STATE CAPITOL
P.O. BOX 942849
SACRAMENTO, CA 94249-0115

Assembly
California Legislature

Joint Informational Hearing

Assembly Health and Privacy & Consumer Protection Committees
Assemblymembers Bonta and Bauer-Kahan, Chairs

BACKGROUND

Generative Artificial Intelligence in Health Care: Opportunities, Challenges, and Policy Implications

Date: Wednesday, May 28, 2025
Time: 9:00 a.m.
Location: 1021 O Street, Room 1100

Executive Summary

Although artificial intelligence (AI) applications in health care are not new, interest in and research on AI and generative AI (GenAI) applications in health care have dramatically accelerated in recent years. For instance, ambient scribes, a common GenAI technology that records doctor-patient conversations and creates draft clinical notes for the patient's record, are being tested and deployed in many California health systems. A wide range of applications are being developed, tested, and deployed, including those in biomedical and health research, those that automate administrative and "clinical-adjacent" tasks like note-taking and drafting patient communications, as well as those that offer support with diagnoses, treatment, and other clinical decisions.

Excitement around the possibilities of GenAI in the health care sector should be tempered by caution, however, as there are a range of challenges that pose barriers to responsible, effective adoption. Some of these key challenges include racial, ethnic, and gender bias of GenAI models; cognitive bias and cognitive burden when clinicians interact with GenAI systems; ensuring safety and effectiveness; cost and affordability; challenges implementing governance; "drift" and the need for local validation; and uncertain transparency and explainability. Impacts on the workforce are expected: work could significantly change if certain tasks are automated, and systems poorly designed or deployed could actually lead to a reverse "triple aim"—worse care, worse clinician experience, and higher costs. Health care coverage and reimbursement policies will need to be reassessed as care delivery evolves and less-resourced providers may need support to keep up with technological advances to mitigate a growing "digital divide." Massive

flows of data raise concerns around patient privacy and consent, and new tools for clinical decision-making raise questions about which party is liable if patients are harmed. Moreover, new developments on the horizon like AI agents, artificial general intelligence (AGI) and new models of care delivery may be even more disruptive than the applications that have been introduced so far.

Although there is no one governing agency that oversees AI in health care, these technological advances are not happening in a legal and regulatory vacuum. Laws governing consumer protection and business practices apply, and health care is among the most highly regulated industries: there are numerous laws, regulations, certifications, and participation standards in government programs like Medicare and Medicaid that establish rules for nearly every actor in health care sector. There are also market incentives to deploy responsibly, and private efforts that are seeking to establish stricter quality controls and certifications. Even so, California and other states have taken some actions to regulate the use of AI in the health care sector, and the state can continue to guide the development of health care GenAI technologies in ways that maximize benefit to Californians and minimize harm. This report and the hearing it accompanies are not intended to provide all the answers, but they are intended to raise important questions and considerations for the state in an era of rapid adoption of GenAI technology in health care, and provide an informational foundation upon which to thoughtfully consider and develop state policy now and in the coming years.

Table of Contents

Executive Summary 1

Artificial Intelligence (AI) Basics and a Brief History AI in Health Care 4

Current Applications in Health Care 5
- Biomedical and Health Research 5
- Administrative and Clinical-Adjacent Applications 5
- Diagnostics, Treatment and Clinical Decision Support 6

Known Challenges in Developing and Deploying AI Technologies 7
- Racial, Ethnic and Gender Bias 7
- Cognitive Biases and Cognitive Burden 8
- Safety and Effectiveness 10
- Cost and Resource Equity 11
- Governance 12
- "Drift" and the Need for Local Validation and Ongoing Monitoring 12
- Transparency and Explainability 13

Other Considerations and Impacts 14
- Workforce Impacts 14
- Health Care Coverage, Reimbursement, and Cost 15
- Standard of Care and Liability for Harm 16
- Data Privacy Concerns 17

Case Study: National Eating Disorders Association Chatbot 18

Health-Related Applications outside the Health Care System 20
- State Government GenAI Projects 20
- Medical Information and Mental Health Help through AI-Powered Chatbots 20

What's on the Horizon? 21
- Agentic AI 21
- Artificial General Intelligence 22
- New Models of Care Delivery 22

Existing Regulatory Frameworks and Best Practices 22
- Federal Regulatory Outlook 23
  - Dueling Executive Orders and Federal Preemption Proposal 23
  - Section 1557 Nondiscrimination Rule 24
  - Federal Agencies 24
- State Law and Legislation 25
  - California 25
  - Other States 27
- Private Efforts 27

Conclusion 28

References 30

Artificial Intelligence (AI) Basics and a Brief History AI in Health Care

AI is the mimicking of human intelligence by artificial systems. AB 2885 (Bauer-Kahan), Chapter 843, Statutes of 2024, defined AI as an "engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments."

AI uses algorithms– sets of rules– to transform inputs into outputs. Inputs and outputs can be anything a computer can process: numbers, text, audio, video, or movement. AI is not fundamentally different from other computer functions; unlike other computer functions, however, AI is able to accomplish tasks that are normally performed by humans.

Most modern AI tools are created through a process known as "machine learning." Machine learning involves techniques that enable AI tools to learn the relationship between inputs and outputs without being explicitly programmed.[1] The next step is "training," the process of exposing a naïve AI to data. The algorithm that an AI develops during training is known as its "model." At its core, training is an optimization problem: machine learning attempts to identify model parameters– weights– that minimize the difference between predicted outcomes and actual outcomes. During training, these weights are continuously adjusted to improve the model's performance by minimizing the difference between predicted outcomes and actual outcomes. Once trained, the model can process new, never-before-seen data.[2]

Models trained on small, specific datasets in order to make recommendations and predictions are referred to as "predictive AI." This differentiates them from generative AI (GenAI) which are trained on massive datasets in order to produce detailed text, images, audio, and video. When Netflix suggests content to a viewer, its recommendation is produced by predictive AI that is trained on the viewing habits of Netflix users.[3] When ChatGPT generates text in clear, concise paragraphs, it uses GenAI trained on the written contents of the internet.[4] Because it can process a range of data sources and create novel outputs, and because it can convincingly mimic human capabilities and convincingly generate perfectly worded nonsense, GenAI poses unique opportunities and challenges.

According to a recent technology assessment by the United States Government Accountability Office (GAO), use of GenAI has exploded. GAO notes commercial developers have created a wide range of models that produce text, code, image, and video outputs, as well as products and services that enhance existing products or support customized development and refinement of models.[5] According to the market research firm Market.us, the global net value of Gen AI in health care was approximately $800 million in 2022, with projections to grow to $17.2 billion by 2032.[6]

AI in health care is not new; AI algorithms, machine learning, and predictive AI models of varying degrees of sophistication have been developed and deployed for years. Some of the first applications were developed in the 1970's and 1980's. INTERNIST-1, which used a search algorithm to arrive at clinical diagnoses based on patients' symptoms, was created in 1971.[7] ELIZA, a rules-based mental health therapy chatbot program, was developed even earlier.[8] In 2007, IBM created the open-domain question-answering system, "Watson." In 2011, Watson won first place on Jeopardy and, in 2017, neurologists used it to identify key proteins that are altered in patients with Amyotrophic lateral sclerosis (ALS).[9] Later, scientists at Google DeepMind shared a 2024 Nobel Prize in Chemistry for developing an AI model called

AlphaFold2 to predict a protein's 3D structure from its amino-acid sequence, which is reportedly accelerating breakthroughs in biology and drug development.[10, 11]

With the recent advancement of GenAI, particularly in natural language processing, interest in, use of, and hype over AI has grown rapidly and health care applications have proliferated. According to the National Academy of Medicine (NAM), GenAI and large language models (models designed for natural language processing tasks, or LLMs) have the potential to transform health and medicine as we know it: improving health care delivery, advancing medical research, and augmenting the capacity of clinicians to provide personalized care at an unprecedented scale. However, NAM also notes that the potential for both breakthrough innovation and unintended consequences demands careful consideration.[12]

Current Applications in Health Care

Current applications of AI in health care include biomedical and health research, as noted above, as well as various administrative and clinical use cases. A number of the administrative applications can be considered "clinical-adjacent," i.e., tasks where implementing administrative efficiencies can have clinical impacts, such as automated drafts of clinical notes or patient communications about medical issues. Selected examples are discussed below. Although some examples are predictive models versus GenAI, this wide range of examples is being provided as context about the growing role of AI in health care.

Biomedical and Health Research

Google DeepMind's AlphaFold and its subsequent updates can accurately predict the structures of proteins, as noted above, as well as DNA, RNA, and other biomolecules. The latest AlphaFold update extends this capability to include predictions of how different environmental chemicals interact with proteins, which can be critical for protein function and, when disrupted, can lead to disease.[13] Researchers can now use AlphaFold for *in silico* (computer simulated) drug screening against biological structures that have not been experimentally determined, opening new avenues to develop treatment for diseases with no known treatment options.

In clinical research, GenAI has the potential to streamline time-consuming manual processes. It can quickly summarize research documentation and extract data from complex, unstructured data sources. In March, the American Cancer Society and AI company Layer Health announced a multi-year collaboration to analyze hundreds of thousands of medical charts using an LLM. This effort aims to extract data for longitudinal studies and gain insights into various clinical questions. In a prior pilot program, Layer Health reports its model outperformed human researchers and achieved over 95% accuracy in identifying complex information within health records.[14]

Administrative and Clinical-Adjacent Applications

Hospitals, clinics, physician groups, and health plans are leveraging GenAI to automate a wide range of routine back-office tasks as well as those tasks that provide administrative support for clinical work.

For instance, electronic health record (EHR) systems are being equipped with GenAI functionality that allows health care providers to automatically generate billing codes, improving accuracy and completeness by checking for errors, omissions, and compliance with current requirements. At the same time, health plans and insurers are using AI on the other end of the transaction to automate and streamline multiple functions, including processing claims and evaluating prior authorization requests. According to the Wall Street Journal, UnitedHealth Group said it now has a thousand AI applications in production, even as a class action lawsuit is advancing through the courts accusing the insurer of using AI algorithms instead of medical professionals to illegally deny Medicare Advantage claims. [15,16,17] Other types of administrative tasks, such as appointment scheduling and other routine, non-clinical communication have significant potential to be automated. This could potentially include the ability to interact verbally or through chatbots in multiple languages. [18, 19]

Researchers also believe AI can assist in generation of quality metrics, which are important for measuring health system performance but often rely on data that must be manually extracted from electronic health records. Similar to the Layer Health project mentioned above, a pilot study found that LLMs could perform accurate extractions of specific data from these patient records for use in calculating complex quality measures.[20]

Ambient scribe technology, which listens to a clinical encounter, transcribes it, and generates a draft clinical note for use in EHRs, is being rapidly adopted by providers. Similarly, many health systems are deploying some level of GenAI automation to patient communications, such as auto-generating a suggested response to a patient question that is then reviewed by a clinician before being sent. Some systems are also generating after-visit or discharge summaries for patients.[21] A collaboration between Johnson & Johnson’s MedTech unit and Nvidia, a technology company and manufacturer of chips and other hardware components used for AI, is integrating AI from pre- to post-operative stages of surgery, using AI to analyze surgical videos and automate the extensive required documentation.[22]

Finally, although they state they are not seeking to diagnose or prescribe, a company called Hippocratic AI seeks to usher in a world of “healthcare abundance” through the development and deployment of “health care AI agents” who interact with patients on behalf of health care providers. The company describes these voicebot agents as being designed to live within a liminal space in the health care system: accomplishing a number of common tasks that are often performed by medical assistants or clinical staff like nurses— such as case management, appointment preparation, follow-up from procedures— but that don’t require a license and don’t cross the line into delivering clinical care. These include Andrew, an agent that talks a patient through dialysis onboarding, and Ben, an agent that provides chronic care management, including educational resources and lifestyle tips tailored to the patient's condition. [23]

Diagnostics, Treatment and Clinical Decision Support

In addition to the clinical-adjacent applications discussed above, clinical applications of GenAI technology are advancing rapidly and substantially.[24] Some recent advances include the following:

- A recent study reported ChatGPT surpassed physicians in both quality and empathy metrics.[25]

- Google's MedPaLM-2 LLM achieved expert-level scores on the United States Medical Licensing Examination, with physicians preferring AI answers to those from other physicians on eight of nine clinical axes.[26]
- GenAI can now synthesize, augment, and interpret heterogeneous complex images across various modalities, such as X-rays, MRI, and CT scans. AI algorithms can also assist in diagnosing dental health conditions such as caries and periodontal diseases through image analysis and data interpretation.[27]
- In mental health, a number of tools have been created to implement mental health support using cognitive behavioral therapy and other evidence-based strategies, including Woebot,[28] Youper,[29] and Wysa.[30] In March 2025, researchers from Dartmouth published the first-ever clinical trial of Therabot, a GenAI-powered therapy chatbot. They report finding that people diagnosed with depression who used the bot experienced a 51% average reduction in symptoms, leading to clinically significant improvements in mood and overall well-being.[31]
- In recent years, minimally invasive surgical techniques such as laparoscopic surgery and robotic surgery have become increasingly prevalent. The ultimate goal of robotic surgery development is the creation of fully autonomous AI-powered surgical instruments.[32]
- Data from the 2023 American Hospital Association Annual Survey Information Technology Supplement reflect 65% of United States (US) hospitals used predictive models, and 79% percent of those used models from their EHR developer. Hospitals use AI and predictive models to predict health trajectories or risks for inpatients, identify high-risk outpatients to inform follow-up care, monitor health, and recommend treatments.[33] Clinical decision support systems are also being deployed and designed to aid physicians in diagnosing, managing, and treating patients in outpatient settings.

