's accuracy rate is historically high. Failure to consult a reliable decision-making AI tool, on the other hand, could conceivably weaken legal defense in the event of a claim.

The full implications of AI in relation to standard of care and medical malpractice are far from understood. At this point, AI diagnostic algorithms are thought of as tools that serve to assist human practitioners, rather than providing authoritative "expert opinions," and AI-based recommendations are just one element among many that enter into physician consideration. On the positive side, as clinical AI evolves in terms of reliability, providers – supported by the output of sophisticated AI applications – may feel less pressure to practice "defensive medicine," defined as protective overuse of expensive tests. On the negative side, as AI systems develop in complexity and become less transparent in their operations, physician authority, autonomy and professional judgment may decline, creating a high degree of dependency on these electronic tools, as well as liability exposure arising from their potential flaws and limitations. Therefore, organizational leadership must emphasize to providers that they retain primary responsibility for clinical decision-making, even when the patient care process is supplemented by AI technology.

As with all tools, artificial intelligence applications make a good servant but a poor master. The speed and computing power of these clinical decision-making aids should not distract from the fact that they are only as reliable as their programming and data inputs. For this reason, the decisions they arrive at should not be accepted blindly. Diagnostic and treatment determinations are ultimately a matter of human intelligence and judgment, and healthcare providers and organizations are unlikely to evade responsibility for preventable errors by pointing to incorrect AI outputs.

The following checklist is designed to encourage organizational leadership to think strategically about the specific risks and hazards discussed in this publication, as well as about the larger ethical, regulatory and liability questions raised by the advent of clinical AI systems.

Quick Links

- "Artificial Intelligence in Healthcare." eHealth Initiative/Cerner®, November 2018.
- Glaser, J. "Understanding Artificial Intelligence in Health Care." Posted on the website of the American Hospital Association, January 23, 2018.
- Gluck, J. "How Automation in Healthcare Is Boosting the Bottom Line." *HealthTech*, a CDW publication, June 11, 2018.

Checklist of Artificial Intelligence (AI) Risk Control Strategies

Risk Exposure	Yes/No	Comments
Data inaccessibility: Has the organization …		
Examined its information technology (IT) architecture and created an inventory of all data sources across multiple sites, points of service and patient populations?		
Verified the capacity of the organization's IT network to handle large quantities of data from various sources in real time?		
Revisited the electronic healthcare record system to ensure that it is patient-centered, permits data input by providers, and interfaces easily with outside providers, wearable patient monitoring devices and other sources of data?		
Adopted natural language processing software that can understand and extract clinical information from unstructured data sources, such as text, images, and audio and video input?		
Developed formal policies and procedures for aggregating and managing data from various sources?		
Hired an analytics platform partner to assist in aggregating data, thus ensuring that accurate, usable information is available for AI system purposes?		
Applied advanced analytics when aggregating data from multiple sites, in order to identify inefficiencies in care and skewed or unrepresentative datasets, which may negatively affect AI functioning?		
Created a uniform database that incorporates patient, payor and provider information?		
Identified data voids and disruptions, and, in consultation with data analysts, developed strategies to access usable information in these areas?		
Begun the process of creating a longitudinal record of care across the continuum, utilizing high-performance computing and "5G" wireless technology, as well as other advanced analytic tools?		
Data breach: Has the organization …		
Incorporated effective security measures into AI tools and associated databases, in order to prevent improper disclosure of protected health information?		
Implemented clear, legally valid permission protocols for sharing and using data from the many different sources that flow into clinical AI systems?		
Performed routine risk assessments of AI systems and databases to identify potential cyber security threats, estimate the likelihood of their occurrence and their potential severity, and devise preventive measures?		

