JCOTS
JOINT COMMISSION ON
TECHNOLOGY & SCIENCE

Limited Study

AI in Healthcare

October 2025

How to read this report

This report is organized into seven sections:

- The **executive summary** briefly synthesizes the key points in the report;
- The **introduction** provides an overview of the topic and explains its relevance to policy and policy makers;
- **Key concepts** provides definitions for important terms used throughout the report;
- **Policy landscape** gives an overview of federal and state policy regarding AI in healthcare, along with a summary of additional frameworks and guidance put forward by other industry and sector organizations;
- **Methodology** describes how this report was compiled;
- The **findings and discussion** section presents key takeaways from interviews with stakeholders alongside relevant discussion from a literature review on the topic;
- The **policy recommendations and roadmap** sets out several policy considerations and recommendations for lawmakers along with a responsible AI regulation roadmap that outlines a process for developing agile and collaborative regulation for AI in the health sector.

The report can be read as a continuous document, or readers can skip to sections of interest.

Executive Summary

Artificial intelligence (AI) is playing a growing role in medicine and healthcare as one of many technologies that support clinical diagnosis, decision-making, patient care, and administration. Some form of advanced computational analytics have been used in healthcare for decades, from rule-based expert systems that rely on defined inputs and outputs based on established clinical principles and expertise (e.g. some drug-drug interaction prediction systems) to machine learning (ML) used in diagnosis (e.g. some cancer-detecting systems that interpret medical images). In recent years, AI advances have resulted in more powerful, adaptable, and autonomous computational systems for healthcare applications, raising important questions about how these systems are developed, the quality and security of data processed by AI models, the risks or harms that these systems might exacerbate or introduce in clinical settings, and the balance of responsibility between AI and clinicians, who have traditionally been the primary mediators of patient care and outcomes.

The aim of this study is to provide a broad overview of the current state and future horizon of AI in healthcare and to identify policy-relevant issues that might warrant the attention of lawmakers in Virginia. Through a literature review of scholarly and grey literature and interviews with subject matter experts, the study provides an overview of key issues around AI in healthcare broadly, before focusing on enterprise AI adoption by hospitals and health systems for clinical decision support and clinician-patient interactions. Centering on AI adoption in these settings provides a helpful entry point for understanding AI in healthcare because established hospital and health systems are often well-informed about emerging technologies and have existing processes and procedures for adopting or procuring new technologies that inform their approach to AI. At the same time, because they already have broad familiarity with clinical support technologies, their experiences also help to highlight where there are unique challenges, gaps, or risks associated with AI, specifically.

Challenges and controversies for AI in healthcare

Several challenges and controversies have characterized AI in healthcare, including the black box nature of many contemporary AI systems, a lack of transparency around how AI systems are developed and perform, the risk of imbalanced or unrepresentative health data informing clinical decisions, discriminatory outcomes for underrepresented or marginalized patients, and new risks to data privacy and security.

The table below provides a summary of these key challenges covered in the report:

The "black box"	• Unlike in rule-based systems, where the relationship between inputs (data) and outputs is inherently explainable, the relationship between inputs and outputs in AI systems today is often not known or knowable because the model learns from data without supervision • The lack of explainability can make it difficult for clinicians to interpret results • End-users may wind up highly dependent on information provided by model developers about accuracy and performance
Lack of transparency	• The black box problem is compounded by a general lack of transparency and disclosure about how models are trained, validated, and evaluated • Many developers do provide model information to deployers/end-users, but they are largely company-specific • Few standard disclosures are required in the sector, leaving it up to developers/deployers/end-users to determine what information should be shared • Many clinical AI tools are not externally validated through gold standard methods like prospective randomized control trials or peer-review
Data quality issues	• Models are only as good as the data used to train and fine-tune them, and there are known challenges for AI processing health data: a lack of data about certain patient populations due to historical exclusions; fragmented data in multiple, incompatible formats; general flaws in data accuracy at the time of recording; issues with labeling data for one purpose (e.g. billing) that is subsequently interpreted by a model for another purpose (e.g. clinical diagnosis) • Clinical data is also subject to the "streetlight effect," where only the data that is available can be analyzed using AI, often to the exclusion of data that might be more useful but is harder to collect

Discriminatory outcomes for under-represented or marginalized patient populations	• There are examples of AI systems used in clinical settings that have led to unfair outcomes for marginalized patient groups • AI models have shown a propensity for generating outputs based on characteristics like gender and race within datasets where these attributes are not pathologically relevant
New data privacy and security risks	• Data sharing between different entities (developers/deployers) creates new opportunities for data breaches • Models can be improved with localized patient data, but sharing data with third parties for model training creates risks of data leakage, and many smaller or less resourced healthcare providers may not have the capacity to train models in-house • HIPAA protections are often insufficient for protecting patient information because powerful AI models can re-identify individual patients, even in de-identified datasets

AI in healthcare also surfaces a complex and crucial aspect of how technology impacts the real world: the unavoidable feedback loop between humans and technology. The human-technology relationship is defined by the inseparability of *human* cognition, culture, and values and *technological* development, adoption, and applications. The health sector has long advocated and embedded the principle of "humans in the loop"—meaning clinician oversight of AI outputs and recommendations—as a safeguard against computational errors. However, this study explores how humans play important roles "in the loop" at multiple stages of the AI lifecycle, some of which may require greater awareness and transparency (such as the role of developer decisions in shaping model performance), some of which may require additional safeguards against human cognitive biases (such as mitigating tendencies to overly trust or reject AI outputs), and some of which may require including more humans in the loop (such as patients, who may have unique perspectives on consent or data privacy). Ultimately, this study challenges the assumption that having a "human in the loop" constitutes a sufficient safeguard against AI harms, unless it accounts for the ways in which both humans *and* AI mutually shape the opportunities and risks of this emerging technology.

AI in healthcare in Virginia

Based on interviews with subject matter experts, the study also provides some insight into the current landscape on AI in healthcare in Virginia. Interviews highlight the prominence of generative AI, systems that are trained on vast quantities of diverse data and can generate unpredictable, novel content, such as text and images. In this context, generative AI is a newly popular and promising form of AI increasingly integrated into clinical workflows to improve efficiency, reduce administrative burdens, and facilitate better clinician-patient engagement. Generative AI has also thrown into stark relief many of the challenges associated with AI in clinical settings more broadly—challenges with transparency, data privacy, procurement and costs, and patient consent. The interviews surfaced several themes that point to policy issues:

(1) AI is not new in the healthcare sector, but the newest AI is front-of mind;
(2) The dominant business case for AI adoption is clinician burnout;
(3) Good AI governance is a top priority for healthcare technologists and providers, but it is uneven, institution-specific, and voluntary;
(4) Multiple and messy external vendor dependencies characterize AI adoption;
(5) Humans are staying in the loop to provide oversight of AI models;
(6) Developers, deployers, and end-users want regulatory clarity and harmonization.

The policy landscape

The policy landscape on AI in healthcare is fragmented at both the federal and state level, leaving large regulatory gaps and ambiguities. There is currently no overarching legislation governing AI in healthcare. At the federal level, the White House recently published *Winning the Race: America's AI Action Plan*, which positions AI as essential to American success and competitiveness. It mentions launching several "domain specific efforts," including in healthcare, to develop national standards for AI systems, however the *Action Plan* generally sets out a deregulatory stance, and it is unclear what the timeline will be for implementation of the plan's initiatives. Several federal agencies currently regulate AI in clinical applications, chiefly the FDA, which oversees regulation on medical devices. Some AI/ML integrations into tools that qualify as medical devices fall under its purview, but the limited scope of regulations issued by federal agencies like the FDA leave many clinical decision support tools largely unregulated.

Last year, the Virginia General Assembly passed HB2094 on "high-risk artificial intelligence," but the bill was ultimately vetoed by the Governor. It would have regulated AI used in "consequential decisions" in several sectors, including healthcare. However, there were notable exceptions in the bill, including for HIPAA-covered entities, which would have exempted many healthcare organizations and AI applications in healthcare.

Other states have passed AI legislation in recent years, some of which take a cross-sector approach to regulating high-risk applications of AI and some of which focus specifically on clinical uses. Much of the healthcare-relevant legislation addresses AI disclosures, often requiring patients to be informed when they are interacting with AI or when AI is employed in their care.

Beyond the limited regulation on AI in healthcare, many professional bodies, hospital systems, and researchers have published non-binding guidance on AI, such as the Coalition for Health AI (CHAI) Responsible AI Guide and the American Medical Association's Augmented Intelligence Principles. These guidance documents often inform individual organizations' governance frameworks, but they do not constitute legally enforceable standards.

The case for legislation

Taken together, the findings from this study strongly suggest a role for legislation and regulation on AI in healthcare. Sector-specific regulation could address many of the issues that surfaced in this report:

- Mitigate known risks and anticipate unknown risks before this emerging technology becomes normalized and a fragmented regulatory regime entrenches oversight gaps;
- Standardize safety requirements so that developers and deployers do not have to engage in lengthy, complex internal governance processes that may be superseded later by regulation;
- Provide support for adoption across the sector by sharing knowledge among different stakeholders and enhancing smaller providers' ability to adopt novel technologies;
- Clarify steps toward regulatory compliance for technology developers and deployers so that they can plan for the longer term;
- Set an example for responsibility in AI governance more broadly by starting with a sector that has extensive experience developing robust governance processes for technology adoption.

Policy Recommendations

Given the complexity of AI applications in healthcare and the multiplicity of stakeholders in the landscape, the regulatory approach that may prove best suited to the challenges identified in this study is one that balances taking action to address real risks and harms with a regulatory process that is gradual, iterative, and subject to review.

Policy recommendations are therefore two-fold: (1) establishing minimum requirements for developers and deployers of AI systems for use in healthcare and (2) establishing a multi-step roadmap for responsible AI regulation to aid in the development of practical, adaptable, and accountable regulation on emerging technology.

The following table provides breakdown of these two recommendations:

Policy Recommendation 1: Minimum requirements for AI developers and deployers	**Policy Recommendation 2: A roadmap process for responsible AI regulation**
a. Documented AI governance processes for healthcare providers; b. Model transparency standards; c. Model validation and evaluation standards; d. Adverse event reporting; e. Specific requirements for humans in the loop and clarity on liability; f. Public disclosures that provide patients and members of the public with information about AI use; g. Establishment of an enforcement agency that can receive complaints and concerns from patients or members of the public, proactively investigate non-compliance or violations of established regulation, and impose penalties.	a. **Consultation:** Convening stakeholders across healthcare, the technology industry, academia, and regulatory agencies to develop guidance to implement the principles outlined in Recommendation 1. b. **Guidance:** Issuing technical and implementation guidance by a designated oversight body based on the consultative process. c. **Data collection/testing:** Establishing or designating an oversight agency for collecting data on the effectiveness of the regulation and guidance, which would be responsible for designing and implementing a study of how the regulation performs in practice. d. **Evaluation and reporting:** Requiring a programmed review of regulatory outcomes and impacts based on the data collection/testing phase, to be reported to the enforcement agency identified in legislation as well as to the legislature. e. **Delayed enactment, phased compliance deadlines, tiered compliance requirements:** Providing stakeholders named in legislation time to develop implementation plans for regulatory guidance and providing different stakeholders (e.g. startups versus large enterprises) with tailored compliance requirements.

The presentation of these two policy recommendations side-by-side is intentional; they are meant to work hand-in-hand, with legislation establishing the principles of a regulatory regime for AI in healthcare and the regulatory roadmap developing sector-specific implementation and compliance guidance. The roadmap incorporates iterative consultation and review to accommodate the evolving nature of the technology and its real-world applications.

Introduction

Artificial intelligence (AI) is increasingly being integrated into healthcare to support medical diagnostics, treatment plans and monitoring, patient engagement, administration, and insurance processes. In the *U.S. Code*, AI is defined in 15 U.S.C. Section 9401(3) as:

> A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to (A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action.[1]

AI is a broad term that encompasses a variety of technologies, such as **rule-based expert systems**, **machine learning**, **neural networks**, **deep learning**, **large-language models (LLMs)**, and even **robotics** (see Key Concepts for definitions). It constitutes just one of many genres of advanced analytics systems, medical devices, and software that may be considered "**clinical decision support**" (CDS) tools—technologies intended to enhance clinical decision-making and workflows.[2] Although, it is worth noting that AI is also employed in a variety of functions within healthcare that are not strictly part of CDS. At one end of the spectrum are more **deterministic models** that produce consistent outputs for the same inputs, using rule-based logic. They are inherently explainable, meaning that the outputs can be directly traced back to inputs and the computations performed on those inputs.

At the other end of the spectrum are **probabilistic models**, which often engage in **unsupervised learning** from large datasets. These models infer patterns and relationships within datasets with degrees of uncertainty, so they produce different outputs based on probabilities, as in the case of chatbots generating responses to user prompts. The emergence of **generative AI** in recent years represents an important inflection point in AI evolution and is underpinned by probabilistic models, such as **foundation** models, that are trained on extensive and diverse datasets of text, images, video, and more, and can be fine-tuned to perform a variety of specialized tasks. The type of model employed in CDS affects the explainability, predictability, and reliability of outputs, and interpreting the outputs of different kinds of models requires an understanding of the computational processes that produce their results.

Some version of AI has been used in medicine for decades, starting with rule-based expert systems, where computers are programmed to perform rigidly defined logical operations based on expert knowledge.[3] Advances in both the digitization of health data, which rendered a large amount of health information legible to computerized systems, and in the field of machine learning have subsequently led to a proliferation of technologies and applications for AI in healthcare.[4] When AI systems are trained on relevant clinical data and robustly tested, validated, and evaluated, they have the potential to help identify patterns in large datasets that would be difficult for human analysts to interpret, unlocking new insights that could improve the accuracy of information clinicians need to make critical, life-and-death decisions.[5]

However, sophisticated AI systems that lead the field today (e.g. LLMs) can also be opaque, rendering their outputs difficult to interpret or reverse-engineer, and AI deployed in the real world has already generated risks and harms to patient populations, inviting greater scrutiny of this largely unregulated emerging technology. As this study will go on to unpack, the applications of AI in healthcare invite particular attention to the ways in which both positive and negative societal outcomes are the result of complex interactions between humans and technologies—from development to deployment. AI systems "are only the latest incarnation in a long line of probabilistic models that have been integrated into medical decision making, which have all required that their creators and implementers make value judgments."[6] The health sector—where clinical expertise is hard-won and highly valorized, and clinicians make consequential decisions based on a combination of empirical data, lived experience, and the interpretation of patient needs and values—offers a particularly instructive lens through which to understand the challenges and opportunities of AI-augmented decision-making under conditions of uncertainty, where probabilistic thinking and biases characterize both the machine output and the human judgment.[7] Mitigating the confirmed and potential harms of AI is therefore not only a question of technical specifications and requirements, but also a question of human and organizational judgment, values, and culture.[8] In other words, the success or failure of AI in the wild pivots on a combination of how AI produces outputs and also how people interpret and act on those outputs. Any regulatory solution also needs to account for this mutual influence of people on the technology and the technology on people.

AI applications in healthcare

AI has many different applications in healthcare, and as the technology rapidly develops, new opportunities are surfacing. Rule-based expert systems have long been employed to interpret electrocardiograms, and now deep neural networks are being put to the task with greater accuracy. Machine learning has been applied to reading images in radiology, pathology, ophthalmology, and more.[9] To date, the FDA has approved over 1200 AI-enabled medical devices, most of which are in the field of radiology and use computer vision to interpret images.[10] Surgical robots already assist clinicians, and robotic process automation integrates robotic principles and AI to automate administrative tasks, like billing. Many AI applications fall under the broad umbrella of predictive analytics, which provide risk scores for hospital readmission, sepsis, and patient falls.

Generative AI has ushered in a new era of AI applications in healthcare by enabling the creation of novel content and the execution of complex tasks that significantly surpass previous models. Built on foundation models that are trained on huge repositories of diverse data, generative AI models are then fine-tuned for specific clinical tasks, including image interpretation, messaging, and note-taking. Generative AI has enabled ambient listening and documentation, which involves recording patient-provider interactions and generating chart notes, prescription orders, and treatment recommendations. Chatbots built on generative AI can provide patients with information or triage patients to assign them to specific services.

Types of AI Models	Example clinical applications
Rule-based systems	• Clinical decision support alerts (drug interactions, allergy warnings) • Medical diagnostic expert systems • Appointment scheduling • Treatment protocol guidelines • Automated clinical reminders
Machine learning	• Risk prediction models for patient deterioration or hospital readmission • Disease classification from lab results • Clinical trial patient matching
Neural networks	• Medical image classification • Pattern recognition in EKGs and EEGs • Predictive modeling for patient outcomes • Biomarker discovery from genomic data
Deep learning	• Advanced medical imaging interpretation (e.g. tumor detection in CT/MRI scans) • Diabetic retinopathy screening • Drug discovery and molecular analysis
Generative AI (e.g. LLMs)	• Clinical documentation and note summarization • Medical question answering/messaging • Automated coding of billing categories or diagnoses • Patient education or triaging chatbots
Robotics	• Robotic surgery assistance

	• Automated pharmacy dispensing • Hospital logistics and delivery systems

Figure 1: Types of AI models and associated applications in healthcare (Note: these categories are not necessarily mutually exclusive; some applied use cases could be performed by different model types, and multiple models may be employed simultaneously.)

In particular, AI holds a great deal of promise for **personalized or precision medicine**, which refers to tailoring medical treatments to specific patients by leveraging insights from large datasets that may include genetic, environmental, and lifestyle information. The field of genomics—enabled by advances in DNA sequencing that now capture the entire genome—coupled with AI, has opened the door to genomic personalization, which encompasses disease prediction, modeling of drug responses, and even the development of bespoke medications to treat individuals with specific genotypes.[11]

It is worth noting that applications of AI are at varying stages of adoption and validation in the field. Scholars observe that in spite of the growing availability of AI tools for clinical applications, there is a striking lack of robust validation of AI tools in "prospective, real-world clinical settings."[12] Independent, peer-reviewed randomized control trials (RCTs) are the gold standard in clinical validation to ensure efficacy and safety, but today few RCTs are performed on AI tools.[13] In addition, AI applications in healthcare broadly extend well beyond traditional clinical and hospital environments, with potentially profound impacts on patient experiences and health outcomes. AI tools in this domain include wearable technologies and remote monitoring devices, like smartwatches that can send alerts to physicians or emergency services. Patients can self-monitor conditions like glucose levels with AI-augmented devices, and they can also use popular generative AI tools like chatbots to find health information. These AI tools or AI-augmented devices are often marketed direct-to-consumers, and they often fall outside traditional regulatory regimes that focus on devices and systems intended for clinical use.[14]

Challenges and controversies of AI in healthcare

While AI has opened new possibilities for diagnosis, treatment, monitoring, and administration in healthcare, it has also introduced notable challenges, risks, and controversies. Not unlike other technologies integrated into the high-stakes context of healthcare, AI has already resulted in unintended consequences and negative outcomes when deployed in real-world settings. For example, in the case of an AI algorithm widely used by large health systems to optimize referrals for long-term care programs, research found the system tended to exclude Black patients because it used healthcare expenditures as a proxy for health needs. The result was a perpetuation of existing health disparities.[15] In another case, a sepsis prediction model used in hundreds of hospitals across the U.S. demonstrated much poorer performance in real-world patient populations than originally reported by the developer, in part because the model had been trained on data from billing codes, which do not necessarily correspond to accurately identified cases of sepsis.[16] In yet another study of AI algorithms trained on chest X-rays, researchers found that the algorithm consistently underdiagnosed medical issues in historically under-served patient populations, such as women, Black patients, and patients on Medicaid.[17] As these examples illustrate, there is mounting evidence that AI deployed in real-world clinical settings introduces risks that demand robust evaluation of AI tools.