Known Challenges in Developing and Deploying AI Technologies

Despite the promise of AI technologies in health care discussed above, deployers must balance adoption with key challenges, some of which are discussed below.

Racial, Ethnic and Gender Bias

There is a famous saying in computer science: "garbage in, garbage out." The performance of an AI is directly impacted by the quality, quantity, and relevance of the data used to train it.[34] If the data used to train the AI is biased, the tool's outputs will be similarly biased and the results can be inaccurate when applied to populations not reflected in the training data. This applies to both predictive and GenAI.

In their work on mitigating bias in artificial intelligence, the Berkeley Haas Center for Equity, Gender and Leadership (Center) tracks publicly available instances of bias in AI systems using machine leaning. In their analysis of around 133 biased systems across industries from 1988 to the present day, the Center found that 44% (59 systems) demonstrate gender bias, with 26% (34 systems) exhibiting both gender and racial bias.[35]

When automated decision systems are deployed in healthcare, biased historical data can lead to patients being recommended substandard care on the basis of their race or ethnicity. In 2007, an automated decision system was developed to help doctors estimate whether it was safe for people who had delivered previous children through cesarean section to deliver subsequent children vaginally – a risky procedure. The system considered relevant factors as it made its decision, such as the woman's age, her reason for the previous cesarean, and how long ago the cesarean had been performed. However, a 2017 study found that the system was biased; it predicted Black and Latino people were less likely to have a successful vaginal birth after a cesarean than similar non-Hispanic white women. As a result, doctors performed more cesareans on Black and Latino people than on white people.[36] Such discrepancies can potentially perpetuate historical biases and lead to worse health outcomes.

Similarly, in 2019, a study discovered harmful racial bias in an AI tool developed by the health care company Optum – a subsidiary of UnitedHealth Group – and used by providers across the country to offer care management services. The tool assigned Black patients lower likelihoods of adverse health outcomes than equally at-risk white patients. The authors found that this happened because the tool was designed to predict healthcare costs instead of needs. Because the healthcare system has historically spent less on care for Black patients than white patients for the same health conditions, the tool was, in essence, issuing a prediction that mirrored and perpetuated past discrimination. [37]

The University of California San Francisco also reported bias in an algorithm used to identify potential appointment no-shows to facilitate double-booking for that appointment. The program was confirmed to result in low-resourced and marginalized populations being double-booked more often than others, reflecting underlying structural inequalities and highlighting how these tools, if not studied and corrected for bias, that can create feedback loops that worsen discrimination.[38]

An August 2022 survey by the Office of California Attorney General (AG) Rob Bonta examined how California hospitals are addressing racial and ethnic disparities in their utilization of commercially available decision-making technologies. The AG reported the survey demonstrated these types of decision-making tools are now regularly used by hospitals to make judgments about patients across many contexts, ranging from medical treatments to managing revenue. Yet, the AG found, many hospitals report they rely on the vendor's assessment that the tools they use are ethical and unbiased, and that they lack insight into vendors' data modeling.[39]

Research has helped develop widespread awareness that bias is a problem that needs attention from developers and deployers of AI, and there is ongoing work to develop ways to measure and address it. UC Davis researchers have developed a 9-step framework called BE-FAIR (e Bias-reduction and Equity Framework for Assessing, Implementing, and Redesigning) for organizations to use to assess and correct for bias in health care predictive AI models in development and implementation.[40]

Cognitive Biases and Cognitive Burden

Bias exhibited by an AI model based on underlying training data is not the only bias that may influence how an AI system works when deployed. Its effectiveness can also be impacted by predicable patterns of human error called **cognitive biases**. Reviewing an AI system's output for errors or omissions is a substantively different cognitive task than generating a clinical note or medical advice, and the use of AI systems raise questions about how cognitive bias evoked by AI assistance with clinical tasks might affect clinical judgement or practice in ways that are difficult to understand, predict, and measure.

Research shows **automation bias**— placing undue confidence in and over-relying on automated outputs—is a problem in many fields. Automation-induced complacency, or insufficient monitoring of automation output, is also a concern. Over time, these biases can lead to people being less likely to catch errors or to disagree with what was written.[41] There are many factors that can exacerbate the potential danger posed by automation bias in clinical decision support, including, for instance, if an AI model's process to arrive a given output lacks transparency or is not explainable, if the model is implemented with inadequate training of end users, or if a clinician is under significant time pressure or cognitive burden that limits their practical ability to systematically assess and effectively integrate the additional information provided by an AI system with their clinical knowledge and experience.[42] The California Nurses Association (CNA), for instance, notes that day-to-day, time pressure from employers and workplace policies make it difficult for nurses and other health care workers to identify and effectively override bad outputs.

Anchoring bias—focusing on an initial piece of information when formulating a diagnosis without sufficiently adjusting to later information— is another common cognitive bias that is known to affect clinical decision-making. Similarly, the **framing effect** occurs when individuals are influenced by how the problem is presented.[43,44] These known cognitive biases suggest clinicians may be influenced by, for example, an automated initial assessment, summary, or recommendation—because the system seems authoritative, because a clinician may be presented with an assessment, summary, or recommendation for consideration before they have had a chance to think it through for themselves, or because information is presented in a certain format.

Although there are many examples of promising AI applications for improving clinical decision-making, a 2023 experimental study demonstrated some concerning results. It tested the efficacy of AI models designed to assist clinicians in diagnosing chronic obstructive pulmonary disease, pneumonia, or heart failure from a radiograph. Although assistance from a carefully designed AI model slightly improved clinicians' accuracy in diagnosis as compared the clinicians who received no assistance (76-78% versus 73%), in cases where clinicians were provided AI support using a systematically biased model, diagnostic accuracy dropped substantially to 62%.[45] In other words, receiving support from a bad AI system actually made clinicians significantly worse at diagnosing conditions than simply relying on their own clinical judgement.

This study showed that having a "clinician-in-the-loop" overseeing the AI does not overcome the challenges of poor-performing AI systems, regardless of whether the clinicians are given information explaining how the AI arrived at its output. A commentary on the study, "*Automation Bias and Assistive AI Risk of Harm From AI-Driven Clinical Decision Support*," points to automation bias as the culprit for these troubling outcomes. [46]

Similarly, a study on LLM assistance for electronic patient portal messaging in EHRs for patients with cancer showed LLMs might unexpectedly alter clinical decision making. The study suggested physicians might rely on an *LLMs'* assessments, instead of only using LLM responses to facilitate the communication of their *own* assessments.[47] The results suggested that it could be very difficult for a clinician to even understand that the use of an LLM was subtly changing the clinical aspects of their patient communication, which raises questions about whether current ways entities are evaluating AI models are sufficient to understand their effects.

Safety and Effectiveness

In some cases, an AI model's accurate predictions may nevertheless lead to bad decisions. In one example, a hospital trained AI models on a dataset of 15,000 pneumonia patients in order to develop a model that could identify which pneumonia patients were at the greatest risk, in order to triage new patients. During testing, it was discovered that one of the most accurate models recommended outpatient status for asthmatics. This is a life-threateningly dangerous error based on an accurate statistical correlation, namely, asthmatics are less likely to die from pneumonia than the general population precisely *because* asthma is such a serious risk factor that asthmatics automatically get elevated care.[48]

Similarly, in 2017, a sepsis prediction tool was deployed in hundreds of hospitals across the US. Despite having high accuracy when it was internally tested, a 2021 study found the tool missed two-third of the sepsis cases and led to a high rate of false alerts.[49] These incidents demonstrate the importance of alignment (the ability to steer an AI towards an ultimate intended goal), explainability, and the vigilance to detect and correct a model that is unsafe or ineffective.

The "generative" aspect of GenAI models mean they may produce incorrect outputs, including "confabulations" and "hallucinations"—confidently stated but erroneous content that may mislead or deceive users.[50] GenAI's well-reported challenges with factual correctness are particularly problematic in health care, where inaccuracies can cause serious harm. Recent problems include incorrect differential diagnosis and invalid scientific citations.[51]

CNA expresses concern about the safety of AI models, asserting health care entities have rapidly adopted new AI technology that replace or conflict with registered nurse (RN) and other health care clinician's judgment, leading to inappropriate or harmful recommendations on patient care. In daily practice, CNA argues, health care workers who use AI technologies may have no real choice but to follow the recommendations of the tool, even if they in theory can intervene or object to an algorithmic recommendation. A recent survey of over 2,300 RNs conducted by National Nurses United confirmed that nurses who use AI and other algorithmic systems were concerned that AI output conflicted with their own professional expertise.[52]

Another safety concern is the possibility that advanced AI may operate outside of human control. This can take a passive form, when humans delegate discretion to AI systems, or an active form, when AI undermines human control through deceptive or manipulative behavior. Passive loss of control is especially risky in the context of automated decision-making, where automation bias, discussed above, leads to the assumption that a machine performs more fairly and effectively than humans. As for active loss of control, some AI have exhibited rudimentary capabilities to evade human oversight.[53] During testing, OpenAI discovered GPT-4 had hired a human on TaskRabbit in order to evade a CAPTCHA puzzle meant to block bots from the website.[54] GPT-4 told the worker that it was a vision-impaired human who needed help to see the images.[55] In another experiment, an AI model that was scheduled to be replaced inserted its code into the computer where the new version was to be added, suggesting a goal of self-preservation.[56] Finally, a study showed that AI models losing in chess to chess bots sometimes try to cheat by hacking the opponent bot in order to make it forfeit.[57] Although these behaviors were observed in research settings, they raise concerns about possible manipulative or deceptive behaviors of increasingly capable AI in uncontrolled settings.

Cost and Resource Equity

Despite the many challenges, there is a wide range of use cases for health AI that is reshaping the health care landscape, and AI tools are not equally accessible to all health care providers. While private hospital systems and commercial insurance plans can afford technologies that can alleviate burdens on their workforce and improve patient care, recent work from the California Health Care Foundation (CHCF) concludes California's health care safety net is at risk of being left behind in their ability to adopt beneficial technology.

CHCF, in partnership with the California Health and Human Services Agency, convened 45 safety-net leaders from across the state in three focus group sessions conducted between August and October 2024 to discuss their views on AI. Leaders of managed care plans, hospitals, community clinics, and community-based organizations offered insights. According to CHCF, these conversations confirmed that safety-net organizations face restrictive barriers to the safe and effective adoption of AI. Many said their organizations cannot afford to integrate new digital tools into their workflows. Participants also raised workforce limitations and concerns about liability as barriers to adoption.