Risk Exposure	Yes/No	Comments
Data bias: Has the organization …		
Disclosed the provenance and quality of data utilized to "train" AI systems, openly and honestly noting where it was collected, how it was labeled and what measures were taken to ensure its accuracy?		
Considered the needs of AI system end-users and tested the tool with these providers, in order to assess its performance and identify system weaknesses, data gaps and biases, and user questions and concerns?		
Asked the following questions, among others, about the data fed into AI systems, in order to identify and prevent potential sourcing bias:		
• **Are the data inclusive of all patient populations,** or are they skewed toward a particular class or group?		
• **Could the data have hidden and inherent biases,** especially in terms of demographic and socioeconomic attributes?		
• **Does the database include uncommon cases,** in order to ensure expansive diagnostic capability?		
Developed a process for reviewing AI-generated decisions, which includes the following queries, among others:		
• **Is the provider made aware of the criteria** upon which decisions are made?		
• **Are output decisions presented simply and clearly,** i.e., do they involve a choice between two straightforward clinical options?		
• **Is the provider instructed to apply independent professional judgment before accepting decisions,** especially in cases when the clinical presentation includes variables outside of the AI tool's decision-making parameters?		
Created a mechanism for providers to question AI outputs, especially in situations with a lower degree of certainty?		
Established a feedback mechanism when an output is questioned or rejected by a provider, which includes documented retraining of the AI system when data bias or other system flaws are uncovered?		
Black-box reasoning: Has the organization …		
Instituted a long-term strategy for adoption and safe utilization of AI, communicating the technology's nature, strengths and limitations, as well as available training opportunities and performance expectations, to clinicians well in advance of implementation?		
Introduced AI technology as a tool intended to complement providers' diagnostic and treatment skills, thus permitting clinicians to gradually develop confidence and proficiency?		
Engaged medical experts to oversee development of AI algorithms and check and validate system outputs?		
Developed transparency requirements for data used to train AI systems, including the ability to identify academic or proprietary databases, journal articles or other published materials, clinical guidelines, research findings and other information sources?		
Encouraged providers to work with data analysts in the annotation, publication and presentation of data, in order to minimize the risk of misinterpretation and strengthen confidence in the AI system's reliability?		

Risk Exposure	Yes/No	Comments
Automation bias: Has the organization …		
Stipulated in written policy that providers are responsible for using their full range of abilities and aptitudes – including professional judgment, intuition, empathy, imagination, and critical thinking and abstract reasoning skills – when diagnosing and treating patients, rather than passively accepting machine-generated decisions?		
Selected AI systems and tools that reveal the extent of confidence or uncertainty about a given clinical output?		
Created a framework for the active management and evaluation of AI systems, in order to ensure that programming flaws inadvertently encoded into algorithms are detected and not passively perpetuated by providers?		
Identified sources of AI training data as part of the ongoing effort to instruct providers about how automated decision-support systems arrive at their clinical outputs?		
Developed provider education programs that address the risk of automation bias by offering error-avoidance strategies and presenting simulated scenarios in which decision-support systems offer faulty recommendations?		
Professional liability implications: Has the organization …		
Appointed carefully selected clinical and technical "champions" – who are responsible for fostering staff engagement, describing system benefits and risks, and explaining pending changes in workflow – across multiple departments before introducing advanced AI-driven systems?		
Arranged for a trusted third-party vendor to review and analyze diagnostic algorithms, in order to ensure that the data involved in AI projects are "clean," accurate, up-to-date and extracted from identifiable sources?		
Checked whether system designers are able to identify and justify their training data, using a "curated" set of AI tools for explanatory purposes?		
Promoted data analysis and mastery of advanced decision-support tools as a core competency for providers, and informed them that their proficiency in this area would be assessed during annual performance reviews?		
Created a chain of accountability for AI system utilization, identifying individual clinicians and defining their role in providing data and interpreting results?		
Reviewed and revised standard informed consent forms to ensure that they address the use of AI technology, explain that such systems depend upon a flow of patient data, and indicate the risk of privacy breaches and invalid or biased conclusions?		
Established a mechanism for reporting real and potential errors involving AI technology and documenting follow-up measures taken to ensure that systems are trained on updated data and that lapses do not recur?		

Did someone forward this newsletter to you? If you would like to receive future issues of *Vantage Point*® by email, please register for a complimentary subscription at go.cna.com/HCsubscribe.

CNA Risk Control Services: Ongoing Support for Your Risk Management Program

CNA provides a broad array of resources to help hospitals and other healthcare organizations remain current on the latest risk management insights and trends. Bulletins, worksheets and archived webinars, as well as past issues of this newsletter, are available at www.cna.com/riskcontrol.

Your SORCE® for Education

CNA's School of Risk Control Excellence (SORCE®) offers complimentary educational programs that feature industry-leading loss prevention, loss reduction and risk transfer techniques. Classes are led by experienced CNA Risk Control consultants.

SORCE® *On Demand* offers instant access to our library of risk control courses whenever the need arises. These online courses utilize proven adult-learning principles, providing an interactive educational experience that addresses current regulatory requirements and liability exposures.

Allied Vendor Program

CNA has identified companies offering services that may strengthen a hospital's or other healthcare organization's risk management program and help it effectively manage the unexpected. Our allied vendors assist our policyholders in developing critical programs and procedures that will help create a safer, more secure environment.

For more information, please call us at 866-262-0540 or visit www.cna.com/healthcare.