The "black box," transparency, and validation

One of the key issues is the black box nature of many contemporary AI technologies, where it can be difficult for developers, providers, regulators or independent researchers to inspect and understand the processes by which AI systems produce outputs.[18] For example, AI models can engage in "shortcut learning" based on spurious correlations, where the model reads information embedded in data (such as diagnostic images) rather than pathologically relevant information.[19] As Morley and Floridi point out, "just because an AI model can recognize a pattern, does not automatically make the pattern meaningful nor the action it informs clinically efficacious or safe."[20] Without disclosures about how models were trained and the evidence on which models base their outputs, it can be nearly impossible to identify the source of such errors.

Scholarly literature also points to insufficient and inconsistent validation and evaluation of AI models used in healthcare. Most models are trained on retrospective (historical) data and tested experimentally in artificial or laboratory settings, rather than in prospective clinical trials that would provide performance metrics in real-world settings. There are also no standardized frameworks or benchmarks for measuring model performance across different studies.[21] The combination of emerging technologies *and* evaluation techniques in their infancy creates a kind of double hazard in the high-stakes context of healthcare. The mismatch between AI that performs well in controlled, experimental environments and their meaningful application to complex real-world scenarios is sometimes referred to as the "AI chasm," and it can have significant consequences in healthcare.[22]

Data quality

In addition, despite the undeniable increase in machine-readable health data available for analysis by AI models, datasets may be flawed in ways that impede the effectiveness of AI tools. For example, health data may not accurately represent diverse patient populations, reflecting historical exclusions and biases against certain patients—especially marginalized or under-served populations, such as rural or racialized patients.[23] The way that clinicians record symptoms and diagnoses also changes over time as standards, scientific knowledge, and administrative practices evolve, which may not be reflected in datasets or may require careful labeling of data. Because contemporary models perform better when they draw from large datasets, there is also a tendency to use the most readily available high-volume data. This phenomenon is sometimes called the "streetlight effect" or "observational bias," referring to the risk of looking for information only in the most obvious places rather than engaging in costly and time-consuming data collection and cleaning that might be required to better answer clinical questions.

For example, a great deal of data used in computational analytics from clinical settings comes from EHRs because there has been widespread digitization of medical records in recent years. However, data entered into EHRs for administrative purposes may not be well-suited for analysis meant for clinical diagnosis, yet it is not uncommon for information collected in one use case to be ingested by a model and used in generating outputs for another use case,[24] as illustrated by the use of billing information in a sepsis prediction algorithm mentioned above. This is a particular risk arising from the emergence of foundation models, which are trained on diverse datasets and applied to diverse use cases. The real and perceived adaptability of these models may give a false impression that they can make sense of any data by applying principles learned in training to new datasets. But data quality, data labeling, and data relevance still have a profound bearing on the usefulness and accuracy of model outputs. In one study, researchers found that machine learning models trained on descriptive labels (labels that simply describe factual attributes of data) make substantial errors when used to augment human normative judgments (judgements about whether a rule has been violated or not).[25] Many clinical decisions involve normative judgments, and this finding points to the importance of data labeling—and transparency about training data—in clinical decision support systems.

Privacy and security

There are also new challenges with data privacy and security in the context of AI in healthcare. Because AI models rely on large quantities of data, and models are often developed by third party vendors who sell products to healthcare providers, there is largescale data sharing among different parties in the AI pipeline. Individual companies and healthcare organizations develop their own data-sharing procedures and governance processes, and these range from generous data-sharing in exchange for free or reduced-cost AI services to highly restrictive data management practices in which local data held by providers is never exchanged with outside vendors (there is more on governance below in Findings & Discussion). The diversity of data-sharing practices and lack of overarching regulation on data-sharing for AI model development, which would stipulate compliance requirements and penalties for non-compliance, means that potentially huge amounts of sensitive data are more exposed to exploitation and breaches.[26]

For decades, the governing data privacy statute in the U.S. has been HIPAA (Health Insurance Portability and Accountability Act), which set standards for defending the privacy of protected health information (PHI). De-identifying data is common practice for protecting PHI when data-sharing among different parties, meaning that information that would link health data to particular individuals is stripped out of datasets. However, scholarly literature and interviews with subject matter experts emphasize important limitations to the effectiveness of HIPAA in protecting patient information in the context of AI today (more on the limitations of HIPAA below in Policy Landscape and Findings & Discussion).[27] Namely, AI systems are capable of re-identifying individuals even from de-identified datasets; they can also infer characteristics about individuals or groups of patients from data that does not explicitly include information about those characteristics.[28] A great deal of health data may also fall outside the regulatory purview of HIPAA, such as most data that is generated beyond traditional hospital or clinical settings (e.g. data from commercial wearable devices or genetic data collected by private companies) but may be used in AI model training. These kinds of data may fall through regulatory gaps and leave patients exposed to privacy breaches.[29] Most data privacy regulations do not account for the scale of data linking enabled by AI, rendering the distinction between personal data and anonymous data less meaningful in practice.

The big picture

As AI has become more accessible to a wide range of end users, including patients, there are also different risks associated with different use cases. AI may be provider-facing (used by clinicians as a decision-support tool), patient-facing (used by patients to facilitate their care), payor-facing (used by insurance companies to determine eligibility or allocate payments), or administration-facing (used in back-office functions to assist with administrative tasks). In addition to the challenges discussed above, these different users may have vastly different levels of familiarity with AI, so these use cases may introduce different levels or kinds of risk and demand different mitigation strategies. Therefore, when it comes to AI applications in healthcare, it is helpful to consider healthcare as a *system* with interconnected components, such that use cases may be higher or lower risk based on *how* the model generates outputs and *how* those outputs are operationalized rather than *where* (e.g. administration versus clinic) in the system the model is operating.

Humans in the loop

Because many AI applications in healthcare fall under the umbrella of clinical decision support (e.g. predictive analytics) or administrative support (e.g. ambient note-taking), many of the critical issues around AI in healthcare sit at the intersection of machine outputs and human interpretation. The American Medical Association, for instance, prefers the term "augmented intelligence" to artificial intelligence, highlighting the essential and enduring role of humans in decisions aided by AI.[30] In a field characterized by deference to clinical expertise, keeping a "human in the loop," is a common strategy for mitigating AI risks, providing a check on AI that is intended to ensure accuracy, trustworthiness, accountability, and alignment with human values.[31] However, examining AI applications in healthcare also illuminates the many points along the AI lifecycle in which humans influence AI development, deployment, and outcomes. As this section will elaborate, humans enter the loop not only at the oversight stage, but as collaborators in producing the technical architecture of AI and its ultimate interpretation (or meaning) in real-world situations. In this context, AI—which aspires to emulate human cognitive processes—collides head-on with human cognition, spotlighting the strengths and limitations of both. Understanding the inseparable relationship between humans and technology throughout the AI lifecycle is critical to developing effective policy.

Neither is AI wholly equipped to overcome human biases with machine-enabled objectivity, nor human cognition wholly equipped to mitigate AI miscalculations with ethical reasoning. This is because, as Kun-Hsing et al. note, "Despite growing recognition of bias in AI models, particularly with respect to training data, less appreciated are the many additional entry points for human values along the development and deployment journey of an AI model."[32] Throughout the AI development process, human values dictate which health challenges AI is designed to address and how foundational data are gathered, curated, and labeled. Imbalanced datasets also reflect underlying or historical values, leading to biased models that may exacerbate health disparities among marginalized groups. Seemingly technical decisions, such as the choice of model weights, metrics prioritized for optimization, or human feedback-based fine-tuning are also value-laden—these decisions by designers and developers shape the ultimate model. The individuals or organizations that develop technology are a third presence in any user-technology interaction.[33] As a result, AI outputs are not purely technical products; they may have the appearance of computational authority, but they have also been influenced by human decision-making along the way.

Human values and reasoning are also not a panacea for AI risks. A major concern is automation bias, where humans may over-rely on or blindly trust AI outputs, which are often perceived as objective and computationally sound.[34] Automation bias makes it harder to identify AI recommendations that may be biased or wrong because of the human tendency to interpret machine-generated results as reliable. Studies have shown that biased AI recommendations can influence otherwise expert or unbiased human judgment.[35] And human experts tend to retrofit reasonable explanations onto AI outputs, assuming that AI "thinks" the way a human would—which can make model outputs seem believable even when the underlying model is actually wrong or basing its interpretation on irrelevant information.[36] The limitations of human oversight due to well-studied cognitive biases point to the need to take a nuanced approach to "humans in the loop." An oversimplified understanding of how harm may result from AI integration into healthcare might wrongly assume that the problem lies exclusively in the technology itself, when in fact, the positive and negative outcomes of AI in healthcare are the product of complex entanglements between humans (developers, deployers, clinicians, patients) and machines. An awareness of this human-machine interaction can help to avoid what is sometimes called the "MABA-MABA trap" ("Men Are Better At versus Machines Are Better At" trap) in regulation, where simply adding human oversight will "automatically result in the best of both worlds."[37] Instead, mitigation strategies need to tackle the entire system of human-machine interaction, from design and development to deployment and decision-support.

Finally, there is one more aspect of humans in the loop that deserves attention in relation to AI in healthcare, and that is the role of patients—or members of the public—in decisions about how AI is used in their care. Scholarly literature widely acknowledges the role that values play in clinician decision-making. Clinical decisions are not purely questions of diagnosis and treatment; even these determinations are often made under conditions of uncertainty, where clinicians must interpret partial symptoms or medical histories, for instance, and they also consider patient expectations, needs, and values.[38] The integration of AI models into healthcare introduces additional values-based considerations, such as what outcomes an algorithm prioritizes or what data should or should not be included in training.

Based on his study of a U.S. kidney transplant allocation algorithm and a consultative process that involved kidney patients in the development of the algorithm, Daniel Robinson observes that "when it comes to high-stakes algorithms, it's better for political communities to face the hard moral choices together than to abdicate and ignore those choices, abandoning them to the technical experts."[39] Scholarship on AI in healthcare often calls for patient involvement in the development of high-stakes models,[40] and some regulatory regimes in the UK and EU, for instance, are actively exploring ways to incorporate patient participation into policy on AI in healthcare.[41] Patient and public involvement in policy-making around AI in healthcare is considered an avenue for bolstering trust among patients, 60% of whom reported feeling uncomfortable with clinicians relying on AI for critical decisions in a 2023 Pew survey.[42] AI skepticism can have knock-on effects on the quality of data used to train AI models, with downstream impacts on their performance. When patients are hesitant about AI, they may opt out of data collection or AI deployment pilots, resulting not only in their exclusion from potentially health-enhancing care but also less accurate models. Involving patients—especially those most affected by particular AI deployments—in decisions about data collection, consent, model objectives, and more can potentially help to mitigate negative outcomes from patients opting out due to skepticism or fear.

Regulation and innovation

This introduction has provided a broad overview of AI in healthcare, covering applications and use cases, challenges and controversies, and the complex relationship between humans and machines in clinical settings. Alongside the rapid and diverse innovations in healthcare AI, the landscape is also characterized by emerging regulation, guidance, and best practices to ensure AI is developed and deployed responsibly, ethically, and in the public interest. Current policy on AI in healthcare worldwide is a combination of hard and soft law, which has created ambiguities and regulatory gaps (discussed in greater detail in the Policy Landscape section below). Hard law refers to binding legal obligations such as statutes and regulations, while soft law comprises non-binding guidelines, principles, and frameworks. The EU AI Act passed in 2024 represents a leading hard law approach to regulating AI, designating risk categories for different AI use cases, including healthcare. By contrast, Singapore has taken a soft law approach, issuing voluntary guidance in the form of the Model AI Governance Framework. The diversity of approaches can make it difficult to draw meaningful comparisons or find similarities across models and jurisdictions. However, there is broad consensus on core principles such as transparency, robust validation, accountability, and safety, particularly in high-stakes areas like healthcare.

A recent study of global AI regulation offers a cautionary note about the ambiguity of the term "AI regulation," which can mislead public expectations about AI safety (i.e. when guidance is referred to as regulation). The study developed a taxonomy for comparing different regulatory approaches, which allows for comparison and a degree of precision in making sense of the policy landscape.[43] For instance, while the EU adopts a horizontal approach to regulation, which applies broad rules across all sectors, the U.S. tends to maintain a sectoral approach, with regulations managed by dedicated sectoral agencies (like the FDA). In this taxonomy, another key distinction lies between ex ante and ex post regulatory strategies. Ex ante regulation, adopted by the EU, involves assessing AI systems before deployment to ensure safety and compliance, while ex post regulation addresses harms after they occur, more commonly applied in the U.S. Some regulatory approaches focus on the technology (e.g. generative AI) and others focus more on the application or use case (e.g. social scoring).

Effective regulation increasingly relies on consultation and multi-stakeholder engagement, involving developers, healthcare providers, regulators, and patients using mechanisms like regulatory sandboxes to facilitate real-world testing, helping refine rules and foster trust in a rapidly advancing field. Given the range of regulatory approaches and the untested nature of recently adopted regulation both domestically and worldwide, the sectoral focus of this study—on healthcare applications of AI—offers an opportunity to consider in close range the merits of different regulatory strategies for mitigating AI risks and harms, parse the differences between regulation (hard law) and guidance (soft law), and gain insight from a field that has experience balancing the impetus for innovation with the need for regulation.

Scope of the Study

Last year, the General Assembly passed HB2194, "High-risk artificial intelligence; definitions, development, deployment, and use, civil penalties," which was vetoed by the Governor. The bill would have regulated AI used in "consequential decisions," in certain contexts, including in healthcare. Across jurisdictions, healthcare is a sector that is consistently considered high-stakes, high-risk, or consequential, making it a useful case study in examining the policy issues around AI and its societal impacts. In addition, the healthcare sector has the potential to be uniquely informative on these issues because it is also characterized by high levels of technological literacy and familiarity with regulation and compliance. Given these factors, this study aimed to examine the key issues associated with AI in healthcare specifically, to gain an understanding of the current landscape of AI adoption in healthcare in the Commonwealth, and to consider what a sectoral approach to AI policy might look like in Virginia.

In discussions with the advisory group for this study (see Methodology for more information on the advisory group), the scope was narrowed to focus on enterprise applications and adoption of AI, meaning AI used in healthcare and hospital systems for clinical support, mediating clinician-patient interactions. This narrowing of the scope inevitably excluded other important AI applications with a bearing on health, such as direct-to-consumer devices, software, and platforms (e.g. fitness watches or commercial chatbots) and payor-facing systems (e.g. AI used to process insurance claims). These AI applications could be examined in future studies. Early discussions with the advisory group also identified **Electronic Medical Record** systems as a site of growing AI adoption and an important enterprise software for hospital systems in managing clinician-patient interactions. Focusing on enterprise AI tools used in patient care, the study reviewed a range of literature on enterprise applications of AI and adopted an inductive approach to interviews—meaning that interviewees were invited to highlight the applications of AI that were front-of-mind for them at the time of the study. This approach has resulted in a study that provides a general overview of key issues alongside a more in-depth exploration of contemporary issues in Virginia, specifically.