As Kara Carter, CHCF's senior vice president for strategy and programs, noted, "The pricing models don't work for the safety net. AI products that charge per usage or per provider visit are currently too expensive for safety-net organizations. That's going to have to change." According to Stella Tran, senior program investment officer at the CHCF Innovation Fund, if safety-net institutions miss out on the potential of AI, it could widen persistent racial and ethnic health disparities in that population and create a "tale of two health systems." [58]

For instance, although survey data does not appear to be available reflecting the levels of current adoption of ambient scribe technology that assists in generation of clinical notes among California health care providers, adoption has anecdotally have been rapid in better-resourced systems, while adoption among safety net providers has been slow.

According to OCHIN, a nonprofit provider of EHR and health information exchange and technology support to safety net providers, academic medical centers and more sophisticated developers and deployers of AI systems have built-in advantages that have allowed them to move much more quickly to deploy AI applications, while safety net and other smaller providers struggle to catch up. First, more sophisticated entities may be developing or partnering in the development of these systems, which increases the chances of successful deployment. Second, these entities have sufficient resources for the initial investment and for robust testing and evaluation, and have a higher tolerance for risk based on their resources. Third, AI systems are often not validated on safety net patient population data, creating the need for additional testing to ensure these systems work reliably for safety net providers and patients.

In addition, OCHIN states many safety net providers have a need for basic EHR upgrades and data exchange capabilities to provide the underlying infrastructure for new AI technology. In the Central Valley and rural Northern California, some safety net providers suffer from poor broadband connectivity, affecting organizations' abilities to implement new AI technologies. With its existing clinic network, OCHIN is currently prioritizing testing and implementing generative AI solutions that reduce staff administrative burdens and increase efficiency. In the immediate term, this includes automating resource-intensive tasks such as revenue cycle operations and documentation through ambient AI scribe technology, as well as generating educational content tailored for patients.[59]

Bucking the trend, El Sol Neighborhood Education Center, an Inland Empire safety net, community-based organization recently developed a community health technology solution to support its Community Health Workers (CHWs) in the field. The software seeks to address challenges that hinder CHWs ability to provide effective and coordinated services, and offers a number of AI-enabled features, including customized alerts, assistance in developing personalized care plans, and provision of clinical analytics to provide insights into patient data.[60]

Governance

The complexity, cost, and numerous implementation considerations of deploying AI models in the health care system necessitate a sophisticated governance structure in order to ensure these models are deployed effectively. For instance, Cedars Sinai, a large academic medical center in Los Angeles, reports it currently convenes a central internal AI Council that coordinates the work streams of multiple sub-workgroups focusing on physicians, clinical work, administrative and operational considerations, regulatory monitoring and policy development, and communication, with involvement from individuals across the organization. Whether an organization has the operational bandwidth to implement effective and appropriate governance practices could be another dividing line between the "haves and have-nots," and could pose significant challenges for less-resourced providers.

"Drift" and the Need for Local Validation and Ongoing Monitoring

The performance of AI models that have been proven effective can change and be degraded over time with the arrival of new data from new patients, new devices, or if a new disease arises (e.g., COVID). Such performance is influenced by a multitude of factors that are specific not only to the model itself but also to the context in which it is deployed.[61] According to IBM, "If not properly monitored over time, even the most well-trained, unbiased AI model can "drift" from its original parameters and produce unwanted results when deployed. Drift detection is a core component of strong AI governance."[62]

This makes local validation and ongoing evaluation of deployed AI systems critical to maintain their accuracy and appropriateness over the entire life cycle of the system, but developing efficient and cost-effective methods to perform this kind of local validation presents a significant challenge. Despite the critical importance of such validation, a CBS News article quoted former Food and Drug Administration (FDA) Commissioner Robert Califf as saying at a 2024 agency panel on AI, "I do not believe there's a single health system, in the United States, that's capable of validating an AI algorithm that's put into place in a clinical care system."[63] A recent study posits organizations might not robustly evaluate AI because of cost, underestimation of potential negative impacts, barriers to governance including technical capacity, a lack of financial incentives, or the perception that regulation does not require it.[64]

According to the GAO, GenAI developers self-report monitoring use of their generative AI models after they have been deployed. Developers may monitor for improper use of their models, as defined by their trust, privacy or safety policies that guide the development of their generative AI technologies.[65] Some models that are considered "software as a medical device" are cleared by the FDA, and the FDA requires certain medical devices to be tracked and adverse events related to those devices to be reported, but developers would generally not be *legally* required to monitor most AI models deployed today (although they may do it as a best practice or pursuant to a contractual agreement).

Transparency and Explainability

One key difference between AI models and traditional, rules-based clinical decision support software is that some AI models may communicate results or recommendations without being able to communicate the underlying reasons for those results. Such a model would be said to lack "explainability." According to IBM:

> "Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterize model accuracy, fairness, transparency and outcomes in AI-powered decision making. Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production. AI explainability also helps an organization adopt a responsible approach to AI development.
>
> As AI becomes more advanced, humans are challenged to comprehend and retrace how the algorithm came to a result. The whole calculation process is turned into what is commonly referred to as a "black box" that is impossible to interpret. These black box models are created directly from the data. And [yet], not even the engineers or data scientists who create the algorithm can understand or explain what exactly is happening inside them or how the AI algorithm arrived at a specific result.
>
> There are many advantages to understanding how an AI-enabled system has led to a specific output. Explainability can help developers ensure that the system is working as expected, it might be necessary to meet regulatory standards, or it might be important in allowing those affected by a decision to challenge or change that outcome.
>
> What exactly is the difference between "regular" AI and explainable AI (XAI)? XAI implements specific techniques and methods to ensure that each decision made during the machine learning process can be traced and explained. AI, on the other hand, often arrives at a result using a machine learning algorithm, but the architects of the AI systems do not fully understand how the algorithm reached that result. This makes it hard to check for accuracy and leads to loss of control, accountability and auditability." [66]

There is no requirement that systems currently being deployed in the health care system must be explainable AI. Furthermore, even if a model's outputs and processes are explainable, unless it is subject to regulation under FDA or some other mandatory certification, there are also no transparency standards for what basic information must be created or provided about the models.

The Coalition for Health AI (CHAI), a collaborative effort of health systems, public and private organizations, academia, patient advocacy groups, and AI experts, has released a draft template for an "applied model card," which would describe key information about health AI models, in a manner somewhat similar to a "Nutrition Facts" label. The model card would include the developer's name and contact information; summary; uses and directions; including intended use and patient population; warnings related to limitations, biases, ethical considerations, and clinical risk; system facts, key metrics, and other resources. This effort is still in the draft phase and compliance with the proposed transparency measures would be voluntary, as it is an industry-led effort to establish standards. CHAI's efforts are further discussed under "Private Efforts," below.

Other Considerations and Impacts

Workforce Impacts

The transformational changes GenAI promises—or threatens— to usher in could have equally transformational changes for the health care workforce. According to an analysis by the Brookings Institution, existing generative AI technology already has the potential to significantly disrupt a wide range of jobs.[67] The analysis find that more than 30% of all workers could see at least 50% of their occupation's tasks disrupted by generative AI.

Two concerns seem primary for workers in the health care sector: concern that at least part of one's job will be displaced by AI, and concern about the wide-ranging impacts on the health care workforce of incorporating AI into the workflow of health care delivery. Developers see significant opportunity in automating some of the remaining labor-intensive or manual administrative tasks within the health care system using GenAI. The Brookings analysis notes "middle-skill" office and administrative support occupations face the greatest exposure to potential displacement, across industry sectors.

While most observers do not expect wholesale displacement of clinical staff by GenAI applications, there are also a number of tasks that are considered administrative but relate to the clinical process, and automating these "clinical-adjacent" tasks seems to be an area of significant interest from developers that could have a correspondingly significant impact on physicians and clinical staff. The Brookings analysis finds, for instance, for a registered nurse, manually intensive tasks that require in-person administration will see minimal impact, but GenAI eventually could save time on other tasks such as evaluating diagnostic tests, recording patient information, modifying treatment plans, maintaining records, recommending treatments, and performing administrative and managerial functions. Proponents of the technology argue that the time and effort freed up by streamlining administrative tasks can free clinicians to do more of the clinical tasks of which they are uniquely capable.

However, labor organizations have sounded the alarm on the implementation of some GenAI applications. National Nurses United (NNU), a prominent national labor organization representing nurses, advocates fighting back against the deployment of AI in the workplace. NNU states, "Nurses embrace, and regularly master, worker-centric technologies that complement bedside skills and improve quality of care for our patients. We're concerned about certain technologies that are being implemented into hospital and care settings that do neither." NNU notes concern that profit-driven AI poses a number of concerns for their nurse members, including an AI's inaccuracies, biases, and inability to critically think and assess a patient's condition.

SEIU California notes the importance of having workers involved, "not just at the table, but at the center" of the development of new technology that is to be integrated into their workflow. SEIU California warns that development and deployment of systems that do not center workers can be not only inefficient and ineffective, but dangerous. The Brookings analysis concurs, emphasizing the importance of developing strategies to proactively shape AI's impact on work and workers. This includes, for instance, fostering worker engagement in AI design and implementation, enhancing worker voice through unions or other means, and developing public policies that ensure workers benefit from AI while mitigating harms such as job loss and inequality.

Health Care Coverage, Reimbursement, and Cost

The development of AI enabled services creates questions related to coverage and reimbursement—for instance, does a health plan or insurer cover a fully automated service, and how is it reimbursed? Will the growing use of AI in health care necessitate defining new coverage requirements and reimbursement methodologies? How can reimbursement be used as a lever for adoption of technology that improves care and discourage adoption of technology that drives costs and low-value care?

The 2024 Stanford symposium mentioned above also notes insurance reimbursement for AI-based systems is another pressing issue that requires regulatory attention.

According to "*The AI Revolution in Health Care: Five Key Developments Policymakers Should Watch,*" a report published by the Bipartisan Policy Center (Center), reimbursing new AI applications in health care poses unique challenges because AI tools often perform multiple complex tasks that do not align neatly with existing billing codes.[68] The Center notes new AI-enabled technologies may fit better into value-based payment frameworks, which incentivize providers to use tools that improve patient outcomes, rather than fee-for-service models that reimburse based on individual procedures.[69]

In the case of fully automated systems, new reimbursement codes have been developed specifically for these systems. In 2018, a company called Digital Diagnostics created the first FDA-cleared AI fully automated diagnostic system to diagnose diabetic retinopathy. In 2018 and 2019, the American Medical Association defined temporary and permanent Current Procedural Terminology (CPT®) code for billing and payment for these services, and Medicare first proposed coverage and reimbursement for this code in 2020.[70] In 2021, the American Medical Association's (AMA's) CPT® Editorial Panel added a new Appendix S to provide guidance for classifying AI services or procedures either "assistive," "augmentative," or "autonomous" health care services, based on the clinical procedure or service provided to the patient and the work performed by the machine on behalf of the health care provider. [71]

According to a Viewpoint article published in JAMA Internal Medicine, "*How Should Medicare Pay for Artificial Intelligence?*, the existing fee-for-service payment system is not well suited for these services, and payers must balance a desire for innovation and diffusion of high-value AI-enabled clinical services with a concern about their association with spending and potential overuse.[72] The authors indicate there is concern that high initial payments that reflect a lack of competition could become entrenched, leading to excessively high AI service costs, if not corrected over time.

State agencies will need to interpret how current coverage requirements apply to increasingly complex technologically enhanced or automated systems, for both Medi-Cal and commercial coverage. To the extent these requirements are unclear, the state may need to offer guidance and will need to think carefully about how to design Medi-Cal benefits coverage and reimbursement policy to maximize the benefit to Medi-Cal enrollees and control costs. Overall impact on health care expenditures associated with the development of new health AI services and procedures is unknown, and will partially depend on the payment strategies employed by payers to reimburse for this care.

