JOINT COMMISSION AND COALITION FOR HEALTH AI (CHAI) GUIDANCE ON

The Responsible Use of AI in Healthcare (RUAIH)

Background

The promise and opportunity of artificial intelligence (AI) tools in healthcare are transformative. AI holds the potential to revolutionize the delivery of healthcare by enhancing the precision of diagnostics, personalizing treatment plans, and predicting patient outcomes with unprecedented accuracy. It can streamline administrative processes, reduce the burden on healthcare professionals, and enable more efficient resource allocation. Furthermore, AI can analyze vast amounts of data to uncover insights previously beyond reach, paving the way for innovative treatments and interventions. With these advancements, AI can significantly improve patient care, mitigate risks, and foster a more resilient and responsive healthcare system. In the context of this document, health AI tools are defined as clinical, administrative, and operational solutions that apply algorithmic methods (predictive, generative, combined methods) to a suite of tasks that are part of direct or indirect patient care (e.g., decision support, diagnosis, treatment planning, imaging, laboratory, patient monitoring), care support services (e.g., clinical documentation, scheduling, care coordination/management, patient communication), and care-relevant healthcare operations and administrative services, (e.g., revenue cycle management, coding, prior authorization, care quality management, etc.).

The transformative opportunity that AI presents is not without risk, however. One of the primary concerns is the potential for AI errors, which could arise from algorithmic biases, data inaccuracies, or unforeseen interactions within the healthcare environment. These errors can lead to misdiagnoses, inappropriate treatment plans, and ultimately, patient harm.

Additionally, the lack of transparency in AI decision-making processes, often referred to as the "black box" problem, poses significant challenges in understanding and trust.

Another risk involves data privacy and security. AI systems require vast amounts of data to function effectively, raising concerns about protecting sensitive patient information. Breaches in data security and uses of patient data in ways that are not anticipated by patients, such as for commercial benefit, could compromise patient confidentiality and trust, leading to legal and ethical ramifications and curb future use of AI-enabled tools.

The rapid pace of AI development can also outstrip the ability of healthcare organizations to keep up with necessary training and updates. This knowledge gap may result in improper use of AI tools, further exacerbating the risk of patient harm. Moreover, overreliance on AI could potentially diminish the role of human judgment in clinical decision-making, leading to depersonalized care and potential ethical dilemmas.

Finally, even where AI-enabled software or medical devices are validated and/or approved by the Food and

Drug Administration (FDA), the integration of AI into existing healthcare systems can be challenging, requiring significant adjustments in clinical workflows, which may disrupt established processes and lead to errors and create resistance among healthcare professionals.

Despite these risks, healthcare organizations can take measures to mitigate them and fully harness the transformative potential of AI. By implementing a process (or guardrails) regarding deployment, validation and testing protocols, and use of AI tools, healthcare organizations can protect patient privacy, maintain robust data privacy and security standards, and reduce the potential for AI errors and breaches for individual products and broadly across the system. Ongoing education and training for healthcare professionals on proper use of AI tools can bridge the knowledge gap and prevent misuse. By fostering a collaborative environment where human judgment and AI tools complement each other, organizations can ensure that AI tools enhance rather than diminish the quality of patient care.

This guidance, a joint effort between Joint Commission and the Coalition for Health AI (CHAI), is an initial, high-level document to help promote a shared understanding of responsible deployment and use of AI tools across healthcare organizations. This helps the industry align elements that enhance patient safety by reducing risks associated with AI error and improving administrative, operational, and patient outcomes by leveraging AI's potential.

Moreover, the proactive adoption of these guidelines underscores the healthcare industry's commitment to ensuring patient safety and high-quality care delivery. By demonstrating a responsible and ethical approach to AI, healthcare organizations can build trust with patients and stakeholders, showing that they prioritize safety, quality care delivery, and data privacy and security. Furthermore, following this guidance can streamline the integration of AI into clinical workflows, facilitating smoother transitions, improving overall efficiency, and providing a standardized approach for consistent implementation across healthcare organizations.

In summary, the benefits of aligning with the Joint Commission/CHAI Guidance on the Responsible Use of AI in Healthcare are multifaceted: enhanced patient safety, improved patient outcomes, enhanced data protection, increased trust, and operational efficiency.