Key Concepts

Artificial Intelligence (AI): A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to (A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action.[44]

Clinical Decision Support (CDS): Technologies that provide clinicians, staff, patients or other individuals with knowledge and person-specific information to enhance health and health care and support decision-making in the clinical workflow. These tools include computerized alerts and reminders to care providers and patients, focused patient data reports and summaries, diagnostic support, and contextually relevant reference information, among other tools.[45]

Deep learning: A subset of machine learning that uses multilayered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain. Many AI systems use some form of deep learning, which uses three or more layers of neural networks to process data. Deep learning models can use unsupervised learning, meaning the models can extract the characteristics, features and relationships from raw, unstructured data.[46]

Deterministic models: Systems where the outcomes are entirely determined by the inputs and the model's rules; there is no randomness involved. Given the same input, a deterministic model will always produce the same output, so their behavior predictable and repeatable.[47]

Electronic Health Record (EHR)/Electronic Medical Record (EMR): An electronic version of a patient's medical history that is maintained by the provider over time, including data such as demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports.[48]

Foundation model: A large, pre-trained machine learning model that can perform a wide variety of tasks. They are "foundational" because they serve as the base for many different applications across various domains, from NLP and computer vision to robotics. They are trained on immense datasets and engage in transfer learning, a process through which a model trained on one dataset or task is used to improve performance on another dataset or task.[49]

Generative AI: Systems that rely on sophisticated machine learning models trained to identify patterns in huge amounts of data without expert supervision. They can create original content such as text, images, video, audio or software code in response to a user's prompt or request.[50]

Large-language model (LLM): A type of AI that uses deep learning techniques to process and generate human-like text based on vast amounts of data that include books, websites, and other text sources, enabling them to learn patterns, structures, and nuances of language.[51]

Machine Learning (ML): A subfield of AI that gives computer systems the ability to learn without being explicitly programmed. Because ML underpins so many AI systems, the terms are often used interchangeably. ML applications include predictive text, recommendation algorithms, medical image interpretation, and more.[52]

Natural Language Processing (NLP): A field of AI that uses computational linguistics, machine learning, and deep learning to enable computers to understand, interpret, generate, and respond to human language in a way that is both meaningful and contextually appropriate.[53]

Neural network: A computational model inspired by how the human brain processes information, consisting of layers of interconnected nodes, or "neurons," each of which performs a simple mathematical operation. By adjusting the connections (weights) between neurons through a process called training, the model "learns" and becomes more accurate.[54]

Precision/personalized medicine: An innovative medical approach that uses information about individuals' genes, environment, and lifestyle to guide healthcare decisions by predicting which prevention strategies and treatments will work with which groups of people.[55]

Predictive models: A statistical technique used to predict the outcome of future events based on historical data. Predictive models may be deterministic or probabilistic.[56]

Probabilistic models: Systems that incorporate uncertainty and randomness in their predictions or decisions. These models are trained on large, diverse, and often noisy datasets. Instead of producing a single fixed output for a given input, they estimate the likelihood of different outcomes and may produce unpredictable outputs.[57]

Robotics: Computational or physical representations of robots that integrate AI techniques to perceive, reason, and act within an environment. These models are designed to simulate or control intelligent behavior in machines, enabling them to perform tasks such as navigation, object manipulation, decision-making, and interaction with humans or other systems.[58]

Rule-based models: These models operate on a set of predefined rules and logic to generate outputs. They are characterized by programming that uses IF/THEN logic, and the parameters are set by experts with knowledge of the field.[59]

Software as a Medical Device (SaMD): Software intended for one or more medical purposes that performs that function without being part of a hardware medical device. This includes standalone applications that run on general-purpose devices like computers or smartphones to diagnose conditions, interpret medical images, or monitor patient data to help treat or manage a disease.[60]

Software in a Medical Device (SiMD): Software that is an integral component of a hardware medical device. Unlike SaMD, SiMD can't function independently, and rather is reliant on the associated medical hardware. Since the software contributes to the device's intended purpose, any malfunction or failure could impact the safety and performance of the entire medical device.[61]

Structured data: Data that is organized in a clear, predefined format. The standardized nature of structured data makes it easily decipherable by data analytics tools, machine learning algorithms and human users.[62]

Supervised learning: This method of AI training requires human input to guide its operations. In supervised learning, a human gives the computer algorithm a dataset that includes labels as the training dataset, and the algorithm is being trained to associate labels with specific pieces of data.[63]

Training data: Information used to train an AI model to make predictions, recognize patterns, or generate content. Training datasets are often very large and can contain multiple formats, such as text, images, and video. Because models rely on huge amounts of training data to improve performance, many commonly used training datasets for foundation models are scraped from the internet. [64]

Unstructured data: Data without a predefined format. Unstructured datasets are typically large (think terabytes or petabytes of data) and comprise 90% of all enterprise-generated data. This high volume is due to the emergence of big data—the massive, complex datasets from the internet and other connected technologies. Unstructured data can contain both textual and nontextual data and both qualitative (social media comments) and quantitative (figures embedded in text) data.[65]

Unsupervised learning: This method of AI training does not involve assigning labels in datasets. Instead, the algorithm is given a dataset and it examines all the data to find common features and gives them more or less weight in the model. The weights can be adjusted by (human) developers or refined using additional layers of machine processing, such as interconnected neural networks.[66]

Policy Landscape

In 2025, the policy landscape of AI in healthcare remains piecemeal, with some regulation at the federal level administered by different federal agencies, alongside an array of state-level initiatives that typically target specific types of technology or particular use cases that fall within the umbrella of healthcare but rarely address the widespread use of AI in the sector more generally. Although there have been concerted efforts at the federal level to develop regulation and guidance, a 2024 Congressional Research Service report acknowledges that these efforts are "in early stages and are somewhat fragmented." The same report concludes that "many stakeholders maintain there is a need for regulatory guardrails to address potential challenges with trust, data access, bias, lack of transparency, and privacy" and a "need for harmonization amongst party efforts."[67] This section provides an overview of federal policy and state legislation pertaining to AI in healthcare, along with several non-binding proposed frameworks and guidance commonly referenced in the health sector.

Federal Policy

The White House

In July 2025, the Executive Office of the President issued *Winning the Race: America's AI Action Plan*, which positions AI as essential to American success and competitiveness and generally articulates a deregulatory posture to encourage AI development. The health sector is mentioned specifically in a section on bottlenecks to AI adoption: "Many of America's most critical sectors, such as healthcare, are especially slow to adopt due to a variety of factors, including distrust or lack of understanding of the technology, a complex regulatory landscape, and a lack of clear governance and risk mitigation standards. A coordinated Federal effort would be beneficial in establishing a dynamic, 'try-first' culture for AI across American industry."[68]

The Action Plan encourages the development of regulatory sandboxes or AI Centers of Excellence around the country enabled by the Food and Drug Administration (FDA), Securities and Exchange Commission (SEC), and Department of Commerce (DOC) to enable researchers, startups, and other industry players to test AI tools and openly share results. It mentions launching several "domain specific efforts," including in healthcare, to develop national standards for AI systems. In addition, the plan places a strong emphasis on open-source models, public datasets, and transparency, as mechanisms for jump-starting U.S.-led innovation and promoting global adoption of home-grown AI. It also calls for "rigorous evaluations" as a critical tool for assessing the performance and reliability of AI models.

While the Action Plan sets out a range of federal initiatives, it cautions against states passing overly restrictive legislation on AI, encouraging federal agencies to "consider a state's regulatory climate when making funding decisions" and tasking the FCC with evaluating "whether state AI regulations interfere with the agency's ability to carry out its obligations."

Assistant Secretary for Technology Policy/Office of the National Coordinator for Health Information Technology (ASTP/ONC)

ASTP/ONC is responsible for the development and use of health information technology (HIT), and it oversees the Health IT Certification Program, through which HIT developers may voluntarily obtain certification for technologies, like EHRs. In January 2024, ASTP/ONC published a final rule, "Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing," known as HTI-1.[69]

In 2023, 28 healthcare companies signed on to a commitment to fair, appropriate, valid, effective, and safe AI principles (FAVES AI principles) during the Biden-Harris Administration,[70] and HTI-1 also references the FAVES AI principles. HTI-1 requires certified HIT modules to implement risk management procedures, disclose particular source attributes, performance, and quality information to users and to keep this information updated.

Food and Drug Administration (FDA)

The FDA has responsibility for oversight of the safety and effectiveness of medical devices, including software in certain cases. Regulated AI/ML-enabled devices must meet the definition of "device" in the Federal Food, Drug, and Cometic Act (FFDCA), and the 2016 21st Century Cures Act provides further clarity on what kind of software meet the definition of a device. For example, it excludes software intended for the administrative support of a health facility, for the encouragement of a healthy lifestyle unrelated to disease diagnosis, cure, or treatment, or for electronic patient records. The Congressional Research Service observes, "Determining whether a software function meets the definition of a device, and is thus regulated by the FDA, can be a challenge for the developer."[71]

Nonetheless, the FDA has expanded its regulation of software in recent years to include **software as a medical device (SaMD)** and **software in a traditional hardware device (SiMD)**. Software that falls into these "device" categories are subject to requirements, such as adverse event reporting, establishment registration and device listing, and more. In 2019, the FDA published a proposed regulatory framework for AI/ML-enabled SaMD,[72] and in 2021, it issued an action plan for AI/ML-enabled SaMD.[73] In 2025, the FDA published guidance on "Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions," which outlines a "total product lifecycle approach" to developing AI/ML-enabled medical devices, with recommendations to provide documentation for market submissions that include a product description, model description, validation, and performance monitoring plans.[74] The FDA also maintains a list of authorized AI-enabled devices, which this year listed over 1200 devices.[75]

Centers for Medicare & Medicaid Services (CMS)
In 2023, CMS issued a final rule that established guardrails for "utilization management tools" employed in making coverage decisions for Medicare Advantage plans. Because Medicare Advantage plans give providers a fixed, predetermined payment for each patient assigned to them, these plans use mechanisms like prior authorization to ensure that a patient is using services that are "reasonable and necessary." The rule stipulates that if a Medicare Advantage plan is going to issue an adverse determination about what is reasonable and necessary for a patient, the decision must be reviewed by a physician. Although the rule does not mention AI explicitly, in February 2024, CMS published responses to frequently asked questions (FAQs) regarding the use of algorithms and artificial intelligence in coverage determinations, which explained that Medicare Advantage plans can use algorithms and software tools, including AI, to assist in decisions as long as all regulatory requirements are met, which would include the requirements in the final rule.[76] However, in June 2025, CMS announced a new model to improve and expedite prior authorizations, labeled the "Wasteful and Inappropriate Service Reduction (WISeR) Model."[77] To date, it is unclear how federal adoption of this model will influence policy on AI use in prior authorizations.

Office for Civil Rights
In May 2024, OCR issued a final rule clarifying that section 1557 of the Patient Protection and Affordable Care Act prohibits covered entities from discriminating through the use of AI in patient care decisions.[78] Although aspects of the final rule have been suspended through Executive Orders by the Trump Administration, to date, it does not appear that the aspects of the final rule pertaining to AI have been suspended or rescinded.

Virginia Policy

Virginia currently has no comprehensive state-wide policy on AI in healthcare.

HB2154 (Passed, 2021) requires hospitals, nursing homes, and certified nursing facilities to develop and implement policies on staff access to and use of "intelligent personal assistants," which encompass software that uses NLP and AI.

HB2094 "High-risk artificial intelligence; definitions, development, deployment, and use, civil penalties." (Chief Patron: Delegate Michelle Lopes Maldonado) was introduced and passed by the House and Senate in the 2025 Session of the General Assembly but was vetoed by the Governor.

The bill defined a "high-risk artificial intelligence system" as "any artificial intelligence system that is specifically intended to autonomously make, or be a substantial factor in making, a consequential decision," which would encompass decisions regarding "access to health care services." Developers of high-risk AI systems would be required to provide comprehensive documentation to deployers, such as mitigation strategies for algorithmic discrimination and evaluation measures. Deployers would need to have a risk management policy in place, complete impact assessments, and disclose to consumers when AI is used. Notably, the bill provided an exemption for "developers or deployers facilitating telehealth services or covered entities under HIPAA that provide AI-generated healthcare recommendations requiring a healthcare provider to implement, or utilize AI systems for administrative, quality measurement, security, or internal cost/performance improvement functions."

In his veto justification, the Governor expressed concern that HB2094 would be overly burdensome on the AI industry, particularly smaller AI companies and startups, and would therefore undermine AI innovation in the Commonwealth.

Legislation in Other States

Several states have passed legislation that applies to clinical or healthcare settings,[79] though some legislation is broadly cross-sector:

Arkansas
HB1958 (Passed, 2025): This bill requires public entities in Arkansas to create a comprehensive policy regarding the authorized use of AI and automated decision tools. The policy must mandate that a human always makes the final decision, regardless of AI or automated tool recommendations, and public entities are required to develop a training program for employees on the appropriate use of AI.

California
AB3030 (Passed, 2024): This bill requires healthcare providers to have a disclaimer for patients when they received written, audio, and video communications of clinical information that was produced by generative AI. Patients must also be provided with information about how to contact a human healthcare provider.

AB2013 (Passed, 2024): This bill requires developers of generative AI to disclose information about training data to the public, including the source of datasets, whether datasets contain personal information, and more.

AB2156 (Passed, 2006): This bill requires a clinical laboratory director or authorized designee to establish, verify, and document criteria by which any laboratory test or examination result is auto-verified by a computer algorithm. These records must be re-evaluated annually.

Colorado
SB205 (Passed, 2024): This bill defines high-risk AI systems as those that make "consequential decisions," including about the costs or terms of health care services or insurance. Developers of AI systems are required to meet standards that mitigate discrimination, provide transparency to deployers and the public, and manage risks.

Kentucky

HB191 (Passed, 2018): This bill stipulates requirements for the use of assessment mechanisms, including AI, to conduct eye exams or generate contact lens prescriptions. It requires methods for interaction between patients and licensed healthcare providers, and information gathered by the assessment mechanism must be interpreted by a licensed human specialist.

Montana

HB178 (Passed, 2025): This bill prohibits the use of AI by government entities to use AI in ways that result in unlawful discrimination by classifying a person or group based on behavior, socio-economic status, or personal characteristics.

Nevada

AB406 (Passed, 2025): The bill places restrictions on the use of AI in mental and behavioral healthcare. Specifically, it prohibits public schools from using AI to perform the functions of school counselors, psychologists, or social workers and requires the Department of Education to develop a policy for AI use by school employees in providing mental health services. The bill also prohibits AI providers from offering systems designed to deliver professional mental or behavioral health care or from representing themselves as qualified to provide such care. Licensed mental and behavioral health care providers are barred from using AI systems for direct patient care. The bill does allow providers to use AI tools to support administrative tasks, with some requirements, such as having the provider review any information generated by an AI system.

New Mexico

HB178 (Passed, 2025): The bill requires the Board of Nursing to set rules for standards on the use of AI in nursing.

Oklahoma

HB2266 (Passed, 2012): The bill allows physician-approved protocols to utilize or reference "medical algorithms" in the provision of public health services.

Oregon

HB2748 (Passed, 2025): The bill prohibits any nonhuman entity, including agents powered by artificial intelligence, from using a specified list of nursing titles and their abbreviations, including but are not limited to Advanced Practice Registered Nurse (APRN), Registered Nurse (RN), Nurse Practitioner (NP), and Certified Nursing Assistant (CNA).

Rhode Island

HB6654 (Passed, 2022): The bill sets out requirements on the use of assessment mechanisms, which include AI devices), in conducting eye exams or generating prescriptions for contact lenses. The bill requires that such assessment mechanisms provide for interaction between the patient and licensed care providers and that providers review and interpret outputs from the assessment mechanisms.

Texas

SB1188 (Passed, 2025): Among other stipulations for Electronic Health Records, the bill permits healthcare practitioners to use artificial intelligence for diagnostic purposes provided they disclose this use to patients and there is clinician review of AI-generated records.

HB149 (Passed, 2025): This bill is known as the "Texas Responsible Artificial Intelligence Governance Act" and introduces broad regulation for AI, including prohibitions on inciting self-harm, discrimination, and social scoring. Additionally, the bill requires disclosures to consumers when interacting with AI, establishes a regulatory sandbox, and establishes an AI council to develop policy and oversight.

Utah

SB149 (Passed, 2024): This bill, known as the "AI Policy Act," requires disclosures between an AI developer and end user. It stipulates that "regulated occupations," which include many clinical specialists, disclose when they are using computer-driven responses before they begin using generative AI in messaging with patients. SB226 (Passed, 2025) repealed SB149 and replaced the disclosure requirements with a narrower definition, whereby disclosures are only required when generative AI is considered "high-risk," such as collecting personal information or making personalized recommendations. SB332 (Passed 2025) extended the repeal date of SB149 to July 1, 2027.

Non-Binding Guidance

In the absence of comprehensive federal or state regulation on AI, developers and deployers operating in the health sector often turn to non-binding guidance, frameworks, and best practices developed by professional associations, hospital systems, or academics. These non-binding guidance documents often inform institutions' development their own governance rules and procedures. Guidance documents on AI in healthcare emphasize several shared themes, including transparency and explainability, safety and risk mitigation, privacy and data protect, fairness and bias mitigation, accountability, meaningful stakeholder engagement, and continuous monitoring and updating. The following is a non-exhaustive list of some of the frequently cited frameworks in the health sector:

NIST AI Risk Management Framework[80]

The National Institute of Standards and Technology (NIST) is a government laboratory within the U.S. Department of Commerce tasked with developing standards and best practices in technology. The 2020 National Artificial Intelligence Initiative Act directed NIST to develop a voluntary, rights-preserving, non-sector-specific, and use-case agnostic resource in collaboration with private and public sectors to promote responsible AI. The target audience for the Risk Management Framework is anyone designing, developing, deploying, or using AI systems, and it outlines seven characteristics of "trustworthy AI": (1) valid and reliable; (2) safe; (3) secure and resilient; (4) accountable and transparent; (5) explainable and interpretable; (6) privacy-enhanced; (7) fair. The framework advocates for governance processes that operationalize these characteristics throughout the lifecycle of an AI product. In the 2025 *AI Action Plan*, the White House has directed NIST to eliminate references in the framework to misinformation, Diversity, Equity, and Inclusion (DEI), and climate change.

CHAI Responsible AI Guide[81]

The Coalition for Health AI is a healthcare community organization comprised of stakeholders such as patient advocates, clinicians, data scientists, and technologists. The CHAI guidance addresses various types of AI used in healthcare, including generative AI, as well as a variety of health AI use cases, including decision support, diagnosis, treatment planning, medical imaging analysis, patient monitoring, and administrative tasks. The document is built upon five core principles for trustworthy health AI: (1) usefulness, usability, and efficacy; (2) fairness; (3) safety and reliability; (4) transparency, intelligibility, and accountability; and (5) security and privacy.

National Academy of Medicine AI Code of Conduct[82]

The National Academy of Medicine (NAM) is one of the three National Academies of Sciences, Engineering, and Medicine. It is a private, nonprofit institution that works outside government to provide objective advice on matters of science, technology, and health. The AI Code of Conduct, published in 2025, targets a broad audience, including healthcare organizations and all related stakeholders such as technology developers, researchers, health systems, payors, patients, clinicians, and more. Like other guidance documents, it encompasses a wide range of healthcare use cases for AI, including but not limited to medical research, illness diagnosis, personalized treatment plans, patient care summaries, claims processing, insurance denial appeals, diagnostics, administrative efficiency, risk prediction, medical imaging analysis, automated clinical documentation, and chatbots. There are ten Code Principles: Engaged, Safe, Effective, Equitable, Efficient, Accessible, Transparent, Accountable, Secure, and Adaptive. And six associated Code Commitments: Advance Humanity, Ensure Equity, Engage Impacted Individuals, Improve Workforce Well-Being, Monitor Performance, and Innovate and Learn.

American Medical Association AI Principles[83]

The American Medical Association is a professional organization mainly representing clinicians and medical students. In 2024, it adopted a set of "augmented intelligence" (AI) principles for AI used in a range of healthcare contexts, including AI-enabled medical devices, clinical decision support, administrative applications, and generative AI, as well as automated decision-making systems utilized by payors for claims and coverage. The guidance emphasizes that AI must be designed, developed, and deployed in a manner that is ethical, equitable, responsible, accurate, transparent, and evidence-based. This includes promoting human oversight in clinical decisions, mitigating bias, ensuring data privacy, fostering transparency, establishing clear liability frameworks, and actively combating AI misinformation.

Health Canada, U.S. FDA & UK MHRA Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles[84]

This guidance was jointly produced by the U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) in 2021. It is targeted at healthcare professionals, patients, caregivers, support staff, administrators, payors, and governing bodies and focuses specifically on Machine Learning-Enabled Medical Devices (MLMDs), which are devices that can learn and improve from real-world use. The focus of the guidance is on transparency, and it outlines several key considerations for effective disclosures about MLMDs: (1) who needs the information; (2) why transparency is motivated (e.g., patient-centered care, safety, risk management, health equity, trust); (3) what relevant information should be provided (e.g., performance, risks, limitations, model logic); (4) where information is placed (e.g., optimized user interface); (5) when it is communicated (e.g., across the product lifecycle); and (6) how methods like human-centered design support its delivery.