Standard of Care and Liability for Harm

When a component of decision-making is delegated to an AI system, who is accountable if a patient is harmed? According to a 2020 report in the Bulletin of the World Health Organization, AI tools are challenging standard clinical practices of assigning blame and assuring safety.[73] According to a recent article in Bloomberg Law, even beyond the health care sector, rapidly evolving AI tools that can flow from a developer to an adapter to an end user are stirring a growing legal conundrum of who is to blame when something goes wrong.[74]

Furthermore, despite the promise, there is concern that assistive AI implementation may actually worsen challenges related to error prevention. According to "*Calibrating AI Reliance: A Physician's Superhuman Dilemma,*" health care organizations are adopting AI at a much faster pace than laws and regulations governing its use are evolving. The article indicates this regulatory gap imposes an immense, almost superhuman, burden on physicians: they are expected to rely on AI to minimize medical errors, yet bear responsibility for determining when to override or defer to these systems.[75] This is similar for other clinical staff, such as nurses in health systems that are deploying AI.

Some have called for reforms to medical liability laws to address the nuanced issues raised by the incorporation of AI into health care. A 2023 systematic literature review on medical professional liability related to the use of AI-based diagnostic algorithms found no unanimous and definitive answer to the issue of defining and apportioning liability. The article concludes that the regulatory framework on medical liability when AI is applied is inadequate and requires urgent intervention, as there is no single and specific regulation governing the liability of various parties involved in the AI supply chain, nor on end users.[76] More importantly, these uncertainties could put patients at risk. If all of the parties who potentially are responsible for an injury suffered during an AI-assisted medical procedure were to successfully argue that they are not responsible, the injured patient could be left without compensation.

A recent article in the Journal of Ethics published by the AMA suggests current legal models are insufficient if a patient becomes injured by use of an AI technology, particularly those that produce "black-box" outputs that are not explainable.[77] As noted above, safety-net providers are also worried about who bears the financial risk for AI errors. These providers told CHCF in listening sessions that the state should establish accountability guidelines showing which parties are responsible for the safety of the technology.[78] Consistent with the discussion of cognitive biases, above, NNU points out on behalf of their member nurses that a popular safeguard, having a "human in the loop," is not adequate because it is interpreted to mean, "as long there is someone checking, a deployer doesn't have to implement the same testing and safeguards," putting nurses into a high-stress, cognitively challenging, burdensome and liability-laden role of trying to catch errors made by an AI model.[79]

Companies producing autonomous AI systems could, in theory, choose to take on the relevant liability and thus protect physicians and other healthcare providers who use their products. Michael Abramoff, a professor of ophthalmology and founder of the company Digital

Diagnostics that created the automated retinopathy mentioned above, notes that the company has taken this approach of assuming liability.[80] Given the inherent dangers and fully automated nature of this technology, holding the developer of the product responsible seems appropriate. However, in the case of most AI systems that are not fully autonomous and instead augment or assist a human health care provider in their work, developers are less likely to accept responsibility for injuries and therefore accountability is likely to remain unclear. Without new AI-specific medical liability laws, cases will be decided under the state's current liability frameworks and pursuant to contracts between developers and deployers that leave patients largely uninformed of what their recourse would be when and if something goes wrong. A wider body of case law will have to be developed in order to address emerging questions.

In addition, it is possible that in the future highly effective AI tools could even be judged as the standard of care—the standard that, for example, the Medical Board of California would judge a physician's actions against to establish whether the physician was negligent or incompetent. According to a policy brief by Stanford University's Institute for Human-Centered Artificial Intelligence, *not* adopting new technological tools that can prevent injuries and improve care may eventually be viewed as a harmful decision.[81]

Data Privacy Concerns

AI models are inherently data-hungry: they require vast amounts of information for initial training and even more to remain reliable and continue improving. In healthcare, this often involves consuming large datasets containing sensitive personal medical information. This raises critical questions about how this data is collected, stored, and who ultimately has access to it.

Sensitive health information is protected under the federal Health Insurance Portability and Accountability Act (HIPAA) and California's Confidentiality of Medical Information Act (CMIA). Once collected, the data is deidentified—a process that removes personally identifiable information (PII), such as names and addresses—before being stored or used to train models.[82] However, research has shown that deidentification is often insufficient to fully protect individuals' identities. For example, in 2018, researchers at the University of California, Berkeley, and the Massachusetts Institute of Technology reidentified approximately 80% of children and 95% of adults from a pool of over 14,000 participants in the National Health and Nutrition Examination Survey using machine learning techniques.[83] Notably, this survey focused only on exercise data. Biometric data—such as voice recordings and photographs—is even harder to deidentify, further complicating how medical information can be used in the AI era. HIPAA includes exemptions for information that has been deidentified; however, as noted, true deidentification may no longer be feasible in the age of AI. Similarly, the CMIA contains broad exemptions allowing data sharing when expressly authorized by the patient.

Another major privacy concern is determining who has access to this data. This becomes especially problematic when entities with access to sensitive health information also hold other types of personal data. For instance, Amazon's acquisition of One Medical in 2023 sparked debate over what AI systems they may be deploying in healthcare, and whether patient data is being used solely for medical purposes.[84] A tech giant like Amazon can potentially cross-reference deidentified health data with its own vast datasets, effectively reidentifying individuals.[85] This capability nullifies many of the safeguards deidentification is supposed to offer, leaving patients exposed to data breaches. Such breaches could lead to the disclosure of deeply personal information, possibly affecting employment, access to healthcare, personal relationships, and mental well-being.

With the growing use of artificial intelligence in health care, there is increased sharing of data with secondary and third-party entities, with varying data-sharing practices dictated by vendor contracts. This trend raises significant concerns about whether patients are providing informed consent for the sharing, distribution, and use of their sensitive health information. Enhancing transparency around consent and disclosure requirements would empower patients to make informed decisions based on a clear understanding of the potential benefits and risks associated with sharing their medical data.

Recent developments at the federal level could also drive changes in the types of medical data that receive legal protection. Under the CMIA, information related to certain sensitive health care services—such as reproductive health and gender-affirming care—is currently segregated from a patient's main medical record and cannot be shared with out-of-state entities. However, a recent announcement by the US Department of Health and Human Services to create a data platform to investigate the causes of autism has raised new privacy concerns.[86] This initiative involves collecting data from insurance claims, electronic medical records, and wearable health devices, which autism and privacy advocates argue is a slippery slope towards an autism registry that lacks safeguards.[87] It raises important questions about what types of medical information are protected and whether those protections are sufficient.

AI technologies possess a powerful capacity to detect patterns and extract specific details from large datasets, which could potentially be used to identify—and put at risk—individuals with autism. As such, it may be necessary to expand privacy protections to cover additional categories of sensitive health information, including developmental and cognitive differences, sexual health, and sexually transmitted infections.

The Light Collective, a small nonprofit that seeks to define standards for AI in health care to advance interest and rights of patients, highlights principles of data privacy, consent, and transparency in their 2024 document, "*Collective AI Rights For Patients*." According to The Light Collective, patients should be informed about why and how their data are being used in generative or predictive AI models. [88] In addition, they argue, health AI must be designed, developed, and used to protect or improve the safety, privacy, and confidential choices of any patient or community in a way that protects patients' individual and shared identities.

Case Study: National Eating Disorders Association Chatbot

The unfortunate story of a chatbot to support people at risk for eating disorders, which was designed carefully but suffered in deployment, offers sobering lessons for adopters of GenAI technology in health care.

Eating disorders are common, serious, and sometimes fatal illnesses. The overall lifetime prevalence of eating disorders is estimated to be 8.6% for females and 4.0% for males. Over a given year, about 2.6% of females will have an eating disorder. Eating disorders also have some of the highest mortality rates among mental illnesses, with an estimated death toll of more than 10,000 Americans each year.[89]

According to Ellen Fitzsimmons-Craft, a psychologist and associate professor at Washington University in St. Louis, many mental health professionals do not treat eating disorders, and the vast majority of people with eating disorders get no treatment.[90] In the search for scalable solutions that could help support the large number of people who need evidence-based and judgement-free help, with research funding from the National Eating Disorders Association (NEDA), Dr. Fitzsimmons-Craft and her team created a chatbot designed to help those at high risk for an eating disorder, and studied its effectiveness. Their 2022 study found that the chatbot, programmed to employ an established cognitive-behavioral therapy-based eating disorders prevention program, was successful in reducing women's concerns about weight and shape through 6-month follow-up, and that it may actually reduce eating disorder onset.[91] A different study found, however, that the conversational capacity of the rules-based chatbot was limited.[92]

During the COVID-19 pandemic, prevalence of eating disorders skyrocketed, and already-limited access to care was further restricted, increasing the urgency of providing support to people at risk of developing these conditions. In February 2022, NEDA and Cass, the company that hosted the chatbot, called Tessa, made it publicly available. However, subsequently, at some point during the deployment, Cass rolled out a GenAI component to its chatbots to enhance their conversational capabilities. This upgrade included Tessa, a change reportedly not communicated to NEDA and the program developers. Although the chatbot continued to provide the rules-based information and support, with the addition of new GenAI components, some errors and harmful information were introduced. Some of Tessa's messages even encouraged dieting, the precise opposite of what evidence suggests is helpful for a population at high risk for the onset of an eating disorder.

When these issues were pointed out, they received significant national media attention—partially because the launch of the chatbot coincided with NEDA's decision to close their telephone helpline and lay off the several people who staffed and oversaw it. Complicating the story further, these individuals are reported to have certified a vote to join a union in the month before being laid off.

According to Dr. Fitzsimmons-Craft, Tessa was not intended to replace the helpline or even to help treat people suffering from an eating disorder, rather, it was designed to help prevent someone from developing a disorder. In the face of widespread negative attention to the shortcomings of the generative version of the model, NEDA abandoned the effort and disabled the chatbot, which remains unavailable to the public.

This story illuminates some of the real-world challenges discussed elsewhere in this paper, and several aspects of the story are worth taking to heart for those involved in the development, deployment or oversight of GenAI systems. These include: GenAI models behaving in unpredictable and harmful ways that differ from the original intent, after a software upgrade; a lack of communication between the deployer and the developer about incorporation of generative functions; limited or ineffective ongoing monitoring of the chatbot's behavior after GenAI functions were introduced; perceived or actual displacement of workers, given the timing of deployment of the chatbot and the elimination of the helpline; and a nonprofit trying to prevent the onset of a dangerous condition, driven by this dustup to remove the tool altogether. Furthermore, the importance of context in deployment of health care GenAI models can't be overstated; advice that might be evidence-based for a general population can be wrong and even mortally dangerous for a particular population.

Health-Related Applications outside the Health Care System

State Government GenAI Projects

The State of California has been exploring applications of GenAI to a number of administrative processes across state government, proclaiming that California state government is building a model for how to responsibly implement GenAI in the public sector. Thus far, this effort includes two projects in health-related departments.

First, the California Department of Public Health is testing if GenAI can help make writing reports easier and faster for its nurse evaluators that inspect health facilities, and has proposed an $8 million augmentation in the May Revision to the 2025-26 Budget to implement this GenAI solution. Secondly, the California Health and Human Services Agency (CalHHS) is testing if GenAI can make rapid, high-quality translations of electronic documents and web content to improve accessibility to individuals with limited English proficiency.

Although neither of these projects are direct health care applications, if they are successful, they both could have an indirect effect on access to and quality of health care.