Joint Commission has been asked by healthcare organizations, stakeholders, and others for guidance on implementing and using AI in healthcare. In September 2024, Joint Commission hosted a meeting with representatives from groups across the healthcare industry to help identify key themes to address in AI guidance. To better understand what hospitals and health systems would find beneficial in AI guidance, Joint Commission conducted a survey of its accredited hospitals and health systems, including critical access hospitals and ambulatory care settings. Joint Commission reviewed the various health AI frameworks developed by key coalitions and groups, including the Coalition for Health AI (CHAI), the National Academy of Medicine's AI Code of Conduct Principles, the NIST AI Risk Management Framework, and the Bipartisan House Task Force Report on Artificial Intelligence. Joint Commission also participated in meetings on AI use in healthcare and reviewed academic research to gather more information on key areas to address in guidance on responsible use of health AI.

Over the past three years, CHAI has worked with healthcare delivery organizations, technology providers, startups, and patient advocacy groups to adapt broad, industry-agnostic AI frameworks into practical, health-specific best practices that can be applied across the sector. Now, in partnership with Joint Commission, CHAI is uniting these efforts to establish a shared view on what responsible AI use in healthcare should look like—providing consistent, actionable guidance for all stakeholders. The below draft guidance was developed based on themes that emerged from these activities.

This document is intended to guide healthcare organizations through what is important for the responsible implementation and use of AI-enabled tools by healthcare organizations. This guidance is not intended to direct the development of AI tools or validate the effectiveness of AI tools themselves. Both Joint Commission and CHAI are focused on the

safe, responsible deployment and operation of AI tools in the healthcare delivery setting. We welcome healthcare organization feedback on this high-level guidance, which will feed into the development and release of a series of community- and resource-informed Responsible Use of AI Playbooks. These playbooks will build on and operationalize this guidance, including recommended baseline controls, examples, and challenges based on community feedback. These playbooks will be practical resources to guide health systems toward aligning with Responsible Use of AI guidance. A voluntary Joint Commission Responsible Use of AI certification program will be developed based on these playbooks.

By design, AI applications for healthcare are advancing rapidly and will continue to develop. Ultimately, it is important to equip healthcare organizations of all sizes with the guidance they need to responsibly implement and use AI tools and the shared understanding of the elements that make up responsible use in healthcare.

Elements of Responsible Use of AI in Healthcare (RUAIH™)

1. AI Policies and Governance Structures
2. Patient Privacy and Transparency
3. Data Security and Data Use Protections
4. Ongoing Quality Monitoring
5. Voluntary, Blinded Reporting of AI Safety-Related Events
6. Risk and Bias Assessment
7. Education and Training

1. AI Policy and Governance Structures

Healthcare organizations should establish policies and procedures for implementing and using AI and a governance structure to manage the responsible use of health AI in their organization, including a mechanism to keep the hospital's governing body updated on uses, outcomes, and potential adverse events.

There should be a formal governance structure responsible for risk-based and organizationally appropriate oversight of health AI tools involved in direct or indirect patient care, care support services, and care-relevant healthcare operations and administrative services. Governance structures should include a designated individual(s) with appropriate technology expertise, ideally in AI if available, to lead implementation and use of AI tools across the healthcare organization.

The AI governance structure does not need to be its own standalone team but should be responsible for aiding in the risk-based management of third-party and internally developed AI tools or AI-embedded tools, including selection, implementation, risk management, lifecycle management, compliance, and oversight, as appropriate. This team could include individuals with the following expertise as appropriate: executive leadership, regulatory/ethical compliance, information technology (IT), safety/incident reporting, relevant clinical/operational expertise, cybersecurity and data privacy needs, and stakeholders reflecting the needs of impacted populations (e.g., staff, providers, patients, caregivers, etc.). The AI team should develop policies and procedures on the review, implementation, and use of AI tools, ethical standards for AI use, safety and risk protocols, data use and privacy practices, and equitable use and access. These AI policies and procedures should be aligned with other related internal policies and with external regulatory and ethical frameworks. They should be regularly reviewed and updated as objectives, strategy, internal policies, and external regulations shift. The fiduciary board of the healthcare organization should be regularly updated on AI use and its outcomes in healthcare.

Rationale: An AI governance structure and policies provide a systematic approach to implementation, evaluation, and use of AI tools. Importantly, governance creates accountability which will help to drive the safe use of AI tools.