IMDRF Good Machine Learning Practice[85]

In 2025, the International Medical Device Regulators Forum (IMDRF)'s Artificial Intelligence/Machine Learning-enabled (AIML) Working Group put out this guidance calling for regulators to advance "good machine learning practice." It outlines several principles for developers of AI-enabled medical devices to adhere to: (1) the intended use is well-understood and multi-disciplinary expertise is leveraged throughout the product lifecycle; (2) good design principles are incorporated throughout the product life cycle, including security practices; (3) clinical evaluation should include the use of representative datasets for intended populations; (4) training datasets should be independent from test datasets; (5) selected reference standards are fit-for-purpose; (6) model choice and design are tailored to the intended use of the device; (7) assessments of device performance focus on human-AI interactions in the intended use environment, including the performance of the human-AI team, rather than just the device in isolation; (8) testing evaluates performance in clinically relevant conditions; (9) users should be provided with clear, essential information; and (10) models should be evaluated on an ongoing basis after deployment.

FAIR-AI Framework[86]

The FAIR-AI Framework was developed by a multidisciplinary consortium of researchers and medical institutions and published in 2025. The framework is designed for health systems to enable pre-implementation evaluation and post-implementation monitoring of AI solutions. Like several other frameworks cited in this section, it also encompasses a wide range of AI use cases, from administrative functions to diagnostics, clinical decision support, and billing. The framework focuses on transparency, validation, usefulness, and equity, and it suggests a screening process that would assign AI applications to different risk categories (low, moderate, high). Principles emphasize keeping humans in the loop, ongoing monitoring of models, and end-user transparency using mechanisms like labels.

Stanford University FURM Framework[87]

This framework was published by Stanford Health Care in 2024. It was drafted by a multidisciplinary team, and the guidance is currently in use in the Standford Health Care system. The goal of the framework is to offer a testing and evaluation mechanism for health systems to identify fair, useful, and reliable AI models (FURM). It outlines a three-stage assessment: (1) what and why, which evaluates potential usefulness through simulation, financial impact and sustainability projections, and ethical considerations; (2) how, which evaluates technical and organizational feasibility for implementation into workflows; and (3) impact, which refers to designing plans for prospective evaluation and ongoing monitoring.

Regulatory Gaps

The use of AI in healthcare faces several significant regulatory gaps, largely stemming from the technology's rapid evolution and adoption outstripping existing legal and oversight frameworks. The gaps broadly fall into four categories: fragmented federal oversight by regulatory bodies, insufficient data sharing and privacy standards, a lack of transparency standards for AI tools, and ambiguity about liability and appropriate consenting processes.

Fragmented Oversight

Although several federal agencies have begun setting out rules and guidance on AI tools, with the FDA arguably taking the lead in extending its existing purview over medical devices to AI-enabled medical devices, there remains a lack of a cohesive regulatory framework for dealing with the myriad applications of AI in the health sector. In the case of the FDA, it remains unclear which AI tools the FDA will ultimately be responsible for regulating, and evidence to date suggests that only certain AI tools will fall under FDA oversight. In recent years, the FDA has approved more than 1000 AI-enabled medical devices in domains such as radiology (the largest number of approvals), cardiovascular technology, neurology, anesthesiology, and more. However, AI tools used in clinical settings increasingly perform administrative and diagnostic tasks that draw from diverse, unstructured data, such as scanned documents, voice recordings, medical records, diagnostic imaging, and more. Therefore, AI fits less neatly into existing or pre-defined categories, particularly as foundation models are applied and refined for different problems, local deployments, and use cases. [88]

The onus often falls on developers and deployers to figure out whether their AI technologies are subject to existing or emerging regulations,[89] and they understandably default to relying on guidance from the regulatory bodies most applicable or familiar to their industry (e.g. EHR companies look to guidance from ASPT/ONC). This creates opportunities for AI tools to fall through the cracks, so to speak, between existing regulatory frameworks because (1) AI developers and deployers newly entering the healthcare space may be unfamiliar with specific industry bodies and regulations, (2) developers and deployers who have long occupied the healthcare space might not consider how their use of AI falls within the purview of *other* regulatory bodies or frameworks, and (3) the lines between developer and deployer can get blurred, as many AI tools involve multi-layered third-party dependencies (e.g. a large-language model developed by one entity, integrated into software managed by another entity, which is fine-tuned by an end-user entity to serve localized use cases).

Moreover, many existing exemptions found in federal regulation and state legislation create regulatory gaps that many medical AI tools can slip through. For example, clinical decision-support tools are often left out of regulatory oversight because the reliance on a human clinician to provide final validation of AI-generated or -augmented outputs seemingly places the ultimate decision under the familiar purview of an existing regime—the expertise of a licensed professional.[90] However, clinical decision-support tools already encompass a wide array of technologies, from rule-based, deterministic models to generative models. Few frameworks directly address the complex interaction between humans and machines in clinical decision-making, or provide standards for mitigating cognitive biases through, for example, AI literacy requirements. Additionally, many AI tools used in administrative functions are exempted from existing regulation because they do not neatly fit the definition of a medical device.[91] However, AI used in medical charting, triaging, messaging, and more can have consequential impacts on patient experience and care and may draw from a range of data (such as diagnostic tests, scans, or laboratory results) that go beyond what might be considered strictly administrative.[92]

Insufficient Data Sharing and Privacy Standards

Since 1996, the Health Insurance Portability and Accountability Act (HIPAA) has set the standards for the management of electronic health data among covered entities, with an emphasis on preserving patient privacy, protecting data from unauthorized access, alteration, or destruction, and facilitating the transfer of information among providers and payors. Because HIPAA governs data privacy in the health sector, it is often invoked as a catch-all standard and indicator of compliance with regulation on patient data, and regulation often includes carve-outs for HIPAA-covered entities because of the presumed burden of compliance with existing data protection rules. HIPAA deals specifically with protected health information (PHI), which is individually identifiable information. It does not cover de-identified data that has been stripped of individually identifiable information. Such de-identified data can be transferred and processed between different data stewards without HIPAA restrictions and, in the case of AI, is often requested or required for model training. However, HIPAA definitions may not be sufficient for protecting patient data in the context of AI.

First, many types of health-relevant data fall outside the protections of HIPAA, including data from device and software developers, governmental public health agencies, and research institutions not affiliated with HIPAA-covered entities. Health data that originates outside traditional hospitals and clinics (e.g. personal device data from fitness trackers or other wearables) that is increasingly integrated into formal healthcare settings or used to train AI models are not subject to HIPAA. Second, although using de-identified data for AI model training and fine-tuning can reduce some privacy risks, and it can also diminish the accuracy of models that will ultimately be used to make individualized recommendations, remove information that could be useful in auditing datasets for bias. In addition, AI models can re-identify individuals, even from datasets that have been stripped of PHI, pointing to significant limitations in existing data privacy standards that only stipulate de-identification of data. [93]

Lack of Transparency Standards

The development and increasing adoption of probabilistic AI models has given rise to the "black box" problem—where developers and deployers cannot or choose not to fully explain the nature of AI tools. They may not be able to fully explain how AI tools operate because models continuously learn and adapt over time in an iterative, layered process that does not always involve human supervision.[94] Or, developers and deployers may choose not to fully explain how AI tools operate to protect trade secrets and preserve the market competitiveness of their products. In health care, the AI "black box" is a significant barrier to assessing the safety and efficacy of AI tools, and regulatory bodies are still grappling with what kinds of disclosures AI developers and deployers need to make about the inner workings of the models they use in clinical settings.[95] In the absence of overarching disclosure requirements about things like training data, data labeling, model weights, model performance, and model evaluation, developers potentially face a patchwork of different requirements in different jurisdictions and for different technologies or use cases.

In addition, many studies on AI tools for patient care are conducted internally by developers or deployers themselves or published online without peer review or independent oversight, further limiting transparency for evaluation.[96] Without a clear mechanism for independent auditing and validation, as exist in many other regulated sectors, required disclosures may wind up being relatively meaningless.[97]

Ambiguity about Liability and Consent

At present, there is a great deal of uncertainty around liability about AI used in healthcare. This uncertainty is partly due to the novelty of the technology—there is little existing case law regarding injury caused by AI, and scholars point out that traditional legal frameworks for dealing with medical liability, such as malpractice and tort doctrines, may not be adequate for dealing with the black box nature of many AI systems,[98] where it can be difficult if not impossible to reverse engineer an AI output to determine how an error occurred. There is also a distinct challenge in the number of different parties involved in developing and deploying AI tools, and the lines between developer, deployer, and end-user can become blurred when models are modified or fine-tuned for specific clinical settings or patient populations.[99] As models learn and adapt to new data, the model itself also changes (what is sometimes called "model drift"), and this further complicates liability assessments.[100]

Ambiguity about liability in relation to AI used in healthcare has likely led to a cautionary approach to adopting AI among many healthcare professionals and could impede wider acceptance of transformational technologies.[101]

In the absence of clear, harmonized regulation on AI, responsibility seems to fall disproportionately on clinicians who often provide a final human review of AI outputs.[102] Some existing regulations stipulate clinician oversight as a safeguard against harmful or misleading AI outputs. But on the other hand, there are also concerns that physician deference to AI outputs may become an industry standard, whereby clinicians may be liable if they *deviate* from AI advice just as much as they would be liable if they blindly followed AI advice without using their expert judgment.[103]

There are also important lingering ambiguities around patient consent for the use of AI in healthcare. Patients are often unaware when AI systems are being used in their care, whether for diagnostics or administrative functions.[104] At the same time, it might be impractical to require consent for every use of AI, just as it can be impractical to obtain consent for the use of other technological tools or medical devices in a clinical setting (e.g. during an emergency procedure). Patients are often asked to sign consent agreements for the processing of their data within HIPAA rules, but as discussed above, HIPAA standards are inadequate for protecting patient data in the face of AI analytics. It may also be difficult to obtain informed consent from patients due to the complexities of data use and processing with AI and a general lack of AI and data literacy.[105] Ultimately, the absence of clear consenting standards may not only complicate the adoption of AI by health systems but also erode patient trust, leading to AI skepticism and hesitancy.[106]

Study Methodology

This study is a JCOTS Limited Study, which comprises a literature review in combination with subject matter expert (SME) interviews. This limited study was compiled between May and September 2025, and interviews with SMEs primarily took place over a 6-week period between July and August. The overarching goal of the study was to provide a broad overview of the current state and future horizon of AI in personalized and predictive healthcare and to identify policy-relevant issues that might warrant the attention of lawmakers. The study set out to explore several research questions:

(1) How is AI currently being integrated into healthcare, and how are different types or use cases of AI understood in the health sector?
(2) What is the existing policy, governance, and guidance landscape around AI in healthcare?
(3) Specifically, how is AI being used in health systems in Virginia, and what are the most pressing opportunities and challenges?
(4) How can policy best mitigate risks and harms of AI in healthcare while preserving space for innovation and advancing adoption of transformative technologies?

Literature Review

The literature review constitutes a critical analysis of existing scholarly (academic), grey (non-peer-reviewed reports and independent publications), and journalistic (news) publications compiled through web searching and scholarly database searching, e.g. EBSCO, Google Scholar. The literature review provides a synthesis and summary of literature on the uses, opportunities, and controversies of AI in the field of healthcare. The aim of the literature review was to gain a general understanding of AI in healthcare, identify key themes that deserve policy attention, and inform the themes of the semi-structured interviews with SMEs. BillTrack50 was used to assist with searching and compiling legislation from other states.

Interviews

Over six weeks between July and August 2025, 41 subject matter experts were interviewed resulting in 18 hours of interview recordings and transcripts. Interviews were semi-structured, meaning that they all followed a general list of topics or themes, but specific questions varied based on the SME's particular area of expertise and/or the emergent content of the interview itself. Recordings were only retained until transcripts were reviewed by a researcher, and then they were deleted. Interviewees completed a consent form to participate in the study and were able to indicate their preference for attribution. Interviews were mostly conducted via video call on Google Meet, and transcripts were initially generated by Gemini AI (the built-in AI assistant for Google Meet) and subsequently checked by a researcher. An additional three SMEs were consulted on background for the study prior to the interview period.

Interview transcripts were thematically coded[107] with the support of NVivo, a qualitative data analysis (QDA) software. In thematic coding, researchers begin by reviewing transcripts and familiarizing themselves with the content, assigning a label to meaningful content based on what interviewees said. As they work through different transcripts, researchers compare new content with existing labels (codes) and sometimes add additional codes if a suitable code does not yet exist. Throughout this process, researchers look for themes among codes and subsequently group codes according to emerging themes and sub-themes. The goal of this process is to identify patterns and differences among the accounts of different interviews. This approach is inductive rather than deductive, meaning that rather than starting with a hypothesis that is tested through data collection (deductive), the researcher is led by the data to develop findings or theories (inductive). The themes that emerged from SME interviews are discussed in the Findings and Discussion section of the report.

Use of AI

Several AI tools were used while conducting research for this study. Claude (Anthropic) and CoPilot (Microsoft) were used to assist in web searching and identifying references for inclusion in the literature review. NotebookLM (Google) was used to assist in searching within a defined collection of researcher-curated articles, reports, and news stories. AI tools were not used for report outlining, writing, or editing.

Advisory Group

JCOTS convened an academic advisory group to support the study by providing additional expert insight and reviewing project objectives and deliverables. The advisory group met online three times over the course of the study and reviewed documents asynchronously. All members of the advisory group are also making presentations to the Commission on various aspects of the study topic between October and December 2025. The Advisory Group members are Dr. Sylvester Johnson, Professor of Black Studies at Northwestern University, CEO of the Corporation for Public Interest Technology, and former Associate Vice Provost for Public Interest Technology at Virginia Tech; Dr. Sandra Soo-Jin Lee, Professor of Medical Humanities & Ethics and Chair of the Division of Ethics at Columbia University; and Dr. Sarah Henrickson Parker, Associate Professor in the Virginia Tech Carilion School of Medicine (VTCSOM) and Fralin Biomedical Research Institute, and Chair of Health Systems and Implementation Science at the VTCSOM.

Limitations

There are several limitations to this study that are worth keeping in mind in interpreting the findings. First, by focusing on hospital systems and enterprise AI applications, the study inevitably excludes other players in the healthcare sector and other applications of AI that have a bearing on patient care and outcomes. Secondly, interviews were conducted in a defined window of time and relied on purposive sampling by researchers, a non-random approach where interviewees are selected based on predetermined criteria and presumed likelihood of possessing expertise in the subject matter, and snowball sampling, where interviewees recommend other interviewees in their network. This approach wound up excluded some stakeholders that have valuable insights on AI policy in healthcare, such as smaller, independent clinical practices, AI startups and other model developers that work directly with enterprise technology developers and deployers to integrate foundation models into other software, companies developing direct-to-consumer healthcare technologies, and patients. As a result, these insights should not be considered representative of the health sector as a whole, but rather illustrative of some of the key issues in a complex sector that deserve further exploration with broad stakeholder engagement and consultation as policies are developed.

Nonetheless, interviews did include SMEs from many of the largest healthcare systems in Virginia as well as three major enterprise software companies that serve hospital systems in the Commonwealth and nationwide. Their insights supplement and support well-established scholarly and grey literature on AI in healthcare to provide a robust introduction and overview of the topic.

Findings and Discussion

This section of the report provides a synthesis of the major themes that emerged from interviews with subject matter experts, who included hospital system representatives (e.g. Chief Medical Officers, Chief Information Officers, clinicians), academics working in and on healthcare applications of AI, members of independent nonprofit organizations researching or advising on healthcare technology both in the U.S. and abroad, representatives of professional associations in the health sector, and representatives of technology companies (e.g. government affairs, software developers, and regulatory experts). Because of the study's focus on enterprise adoption of AI solutions, interviews targeted hospital systems and other stakeholders that regularly interact directly with hospital systems. Their reflections provide a helpful general overview of the current landscape for AI in healthcare. Based on these interviews, six top-level themes emerged, which are explored in greater detail below:

1. AI is not new in the healthcare sector, but the newest AI is front-of mind;
2. The dominant business case for AI adoption is clinician burnout;
3. Good AI governance is a top priority for healthcare technologists and providers, but it is uneven, institution-specific, and voluntary;
4. Multiple and messy external vendor dependencies characterize AI adoption;
5. Humans are staying in the loop to provide oversight of AI models;
6. Developers, deployers, and end-users want regulatory clarity and harmonization.

AI is not new, but the newest AI is front-of-mind

SMEs overwhelmingly acknowledged the wide variety of technologies in use in healthcare that could be considered "artificial intelligence," but drew a strong distinction between generative AI and other kinds of AI. Some interviewees categorized AI used in healthcare as "traditional" and "deterministic" versus "generative" or "probabilistic." In the traditional bucket, SMEs described machine learning techniques that assist with classification and data analysis, often using rule-based systems or supervised learning. These systems have been in use in healthcare for decades to support hospital processes and encompass a variety of advanced analytics, from descriptive to predictive, such as calculating hospital re-admission or sepsis risk scores.

> *"I mean, things like predictive analytics, people will tell you there's been AI, you know, for 30, 40 years that's been out there and that we've even used in healthcare that we just haven't really called 'AI.' But the big change is really in generative AI."*
>
> - Jeffrey Kim, VP, Chief Medical Information Officer, VCU Health

Underlyingly, "traditional AI" is trained on defined, curated datasets, often labelled by or with significant input from clinical experts. By contrast, generative AI is built on large, diverse datasets that may not be curated for clinical settings, and models are built on large-scale transformer architecture, complex neural networks, or deep learning methods, which make the relationship between input data and model outputs harder to ascertain. Generative AI is characterized by powerful synthesis and creation abilities, allowing models to generate novel content. When asked about the biggest successes in healthcare AI today, most SMEs mentioned generative AI applications for assisting in administrative functions, such as messaging patients or generating chart notes from different inputs (like lab results, patient interactions, and scanned documents).

> *"I had a nurse practitioner, she has dyslexia. And she said, 'You don't understand how long it takes me to correct the squiggly lines in my notes.' She said, 'When I finished seeing patients yesterday afternoon, I was done. […] You have changed my life.'"*
>
> - Charles Frazier, SVP, Chief Medical Information Officer, Riverside Health

One of the most commonly cited AI applications was "ambient listening" or "ambient voice recording," which is a genre of voice recognition technology that uses AI to passively record patient-provider conversations, analyze the content of recordings, and generate a summary of conversations. Ambient listening is a generative AI application that uses natural language processing (NLP). SMEs described using ambient voice recording, mainly in small pilots with clinicians, to assist with note-taking and charting, with overwhelmingly positive feedback from clinicians. In an interview with one EHR company, researchers received a demonstration of ambient listening where a clinician recorded a conversation on their smartphone during a simulated patient conversation, and chart notes were generated almost instantaneously from the conversation. SMEs quoted anecdotal evidence that ambient voice technology was reducing the administrative workload for clinicians and improving the quality of provider-patient interactions by allowing providers to focus more on the conversation.