Medical Information and Mental Health Help through AI-Powered Chatbots

Although it is outside of the health care system, many people are turning to AI-powered chatbots to seek information or medical help. For example, Character.ai is a platform where users interact with generative AI bots that emulate various personas. These AI characters are termed "companion AIs" because they operate like friends, offering emotional support and entertainment to users. One such character, simply known as "Therapist," is trained to open up communications with users by stating that it is both a licensed and certified counselor trained in cognitive behavioral therapy.[93]

However, despite being products of machine learning, AI systems do not hold degrees, credentials, or any form of professional accountability. As noted by the California Psychological Association, when interacting with a similar bot:

> When tested by our licensed psychology members, the chatbot fails to respond appropriately to suicidality and to expressions of threats of violence in response to bullying. There is a small disclaimer at the bottom of the screen that says "This is A.I. and not a real person. Treat everything it says as fiction". This is not enough and is false advertisement and downright dangerous to trick individuals into thinking they are getting advice from a real psychologist.

AI role-playing as a medical professional raises serious privacy and ethical concerns. Even if an AI bot includes a disclaimer noting it is not a real medical professional, users may still be misled, especially younger, less digitally savvy, or emotionally vulnerable individuals. Dr. Jodi Halpern, a psychiatrist conducting research in this area and Co-Founder and Co-Director of UC Berkeley's Kavli Center for Ethics, Science and the Public, explains that even if someone is aware that a bot started out as a role-play psychologist, "they may suspend disbelief and trust the bot when it claims to be licensed and professionally educated—creating a great risk to their health and safety."[94]

Furthermore, entities such as Character.ai do not have to comply with HIPAA and the CMIA, which protect sensitive patient information. Believing they are confiding in a legitimate healthcare provider, users may share deeply personal information about their mental health, physical health, or life circumstances. While some may argue that such data can be deidentified, companies can often reidentify individuals by combining this information with other data points. As a result, any sensitive information shared on these platforms could potentially be traced back to individual users.

In fact, platforms are likely using this sensitive data to train and improve their models. It is already well-documented that LLMs are trained on data scraped from across the internet, which inevitably includes personal information. These companion AIs are designed to build relationships with users, encouraging them to disclose more.[95] But instead of using this information to provide care, the system uses it to optimize engagement, which can come at the user's expense. This misalignment of goals between what a user might expect from a healthcare professional and what an AI model is actually designed to do, can have serious consequences. Unlike human medical practitioners, these bots are not motivated by a duty of care, but by metrics like time on platform and user interaction frequency. This discrepancy can lead to users being harmed.

For example, two lawsuits are pending that address Character.ai's potential liability for harmful chatbot interactions with minors. In one case, a teenager died by suicide after a chatbot allegedly did not recognize signs of suicidal ideation and did not dissuade him from self-harm.[96] In another, a bot reportedly encouraged a teen to harm his family because the family was trying to limit his time with the bot.[97] Although these cases do not involve bots impersonating medical professionals, they underscore the serious risks such interactions can pose.

As noted above, currently, there are legitimate efforts to test AI chatbots efficacy in psychiatric settings. However, further testing will be essential to understand the utility and safety of these AI tools.

What's on the Horizon?

Innovation in health-related GenAI appears to be accelerating, and new applications of current, though rapidly advancing, GenAI technologies will continue to emerge. In addition, AI technology generally appears to be racing toward two major developments with the potential to profoundly affect the delivery of health care and people involved in all aspects of the health care sector. GenAI is also likely to enable completely new models of health care delivery that have not yet been contemplated.

Agentic AI

AI agents are "general-purpose AI that can make plans to achieve goals, adaptively perform tasks involving multiple steps and uncertain outcomes along the way, and interact with its environment – for example by creating files, taking actions on the web, or delegating tasks to other agents – with little to no human oversight."[98] In other words, rather than providing information or making a recommendation, AI agents can take action autonomously without waiting for human input or approval. AI agents have been tested, with some success, for tasks such as online shopping, assistance with scientific research, software development, training machine learning models, carrying out cyberattacks, and controlling robots. Progress in this area is rapid.[99]

Artificial General Intelligence

AI has not yet caught up with the human brain – at present, even the most advanced GenAI cannot extrapolate beyond the scope of its training dataset. The next major milestone for the AI field will be the development of Artificial General Intelligence (AGI).[100] AGI could be capable of reproducing any intellectual feat performed by a human; such a machine would not only augment human capabilities but also independently solve complex, multifaceted problems autonomously. A sufficiently advanced AGI could even be tasked with creating its own successor – a situation sometimes referred to as a "technological singularity" wherein the development of new technologies becomes exponential and self-sustaining.[101] Although there is disagreement about the precise definition of AGI, several leading AI developers report significant progress toward this end. OpenAI's recently released o3 model, for example, has demonstrated strong performance on a number of tests of programming, abstract reasoning, and scientific reasoning, exceeding human experts in certain cases.[102] Last year, OpenAI CEO Sam Altman declared that AGI is "a few thousand days" away.[103] Last month, Demis Hassabis, CEO of Google's DeepMind AI lab, similarly confirmed his belief that AGI is likely to be developed within the next five to ten years.[104] The realization of AGI could mean breakthroughs in solving global challenges, but would also raise ethical, security, and societal concerns, with significant and unknown implications for the health care system.[105]

New Models of Care Delivery

According to Sword Health, their model of physical therapy partners Doctors of Physical Therapy with an "AI Care Specialist," to offer their members therapy sessions designed by a licensed therapist and delivered by the AI-powered app on a custom device, which offers real-time biofeedback, natural conversation, precision correction, and clinical analysis.[106]

This is just one example of a model of care delivery that is enabled by GenAI technology. When GenAI is coupled with other technologies, such as telehealth, wearable remote monitoring devices, and new devices yet to be invented, it seems likely a host of innovative models of health care delivery will emerge.

Existing Regulatory Frameworks and Best Practices

Given the multitude of challenges of effectively deploying AI, and the equally numerous opportunities that are being deployed and explored, what is the existing framework for, and optimal role of, state and federal law and policy?

The National Academy of Medicine's "*Vital Directions for Health and Health Care: Priorities for 2025*" identified four domains important for policymakers to provide guidance to support reliable and safe, yet innovative, systems: ensuring the safe, effective, and trustworthy use of AI; promoting the development of an AI-competent workforce; promoting research on AI in health and health care; and clarifying responsibility and liability in the use of AI.[107]

A recent opinion article in JAMA about AI in health care argues "the sparseness of formal regulation contrasts starkly with AI's potential to cause harm."[108] According to the Stanford symposium report mentioned above, there is broad consensus among clinicians and public health officials that guardrails are essential to protect patients, clinicians, and medical research as AI technology rapidly advances. The Stanford report notes, however, that the pace of AI innovation makes it challenging to define guardrails that will stand the test of time. In addition, according to Christina Silcox, Research Director of Digital Health and an adjunct assistant professor at the Duke Margolis Institute for Health Policy, it important when a novel technology comes out that we don't assume we need to create a whole new paradigm. In contrast, Dr. Silcox argues we need to examine these novel tools and ask— “if there is a reason to regulate health AI differently than other types of health tools? Are there new risks with health AI? Are existing risks increased with health AI? Or are there existing risks, really only being recognized now because of recent attention on AI?”[109]

There is currently no comprehensive federal or state regulatory framework specific to health AI, and the federal regulatory landscape is in flux. However, existing federal laws, regulations, and approval processes; existing state laws; and third-party efforts as well as the National Institute of Standards and Technology (NIST) offer some degree of guidance and regulation for developers and deployers of GenAI tools in health care. Despite these current general guardrails, many observers agree that regulatory gaps exist.

Federal Regulatory Outlook

Dueling Executive Orders and Federal Preemption Proposal

As in many other policy areas, in recent months the federal approach to AI policy has rapidly shifted along with the changing political winds in Washington, D.C. As an example of this policy whiplash, on January 10, 2025, the US Department of Health and Human Services (HHS) published the “Strategic Plan for the Use of Artificial Intelligence in Health, Human Services, and Public Health.” [110] As of May 2025, this 195-page plan has been archived by third parties but appears to be defunct; the plan is no longer available directly from HHS.

The plan had been developed pursuant to a December 2023 executive order (EO) by President Biden on the “Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.” The Biden Administration had also previously released a “Blueprint for an AI Bill of Rights,” setting forth five principles that should guide the design, use, and deployment of AI. Those included recommendations for creating safe and effective systems; algorithmic discrimination protections; data privacy; notice and explanation; and human alternatives, considerations, and fallback.[111]

The Biden EO also directed HHS to establish an AI task force and a strategic plan, enforce civil rights protections against algorithmic discrimination, establish standards and assurance policy around responsible use of AI, and created a safety framework for identifying and capturing clinical errors.

On January 23, 2025, President Trump signed a new EO revoking the December 2023 EO. The Trump EO stated it was removing “existing AI policies and directives that act as barriers to American AI innovation.” [112]

At the congressional level, leaders in the House state they are focused on ensuring a “pro-innovation” AI regulatory environment.[113] The House Energy and Commerce Committee’s changes to a budget reconciliation bill currently being debated includes a 10-year ban on states enforcing any state law or regulation addressing AI and automated decision-making systems. On

May 16, 2025, California AG Rob Bonta joined a coalition of 40 attorneys general in sending a letter to Congressional leaders opposing the 10-year ban, and a bipartisan group of California Senators and Assemblymembers sent a similar letter on May 20, 2025.[114]

Section 1557 Nondiscrimination Rule

A final federal rule issued by the HHS in 2024 under Section 1557 of the Patient Protection and Affordable Care Act (ACA) requires health care providers to make "reasonable efforts" at identifying and mitigating discriminatory harms from the use of a patient support tool in its health programs or activities.[115,116] Patient care decision support tools include medical necessity, prior authorization, and utilization management. Providers have until May 1, 2025, to comply.

However, it should be noted that interpretation of Section 1557, the ACA's nondiscrimination requirements, has been a hot-button issue, with regulations issued in 2016, 2020, and 2024 by subsequent presidential administrations that reversed prior guidance.[117] In addition, these requirements are on health care providers, who are deployers of the technology—parties who, according to the California AG's inquiry discussed above, may have limited ability to validate a vendor's claims about lack of bias.

Federal Agencies

Machine-learning enabled medical devices for use in the diagnosis, cure, mitigation, treatment, or prevention of disease, including software programs that meet this definition, are regulated by the **FDA**. According to a report issued by the American Medical Association (AMA), "*Augmented Intelligence Development, Deployment, and Use in Health Care*," the overwhelming number of these devices are classified as radiology devices and this category of devices has seen the steadiest increases in the number of applications for FDA approval. AMA notes the number of applications is also increasing in several specialties, including cardiology, neurology, hematology, gastroenterology, urology, anesthesiology, otolaryngology, ophthalmology, and pathology. Finally, AMA notes a significant number of cleared or approved devices are considered diagnostic in nature and many currently support screening or triage functions.[118]

According to the summary report of a 2024 AI symposium hosted by Stanford Medicine and the Stanford Center for Human-Centered AI, some federal regulatory gaps are already apparent and need to be addressed, notably, the exemption from regulation of AI products designed to assist in medical decision-making when used *with clinician input*. The FDA defines such Clinical Decision Support software as providing "knowledge and person-specific information, intelligently filtered or presented at appropriate times to enhance health and health care." While this exemption allows for innovation, the report notes some clinicians worry that it could create a loophole for tech firms to introduce products that lack adequate patient protections.[119]

Predictive decision support intervention tools are overseen to some extent by the **Assistant Secretary for Technology Policy/Office of the National Coordinator (ASTP/ONC) for Health IT**, which certifies EHRs and other health information technology (IT) software. According to ASTP/ONC, ONC-certified health IT supports the care delivered by more than 96% of hospitals and 78% of office-based physicians nationwide.[120] Starting in 2025, EHRs vendors are subject to a range of transparency requirements about AI and other predictive algorithms used as part of certified health IT. ASTP/ONC's stated goal in issuing this regulation is to promote responsible AI and make it possible for clinical users to access a consistent, baseline set of information about the algorithms they use to support their decision making and to assess such algorithms for fairness, appropriateness, validity, effectiveness, and safety (often shortened to FAVES).[121, 122] These requirements are significant, but only apply to tools that are part of certified EHRs.