2. Patient Privacy and Transparency

Healthcare organizations should have policies in place regarding data access, use, and protection as well as consumer transparency disclosures or education regarding AI-enabled tools.

Protecting the privacy of patient data is critical for all facets of healthcare operations, particularly AI, which relies on large datasets for optimal function. Additionally, uptake of AI tools relies in part on consumer confidence in providers and AI tools, which will erode without adequate privacy protection or if patients are unaware of the use of AI in healthcare decisions. Healthcare organizations must ensure that patient data is protected from unauthorized use or release.

To ensure consumer confidence and support, data that could potentially publicly identify an individual(s) must be protected from unauthorized disclosure. Organizations should institute policies and procedures to protect the privacy of patient data, in accordance with applicable state and federal laws and regulations.

Furthermore, to build consumer confidence, support, and adoption of AI tools, organizations should develop a mechanism to disclose and educate patients and their families on the use and benefit of these tools. Hospital AI policies should address transparency for both hospital staff and patients, including how AI tools are used. When appropriate, patients should be notified when AI directly impacts their care and how their data may be used in the context of AI. Where and when relevant, consent should be obtained.

Rationale: Hospitals and health systems have a wealth of healthcare data. AI tools may rely on patient data for training datasets, produce entirely new datasets built on patient data, or provide information in a patient's data record. To protect patient privacy and ensure consumer confidence, healthcare organizations must ensure the data privacy and the transparency of AI use to patients and staff.

3. Data Security and Data Use Protections

Healthcare organizations have specific obligations to protect data from unauthorized access or theft. AI tools and the large datasets they rely on and produce emphasize this need. Healthcare organizations should take steps to promote data security and establish requirements within their data use agreements to limit the permissible uses of exported data. This applies to employees, contractors, or third-party vendors who may have access to patient data or other sensitive information in the healthcare organization, including datasets used for algorithm training.

When deploying or procuring AI, each organization must ensure that all uses of patient information comply with HIPAA's Privacy, Security, and Breach Notification Rules. If protected health information (PHI) is involved, organizations should execute appropriate Business Associate Agreements, apply the "minimum necessary" standard, and maintain safeguards commensurate with risk. When data are properly de-identified under HIPAA (via Safe Harbor or Expert Determination) such that there is no reasonable basis to identify an individual, HIPAA's requirements for PHI no longer apply to that dataset; however, organizations should still apply strong protections and contractual guardrails. In those cases, use the elements below—on security practices and data-use agreements—as practical guidance, and continue to watch for re-identification risks and any applicable state or contractual obligations. This is especially pertinent in the context of AI tools, where deidentified data may be used to train, tune, or test AI tools, and data misuse can have major implications.

Data theft and unauthorized use of data will jeopardize patient privacy and put organizations at risk. Elements for protecting data should at least include the following:

- Encryption: Encrypt all patient data, both in transit and at rest, to prevent unauthorized access.
- Access Controls: Implement strict access controls to ensure that only authorized personnel have access to sensitive data. This includes regularly auditing access logs.

- Regular Security Assessments: Conduct regular security assessments and vulnerability scans to identify and address potential security risks and mitigate known vulnerabilities and risks.
- Incident Response Plan: Develop and maintain an incident response plan to address data breaches and other security incidents promptly and effectively.

Healthcare organizations should consider including the following elements in data use agreements:

- Permitted Uses: Clearly define the permissible uses of exported data in data use agreements. Prohibit the use of data for purposes other than those explicitly stated in the agreement. Organizations should consider expanding permissible/prohibited uses to include rights around model outputs, local performance data, and monetization of data.
- Data Minimization: Ensure that only the minimum necessary data is exported and used for the specified purposes.
- Prohibition of Re-identification: Explicitly prohibit the re-identification of de-identified data in data use agreements.
- Third-Party Obligations: Require third-party vendors to comply with all data security and privacy requirements outlined in the agreement, including encryption, access controls, and regular security assessments.
- Audit Rights: Reserve the right to audit third-party vendors for compliance with the data use agreement and impose penalties for non-compliance.

Healthcare organizations should consider adopting Joint Commission's Responsible Use of Health Data (RUHD™) framework, which promotes appropriate guardrails for the secondary use of health data, and includes having an oversight structure, patient transparency, algorithm validation, permitted uses, data controls, and a deidentification process.