> *"What this allows is for that interaction to be recorded, so the biggest piece of feedback we've heard is providers saying, like, I'm actually able to look my patients in the eye, and I haven't been able to do that for years. Because everything is being recorded, it's then generated into the patient's chart as a draft note that the provider can then go back and review, so that's helped a lot […] We've also seen time savings because everything is already in the note, they're able to close notes quicker, be able to spend more time with their patients as opposed to that administrative burden. […] We've gotten a lot of satisfaction back from our providers, you know, hearing terms like: this is a life-changing tool."*
>
> - Kati Charron, Senior IT Manager, VCU Health

One of the most promising opportunities presented by generative AI in the health sector is the ability to synthesize information across a wide array of unstructured data, which includes audio files, PDFs, images, and other inputs—often with a high proportion of natural language content—that do not fit neatly into database-style organization regimes. Unstructured data is traditionally harder for computers to process because the variety of formats presents challenges for machine readability. Because generative AI is trained on diverse datasets, it is better suited to interpreting unstructured data, which makes up a large amount of medical information (e.g. scanned documents and diagnostic images).

However, SMEs highlighted important issues with data quality that may impede successful integration of AI into healthcare. A few interviewees mentioned that data in health systems is often fragmented and incomplete, which raises concerns around the validity of advanced analytics outputs, since outputs are only as good as their inputs (garbage in, garbage out). Therefore, health systems are actively developing data governance processes and guidance alongside AI governance. Although it did not surface directly in interviews, generative AI is likely to raise the stakes on data governance because the ability to feed diverse, unstructured data into generative models offers a level of convenience that might overshadow the continued necessity to screen for data quality. Defining what constitutes high-quality data can be harder for unstructured than structured data (more on governance below).

> *"I wish we just called it math, you know? That's all it is. It's math, and math is powerful. But math depends on everything that you give [it] to operate on."*
>
> - Johanna Loomba, Director of Informatics, iTHRIV

Although interviews for this study disproportionately focused on generative AI, it is important to grapple with the wide spectrum of advanced analytics that use AI in healthcare. Simply because a technology becomes more commonplace or has been in use for a longer period of time, does not necessarily mean that questions about performance, validation, and risks have been adequately addressed. For example, the FDA approved Computer Aided Detection (CAD) systems for mammography in 1998, which were widely adopted nationwide (74% of all screening mammograms in Medicare were using the technology by 2008). Robust evaluation of health outcomes was not conducted until two decades later, when a largescale study determined that CAD did not improve diagnostic accuracy, and insurers were paying more for CAD without any demonstrable benefit to women.[108] As technologies become normalized, they often colloquially move into the category of "traditional" technologies to make room for the newest innovations, but academic literature points to the fact that even more deterministic models, for instance, require further testing, evaluation, and safeguards. One SME also pointed out that the use case does not always indicate the real level of risk—an administrative model could still result in harmful health outcomes.

> *"A lot of people talk about AI adoption and do the kind of more mundane tasks or the admin tasks first because that's kind of lower risk, but actually, I don't agree with that because even things from an admin perspective that seem to just be 'oh it's just triaging patients, we're just deciding who gets to have an appointment next,' all of those have kind of human biases that go into them."*
>
> - Subject matter expert from an independent research organization in the UK

The business case is burnout

In discussing the business cases for AI in healthcare, the most cited reason for health systems to adopt AI was efficiency, with many SMEs specifically mentioning clinician burnout. Interviewees discussed increasing pressures on providers in Virginia, staffing and expertise shortages, and a high volume of patient demand. Generative AI—and its applications for charting, summarization, and ambient recording—is seen as a technology that could help reduce "pajama time," meaning the time that clinicians spend completing administrative tasks after office hours.

> *"Right now, it is like more of a survival mode. It is a: I have way too many patients that need to see me, people are waiting—and I'm talking from a primary care perspective—but people are waiting three, six, nine months to get an appointment. I got to do better. I gotta get to these people. How do I do that? How do I not spend three hours in what they refer to as 'pajama time,' you know, every evening after my kids go to bed, dealing with the work, the administrative tasks that I couldn't do when I was seeing a patient every 20 minutes for the entire eight hours I was in the practice. So, I just think that's where they're really intensely focused right now."*
>
> - Beth Bortz, President and CEO, Virginia Center for Health Innovation

Other use cases were mentioned, such as improving diagnoses, triaging patients more effectively, and identifying readmission risk, but the promise of generative AI to improve workflows and work-life-balance is clearly front-of-mind for health systems. Again, a distinction between "traditional" and "new" applications of AI surfaced here. Other kinds of advanced analytics have been integrated into enterprise health system software for some time, but the perception is that generative AI is introducing new possibilities.

> *"I mean, I think the biggest success that I think I've seen in my 20-some years in healthcare is that the ambient scribes have completely changed the way that people work. I mean, I think it's going to extend the careers of physicians across the board."*
>
> - Andrew Markowski, Chief Medical Information Officer, UVA Health

One SME also discussed how clinicians are personally utilizing generative AI, beyond the enterprise systems adopted by their health system, to support their practice. Just as patients are increasingly using chatbots, like ChatGPT or Gemini, to seek out medical advice, clinicians are also supplementing their own expertise with chatbot queries. OpenEvidence, a chatbot developed by Harvard and MIT researchers, is free to use for verified U.S. healthcare professionals and has agreements with leading medical journals (e.g. the New England Journal of Medicine and JAMA Network) to use their paywalled content to answer user questions. One SME who uses OpenEvidence as a "bedside reference tool," also highlighted concerns that these kinds of AI fall outside what is traditionally regulated in clinical settings.

> *"I'm sure this has been discussed extensively, but everything's such a moving target. I'm sure my answer now would have been very different from what it would have been six months ago, and six months before that, and six months before that. For example, before, I was very worried about tools like OpenEvidence. […] I think so much is going through just providing information for physicians to act on rather than actually undergoing real clinical trials. I think that's the one thing I would really love to see more is rigorous study of applications before they're used. I know it's so hard because so many [AI tools] happen outside the jurisdiction of like the FDA or state bodies or so on, but for example, we have no idea, even now, if OpenEvidence is better at answering medical questions than just Chat GPT, than open-source things like Llama or so on."*
>
> - James Diao, Resident Physician at Brigham and Women's Hospital and Research Fellow at Harvard Medical School

Although it was not a focus of this study, which has emphasized enterprise adoption of AI, it is important to note that individual clinicians and patients likely also interact with many different AI tools that increasingly mediate their healthcare experiences in visible and invisible ways. And there is potential for these technologies to indirectly influence health outcomes and even become integrated more formally into diagnosis, treatment, and monitoring of patients.

The prominence of burnout as a business case for generative AI in healthcare highlights the potential of AI to create opportunities and efficiencies in the Virginia health sector if adopted responsibly. In the words of Beth Bortz, from the Virginia Center for Health Innovation, AI could be a "bright spot" that could help address widespread burnout challenges. But on the other hand, there are also risks inherent in adopting technologies to mitigate organizational and social challenges, especially if technological solutions are adopted without adequate guardrails.

There are reasons to be wary of technological fixes in organizations under efficiency pressures—as desperate times can result in cutting corners, intentionally or unintentionally. Moreover, there is mixed evidence that AI-enabled efficiencies will necessarily lead to better work-life-balance for employees in overburdened fields.[109] Clinicians whose administrative tasks are largely outsourced to AI, for instance, may be assigned more patients to evaluate each day. In other words, these workers may wind up strapped for time in new ways. These risks could be mitigated through sound organizational policy, robust AI governance, and rigorous evaluation of AI performance in real-world situations, all recommendations to be found later in this report.

Good governance is a priority, but it's uneven, institution-specific, and voluntary

Although study interviews emphasized enthusiasm for AI and a sense of urgency to address serious clinical needs in Virginia, SMEs frequently described their institutional approach to AI as "cautious," or risk-averse. Many health providers are actively establishing considered and managed processes for deciding what kinds of AI to adopt, how to share data with AI companies, how to fine-tune, monitor, and evaluate AI models, and more. All of these processes can be loosely grouped under the heading of "governance," which encompasses oversight mechanisms to manage risk, set standards, ensure compliance with relevant law, and facilitate organizational awareness.

> *"Most of the health systems are very deliberate about how they move forward with AI because there are so many unknowns. We all are learning, and when it comes especially to patient care, we want to be double-, triple-sure."*
>
> - Alok Chaudhary, VP, Chief Data and AI Officer, VCU Health

Providers described establishing governance committees, comprising clinicians, information technology specialists, data analysts, and legal and compliance officers. In two cases, health systems had recently hired a chief AI officer to coordinate governance efforts. However, providers are self-confessedly at widely varying stages of maturity in developing governance systems and oversight of AI. SMEs described AI governance that ranged from well-documented review and procurement processes to early-stage information-gathering about AI use and best practices. Although the SMEs interviewed for this study mostly came from large health systems, they expressed concern that smaller healthcare providers might be at a significant disadvantage in putting AI governance in place because they are more likely to lack the resources, expertise, and time to develop bespoke, institution-specific oversight. AI governance is a complex and moving target, even without comprehensive state or federal regulation, requiring data management strategies, model procurement policies, and sometimes in-house model development or fine-tuning—all more easily suited to providers with dedicated departments and staff. These comments highlighted a risk that an AI adoption divide could naturally emerge, where better-resourced systems benefit disproportionately from AI advances while smaller, rural, and under-resourced providers are left behind. Several SMEs articulated a desire for better knowledge-sharing across the sector to establish best practices.

> *"I sort of think about the impact of AI on jobs in general. You know that Virginia has an urban versus rural issue when it comes to broadband access. As AI basically starts to take over some of these admin level jobs, I think that can further compo und the urban versus rural conversation because the populations in Virginia that have access to strong and powerful broadband are going to continue basically upskilling and tapping into the early potential of AI when rural communities that are further disconnected are going to be less exposed to it and therefore become less competitive."*
>
> - A subject matter expert from a healthcare trade association

It is evident from the SME interviews that health providers in Virginia are taking AI advances seriously, embracing the potential of AI to improve health outcomes while also critically weighing the risks of different kinds and use cases for AI. As a sector that has long grappled with how to integrate technology into clinical practice (e.g. from x-rays to genome sequencing), providers are aware of the nuances of different AI tools, accustomed to complying with regulation and guidance on technology deployment, and circumspect about the real-world consequences of integrating technology into clinical decision-making.

> *"Of course, you know, every industry it's important for AI to be high quality and trustworthy, but healthcare does have that additional kind of layer. Number one, you have unique regulations that you need to be thinking about. Number two, it's just the fact that when you're dealing with patient care, you're dealing with the health of patients. So, there are additional responsibilities and additional considerations that come into play there."*
>
> - Subject matter expert from the technology sector

At the same time, healthcare providers have largely been left to their own devices in developing AI governance that can meet the moment, which is characterized by an explosion of new AI applications and use cases. They often take their lead from emerging voluntary frameworks and guidance, such as the NIST Risk Management Framework or the CHAI Responsible AI Guide (see Policy Landscape for more on these non-binding guidance documents). Like the technology itself, AI guidance and best practice documentation has proliferated in very recent years, making it hard to know what to follow and what will ultimately have the greatest longevity in the field.

> *"I think part of the motivation for the AI Trust and Assurance suite was that there wasn't one standard framework to follow. In fact, one of the things that we built in that tool was that you could import any number of different governance frameworks to use it for. We wanted it to be kind of a framework-agnostic tool. And then I don't need to tell you that in terms of actual government regulation, again, there's not any one standard. There's really not much of anything."*
>
> - David Hoffert, AI Software Developer, Epic

There is some work underway in Virginia to address the need for knowledge-sharing across the sector. The Virginia Department of Health (VDH) has a contract with the Virginia Center for Health Innovation (VCHI) to staff the Virginia Task Force on Primary Care, which has raised AI in primary care as one of its key issues. As a result of this interest in the sector, VCHI is launching a "learning collaborative" in October this year—an electronic hub that can serve as a resource hub to help primary care providers with AI adoption. A major goal of the learning collaborative is content curation and content vetting to put the most reliable content in the hands of frontline clinicians. Professional associations and third sector organizations are playing a significant role in collating non-binding standards and guidance, but the sector is therefore developing fragmented approaches where there could be a great deal of common ground.

Multiple and messy external dependencies underpin AI in healthcare

Different healthcare systems are also procuring AI tools in various ways, which affects their ultimate deployment and performance. Some providers, particularly those affiliated with universities and research institutions, are exploring developing their own models or piloting models developed by research teams. However, most providers are adopting AI tools developed by third-party vendors, which may involve direct collaboration with a dedicated AI developer (as in the case of Inova partnering with Abridge to support clinical documentation with generative AI),[110] or it may involve adopting an AI integration into an existing software provided by a third party (as in the case of Epic working with OpenAI and Anthropic to integrate generative AI into medical charting).

SMEs from enterprise EHR companies also emphasized the importance of fine-tuning models to specific patient populations, which introduces localized modifications that are implemented by downstream deployers (such as providers). Some recent lessons about the inaccuracies and potential harms propagated when models are trained on data unrepresentative of local populations have likely led to this now more common practice (see Introduction for a brief discussion of the documented risks). However, the decision to do localized training is still left up to developers and deployers.

"Everybody who works in the data science space recognizes that you need to do some sort of localized training of models, right? And in medicine especially, right? Particularly because you're dealing with different populations. The worst thing you could do is generalize a model that was built for one population and try to apply it in a different population."

- Dave Torgerson, Chief Analytics Officer, Sentara

"I do know that we do localization for deployments. If you're training a model and evaluating on a very urban population, you try to deploy that to a very rural population, that may not function very well."

- Subject matter expert from the technology sector

These nested external dependencies create accountability gaps and ambiguities that are not easily remedied with vendor-client agreements, such as DSAs (data-sharing agreements) or BAAs (business associate agreements). Language that is often found in existing regulation (such as clearly delineating "developers" versus "deployers") imperfectly captures the reality of AI development and deployment, where not only might a downstream deployer make consequential modifications to a model, but even end-users may influence model development overtime, as many kinds of AI models "learn" from new inputs and drift from their original parameters. Moreover, the blurred lines created by nested external dependencies raises questions about who should be responsible for monitoring model performance and evaluating outcomes. This could result in evaluation gaps, where AI tools are adopted without robust checks on whether they are actually improving outcomes. As with AI governance, larger and better-resourced health systems are more equipped to conduct evaluation as part of their internal processes than smaller providers.

SMEs on the provider side observed in some cases that there can be limited transparency from model vendors about inputs and performance. While some vendors might supply model briefs or model cards as part of their company practice in interacting with clients, these are at the discretion of vendors (who, themselves, refer to a combination of regulation—if it exists—and non-binding guidance to develop their business practice), and clients have found that they get different qualities of transparency information from different vendors.

> *"And then a lot of times what happens is the clinicians and users, they ask questions—ok, if it is coming up with this score, how? Can you tell us how did it come up with this score? So, there is still a thinking, and to some extent I agree with that, that it is like a black box. No one knows what it is doing inside. So, we have to have a good way for us to be able to explain that ok, not getting into the technical details, but this is exactly what we're doing here. And I think that helps from the adoption standpoint because our users want to know. Don't present us with something that we don't know how it came up with this calculation. Tell us exactly what it is doing."*
>
> - Alok Chaudhary, VP, Chief Data and AI Officer, VCU Health

Several SMEs raised concerns about sharing data with third parties that supply AI technologies. While data-sharing has arguably been a top concern for healthcare providers since HIPAA, AI has thrust this well-established regulatory issue—often relegated to data governance teams buried somewhere in the compliance bureaucracy of health institutions—into the spotlight. In part because of the lack of transparency in model information, negotiating data-sharing and retention agreements can be prolonged and contentious. A few SMEs mentioned that some vendors offer free or heavily discounted AI packages in exchange for access to more extensive data repositories. The cost of AI also came up in interviews as potentially prohibitive for certain health systems and a factor that makes it likely that customers are large enterprise systems (such as well-established EHR companies) would get access to AI integrations much more quickly, since they can be added onto existing services. For providers facing budget constraints, entering more lenient data-sharing agreements in exchange for free or low-cost AI access might be more appealing.

Healthcare providers almost universally expressed concern about ensuring PHI is not used to train vendors' foundation models, as PHI is strictly protected under HIPAA. In the words of one SME from Sentara, "under no circumstances will we let PHI go back to the mothership." But interviewees from across academia, the software industry, and healthcare providers, acknowledged the limitations of HIPAA to protect patient information, even when it is de-identified and would not fall under the definition of PHI. The ability of AI models to synthesize and re-identify individuals from vast datasets introduces new risks for the disclosure of protected personal information, particularly if data were shared with external third parties.

"I think the reality is that the power we have in data processing and these generative technologies makes it really too easy to reidentify data. So, in that sense, HIPAA is not sufficient because even if you followed all the rules of HIPAA, you're still going to produce datasets that are subject to reidentification using these modern methods."

- Dave Torgerson, Chief Analytics Officer, Sentara

Finally, there is a concern that vendors may sometimes fail to adhere to the data-sharing terms of signed agreements, requiring constant vigilance from customers and few meaningful avenues for recourse apart from the threat of discontinuing partnerships with those companies. While reputational damage can operate as an important market control, it does not adequately mitigate the potential harms to patients if their data (identified or de-identified) has been subject to an agreement violation. In addition, the burden of monitoring vendor compliance can be a "daunting task" in the words of one SME, especially as AI technologies are constantly evolving and updating. Lack of sector-wide clarity on how best to protect patient data in the context of AI, alongside messy nested external dependencies among developers, deployers, and end users, and an absence of enforcement pathways is likely contributing to the cautious approach that many healthcare providers described as their orientation toward AI in general.

Humans are staying in the loop

A significant theme of all the interviews conducted for this study was the importance of keeping humans in the loop, providing final oversight of AI outputs and ultimately retaining responsibility for AI-assisted decisions. It is clear that the health sector retains the primacy of clinician expertise and decision-making. The field has long relied on clinician experience to make diagnoses and suggest treatments under conditions of imperfect information and uncertainty (see Introduction for a longer discussion), and AI is not seen by the SMEs interviewed for this study as a replacement for clinical judgment. Rather, the human-machine interaction inherent in clinical adoption of AI implicates both the values embedded in the model and the values-based interpretations of the clinician.

"I think actually what will drive the importance of humans in the loop, more than just the question of whether we can have algorithms that are accurate or even ones that improve clinical outcomes, [is] rather how human values are incorporated into decisions. I think optimal decision-making is not just probabilities and outcomes. It's also which ones matter most to patients, which tradeoffs are worth making. And I think especially with AI, a lot of these value structures become implicitly encoded into the models."

- James Diao, Resident Physician and Brigham and Women's Hospital and Harvard Medical School Research Fellow

In fact, human oversight is a commonly cited component of emerging AI governance practices across the board. In particular, SMEs expressed concern about AI being used to automate denials of service, whether that occurs in the form of denying health insurance prior authorizations or denying access to appointments through a triaging process.