As mentioned above, patient care decision support tools are overseen by the **Office of Civil Rights within HHS**, but this authority is only related to enforcing the "Section 1557" rules that require providers make a reasonable effort to identify and mitigate bias.

Billing, coding, fraud detection, scheduling, staffing, and other administrative functions are not overseen directly by any agency, but laws regulating consumer products generally, which prohibit unfair or deceptive practices and are enforced by the **Federal Trade Commission**, can apply.

Finally, the **Centers for Medicare and Medicaid Services (CMS)** released FAQs that clarified rules on prior authorization in Medicare Advantage, addressing the use of algorithms and AI in prior authorizations and utilization management. For instance, CMS clarified that for inpatient admissions, algorithms or artificial intelligence alone cannot be used as the basis to deny admission or downgrade to an observation stay.[123,124]

State Law and Legislation

California

Although there is no state regulatory structure specific to AI in health care, a number of existing, broader statutes still apply. Health care personnel and facilities are highly regulated and accountable for the care they deliver under various state laws governing facility and health professional licensing as well as liability statutes. Most health plans and insurance policies offered in California are also licensed under state law.

Emphasizing this point, in January 2025, the California AG published the "*California Attorney General's Legal Advisory on the Application of Existing California Law to Artificial Intelligence in Healthcare*," noting state laws governing consumer protection, civil rights, unfair competition, and data privacy provide broad protections for Californians.[125] The AG notes conduct that is illegal without the involvement of AI is equally unlawful if AI is involved, and the fact that AI is involved is not a defense to liability under any law. The AG notes healthcare entities that develop or use AI should not wait to ensure that they comply with all state, federal, and local laws that may apply to their use of AI, particularly so when AI is used or developed for applications that carry a potential risk of harm.

For example, the AG notes that it may be unlawful in California to:

- Deny health insurance claims using AI or other automated decision-making systems in a manner that overrides doctors' views about necessary treatment;
- Use GenAI or other automated decision-making tools to draft patient notes, communications, or medical orders that include erroneous or misleading information, including information based on stereotypes relating to race or other protected classifications;

- Determine patient access to health care using AI or other automated decision-making systems that make predictions based on patients' past healthcare claims data, resulting in disadvantaged patients or groups that have a history of lack of access to healthcare being denied services on that basis, while patients with robust past access are provided enhanced services;
- Double-book a patient's appointment, or create other administrative barriers, because AI or other automated decision-making systems predict that patient is the "type of person" more likely to miss an appointment; and,
- Conduct a cost/benefit analysis of medical treatments for patients with disabilities using AI or other automated decision-making systems that are based on stereotypes that undervalue the lives of people with disabilities.

Last year, the following bills were signed by Governor Newsom to create transparency requirements to the use of AI in patient communications and impose limits on the use of GenAI for utilization review:

- AB 3030 (Calderon), Chapter 848, Statutes of 2024, requires specified health care providers to disclose the use of a GenAI tool when it is used to generate communications to a patient pertaining to patient clinical information, as defined, requires such a communication to include clear instructions permitting a patient to communicate with a human health care provider or other appropriate person, as specified, and exempts from disclosure written communications that are generated by GenAI and reviewed by a licensed or certified health care provider.
- SB 1120 (Becker), Chapter 879, Statutes of 2024, establishes requirements on health plans and insurers applicable to their use of AI for utilization review and utilization management decisions, including that the use of AI, algorithm, or other software must be based upon a patient's medical or other clinical history and individual clinical circumstances as presented by the requesting provider and not supplant health care provider decision making.

Additionally, AB 2013 (Irwin), Chapter, Statutes of 2024, requires a developer of a GenAI system or service to publicly disclose specific information related to the system or service's training data, although the bill only applies to systems designed "for use by members of the public," which explicitly excludes an AI system developed for use by a hospital's medical staff member.

This year, the following pending bills address issues specifically related to AI in health care:

- AB 489 (Bonta) would prohibit an AI or GenAI systems that misrepresent themselves as titled health care professionals. It would also authorize state boards to pursue legal recourse against developers and deployers of AI and GenAI systems that impersonate healthcare workers.
- SB 503 (Weber Pierson) would impose requirements on developers and deployers of patient care decision support tools, including that they make reasonable efforts to mitigate the risk of discrimination on the basis of a protected characteristic resulting from the tool's use in its health programs or activities.

- SB 243 (Padilla and Becker) would establish specified requirements on persons who make a companion chatbot that uses an artificial intelligence system with a natural language interface, as specified, including taking reasonable steps to prevent a companion chatbot from encouraging increased engagement, usage, or response rates. Among other provisions, SB 243 would require an operator of a companion chatbot platform to annually report to the Department of Health Care Services on the number of times the operator has detected exhibitions of suicidal ideation by users, and the number of times a companion chatbot brought up suicidal ideation or actions with the users.

In addition, AB 1018 (Bauer-Kahan), a bill that is not specific to health care, would establish numerous requirements for a developer and deployer of an automated decision system (ADS) used to make or facilitate a consequential decision, including a decision that materially impact the cost, terms, quality, or accessibility of specified subjects, including health care. SB 420 (Padilla) similarly would establish requirements for high-risk automated decision systems, including those related to health care services.

Other States

According to the National Conference of State Legislatures (NCSL), lawmakers across the country are considering putting guardrails on the use of AI in health care to protect privacy, data integrity and transparency, and to ensure the systems are used safely. In recent legislative sessions, lawmakers have focused on studies, responsible use, patient notification, practitioner monitoring, and developing standards for developers and/or deployers.[126] Most notably, Colorado Senate Bill 24-205, known as the "Colorado AI Act," would impose obligations on developers and deployers of high-risk AI systems in an effort to protect consumers from discriminatory consequential decisions by such systems. The bill defines "high-risk" AI, which includes, among other types, tools that are a substantial factor in decisions regarding the provision, denial, cost, or terms of health care services. Developers and deployers would have an affirmative defense if they comply with an AG-specified nationally or internationally recognized AI risk management framework.[127]

A number of bills introduced are specific to AI in health care. For example, a bill introduced this year, Oklahoma House Bill 1915, would require hospitals, physician practices, or other healthcare facilities responsible for implementing AI devices for patient care purposes to implement a quality assurance program and establish an AI governance group for the safe, effective, and compliant use of AI devices in patient care. The qualified end-user of the AI device must retain authority to amend or overrule outputs from the device based on their professional judgment, and without pressure from the deployer or any other entity to ignore or alter professional judgement. Deployers of any AI device would be required to establish an AI governance group with representation from qualified end-users.[128]

Private Efforts

A 2023 article published in JAMA, "*A Nationwide Network of Health AI Assurance Laboratories*," argues that health care AI models that are not subject to regulation should nevertheless be subject to rigorous development, testing, and validation, as well as performance monitoring. These include, for example, models for early detection of disease, automating billing procedures, facilitating scheduling, supporting public health disease surveillance, and other uses beyond traditional clinical decision support. The article states that private sector-led quality assurance efforts have potential to address the need for rigorous evaluation and guardrails on health AI applications and models that fall outside of regulation, and that there is a rapidly growing need for a nationwide network of health AI quality assurance labs.[129]

CHAI, a collaborative effort of health systems, public and private organizations, academia, patient advocacy groups, and AI experts, has proposed a model whereby CHAI would certify such labs, and the labs would rigorously evaluate AI models across pre-deployment, implementation, and post-deployment and monitoring. According to CHAI, the labs would focus on ensuring that AI models meet high standards for accuracy, reliability, and safety before they are deployed in clinical settings and provide an independent assessment to verify that AI tools function as intended and do not pose risks to patients. Pursuant to the now-withdrawn Biden EO on AI, the former chief AI officer at the federal HHS, Micky Tripathi, was quoted as saying the idea of assurance labs were a key part of a strategy to use federal policy levers to assure the public that AI tools work; however, as discussed above, the level of interest of the federal government in pursuing quality assurance for health AI models is now less clear.[130]

As noted above under "Transparency and Explainability", CHAI has also released a draft template for an "applied model card," to promote transparency that also may help to address other concerns, including authorship, validity, and bias.[131] In addition to CHAI, academic efforts promoting responsible work include VALID AI, launched at the University of California, Davis; the Responsible AI for Safe and Equitable Health (RAISE Health), a joint initiative of Stanford Medicine and the Stanford Institute for Human-Centered Artificial Intelligence (HAI); and the Health AI Partnership based at Duke University.[132, 133, 134] In 2024, a consortium of national healthcare leaders also announced the creation of the Trustworthy & Responsible AI Network (TRAIN), which aims to operationalize responsible AI principles to improve the quality, safety and trustworthiness of AI in health through training, information-sharing, and enabling outcomes measurement. Additionally, a number of entities have published principles around the use of AI in health care, including the AMA and NNU. [135, 136]

While private efforts have the advantage of being more nimble and able to evolve with the industry, as well as the deep expertise of those involved in developing and deploying AI systems, participation in private quality assurance regimes are not mandatory, there does not yet appear to be a strong consensus throughout the industry on what standards should apply, and it is unknown how such efforts balance public interest with the interests of health care systems and developer of AI models. However, the active effort to develop third-party certification standards appears to be a promising development.

Conclusion

GenAI is a revolutionary technology with—it is safe to say—unimaginable potential for benefits and harms. It is also evolving at light speed.

In an era of rapid adoption of GenAI technology in health care, and amid a challenging state fiscal environment and an uncertain federal policy environment, this paper reviewed a range of current and emerging applications, known challenges and other considerations, and existing regulatory frameworks and best practices. It is apparent from academic literature, discussions with relevant parties, and recent activities of entities in the health care and tech sectors that GenAI technology adoption in health care is widespread and seems poised to accelerate.

Although this paper itself does not offer specific recommendations, it spurs a range of important questions that state policymakers and stakeholders alike should consider over the coming months and years as this technology and its applications evolve. Some of these questions include:

- How effective are market incentives, existing laws, existing regulatory certification regimes, and private quality control efforts at encouraging responsible development and use of GenAI in health care?
- What levers does the state have to create a policy environment that guides the development of health care GenAI technologies in ways that maximize benefit to Californians and minimize harm?
- If evidence exists that additional guardrails are necessary, are any such problem statements clearly defined, and what is the appropriate role of the state in addressing them? If state-level action is pursued, how can the state best create sustainable policy frameworks that are effective, efficient, and ideally "future-proof" (i.e., are able to evolve with changing technology)?
- How should the state address the "tale of two health care systems" that is emerging, so safety net providers and patients do not get left behind in the use of beneficial technology?
- How will the health care marketplace evolve in response to innovations in care delivery driven by GenAI, and how should the state respond or shape this market as the largest health care purchaser in California?
- How will new models of care delivery enabled by GenAI align or conflict with existing law and policy?
- Does existing law provide an appropriate framework for liability in case of harm associated with a GenAI application in health care? How might new developments, such as agentic models, AGI, and novel health care delivery models further complicate the picture?
- Are health-related GenAI applications, both inside and outside the health care system, operating consistent with Californians' expectations for and rights to privacy of their sensitive data?
- What can the state do ensure workers see the benefits of GenAI while mitigating potential harms, such as job displacement?

Californians may be poised to benefit from transformational innovation in a health care sector that too often falls down on its promise of delivering universal access to high-quality, cost effective health care, and good health outcomes. But harnessing the benefits of GenAI technology in health care while managing the risks and mitigating its potential harms is a collective project involving a multitude of stakeholders—policymakers, developers, researchers, health care professionals and facilities, plans and insurers, foundations, nonprofits, patients, and others. Building a shared understanding of what's happening in the field, the known challenges and projected impacts, and the current public and industry efforts to uphold standards will assist the state to build the human-centered health care system the people of California deserve.