Rationale: Protecting data security and establishing clear requirements within data use agreements are crucial for safeguarding patient privacy, maintaining consumer confidence in healthcare organizations, and protecting healthcare organizations from risk.

4. Ongoing Quality Monitoring

Healthcare organizations should have a process to monitor and regularly evaluate the safe performance of AI-enabled clinical tools.

During procurement, healthcare organizations should request information from developers/vendors on how AI tools were tested and validated for their intended use, whether they are willing to tune and/or validate a sample that is representative of the deployment context, and how relevant biases were evaluated.

Additionally, once deployed, ongoing, risk-based, and context-appropriate quality monitoring of AI tools in healthcare is essential for several reasons. AI algorithms may have the capacity to learn and adapt over time, data inputs can change or drift over time, and AI tools and their underlying algorithms may be updated periodically. This means AI tool outcomes and performance can change. The dynamic nature of AI tools necessitates ongoing evaluation to ensure that the tools continue to deliver accurate, reliable, and safe results.

AI tools are often developed outside the healthcare organization where they are implemented and may not undergo consistent external review, especially in the local healthcare organization setting. As a result, internal or local monitoring is critical to identify any issues during deployment and use. Necessary and available monitoring resources may come from multiple sources and may vary based on the needs and context of an organization. For example, they may come internally from the healthcare organization, be obtained externally in agreements with third-party vendors, or be part of vendor-agreed-upon support or tooling for AI tools that are device-embedded or system-native (e.g., electronic health record [EHR] native solutions, monitoring dashboards provided by vendors of third-party AI tools).

Note that monitoring responsibility should be discussed as part of third-party procurement and contracting.

Additionally, ongoing monitoring by the user organization helps to identify and mitigate any biases in AI algorithms and is necessary to ensure safe, high-quality outcomes for patients.

Healthcare organizations should develop comprehensive policies that outline the process(es) and identify the responsible parties for monitoring and evaluating AI tools locally. This includes the following:

- Regularly validating and testing AI tools for relevant performance and reliability.
- Evaluating the quality and reliability of AI data, such as comparing AI outputs to a set of known parameters.
- Assessing use-case relevant outcomes and confidence in AI tool outputs.
- Ensuring AI tools rely on up-to-date data.
- Developing an AI dashboard.
- Creating a process for reporting adverse events or ongoing errors to leadership and vendors, as appropriate.

Ongoing post-deployment monitoring should be risk-based and scaled to your setting. Start by asking: How close is this tool to patient care decisions, and what could go wrong if it performs poorly? Tools that inform or drive clinical decisions should be checked more often; tools that simply help with administrative or documentation tasks can be checked periodically. When possible, use structures you already have (quality, patient safety, compliance) rather than creating something new. Additionally, establish clear feedback channels between third-party vendors and those responsible for monitoring the AI tool so that relevant parties stay informed about model changes or updates that might require an unplanned performance check, and so that issues noted locally can be effectively communicated to the vendor.

Rationale: By adopting rigorous monitoring practices, healthcare organizations can safeguard patient privacy, maintain consumer confidence, enhance the overall quality of care, and protect the organization from risk.

5. Voluntary, Blinded Reporting of AI Safety-Related Events

Healthcare organizations should have a process for the voluntary, blinded reporting of AI-safety related events to monitor and regularly evaluate the safe performance of AI tools.

Organizations should engage in confidential, blinded reporting of AI-related safety events to an independent organization that can share information to the field. Voluntary reporting will reduce the potential for stifling regulatory burden that could limit the potential innovations that AI can deliver to healthcare, while providing opportunities for learning and quality improvement across healthcare organizations. Adverse event reporting also allows for sharing critical information without compromising patient privacy or releasing an organization's identity.

Existing structures for reporting and assessing safety and quality issues, both within and outside the healthcare organization, may be useful. Examples may include Joint Commission's sentinel-event process and confidential reporting to federally listed Patient Safety Organizations (PSOs), among several others. When possible, organizations should use existing structures to track and report AI-related incidents and consider updating the kinds of incidents being tracked and reported. AI tools may contribute to a near miss or harm, such as unsafe recommendations, major performance degradation after an update, or biased outputs. Organizations should treat these events like patient safety events by capturing them in internal incident systems and where appropriate, submitting de-identified details through a PSO or other existing channels (and use FDA pathways if the AI is a regulated device). Confidential, blinded sharing through an independent entity enables pattern-finding and rapid learning across institutions while protecting identities and PHI. Efforts such as CHAI's upcoming Health AI Registry

illustrate how voluntary reporting can advance quality improvement and accountability without imposing a new regulatory burden. These processes should involve a designated individual(s) focused on AI quality/safety-related issues.