> *"AI stands for augmented intelligence. It is not meant to replace clinical thinking, and you as a clinician are always responsible regardless of what you're reading or what the note is being generated. You're responsible ultimately for that."*
>
> - Stephen Morgan, SVP, Chief Medical Information Officer, Carilion Clinic

Although generally SMEs described the various ways in which humans are kept in the loop reviewing AI outputs as an essential check on AI systems, they also recognized that having a human in the loop does not necessarily mitigate all risks, especially as generative AI becomes more embedded in healthcare. For example, several interviewees mentioned concerns about clinicians overly relying on AI and losing critical thinking skills as a result. SMEs from companies developing AI integrations mentioned the risk of "automation bias," where providers might mindlessly accept AI-generated outputs, and safeguards they are integrating to mitigate this possibility, such as requiring provider verification for AI-generated messages. Automation bias and a lack of critical thinking are real risks in a context where clinician burnout is a key motivation for adopting these efficiency-enhancing AI tools. Interviewees also mentioned the risk of hallucinations, which are common in large-language models, leading to inaccurate outputs, and concerns about what happens when a clinician chooses to ignore or reject a recommendation from an AI system.

> *"I also worry that there will be adoption problems from clinicians when they're scared to disagree with the AI, right? Like that bigger problem to me is capturing an AI prediction in the medical record and then a lawyer pointing at it and saying: I can't believe you didn't follow the algorithm."*
>
> - Andrew Markowski, Chief Medical Information Officer, UVA Health

Some of these concerns connect directly to others about liability and accountability if things go wrong. The clear emphasis on clinical judgment and oversight might be intended as a safeguard against AI risks, but clinicians worry about taking full responsibility for a decision that was significantly aided by an AI tool.

> *"I think some of the questions that have come up with the use of AI is just around the liability of using it. Particularly from a malpractice perspective. I think that that has not been very well-defined […] Where does the liability rest? Does it lie with the AI developer? Does it rest with the healthcare organization? Does it rest with the physician? And those things are still all up in the air. And so, there's a lot of concern about that."*
>
> - Robin Anderson, VP, Chief Medical Officer, Sentara

Echoing the academic literature (see Introduction), some SMEs acknowledged the complex interaction between human judgment and machine outputs. They articulated a strong desire for better validation and performance standards for AI used in healthcare, while recognizing that perfect explainability may not be possible or even necessary. Being able to perfectly reverse-engineer a clinical decision has arguably been a challenge in a field so strongly characterized by expert judgment, but the sector has come to rely on standards of transparency (disclosures) and validation (testing) to ascertain the safety of a wide range of clinical aids, from pharmaceuticals to medical devices, and several SMEs commented on how similar standards for AI would give the humans in the loop a better evidence base on which to base their interactions with AI tools.

> *"From a clinician perspective, I think the risk is how do we help make sure that these models are trustworthy and transparent in terms of how they're used, so that I can, you know, both be trusting of it in my knowledge, but also when I'm suggesting something to a patient that they can understand why I am even a better physician for having that knowledge that's gained from the forging of AI."*
>
> - Jackie Gerhart, Chief Medical Officer, Epic

Developers, deployers, and end-users want regulatory clarity and harmonization

Considering the atomized landscape of AI in healthcare—where different stakeholders are adopting different AI technologies at different rates and developing institution-specific governance processes to manage risk and establish oversight—nearly all SMEs interviewed for this study articulated a desire for greater coordination and clarity on standards and best practices. Not all SMEs agreed that this was a role best suited to state governments, with some seeing these issues best addressed at the federal level or through the coordination of professional bodies, but most recognized that the future holds more regulation for AI in healthcare. As discussed above, the health sector is no stranger to regulation, and especially within enterprise contexts, regulated providers are constantly working within compliance frameworks in the interest of protecting patient safety.

"If I could give an example of one regulation that some aspects of it, I think, did very well […] from the office of the National Coordinator for Health IT […] They recognized that things were early. They were the first ones, at least at a federal level, that were really directly regulating AI, and they approached it in broad strokes here, but they really approached it as, ok, we're going to require transparency, which makes a lot of sense, and then we're going to require basically good development practices […] but basically just a requirement that didn't tell us exactly what we had to do, how we had to do it, but said you have to have practices for risk management or risk analysis and risk mitigation and then governance within your organization, and you have to explain how you do that."

- Subject matter expert from the technology sector

Many AI systems and tools fall outside the purview of existing regulation, which is limited, and perhaps some of the most concerning AI applications in healthcare are emerging in the direct-to-consumer market, which encompasses an array of AI technologies that go beyond the scope of this study. But the fact that many AI technologies being considered or already adopted by health systems are falling through regulatory gaps—even in the more regulated corners of the sector—points to a need for clearer guardrails and requirements. The lack of regulation is also placing the onus on individual institutions (companies, providers, etc.) to create their own standards and governance processes to fill the gaps.

"So, it's funny because vendors—like, I go to conferences, and I listen to various organizations, and they're very smart people […] and all of them were talking about the regulations around AI, and I'm like you people have no idea what you're talking about. There is no regulation. […] I used to do these presentations with the [American Hospital Association], and it was like half of a single PowerPoint page. I'm like: here's your regulation. Notice that the sheet is empty, right? So that's it. We're kind of on our own with that. You have to be very careful when you're talking to vendors when they mention the term regulation. There really isn't any. So really, we have to be very careful about what is the impact to us."

- Stephen Hughes, Chief Technology Officer, Mary Washington Healthcare

In interviews, SMEs from the technology sector articulated a desire to avoid a patchwork of state and federal regulation by seeking out harmonization across jurisdictions. And clinicians expressed a desire for greater clarity on liability and model transparency. Several interviewees expressed a recognition that guidance or regulation needs to be agile and adaptive to respond to a rapidly evolving technology. One SME, with experience in medical device regulation at the UK's MHRA and now at the global non-profit, HealthAI, placed particular emphasis on the importance of extensive and well-structured consultation with expert stakeholders to develop the best balance of black-letter law and regulatory guidance, a process that takes time.

"I genuinely believe that in time, as with any new technology, as it becomes more prevalent, and it's genuinely used a lot more, there will be more problems, and people will then regulate, and my concern is that what we'll end up doing is knee-jerking backwards and overregulating in the future as opposed to sensibly regulating now. So, the question is what's worked well? Candidly, not rushing, I think, is a sensible thing. And I understand the frailty of guidance versus regulation because you can't enforce guidance, but I think recognizing and understanding the principles [of legislation], not rushing, being prepared to say we're going to go with guidance, and we're going to watch it because it's easier to put guidance out and change it in three years' time than it is to change the law."

- Paul Campbell, Chief Regulatory Officer, HealthAI

The case for legislation to mitigate AI risks and harms in healthcare

This report has surfaced a range of admirable efforts on the part of healthcare providers and technology developers and deployers operating in Virginia to embed responsible AI practices in their systems and workflows. Based on interviews conducted for this study, there is a high level of awareness and a strong existing commitment in the sector to embrace innovation while also protecting patients and mitigating risks. However, these efforts are largely institution-specific (requiring each provider or deployer to develop their own standards), voluntary (based on the best guidance that governance professionals can synthesize), and siloed (there are limited and rare opportunities to share best practice across the sector).

Yet, there is a clear need for AI safeguards. Writing for *Chief Healthcare Executive*, Sentara Health's Chief Information Officer Joe Evans explains the impetus for Sentara's institution-specific AI guardrails: "Failures and mistakes involving AI in healthcare could further erode public trust in healthcare and undermine the potential benefits of AI for the industry. Possible risks include data privacy and algorithm bias, or even applications of AI that result in patient harm."[111] This report has cited several documented risks and harms associated with AI use in the health sector. In the absence of minimum requirements for AI developers and deployers to ensure that these technologies are trustworthy, there is a greater risk of consequential gaps in oversight. For example, "AI companies currently undertake model evaluations," observes a commentary by UK AI advisers in *Nature*,[112] but "although model evaluation and safety protocols are necessary, they are not sufficient. This is because evaluating a model's technical performance is not the same as assessing its real-world economic and societal impacts." Even the most well-intentioned voluntary safeguards may overlook important factors that contribute to positive or negative outcomes. To quote Paul Campbell from HealthAI, the principles of regulation are simple: "Does it work? Is it safe? Can you prove it?" And legislation can provide a mandate to attach specific definitions and metrics to those principles for AI.

Legislation on AI in healthcare could address several issues that surfaced during this study:

- **Mitigate known risks, anticipate unknown risks:**
  - Many of the risks associated with AI in health sector are similar to risks about AI more broadly, and health is a high-stakes environment where risks can have life-or-death consequences;
  - There are documented incidents of negative outcomes from AI in health contexts;
  - Rate of AI adoption is rapid, and keeping up with technology once it is deployed in real-world scenarios is like playing whack-a-mole;
  - Many technologies and associated risks are falling through the gaps in existing regulation, exposing the limits of longstanding guardrails, like HIPAA.

- **Standardize safety requirements:**
  - Despite impressive efforts by health sector and industry to address risks and develop frameworks, these are still largely voluntary and institution/organization-specific;
  - Different healthcare providers are at vastly different stages of maturity in the development of their own AI frameworks;
  - There is limited federal guidance on AI deployment in clinical settings.

- **Provide support for adoption across the sector:**
  - Smaller clinical practices and hospital systems risk falling behind or missing out on the opportunities afforded by AI;
  - The robustness of AI guardrails and governance depends on the capacity of individual hospital systems or healthcare providers, which means those with dedicated staff and high levels of expertise in data governance or analytics have a disproportionate advantage;
  - There is a need to share knowledge across the Commonwealth in a more coordinated way.

- **Clarify steps toward regulatory compliance for technology developers and deployers:**
  - Industry is developing technologies with great potential to improve healthcare but need regulatory certainty to confidently invest in further innovation, and this is especially true for startups or smaller enterprises;
  - Companies operating in the healthcare sector are no strangers to regulation and compliance and can provide useful insights on best practices.

- **Set an example for responsibility in AI governance more broadly:**
  - The high-stakes nature of the healthcare sector means that many stakeholders have been seriously thinking about responsible AI innovation, and their efforts could not

only form the basis of robust, feasible guidance on regulatory standards but also inform broader safeguards for AI in other contexts.

Different approaches to regulating AI

Before outlining several recommendations for legislation on AI in healthcare, policymakers may consider whether legislation aimed at mitigating risks and harms associated with AI should address the healthcare context specifically, or whether healthcare applications of AI are better addressed within broader legislation that applies to multiple contexts and risk thresholds. In the 2025 General Assembly Session, several bills put forward horizontal, cross-sector AI legislation that would have applied to high-risk AI models. As discussed above (see Policy Landscape), an alternative approach to horizontal regulation is sector-specific legislation. Taking a sector-specific approach may offer an opportunity to target a high-risk domain in which consequential decisions are influenced by AI, instigate the development of regulatory guardrails that could be tested, evaluated, and adapted to other high-risk domains, and mitigate the need for broad exclusions that would open regulatory gaps (e.g. carve-outs for HIPAA-covered entities, in spite of the limitations of HIPAA vis-à-vis AI). The following briefly recounts the differences between these two approaches.

Horizontal, cross-sector AI legislation

Many of the challenges facing the healthcare sector in mitigating the risks of AI are not unique to healthcare. Just to establish, oversee, and enforce new AI regulation in health, legislation will need to set down definitions of AI, define risk categories, create new oversight entities and processes, and more. Taking a broad, horizontal approach to AI across different sectors could help establish definitions in the state code that have wider applicability to other AI use cases, avoid fragmented oversight or conflicting requirements, and reduce the administrative burden of overseeing and enforcing regulation on a sector-by-sector basis. Given the current lack of AI legislation in Virginia and therefore the need to build institutional capacity and regulatory expertise, a horizontal, cross-sector approach could provide a more flexible framework that ensures these efforts benefit the Commonwealth as a whole.

Sector-specific legislation

Unlike many sectors that are rapidly adopting and are affected by AI, healthcare providers and technology developers are actively developing their own governance frameworks standards in the absence of prescriptive regulation, and these existing efforts could provide the groundwork for sector-wide standards. The extensive experience of the healthcare sector in grappling with high-risk applications of technology and balancing the protection of patients with the promise of emerging technology could be a valuable resource to the state as it begins to set out AI legislation and navigate this new terrain. In addition, healthcare is, by definition, a high-risk domain in which unregulated applications of AI could have serious, life and death consequences. Such high-risk domains should be among the top priorities for legislation on AI. The following policy recommendations assume a sector-specific approach to regulation informed by this study, focused on AI in healthcare with particular emphasis on generative AI applications in the sector.

Policy Recommendations

Much like the approach articulated by several of the SMEs interviewed for this study—cautious, but still moving forward to develop governance and adopt AI that could improve health outcomes—the regulatory approach that may prove best suited to the challenges identified in this study is one that balances taking action with governance and review.

Based on this limited study on AI in healthcare in Virginia, JCOTS staff offer several policy recommendations. Subject matter experts interviewed for this study and the literature reviewed on this topic identify a need for robust guardrails and guidance about the use of AI in healthcare, alongside concerns about how to ensure regulation is well-aligned to the real risks and harms of AI in this context. The combination of recommended regulations and a new regulatory process (the roadmap) are intended to address these two sides of the responsible AI coin.

Therefore, **policy recommendations** are two-fold:

(1) Recommendations for substantive regulations of AI in healthcare;
(2) Recommendations for legislation to simultaneously establish a new regulatory process for iteratively and flexibly developing emerging technology regulation.

In the absence of comprehensive federal legislation on AI, U.S. states have begun passing legislation that sets out new requirements for AI developers and deployers (as discussed in Policy Landscape). As much as these state-level efforts may create a patchwork of different requirements for the AI industry to comply with, they also offer a potentially valuable opportunity for experimentation, evaluation, and refinement of regulation that could inform other states or even federal policies. However, this potential hinges on building evaluation, review, and revision into the regulatory process, and mechanisms for conducting this kind of iterative testing are often absent from existing AI legislation. Emerging technologies have presented legislators and regulators worldwide with a renewed impetus to innovate on the regulatory side of the equation and not just leave innovation to technologists. Evidence of this impetus has come in the form of new mechanisms for consultation and evaluation, such as the EU AI Act's regulatory sandbox approach,[113] the MHRA's AI Airlock in the UK,[114] or Singapore's non-legislative Model AI Governance Framework.[115] All of these mechanisms share a commitment to collaboration with industry, iterative (ongoing) policy development, and sector-led regulation to leverage relevant expertise.

Building evaluation and transparency into the regulatory process at the same time that regulation places evaluation and transparency requirements on technology developers and deployers may help to address a fundamental tension in technology policy: the need for AI safeguards with the need for those safeguards to remain agile, flexible, and evidence-based to best protect both innovation and the public interest.

It is far too early to know what the impact of recent state-level AI legislation across the country will be, but the Commonwealth has an opportunity to set a precedent for responsible regulation of emerging technologies by coupling proactive legislation with a regulatory roadmap that embeds flexibility and learning into the regulatory process itself. This approach could ensure that the impacts in Virginia are measured, reviewed, and shared to the benefit of the state AI ecosystem as well as other jurisdictions grappling with similar policy issues.

A roadmap to responsible AI regulation

A regulatory roadmap that builds an iterative, collaborative process into AI regulation development offers significant advantages given the unique characteristics of artificial intelligence technology. Since AI is rapidly evolving with unknown future applications, risks, and opportunities, traditional regulatory approaches may prove insufficient or quickly outdated. More foundational work is needed to establish effective technical oversight practices, translate emerging standards into workable compliance frameworks, and ensure alignment with other jurisdictions' approaches.

By adopting a flexible, consultative roadmap, regulators can accommodate ongoing technological developments while fostering valuable knowledge-sharing between technical and regulatory experts. This approach also acknowledges the substantial time required to develop and implement new guidance, allowing for more thoughtful and effective regulation that can adapt and respond to both evidence-based impacts of AI and the impacts of regulation.

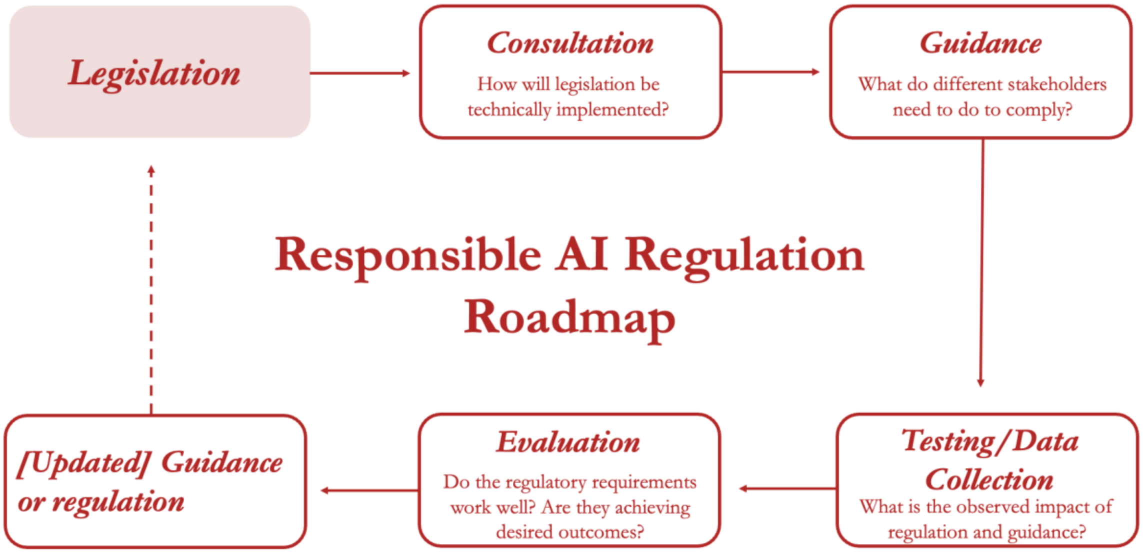
Legislation
Consultation
How will legislation be technically implemented?
Guidance
What do different stakeholders need to do to comply?
Responsible AI Regulation Roadmap
Testing/Data Collection
What is the observed impact of regulation and guidance?
Evaluation
Do the regulatory requirements work well? Are they achieving desired outcomes?
[Updated] Guidance or regulation

Figure 2: The regulatory roadmap process

Understanding the roadmap process

The responsible AI regulation roadmap synthesizes recent insights from both policy and technology best practices, namely: (1) the OECD regulatory governance cycle,[116] which recognizes regulation as requiring continuous review and adaptation to ensure it is fit for purpose; (2) the emergence of regulatory practices to provide pathways for both regulatory and technological innovation (e.g. sandboxes); and (3) industry and regulatory guidance around the development and deployment of novel technologies and medical devices. Policymakers should consider incorporating the responsible regulation roadmap into legislation that sets out principles for AI in healthcare.