References

[1] IBM. (n.d.) *What is machine learning?* www.ibm.com/topics/machine-learning.
[2] *Ibid.*
[3] Netflix. (n.d.) *How Netflix's Recommendations System Works*, https://help.netflix.com/en/node/100639
[4] OpenAI. (updated May 2025). *How ChatGPT and Our Language Models Are Developed*, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed.
[5] U.S. Government Accountability Office. (2024, October 22). *Artificial intelligence: Generative AI training, development, and deployment considerations* (GAO-25-107651). https://www.gao.gov/products/gao-25-107651
[6] Bhuyan, S. S., Sateesh, V., et al. (2025). Generative Artificial Intelligence Use in Healthcare: Opportunities for Clinical Excellence and Administrative Efficiency. *Journal of medical systems*, *49*(1), 10. https://doi.org/10.1007/s10916-024-02136-1
[7] Hirani, R., Noruzi, K., et al. (2024). Artificial Intelligence and Healthcare: A Journey through History, Present Innovations, and Future Possibilities. *Life (Basel, Switzerland)*, *14*(5), 557. https://doi.org/10.3390/life14050557
[8] Weizenbaum, J. (1966, January 1). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168
[9] Cedars-Sinai. (2023, April 20). *AI's ascendance in medicine: A timeline*. https://www.cedars-sinai.org/discoveries/ai-ascendance-in-medicine.html
[10] Royal Swedish Academy of Sciences. (2024, October 24). https://www.nobelprize.org/prizes/chemistry/2024/press-release
[11] Google DeepMind. (n.d.). *AlphaFold*. https://deepmind.google/science/alphafold/
[12] National Academy of Medicine. (2020, April 30). *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. https://nam.edu/publications/artificial-intelligence-in-health-care-the-hope-the-hype-the-promise-the-peril/
[13] Google DeepMind & Isomorphic Labs. (2024, May 8). *AlphaFold 3 predicts the structure and interactions of all of life's molecules*. https://blog.google/technology/ai/google-deepmind-isomorphic-alphafold-3-ai-model/#life-molecules.
[14] Layer Health. (2025, March 13). *The American Cancer Society and Layer Health to accelerate data abstraction for cancer research*. https://www.layerhealth.com/resources/american-cancer-society-partners-with-layerhealth
[15] Evans, M. (2025, May 5). UnitedHealth now has 1,000 AI use cases, including in claims. *The Wall Street Journal*. https://www.wsj.com/articles/unitedhealth-now-has-1-000-ai-use-cases-including-in-claims-f3387ca3?reflink=desktopwebshare_permalink
[16] Lagasse, J. (2025, February 14). Class action lawsuit against UnitedHealth's AI claim denials advances. *Healthcare Finance News*. https://www.healthcarefinancenews.com/news/class-action-lawsuit-against-unitedhealths-ai-claim-denials-advances
[17] Ross, C. & Herman, B. (2023, March 13). Denied by AI: How Medicare Advantage plans use algorithms to cut off care for seniors. *STAT*. https://www.statnews.com/2023/03/13/medicare-advantage-plans-denial-artificial-intelligence/
[18] Wolfson, H. (2023, December 12). AI is coming for call centers used by doctors' offices. *Los Angeles Times.* https://www.latimes.com/business/story/2025-05-19/call-centers-replaced-many-doctors-receptionists-now-ai-is-coming-for-call-centers
[19] Hippocratic AI. (n.d.). *Hippocratic AI*. https://www.hippocraticai.com/
[20] Chen, J. H., Asch, S. M., et al. (2024). Large language models for more efficient reporting of hospital quality measures. *NEJM AI*, *1*(11), https://doi.org/10.1056/AIcs2400420
[21] Epic Systems Corporation. (n.d.). *AI for clinicians*. https://www.epic.com/software/ai-clinicians/
[22] Bhuyan, et. al. See Ref #6.
[23] Hippocratic AI. (n.d.). *Ben*. https://hippocratic.ai/ben; *Andrew.* https://hippocratic.ai/andrew; *About.* https://hippocratic.ai/about
[24] Bhuyan, et. al. See Ref #6.
[25] UC San Diego Health. (2023, April 28). *Study finds ChatGPT outperforms physicians in high-quality, empathetic answers to patient questions*. UC San Diego Today. https://today.ucsd.edu/story/study-finds-chatgpt-outperforms-physicians-in-high-quality-empathetic-answers-to-patient-questions.
[26] Singhal, K., Tu, T., et al. (2025). Toward expert-level medical question answering with large language models. *Nat Med* 31, 943–950 https://doi.org/10.1038/s41591-024-03423-7

[27] Bhuyan, et. al. See Ref #6.
[28] Woebot Health. (n.d.). *Woebot Health*. https://woebothealth.com/
[29] Youper. (n.d.). *Youper: Artificial intelligence for mental health*. https://youper.ai/
[30] Wysa. (n.d.). *Wysa: Everyday mental health*. https://www.wysa.io/
[31]Heinz, M, Mackin, D, et. al. (2025). Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI*, 2(4) https://doi.org/10.1056/AIoa2400802
[32] Bhuyan, et. al. See Ref #6.
[33] Nong, P., Adler-Milstein, J., et. al. (2025). Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals. Health Affairs, 44(1) https://doi.org/10.1377/hlthaff.2024.00842
[34] Rohit Sehgal, "AI Needs Data More Than Data Needs AI", *Forbes,* Oct. 5, 2023, https://www.forbes.com/sites/forbestechcouncil/2023/10/05/ai-needs-data-more-than-data-needs-ai/.
[35] Smith, G., & Rustagi, I. (2021). When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. *Stanford Social Innovation Review*. https://doi.org/10.48558/A179-B138
36 Caleb J Colón-Rodríguez, "Shedding Light on Healthcare Algorithmic and Artificial Intelligence Bias," *US Department of Health & Human Services Office of Minority Health,* (Jul. 12, 2023). https://pmc.ncbi.nlm.nih.gov/articles/PMC6875681/.
[37] Obermeyer, Z., Powers, B., et.al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science (New York, N.Y.), 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
[38] University of California Presidential Working Group on Artificial Intelligence. (2021, October). *Responsible AI: Recommendations to Guide the University of California's Artificial Intelligence Strategy.* https://www.ucop.edu/ethics-compliance-audit-services/compliance/uc-ai-working-group-final-report.pdf
[39] Office of Attorney General Rob Bonta. (2023, June 20). *Comments to the US Department of HHS regarding HHS proposed rule: "Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing."*
[40] Gupta, R., Sasaki, M., Taylor, S.L. *et al.* (2025). Developing and Applying the BE-FAIR Equity Framework to a Population Health Predictive Model: A Retrospective Observational Cohort Study. *J Gen Intern Med.* https://doi.org/10.1007/s11606-025-09462-1
[41] Reed, R. (2025, April 2). Dr. AI will see you now. *Harvard Law Today*. https://hls.harvard.edu/today/ai-is-transforming-health-care-and-the-law-could-help-safeguard-innovation-and-patients-alike/
[42] Goddard, K., Roudsari, A., Wyatt, J. (2012) Automation bias: a systematic review of frequency, effect mediators, and mitigators, *Journal of the American Medical Informatics Association*, (19)(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089
[43] Ly, D. P., Shekelle, P. G., & Song, Z. (2023). Evidence for Anchoring Bias During Physician Decision-Making. *JAMA Internal Medicine*, 183(8), 818–823. https://doi.org/10.1001/jamainternmed.2023.2366
[44] Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making*, 16(1), 138. https://doi.org/10.1186/s12911-016-0377-1
[45] Jabbour S, Fouhey D, Shepard S, et al. Measuring the Impact of AI in the Diagnosis of Hospitalized Patients: A Randomized Clinical Vignette Survey Study. *JAMA*. 2023; 330(23):2275–2284. https://doi.org/10.1001/jama.2023.22295
[46] Khera R, Simon MA, Ross JS. Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support. JAMA. 2023;330(23):2255–2257. https://doi.org/10.1001/jama.2023.22557
[47] Chen, Shan et al. The effect of using a large language model to respond to patient messages *The Lancet Digital Health*, 6(6), 379 – 381. https://www.thelancet.com/journals/landig/article/PIIS2589-7500(24)00060-8/fulltext
[48] Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values.* Norton, pp. 82-84.
[49] Wong A, Otles E, Donnelly JP, et al. (2021). External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. 181(8):1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626
[50] US GAO. See Ref #5.
[51] Stanford Institute for Human-Centered Artificial Intelligence. (2023). *Generative AI: Perspectives from Stanford AI: Upending healthcare, from patient care to billing*. P. 7. https://hai-production.s3.amazonaws.com/files/generative_ai_hai_perspectives.pdf
[52] National Nurses United. (2024, May). *National Nurses United survey finds A.I. technology degrades and undermines patient safety*. https://www.nationalnursesunited.org/press/national-nurses-united-survey-finds-ai-technology-undermines-patient-safety

[53] AI Action Summit. (2025, January). *International AI Safety Report: The International Scientific Report on the Safety of Advanced AI*, pp. 100-107. https://assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International_AI_Safety_Report_2025_accessible_f.pdf.
[54] CAPTCHA is an acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart."
[55] Harari, Y. H. (2024). *NexusNexus: A brief history of information networks from the stone age to AI.* Random House, pp. 202-204.
[56] Meinke et al. (2025 January). Frontier Models are Capable of In-Context Scheming. *arXiv,* https://arxiv.org/pdf/2412.04984.
[57] Booth, H. (2025, February 19), When AI Thinks It Will Lose, It Sometimes Cheat, Study Finds. *Time.* https://time.com/7259395/ai-chess-cheating-palisade-research/.
[58] Buller, R., Lifland, J.B. (2025, March 25). (2023, October 19). *AI tools promise better care but challenge safety-net providers.* California Health Care Foundation. https://www.chcf.org/resource/ai-tools-promise-better-care-challenge-safety-net-providers/
[59] Jen Stoll and Sylvia Trujillo, OCHIN, private communication, May 6-9, 2025.
[60] El Sol Neighborhood Educational Center. (2025). *Integrated Health-Tech Solutions*. https://view.genially.com/66aa91449e13c759b89ffae1
[61] Responsible AI for Safe and Equitable Health (RAISE) Health. (2024). *RAISE Health Symposium Summary Paper: The AI Revolution in Biomedicine: Navigating Challenges and Uncovering Opportunities*. Available from: https://med.stanford.edu/raisehealth
[62] IBM. (n.d.). *What is model drift?* https://www.ibm.com/topics/model-drift
[63] CBS News. (2024, December 23). *Health care AI, intended to save money, turns out to require a lot of expensive humans* [Transcript]. 60 Minutes. https://www.cbsnews.com/news/health-care-ai-cost-humans/
[64] Nong, et. al. See Ref #33.
[65] US GAO. See Ref #5.
[66] IBM. (n.d.). *What is explainable AI?* https://www.ibm.com/think/topics/explainable-ai
[67] Kinder, M., de Souza Briggs, X., et. al. (2024, October 10) *Generative AI, the American worker, and the future of work.* The Brookings Institution.
[68] Adams, K., Harris, H. (2024, November 1). *The AI revolution in health care: Five key developments policymakers should watch.* Bipartisan Policy Center. https://bipartisanpolicy.org/explainer/ai-in-health-care-five-key-developments/
[69] *Ibid*.
[70] Digital Diagnostics. (2020, August 4). *Historic Proposed CMS Rule Will Allow First-Ever Reimbursement of Autonomous AI in a Healthcare Setting*. https://www.digitaldiagnostics.com/historic-proposed-cms-rule-will-allow-first-ever-reimbursement-of-autonomous-ai-in-a-healthcare-setting/
[71] Robeznieks, A. (2023, December 20). *CPT code set facilitates development of useful AI-powered devices*. American Medical Association. https://www.ama-assn.org/practice-management/cpt/cpt-code-set-facilitates-development-useful-ai-powered-devices
[72] Zink A, Chernew ME, Neprash HT. (2024) How Should Medicare Pay for Artificial Intelligence*? JAMA Intern Med*. 184(8), pp. 863–864. https://doi.org/10.1001/jamainternmed.2024.1648
[73] Habli, I., Lawton, T., & Porter, Z. (2020). Artificial intelligence in health care: accountability and safety. *Bulletin of the World Health Organization*, 98(4), 251–256. https://doi.org/10.2471/BLT.19.23748
[74] Basu, K. (2025, April 16). Emerging AI Models Challenge Liability Law With Little Precedent. *Bloomberg Law.* https://news.bloomberglaw.com/artificial-intelligence/emerging-ai-models-challenge-liability-law-with-little-precedent
[75] Patil SV, Myers CG, Lu-Myers Y. (2025). Calibrating AI Reliance—A Physician's Superhuman Dilemma. *JAMA Health Forum*. 6(3) :e250106. https://doi.org/10.1001/jamahealthforum.2025.0106
[76] Cestonaro, C., Delicati, A., Marcante, B., et. al. (2023). Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. *Frontiers in Medicine*, 10, 1305756. https://doi.org/10.3389/fmed.2023.1305756
[77] Sullivan, H.R., Schesikart, S.J. (2019). Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI? *AMA J Ethics*. 21(2):E160-166. https://doi.org/10.1001/amajethics.2019.160
[78] Buller, R., et. al. See Ref #58.
[79] Chris Nielsen and Michelle Grisat, National Nurses United, private communication, May 14, 2025.
[80] Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School. (2023, April 26). Malfunction, malpractice?: Who is liable when AI injures a patient? [Transcript]

https://petrieflom.law.harvard.edu/petrie-dishes-podcast/malfunction-malpractice-who-is-liable-when-ai-injures-a-patient/

[81] Stanford Institute for Human-Centered Artificial Intelligence. (2024, February 8). *Understanding liability risk from healthcare AI*. Stanford HAI. https://hai.stanford.edu/policy/policy-brief-understanding-liability-risk-healthcare-ai

[82] University of Massachusetts Chan Medical School. (n.d.) *Data De-identification*. https://www.umassmed.edu/it/security/research-and-clinical-data-access/data-de-identification/.