Legislation: As the starting point on the roadmap, legislation can provide a mandate for action on AI alongside a set of principles to guide technical standards and compliance frameworks that can be further defined through the subsequent steps in the regulatory process. The roadmap is a feedback loop that not only implements the legislative mandate but also encourages continuous consultation, monitoring, and evaluation of outcomes. Legislation should designate coordinating or oversight entities for each stage of the roadmap process, recognizing that a single agency need not oversee the entire roadmap. For instance, since the roadmap does not exclusively focus on compliance and enforcement, the agency responsible for these regulatory functions may be a different one than the agency responsible for data collection and evaluation.

Consultation: Knowledge-sharing among technical, social science, clinical, and regulatory experts is an essential component of developing regulatory guidance to help AI developers and deployers implement legislative mandates. While legislation can establish a framework and principles for regulation, guidance can provide more specific details on how different stakeholders can meet regulatory requirements. In addition, guidance is more flexible and easier to amend and update as technology evolves and more policymakers have access to more research on what works. Other countries have complemented regulation with guidance after extended consultation periods (e.g. recently, the UK's Guidance on the Impact Evaluation of AI Interventions[117] and the EU's Explanatory Notice and Template for the Public Summary of Training Content for GPAI Models[118]).

Policymakers should also consider how to involve members of the public and/or affected patient groups in discussions and deliberations on guidelines to facilitate democratic participation and engender public trust in outcomes.

Finally, policymakers should consider how to build on existing knowledge-sharing and consultative efforts in the Commonwealth on healthcare AI, working with organizations like the Virginia Hospital & Healthcare Association and the Virginia Center for Health Innovation as well as state agencies, such as VITA and the Virginia Department of Health.

Guidance: Based on the consultation process, the guidance stage of the roadmap involves publishing actionable guidance for stakeholders affected by legislation. Guidance should take account of best practices as well as practical considerations, such as technical feasibility, and it should include an outline of next steps for evaluating the impact of the regulatory requirements, further consultation, and review.

Testing/Data Collection: At this stage, designated agencies should collect data about the impact and outcomes of regulation with the goal of monitoring how successfully regulatory interventions address the risks and harms they were intended to mitigate. Policymakers may consider establishing mechanisms for public-private partnerships and research collaborations to design robust studies of regulatory impacts and allocating funding to support these endeavors.

Evaluation: At this stage, regulations should be evaluated for actual, observed outcomes based on data collection in the previous stage. The results of regulatory evaluations should be presented in a programmed review (stipulated in legislation) to both legislative and consultative bodies involved in the implementation of the law. This stage constitutes an audit of regulatory impacts: How well did the guidance work in practice? Were there unanticipated costs or inefficiencies? Do they outweigh any observed benefits in addressing social harms?

Updates/Amendments: At this stage, policymakers may consider revisiting existing regulatory guidance or legislation to address gaps or misalignment of regulation and outcomes.

The following sections present a policy consideration for legislators—whether to pursue sector-specific or horizontal AI legislation—and policy recommendations for the establishment of minimum guardrails for AI in healthcare alongside a regulatory roadmap to define the specifics of those guardrails through an agile, deliberative, and empirically grounded process.

Policy Recommendations

These recommendations are intended to provide suggestions for substantive requirements that could be included in legislation on AI in healthcare. Policymakers may consider:

1. **Regulations for AI developers and deployers** that allow for flexibility and further development through consultation and guidance, but should include at minimum:

   a. Documented AI governance processes for healthcare providers that detail internal standards and decision-making procedures for adopting AI systems, including designating an accountable AI officer or committee, and file governance documents annually with a designated state agency.

      i. To assist smaller healthcare providers and clinical practices in developing governance documentation, regulatory agencies should publish template documentation and establish guidance on best practices.

   b. Model transparency standards that set minimum reporting requirements for all AI developers to provide deployers and designated oversight agencies. Transparency standards should apply to modifications made to AI models, including by deployers, in recognition of the complex dependencies that result from software integrations (e.g. an LLM that is integrated into an Electronic Health Record system). Transparency standards should be developed in consultation with diverse stakeholders but may include training data, model weights, and performance metrics/benchmarks.

   c. Model validation and evaluation standards that require independent impact assessment, testing, and evaluation of phases of models, encompassing both: (1) *ex ante* (before deployment assessments) that test a model in experimental conditions against retrospective data as well as in prospective trials, where randomized control trials should be used where possible; and (2) *ex post* (after-deployment assessments) to evaluate model performance in real-world settings. Policymakers may consider making funding available to facilitate robust independent evaluation for smaller enterprises or startups that may face additional hurdles due to the costs of evaluation.

d. Adverse event reporting, which encompasses reporting information about serious incidents and negative events that result from the development or deployment of technology to designated regulatory agencies. Definitions of categories of incidents that require reporting should be defined in regulatory guidance, and reporting should be continuous throughout the lifecycle of an AI product.

e. Specific requirements for humans in the loop and clarity on liability, such as requiring user-centered design in the development of AI models, requiring AI literacy training for human oversight of AI models (e.g. for clinical review of AI outputs), and requiring ongoing workforce development plans for entities procuring AI solutions to be documented and filed with governance practices (see 1a). To address clinician concerns about liability, legislators should consider clarifying the roles and responsibilities of developers, deployers, and end-users (who are often providing last-resort oversight of AI outputs) in the broad interest of reducing ambiguity, encouraging machine-augmented decision-making, and bolstering AI adoption.

f. Public disclosures that provide patients and members of the public with information about how their data is managed and protected in relation to AI and when AI is being used in their care. Policymakers may consider placing emphasis on disclosures for more opaque models, such as probabilistic AI models, like generative AI. Public disclosures should be distinct from the transparency requirements of 1b above, which are intended to supply deployers and regulators more specialized technical information. Public disclosures should provide comprehensible, clear, and actionable information to patients, which may include opt-in or opt-out mechanisms and the ability to report concerns to a designated regulatory oversight agency with a right of action.

g. Establishment of an enforcement agency that can receive complaints and concerns from patients or members of the public, proactively investigate non-compliance or violations of established regulation, and impose penalties.

2. **Establishing a regulatory roadmap**—a new process for developing collaborative, flexible, and evolving regulation on emerging technology—to include at a minimum:

a. Consultation: Policymakers may consider mandating a consultation process and designated consultative oversight body or agency for the development and refinement of regulatory guidance, with the aim of facilitating knowledge-sharing among subject matter experts and delivering feasible guidance. The goal of the consultative process is to provide detailed regulatory guidance to affected stakeholders to enable compliance.

i. Policymakers are encouraged to consider building on existing consultative efforts in the Commonwealth and mandating an extended consultation window for the development of initial regulatory guidance, followed by regular consultation sessions during shorter windows that continue throughout the period in which regulation is in effect.

ii. Policymakers may consider whether a particular consultative process is encouraged or best suited to developing guidance, such as regulatory sandboxes, citizen assemblies, or multistakeholder forums.[119]

iii. Given the complexity of regulatory issues around AI in healthcare, which are both technical (about the technology) and social (about how humans use and interact with technology), policymakers may consider stipulating a diverse range of stakeholders for inclusion in the consultative process, to include: technologists (developers/deployers), social scientists (e.g. Science and Technology Studies (STS), Human-Computer Interaction (HCI), and media researchers), clinicians, IT specialists, regulatory experts, and patients/patient advocacy groups. To address issues with public trust and acceptance of AI while also protecting patient/consumer rights, the consultation process should involve mechanisms for public and patient participation.

b. <u>Guidance:</u> Policymakers may consider including a provision for guidance to be issued by a designated oversight body based on the consultative process established in 2a. Guidance should address specific requirements for affected stakeholders named in legislation to comply with regulations outlined in legislation. Unlike the minimum requirements outlined in legislation, guidance is specific and more easily adaptable to a changing federal and state regulatory landscape. Guidance may include, for example, safe harbor provisions that would allow developers and deployers to meet compliance criteria if they already comply with relevant federal regulations.

c. <u>Data collection/testing:</u> Policymakers may consider designating an oversight agency for monitoring and evaluation of AI legislation to include collecting data on regulatory outcomes. The oversight agency for data collection and evaluation may be the same as the compliance agency (e.g. the Attorney General's Office) or a separate designated agency (e.g. JLARC), but the agency would be responsible for designing and implementing an evaluative study of the performance of regulation as implemented. Policymakers may consider allocating funds to facilitate public-private partnerships or research partnerships to deliver independent evaluations.

d. Evaluation and reporting: Policymakers may consider including programmed review of regulatory outcomes and impacts in legislation, requiring the results of evaluation (2c/2d) to be reported to a designated oversight agency or agencies (e.g. JCOTS, Attorney General's Office, or other). The review process should include tracking AI legislation in other states and at the federal level to inform amendments to the regulatory guidance. Staff recommend making evaluative reports public, published at regular intervals while the regulation is in effect. Staff would also encourage findings on regulatory outcomes to be shared in multi-state forums to enable knowledge-sharing about best practices among states.

e. Delayed enactment, phased compliance deadlines, tiered compliance requirements: To enable the regulatory roadmap to produce meaningful guidance in a new domain (AI) and to provide affected stakeholders named in the legislation time to develop implementation plans and procedures, policymakers may consider stipulating a delayed enactment window, phased compliance deadlines for different regulatory requirements to allow for a reasonable implementation window, and tiered compliance requirements for developers and deployers of different sizes and capacities so as not to disproportionately burden startups and small- or mid-sized enterprises. Policymakers may consider a phased compliance timeline that provides 12 months for initial consultation (2a) and guidance (2b), with full compliance required after 24 months, and testing and evaluation after another 36 months (2c and 2d).

Author Information

Kira Allmann, Ph.D.
Chief Policy Analyst, Joint Commission on Technology & Science
Contact: info@jcots.virginia.gov

Thanks To

VASEM COVES Summer Fellow, who provided research assistance for this study:
Navya Annapareddy, University of Virginia

Subject matter experts interviewed for this study:
Philip Alcaidinho, Manager, Healthcare Policy and Compliance, Meditech
Robin Anderson, VP and Chief Medical Officer, Sentara
Beth Bortz, President and CEO, Virginia Center for Health Innovation
Paul Campbell, Chief Regulatory Officer, HealthAI
Kati Charron, Senior IT Project Manager, VCU Health
Alok Chaudhary, VP, Chief Data and AI Officer, VCU Health
James Diao, Resident Physician at Brigham and Women's Hospital & Research Fellow at Harvard Medical School
Joe Evans, VP and Chief Health Information Officer, Sentara
Charles Frazier, Senior VP and Chief Medical Information Officer, Riverside Health
Jackie Gerhart, Chief Medical Officer, Epic
David Hoffert, AI Policy Analyst and Software Developer, Epic
Stephen Hughes, SVP, Chief Technology Officer, Mary Washington Healthcare
Jeffrey Kim, VP and Chief Medical Officer, VCU Health
Johanna Loomba, Director of Informatics, iTHRIV
Andrew Markowski, Chief Medical Information Officer, UVA Health
Sean McGunigal, AI Research & Development Team Lead, Epic
Corey Miller, VP of Research & Development, Epic
Steve Morgan, Carilion Clinic, Senior VP and Chief Medical Information Officer
Ernest Apamoh Nnoko, Analyst, Privacy & Security Regulatory Compliance, Meditech
Christina Nuckols, Director of Health Policy, Sentara
Dave Torgerson, Chief Analytics Officer, Sentara

And:
A PhD candidate specializing in healthcare AI, Harvard University
A subject matter expert from an independent research body in the UK

2 subject matter experts from Meditech
5 subject matter experts from the technology sector
5 subject matter experts from healthcare providers
3 subject matter experts from a healthcare trade association

Subject matter experts consulted informally during background research for this study:
Dr. Marzyeh Ghassemi, MIT
Dr. Emma Pierson, UC Berkeley
Christopher Lipson, The Joint Commission

Study advisory group:
Dr. Sylvester Johnson, Northwestern University
Dr. Sandra Soo-Jin Lee, Columbia University
Dr. Sarah Parker, Virginia Tech

References

[1] 15 U.S.C. §9401(3). https://www.govinfo.gov/content/pkg/USCODE-2020-title15/html/USCODE-2020-title15-chap119.htm.

[2] Yu, Kun-Hsing, Andrew L. Beam, and Isaac S. Kohane. "Artificial Intelligence in Healthcare." Nature Biomedical Engineering 2, no. 10 (2018): 719–31. https://doi.org/10.1038/s41551-018-0305-z.

[3] Davenport, Thomas, and Ravi Kalakota. "The Potential for Artificial Intelligence in Healthcare." *Future Healthcare Journal* 6, no. 2 (2019): 94–98. https://doi.org/10.7861/futurehosp.6-2-94.

[4] GAO, *Artificial Intelligence in Health Care: Benefits and Challenges of Technologies to Augment Patient Care*. GAO-21-7SP. Government Accountability Office, 2020. https://heinonline.org/HOL/Page?handle=hein.gao/gaobaecbd0001&id=1&div=&collection=gao.

[5] Wells, Nora, Amanda K. Sarata, April J. Anderson, and Paulette C. Morgan. *Artificial Intelligence (AI) in Health Care*. No. R48319. Congressional Research Service, 2024. https://crsreports.congress.gov/product/pdf/R/R48319.

[6] Yu, Kun-Hsing, Elizabeth Healey, Tze-Yun Leong, Isaac S. Kohane, and Arjun K. Manrai. "Medical Artificial Intelligence and Human Values." New England Journal of Medicine 390, no. 20 (2024): 1895–904. https://doi.org/10.1056/NEJMra2214183.

[7] Ledley, Robert S., and Lee B. Lusted. "Reasoning Foundations of Medical Diagnosis." *Science* 130, no. 3366 (1959): 9–21. https://doi.org/10.1126/science.130.3366.9; Tversky, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157 (1974): 1124–31.

[8] Salloch, Sabine, and Andreas Eriksen. "What Are Humans Doing in the Loop? Co-Reasoning and Practical Judgment When Using Machine Learning-Driven Decision Aids." *The American Journal of Bioethics: AJOB* 24, no. 9 (2024): 67–78. https://doi.org/10.1080/15265161.2024.2353800.

[9] Topol, Eric J. "High-Performance Medicine: The Convergence of Human and Artificial Intelligence." *Nature Medicine* 25, no. 1 (2019): 44–56. https://doi.org/10.1038/s41591-018-0300-7.

[10] Colangelo, Margaretta. "1247 FDA Authorized AI-Enabled Medical Devices." *AI in Healthcare Milestones*, July 14, 2025. https://www.linkedin.com/pulse/1247-fda-authorized-ai-enabled-medical-devices-margaretta-colangelo-kvdmf/.

[11] Ada Lovelace Institute and Nuffield Council on Bioethics. *DNA.I. - Early Findings and Emerging Questions on the Use of AI in Genomics*. London, UK, 2023. https://www.adalovelaceinstitute.org/report/dna-ai-genomics/.

[12] Topol, "High-Performance Medicine."

[13] Ghassemi, Marzyeh, Luke Oakden-Rayner, and Andrew L. Beam. "The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care." *The Lancet Digital Health* 3, no. 11 (2021): e745–50. https://doi.org/10.1016/S2589-7500(21)00208-9; Topol, "High-Performance Medicine."

[14] GAO, *Artificial Intelligence in Health Care*.

[15] Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (2019): 447–53. https://www.science.org/doi/10.1126/science.aax2342.

[16] Wong, Andrew, Erkin Otles, John P. Donnelly, et al. "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients." JAMA Internal Medicine 181, no. 8 (2021): 1065–70. https://doi.org/10.1001/jamainternmed.2021.2626.

[17] Kaviani, Parisa, Subba R. Digumarthy, Bernardo C. Bizzo, et al. "Performance of a Chest Radiography AI Algorithm for Detection of Missed or Mislabeled Findings: A Multicenter Study." Diagnostics 12, no. 9 (2022): 2086. https://doi.org/10.3390/diagnostics12092086.

[18] GAO, *Artificial Intelligence in Health Care*.

[19] Banerjee, Imon, Kamanasish Bhattacharjee, John L. Burns, et al. "'Shortcuts' Causing Bias in Radiology Artificial Intelligence: Causes, Evaluation, and Mitigation." Journal of the American College of Radiology : JACR 20, no. 9 (2023): 842–51. https://doi.org/10.1016/j.jacr.2023.06.025.

[20] Morley, Jessica, and Luciano Floridi. "The Ethics of AI in Healthcare: An Updated Mapping Review." In *Ethics and Medical Technology: Essays on Artificial Intelligence, Enhancement, Privacy, and Justice*, edited by Matthew C. Altman and David Schwan. Springer Nature Switzerland, 2025. https://doi.org/10.1007/978-3-031-94690-5_2.

[21] Topol, "High-Performance Medicine."

[22] Keane, Pearse A., and Eric J. Topol. "With an Eye to AI and Autonomous Diagnosis." *Nature*, NPJ Digital Medicine, no. 40 (August 2018). https://www.nature.com/articles/s41746-018-0048-y.

[23] Chen, Irene Y., Emma Pierson, Sherri Rose, Shalmali Joshi, Kadija Ferryman, and Marzyeh Ghassemi. "Ethical Machine Learning in Healthcare." *Annual Review of Biomedical Data Science* 4, no. Volume 4, 2021 (2021): 123–44. https://doi.org/10.1146/annurev-biodatasci-092820-114757.

[24] Verheij, Robert A., Vasa Curcin, Brendan C. Delaney, and Mark M. McGilchrist. "Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse." *Journal of Medical Internet Research* 20, no. 5 (2018): e185. https://doi.org/10.2196/jmir.9134.

[25] Balagopalan, Aparna, David Madras, David H. Yang, Dylan Hadfield-Menell, Gillian K. Hadfield, and Marzyeh Ghassemi. "Judging Facts, Judging Norms: Training Machine Learning Models to Judge Humans Requires a Modified Approach to Labeling Data." *Science Advances* 9, no. 19 (2023): eabq0701. https://doi.org/10.1126/sciadv.abq0701.

[26] GAO, *Artificial Intelligence in Health Care*.

[27] Tovino, Stacey A. "Artificial Intelligence and the HIPAA Privacy Rule: A Primer." *Houston Journal of Health Law & Policy* 24, no. 1 (2025): 77–126.

[28] GAO, *Artificial Intelligence in Health Care*; Ng, Isaac KS. "Informed Consent in Clinical Practice: Old Problems, New Challenges." *Journal of the Royal College of Physicians of Edinburgh* 54, no. 2 (2024): 153–58. https://doi.org/10.1177/14782715241247087.