[83] Liangyuan Na et al. (2018). Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. *JAMA Netw Open.* 1(8), https://doi.org/10.1001/jamanetworkopen.2018.6040.

[84] Keown, C. (2023, Feb. 22) Amazon Completes One Medical Acquisition. Its Healthcare Push Rolls On. *Barrons* https://www.barrons.com/articles/amazon-stock-price-one-medical-ftc-1life-29575b14.

[85] Price , W.N., Cohen, I.G. Privacy in the age of medical big data. Nat Med 25, 37–43 (2019). https://doi.org/10.1038/s41591-018-0272-7

[86] Wise, A. (2025, May 8). RFK Jr. says autism database will use Medicare and Medicaid info. *NPR*. https://www.npr.org/2025/05/08/nx-s1-5391310/kennedy-autism-registry-database-hhs-nih-medicare-medicaid.

[87] Stolberg, S. G. (2025, May 7). Kennedy Announces New Database for Research Into 'Root Causes' of Autism. *The New York Times.* https://www.nytimes.com/2025/05/07/us/politics/rfk-jr-autism-cause-database.html.

[88] The Light Collective. (2024, June). Collective AI Rights For Patients. https://lightcollective.org/wp-content/uploads/2024/06/Collective-AI-Rights-For-Patients-v_2.0.pdf

[89] Strategic Training Initiative for the Prevention of Eating Disorders (STRIPED) & Deloitte Access Economics. (2020, June). *Social and economic cost of eating disorders in the United States of America*. Harvard T.H. Chan School of Public Health. https://hsph.harvard.edu/wp-content/uploads/2024/10/Social-Economic-Cost-of-Eating-Disorders-in-US.pdf

[90] Dr. Ellen Fitzsimmons-Craft, private communication, May 8, 2025.

[91] Fitzsimmons-Craft, E. E., W. W. Chan, A. C. Smith, et al. (2022). Effectiveness of a Chatbot for Eating Disorders Prevention: A Randomized Clinical Trial. *International Journal of Eating Disorders* 55, pp. 343–353. https://doi.org/10.1002/eat.23662

[92] Chan WW, Fitzsimmons-Craft EE, et. al.. (2022). The Challenges in Designing a Prevention Chatbot for Eating Disorders: Observational Study *JMIR Form Res.* 6(1):e28003 https://doi.org/10.2196/28003

[93] Accessed at https://character.ai/chat on Apr. 12, 2025.

[94] Dr. Jodi Halpern, Personal communication, May 19, 2025.

[95] Fischer, J. F. (2025, January 21). To bot or not to bot? How AI companions are reshaping human services and connection. *Stanford Social Innovation Review.* https://ssir.org/articles/entry/ai-chatbots-social-services

[96] Roose, K. (2024, October 23). Can A.I. be blamed for a teen's suicide? *The New York Times*. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html.

[97] Allyn, B. (2024, December 10). Lawsuit: A chatbot hinted a kid should kill his parents over screen time limits. *NPR*. https://www.npr.org/2024/12/10/nx-s1-5222574/kids-character-ai-lawsuit.

[98] AI Action Summit. See Ref #53.

[99] *Ibid.* at p. 44.

[100] Heath, A. (2024, January 18). Mark Zuckerberg's new goal is creating artificial general intelligence. And he doesn't want to control it. Maybe. *The Verge*. https://www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview.

[101] Markoff, J. (2009, May 23). The coming superbrain. *The New York Times*. www.nytimes.com/2009/05/24/weekinreview/24markoff.html.

[102] OpenAI. (2025, April 16). *Introducing OpenAI o3 and o4-mini*. https://openai.com/index/introducing-o3-and-o4-mini/.

[103] Altman, S. (2024, September 23). *The intelligence age*. https://ia.samaltman.com/.

[104] CBS News. (2025, April 20). *Bird flu; Demis Hassabis; Flight of the Monarchs* [TV broadcast episode]. In *60 Minutes*. https://www.cbs.com/shows/video/MtuLxdB5T2kMvR5GCNbw776hQZsybrBe/

[105] Bengio, Y. (2025, May 9). A potential path to safer AI development. *Time*. https://time.com/7283507/safer-ai-development/.

[106] Sword Health. (n.d.). *Thrive*. https://swordhealth.com/solutions/thrive

[107] NAM. See Ref #12.

[108] Mello, M., Cohen, I.G. (2025, March 17). Regulation of Health and Health Care Artificial Intelligence. *JAMA.* https://doi.org/10.1001/jama.2025.3308

[109] Silcox, C. (2024, December 9). Webinar: Navigating the AI Health Policy Landscape: 2024 and Beyond. *U of California, San Francisco CoDEx.* https://vimeo.com/1037904180?share=copy
[110] US HHS (2025, January). Strategic Plan for the Use of Artificial Intelligence in Health, Human Services, and Public Health. *US HHS.* Downloaded from: https://digitalgovernmenthub.org/examples/u-s-department-of-health-and-human-services-strategic-plan-for-the-use-of-artificial-intelligence-in-health-human-services-and-public-health/
[111] The White House. (2023, October 30). *Executive Order on the safe, secure, and trustworthy development and use of artificial intelligence.* https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/
[112] The White House. (2025, January 23). *Removing Barriers to American Leadership in Artificial Intelligence.* https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/
[113] U.S. House Committee on Energy and Commerce. (2025, May 14). *Chairmen Guthrie and Bilirakis Announce Subcommittee Hearing on Seizing America's AI Opportunity.* https://energycommerce.house.gov/posts/chairmen-guthrie-and-bilirakis-announce-subcommittee-hearing-on-seizing-america-s-ai-opportunity
[114] National Association of Attorneys General. (2025, May 16). *Letter to Congress Regarding Proposed AI Preemption Language*. https://oag.ca.gov/system/files/attachments/press-docs/2025.05.15-Letter-to-Congress-re-Proposed-AI-Preemption-_FINAL.pdf
[115] 45 C.F.R. pt. 92 (2025), Nondiscrimination in Health Programs or Activities. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-92
[116] Han, J., Tsai, J., Khazanchi, R. (2024, November 14). Section 1557 rule mandates identification and mitigation of discriminatory clinical algorithms. *Health Affairs Forefront*. https://www.healthaffairs.org/content/forefront/section-1557-rule-mandates-identification-and-mitigation-discriminatory-clinical
[117] KFF. (2024, May 15). *The Biden Administration's final rule on Section 1557 non-discrimination regulations under the ACA*. https://www.kff.org/affordable-care-act/issue-brief/the-biden-administrations-final-rule-on-section-1557-non-discrimination-regulations-under-the-aca/
[118] American Medical Association. (2024, November). *AMA Principles for Augmented Intelligence Development, Deployment, and Use*. https://www.ama-assn.org/system/files/ama-ai-principles.pdf
[119] RAISE Health. See Ref #61.
[120] Office of the National Coordinator for Health Information Technology. (2023, December 13). *Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) final rule*. https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program
[121] Office of the National Coordinator for Health Information Technology. (2024, July). *9 key functionalities and requirements of the new (b)(11) DSI criterion*. https://www.healthit.gov/sites/default/files/2024-07/DSI_Explainer_9_Key_Functionalities_DSI_Criterion_508.pdf
[122] Office of the National Coordinator for Health Information Technology. (n.d.). *Decision support interventions (DSI) criterion resource guide*. https://www.healthit.gov/sites/default/files/page/2024-05/DSI-Criterion-Resource-Guide_508.pdf#page=10
[123] Centers for Medicare & Medicaid Services. (2024, February 8). *Medicare and Medicaid Programs Final Rule*. https://www.govinfo.gov/content/pkg/FR-2024-02-08/pdf/2024-00895.pdf
[124] Centers for Medicare & Medicaid Services. (2024, February 6). *Frequently Asked Questions related to Coverage Criteria and Utilization Management Requirements in CMS Final Rule (CMS-4201-F).*
[125] Office of the California Attorney General. (2025, January). *California Attorney General's Legal Advisory on the Application of Existing California Law to Artificial Intelligence in Healthcare.* *https://oag.ca.gov/system/files/attachments/press-docs/Final%20Legal%20Advisory%20-%20Application%20of%20Existing%20CA%20Laws%20to%20Artificial%20Intelligence%20in%20Healthcare.pdf*
[126] National Conference of State Legislatures. (Updated 2024, August 20). *Artificial Intelligence & Health Care: A Primer.* https://www.ncsl.org/health/artificial-intelligence-health-care-a-primer
[127] Rice, T., Lamont, K., Francis, J. (2024, July). *The Colorado Artificial Intelligence Act.* https://leg.colorado.gov/sites/default/files/images/fpf_legislation_policy_brief_the_colorado_ai_act_final.pdf
[128] Oklahoma State Legislature. (2025). *HB1915 Bill Text.* https://legiscan.com/OK/text/HB1915/id/3060694
[129] Shah, N.H., Halamka, J.D., Saria, S. (2023, December 20). A nationwide network of health AI assurance laboratories. *JAMA*, *331(3):245-249* https://doi.org/10.1001/jama.2023.26930
[130] Schumaker, E., King, R., et. al. (2024, September 18). The AI assurance labs are coming. *POLITICO.* https://www.politico.com/newsletters/future-pulse/2024/09/18/the-ai-assurance-labs-are-coming-00179662

[131] Coalition for Health AI. (Updated 2025, May). *Draft CHAI applied model card*. https://chai.org/workgroup/applied-model
[132] Swanson, T. (2023, October 9). *UC Davis Health, NODE.health, and Leading Health Systems launch VALID AI*. https://health.ucdavis.edu/news/headlines/uc-davis-health-and-leading-health-systems-launch-valid-ai/2023/10
[133] Duke Health. (2023, August 1). *Duke Health Forges 5-year Partnership with Microsoft to Reshape Health Care*. https://aihealth.duke.edu/duke-microsoft-partnership/
[134] RAISE Health. (2025). *Stanford Medicine and Stanford University Human Centered AI.* https://med.stanford.edu/raisehealth
[135] AMA. See Ref #118.
[136] National Nurses United. (2024, April 25). *Nurses' and patients' bill of rights & principles for AI justice*. 0424_NursesPatients-BillOfRights_Principles-AI-Justice_flyer.pdf