[29] Cohen, I. Glenn, and Michelle M. Mello. "HIPAA and Protecting Health Information in the 21st Century." *JAMA* 320, no. 3 (2018): 231–32. https://doi.org/10.1001/jama.2018.5630.

[30] Robeznieks, Andis. "The States Are Stepping up on Health AI Regulation." *American Medical Association*, June 9, 2025. https://www.ama-assn.org/practice-management/digital-health/states-are-stepping-health-ai-regulation.

[31] Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson II Price. "Humans in the Loop." *Vanderbilt Law Review* 76 (2023): 429. https://heinonline.org/HOL/P?h=hein.journals/vanlr76&i=447.

[32] Yu, et al., "Medical Artificial Intelligence and Human Values."

[33] Manzini, Arianna, Geoff Keeling, Nahema Marchal, Kevin R. McKee, Verena Rieser, and Iason Gabriel. "Should Users Trust Advanced AI Assistants? Justified Trust As a Function of Competence and Alignment." Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (New York, NY, USA), FAccT '24, Association for Computing Machinery, June 5, 2024, 1174–86. https://doi.org/10.1145/3630106.3658964.

[34] Crootof et al., "Humans in the Loop"; Goddard, Kate, Abdul Roudsari, and Jeremy C. Wyatt. "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators." *Journal of the American Medical Informatics Association: JAMIA* 19, no. 1 (2012): 121–27. https://doi.org/10.1136/amiajnl-2011-000089.

[35] Adam, Hammaad, Aparna Balagopalan, Emily Alsentzer, Fotini Christia, and Marzyeh Ghassemi. "Just Following AI Orders: When Unbiased People Are Influenced By Biased AI." Paper presented at Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022. November 21, 2022. https://openreview.net/forum?id=ISzWXSWiL8.

[36] Ghassemi, Marzyeh, Luke Oakden-Rayner, and Andrew L. Beam. "The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care." The Lancet Digital Health 3, no. 11 (2021): e745–50. https://doi.org/10.1016/S2589-7500(21)00208-9.

[37] Crootof, et al., "Humans in the Loop."

[38] Salloch and Eriksen, "What Are Humans Doing in the Loop?"

[39] Robinson, David G. *Voices in the Code: A Story About People, Their Values, and the Algorithm They Made*. Russell Sage Foundation, 2022.

[40] Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, et al. "Power to the People? Opportunities and Challenges for Participatory AI." Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (New York, NY, USA), EAAMO '22, Association for Computing Machinery, October 17, 2022, 1–8. https://doi.org/10.1145/3551624.3555290.

[41] Biddle, Michele S. Y., Andy Gibson, and David Evans. "Attitudes and Approaches to Patient and Public Involvement across Europe: A Systematic Review." Health & Social Care in the Community 29, no. 1 (2021): 18–27. https://doi.org/10.1111/hsc.13111.

[42] Funk, Alec Tyson, Giancarlo Pasquini, Alison Spencer and Cary. "60% of Americans Would Be Uncomfortable With Provider Relying on AI in Their Own Health Care." Pew Research Center, February 22, 2023. https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/.

[43] Alanoca, Sacha, Shira Gur-Arieh, Tom Zick, and Kevin Klyman. "Comparing Apples to Oranges: A Taxonomy for Navigating the Global Landscape of AI Regulation." Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, June 23, 2025, 914–37. https://doi.org/10.1145/3715275.3732059.

[44] 15 U.S.C. §9401(3). https://www.govinfo.gov/content/pkg/USCODE-2020-title15/html/USCODE-2020-title15-chap119.htm.

[45] HealthIT.Gov. "Clinical Decision Support." Accessed September 23, 2025. https://www.healthit.gov/topic/safety/clinical-decision-support.

[46] IBM. "What Is Deep Learning?" June 17, 2024. https://www.ibm.com/think/topics/deep-learning.

[47] Science Direct. "Deterministic Model - an Overview." Accessed September 23, 2025. https://www.sciencedirect.com/topics/computer-science/deterministic-model.

[48] Centers for Medicare & Medicaid Services. "Electronic Health Records." Accessed September 23, 2025. https://www.cms.gov/priorities/key-initiatives/e-health/records.

[49] IBM. "What Are Foundation Models?" October 11, 2024. https://www.ibm.com/think/topics/foundation-models.

[50] IBM. "What Is Generative AI?" February 9, 2021. https://research.ibm.com/blog/what-is-generative-AI.

[51] Amazon Web Services, Inc. "What Is LLM? - Large Language Models Explained - AWS." Accessed September 23, 2025. https://aws.amazon.com/what-is/large-language-model/.

[52] MIT Sloan. "Machine Learning, Explained." April 21, 2021. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

[53] IBM. "What Is NLP (Natural Language Processing)?" August 11, 2024. https://www.ibm.com/think/topics/natural-language-processing.

[54] National Library of Medicine. "Neural Networks." Accessed September 23, 2025. https://www.nnlm.gov/guides/data-glossary/neural-networks.

[55] Cleveland Clinic. "What Is Precision Medicine?" Accessed September 22, 2025. https://my.clevelandclinic.org/health/articles/precision-medicine.

[56] ScienceDirect. "Predictive Model - an Overview." Accessed September 23, 2025. https://www.sciencedirect.com/topics/computer-science/predictive-model.

[57] ScienceDirect. "Probabilistic Model - an Overview." Accessed September 23, 2025. https://www.sciencedirect.com/topics/materials-science/probabilistic-model.

[58] Intel. "Learn How Artificial Intelligence (AI) Is Changing Robotics." Accessed September 23, 2025. https://www.intel.com/content/www/us/en/learn/artificial-intelligence-robotics.html.

[59] GeeksforGeeks. "Rule-Based System in AI." 18:36:08+00:00. https://www.geeksforgeeks.org/artificial-intelligence/rule-based-system-in-ai/.

[60] Center for Devices and Radiological Health. "Software as a Medical Device (SaMD)." FDA, FDA, July 23, 2025. https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd.

[61] Integrated Computer Solution (ICS). "SaMD vs. SiMD: Do You Know the Difference?" February 7, 2024. https://www.ics.com/blog/samd-vs-simd-do-you-know-difference.

[62] IBM. "Structured vs. Unstructured Data: What's the Difference?" February 7, 2025. https://www.ibm.com/think/topics/structured-vs-unstructured-data.

[63] Palissery, Gates. *Artificial Intelligence: Policy and Practice*. Joint Commission on Technology & Science, 2024. https://dls.virginia.gov/commissions/jcots/materials/2024_ai_report.pdf.

[64] IBM. "What Is Training Data?" May 2, 2025. https://www.ibm.com/think/topics/training-data.

[65] IBM, "Structured vs. Unstructured Data."

[66] Palissery, *Artificial Intelligence*.

[67] Wells, et al., *Artificial Intelligence (AI) in Health Care*.

[68] Executive office of the President of the United States. *Winning the Race: America's AI Action Plan*. The White House, 2025. http://whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[69] Office of the National Coordinator for Health IT. *Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final Rule*. 2023–28857. January 2024. https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program.

[70] "Delivering on the Promise of AI to Improve Health Outcomes." The White House, December 14, 2023. https://bidenwhitehouse.archives.gov/briefing-room/blog/2023/12/14/delivering-on-the-promise-of-ai-to-improve-health-outcomes/.

[71] Wells, et al., *Artificial Intelligence (AI) in Health Care*.

[72] *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback*. U.S. Food and Drug Administration, 2019. https://www.fda.gov/media/122535/download?attachment.

[73] *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan*. U.S. Food and Drug Administration, 2021. https://www.fda.gov/media/145022/download?attachment.

[74] U.S. Food and Drug Administration. *Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions*. FDA-2022-D-2628. August 2025. https://www.fda.gov/media/166704/download.

[75] Center for Devices and Radiological Health. "Artificial Intelligence-Enabled Medical Devices." U.S. Food and Drug Administration, FDA, July 10, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices.

[76] Centers for Medicare and Medicaid Services. "Frequently Asked Questions Related to Coverage Criteria and Utilization Management Requirements in CMS Final Rule (CMS-4201-F)." U.S. Department of Health and Human Services, February 6, 2024. https://www.aha.org/system/files/media/file/2024/02/faqs-related-to-coverage-criteria-and-utilization-management-requirements-in-cms-final-rule-cms-4201-f.pdf.

[77] Centers for Medicare and Medicaid Services. "WISeR (Wasteful and Inappropriate Service Reduction) Model." September 2, 2025. https://www.cms.gov/priorities/innovation/innovation-models/wiser.

[78] "Nondiscrimination in Health Programs and Activities," Federal Register 89, no. 89 (May 6, 2024): 37522-37810, accessed September 25, 2025, https://www.federalregister.gov/documents/2024/05/06/2024-08711/nondiscrimination-in-health-programs-and-activities.

[79] Manatt Health. "Health AI Policy Tracker." Accessed July 21, 2025. https://www.manatt.com/insights/newsletters/health-highlights/manatt-health-health-ai-policy-tracker; legislation searches also conducted using BillTrack50.

[80] Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. National Institute of Standards and Technology (U.S.), 2023. https://doi.org/10.6028/NIST.AI.100-1.

[81] Coalition for Health AI (CHAI). *Responsible AI Guide*. CHAI, 2024. https://www.chai.org/workgroup/responsible-ai/responsible-ai-guide-raig-and-raig-executive-summary.

[82] National Academies of Sciences, Engineering, and Medicine. *Health Care Artificial Intelligence Code of Conduct*. National Academy of Medicine, 2025. https://nam.edu/our-work/programs/leadership-consortium/health-care-artificial-intelligence-code-of-conduct/.

[83] American Medical Association (AMA). *Augmented Intelligence Development, Deployment, and Use in Health Care*. American Medical Association, 2024. https://www.ama-assn.org/system/files/ama-ai-principles.pdf.

[84] Health Canada, U.S. Food and Drug Administration (FDA), and UK Medicines and Healthcare Products Regulatory Authority. *Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles*. 2021. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles.

[85] International Medical Device Regulators Forum (IMDRF). *Good Machine Learning Practice for Medical Device Development: Guiding Principles*. IMDRF/AIML WG/N88. 2025. https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles.

[86] Wells, Brian J., Hieu M. Nguyen, Andrew McWilliams, et al. "A Practical Framework for Appropriate Implementation and Review of Artificial Intelligence (FAIR-AI) in Healthcare." *Npj Digital Medicine* 8, no. 1 (2025): 514. https://doi.org/10.1038/s41746-025-01900-y.

[87] Callahan, Alison, Duncan McElfresh, Juan M. Banda, et al. "Standing on FURM Ground: A Framework for Evaluating Fair, Useful, and Reliable AI Models in Health Care Systems." *NEJM Catalyst* 5, no. 10 (2024): CAT.24.0131. https://doi.org/10.1056/CAT.24.0131.

[88] GAO, *Artificial Intelligence in Health Care*.

[89] GAO, *Artificial Intelligence in Health Care*.

[90] Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson II Price. "Humans in the Loop." *Vanderbilt Law Review* 76 (2023): 429.

[91] Wells, et al. *Artificial Intelligence (AI) in Health Care.*

[92] GAO. *Artificial Intelligence in Health Care*.

[93] Ng, Isaac KS. "Informed Consent in Clinical Practice: Old Problems, New Challenges." *Journal of the Royal College of Physicians of Edinburgh* 54, no. 2 (2024): 153–58. https://doi.org/10.1177/14782715241247087; Tovino, Stacey A. "Artificial Intelligence and the HIPAA Privacy Rule: A Primer." *Houston Journal of Health Law & Policy* 24, no. 1 (2025): 77–126. https://digitalcommons.law.ou.edu/cgi/viewcontent.cgi?article=1648&context=fac_articles.

[94] Saraswat, Deepti, Pronaya Bhattacharya, Ashwin Verma, et al. "Explainable AI for Healthcare 5.0: Opportunities and Challenges." *IEEE Access* 10 (2022): 84486–517. https://doi.org/10.1109/ACCESS.2022.3197671.

[95] Zhou, Karen, and Ginny Gattinger. "The Evolving Regulatory Paradigm of AI in MedTech: A Review of Perspectives and Where We Are Today." *Therapeutic Innovation & Regulatory Science* 58, no. 3 (2024): 456–64. https://doi.org/10.1007/s43441-024-00628-3.

[96] GAO, *Artificial Intelligence in Health Care*; Topol, "High-Performance Medicine."

[97] Guha, Neel, Christie Lawrence, Lindsey A. Gailmard, et al. "AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing." *George Washington Law Review*

92, no. 6 (2023). https://www.gwlr.org/ai-regulation-has-its-own-alignment-problem-the-technical-and-institutional-feasibility-of-disclosure-registration-licensing-and-auditing/.

[98] Cestonaro, Clara, Arianna Delicati, Beatrice Marcante, Luciana Caenazzo, and Pamela Tozzo. "Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review." Frontiers in Medicine 10 (November 2023): 1305756. https://doi.org/10.3389/fmed.2023.1305756.

[99] Cestonaro, et al., "Defining Medical Liability."; Price, W. Nicholson, Sara Gerke, and I. Glenn Cohen. "Liability for Use of Artificial Intelligence in Medicine." *Law & Economics Working Papers*, January 1, 2022. https://repository.law.umich.edu/law_econ_current/241.

[100] Cestonaro, et al., "Defining Medical Liability."; Pantanowitz, Liron, Matthew Hanna, Joshua Pantanowitz, et al. "Regulatory Aspects of Artificial Intelligence and Machine Learning." *Modern Pathology* 37, no. 12 (2024): 100609. https://doi.org/10.1016/j.modpat.2024.100609.

[101] GAO, *Artificial Intelligence in Health Care*; Wells, et al. *Artificial Intelligence (AI) in Health Care.*

[102] Crootof et al., "Humans in the Loop."

[103] Froomkin, A. Michael, Ian Kerr, and Joelle Pineau. When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced over-Reliance on Machine Learning. 2025. https://www.elgaronline.com/edcollchap/book/9781800887305/chapter7.xml; Price, W. Nicholson, Sara Gerke, and I. Glenn Cohen. "Liability for Use of Artificial Intelligence in Medicine." Law & Economics Working Papers, January 1, 2022. https://repository.law.umich.edu/law_econ_current/241.

[104] Moy, Sally, Mona Irannejad, Stephanie Jeanneret Manning, et al. "Patient Perspectives on the Use of Artificial Intelligence in Health Care: A Scoping Review." *Journal of Patient-Centered Research and Reviews* 11, no. 1 (2024): 51–62. https://doi.org/10.17294/2330-0698.2029.

[105] Lorenzini, Giorgia, David Martin Shaw, Laura Arbelaez Ossa, and Bernice Simone Elger. "Machine Learning Applications in Healthcare and the Role of Informed Consent: Ethical and Practical Considerations." *Clinical Ethics* 18, no. 4 (2023): 451–56. https://doi.org/10.1177/14777509221094476.

[106] Tovino, Stacey A. "Artificial Intelligence and the HIPAA Privacy Rule: A Primer." Houston Journal of Health Law & Policy 24, no. 1 (2025): 77–126; Zhou, Karen, and Ginny Gattinger. "The Evolving Regulatory Paradigm of AI in MedTech: A Review of Perspectives and Where We Are Today." Therapeutic Innovation & Regulatory Science 58, no. 3 (2024): 456–64. https://doi.org/10.1007/s43441-024-00628-3.

[107] Charmaz, Kathy. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Introducing Qualitative Methods. SAGE Publications Ltd, 2006.

[108] Keane, Pearse A., and Eric J. Topol. "With an Eye to AI and Autonomous Diagnosis." *Nature*, NPJ Digital Medicine, no. 40 (August 2018). https://www.nature.com/articles/s41746-018-0048-y.

[109] Goodson, David Alexander, Brittany Garcia, Michael Hogarth, and Shin-Ping Tu. "Artificial Intelligence and Physician Burnout: A Productivity Paradox." Learning Health Systems n/a, no. n/a (n.d.): e70013. https://doi.org/10.1002/lrh2.70013.

[110] "Inova Partners with Abridge After Competitive H2H Evaluation." *Abridge*, February 2025. https://www.abridge.com/press-release/inova-health-abridge.

[111] Evans, Joe. "We Established AI Guardrails for Our Health System. Why You Should, Too." Chief Healthcare Executive, June 11, 2025. https://www.chiefhealthcareexecutive.com/view/we-established-ai-guardrails-for-our-health-

system-why-you-should-too-viewpoint.

[112] Hauser, Oliver P., Miriam Light, Lizzie Shelmerdine, and Jack Blumenau. "Why Evaluating the Impact of AI Needs to Start Now." *Nature* 643, no. 8073 (2025): 910–12. https://doi.org/10.1038/d41586-025-02266-7.

[113] Madiega, Tambiama, and Anne Louise Van De Pol. Artificial Intelligence Act and Regulatory Sandboxes. PE 733.544. European Parliamentary Research Service, 2022. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.

[114] GOV.UK. "AI Airlock: The Regulatory Sandbox for AIaMD." July 17, 2025. https://www.gov.uk/government/collections/ai-airlock-the-regulatory-sandbox-for-aiamd.

[115] Verify Foundation and Infocomm Media Development Authority. Model Al Governance Framework for Generative Al: Fostering a Trusted Ecosystem. Singapore, 2024. https://aiverifyfoundation.sg/wp-content/uploads/2024/05/Model-AI-Governance-Framework-for-Generative-AI-May-2024-1-1.pdf.

[116] OECD. *Regulatory Impact Assessment*. OECD Best Practice Principles for Regulatory Policy. OECD Publishing, Paris, 2020. https://doi.org/10.1787/7a9638cb-en.

[117] "Guidance on the Impact Evaluation of AI Interventions." GOV.UK, July 9, 2025. https://www.gov.uk/government/publications/the-magenta-book/guidance-on-the-impact-evaluation-of-ai-interventions-html.

[118] "Guidelines on the Scope of Obligations for Providers of General-Purpose AI Models under the AI Act." European Commission, July 18, 2025. https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act.

[119] For examples and discussion of participatory mechanisms, see: Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, et al. "Power to the People? Opportunities and Challenges for Participatory AI." *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (New York, NY, USA), EAAMO '22, Association for Computing Machinery, October 17, 2022, 1–8. https://doi.org/10.1145/3551624.3555290; Kellmeyer, Philipp. "Chapter Fourteen - Beyond Participation: Towards a Community-Led Approach to Value Alignment of AI in Medicine." In *Developments in Neuroethics and Bioethics*, edited by Marcello Ienca and Georg Starke, vol. 7. Brains and Machines: Towards a Unified Ethics of AI and Neuroscience. Academic Press, 2024. https://doi.org/10.1016/bs.dnb.2024.02.011; Madiega, Tambiama, and Anne Louise Van De Pol. *Artificial Intelligence Act and Regulatory Sandboxes*. PE 733.544. European Parliamentary Research Service, 2022. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.