Rationale: This system promotes the dissemination of knowledge across the industry, helping healthcare providers stay informed about potential risks and best practices. By reducing the potential for introducing new regulatory requirements, this approach encourages innovation and safe integration of AI technologies into healthcare. Voluntary, confidential reporting of AI-related adverse events fosters a learning health system and emphasizes the importance of keeping humans in the decision-making process.

6. Risk and Bias Assessment

Healthcare organizations should implement a process to identify and address risks and biases in healthcare AI tools, when possible, especially those that may pose a threat to patient safety or limit access to care.

Healthcare organizations must establish processes to categorize and document the potential risks and biases of health AI tools across relevant domains. Prospective and observed risks and biases of AI tools and systems can impact the overall value of the AI tool and its performance and outcomes. Healthcare organizations should request information from vendors on known risks, biases, and limitations of the AI tool. They should also consider asking how bias was evaluated and for which populations. An AI Model Card, such as CHAI's Applied Model Card, is a great way to collect this information and can be adapted to help monitor bias and risks post-deployment.

Bias can exist at any stage of the AI system lifecycle and can be due to several factors, including but not limited to data, the design and features of the underlying models, the methods of training/testing, and use of the AI tool. It is therefore important that in addition to vendor-reported information, healthcare organizations internally, or through partnerships with their third-party vendors, also check for biases when validating local data and post-deployment as part of ongoing monitoring. For instance, if an AI tool was developed using data from primarily younger, healthy patients, use of this tool with older populations may lead to poorer performance and error in care for a subset of patients. This obvious example is easier to identify and mitigate. Other scenarios can be elusive and lead to unrecognized risks to patients, administrative burden, or operational inefficiencies.

Healthcare organizations using AI tools should have a process to evaluate and address use-case relevant biases when possible. Evaluation should occur before deploying the tool and continue as the AI tool is being used to identify any biases in outcomes that may not have been detected initially. This should include using representative training and validation datasets, when possible, and regularly monitoring and/or auditing AI systems to identify and address biases.

To address potential biases, healthcare organizations should do the following:

- Determine whether tools were developed using fit for purpose and appropriately representative training datasets.
- Determine whether the AI tools have undergone bias detection assessments during development.
- Determine whether the algorithms are tested for the specific populations they serve and ensure they are appropriately tuned and/or tested on local data.
- Regularly monitor and audit AI tools to identify and mitigate or manage biases when appropriate.

Rationale: AI tools may not perform well in specific settings or disease conditions or may not be generalized to larger populations if AI training data lacks diversity or bases predictive output on biased associations. This can lead to safety errors, misdiagnoses, administrative burden, operational inefficiencies, compromised quality, and organizational risk.

7. Education and Training

Healthcare organizations should provide basic education and training tools to healthcare providers, ensuring they understand the benefits of AI and can be partners in protecting against potential risks. Healthcare organizations should train clinicians and staff who will leverage AI-enabled technology on the proper use of AI tools and any limitations or guidelines for use.

As part of responsible implementation of AI tools, healthcare organizations should provide basic and use-case-specific education and role-specific training to users of health AI tools, which may include providers and staff. Healthcare organizations should, at a minimum, define and document how users of the AI system will be given relevant AI tool and system documentation and role-specific training to ensure that AI systems are used and monitored appropriately, safely, and effectively. Staff and providers should know where to gain access to relevant information about the AI tool and its use and what the organization's AI policies and procedures are.

In addition to AI-system specific training and user education, healthcare organizations should consider education initiatives focused on AI literacy and change management to upskill all staff and promote safe, informed adoption. AI literacy initiatives should provide a foundational framework for understanding basic principles of AI and machine learning and their uses, risks, and benefits. General education may also include establishing a common terminology around AI tools and the organization's policies and procedures guiding AI use.

Healthcare organizations should evaluate when clinicians and staff need to be trained on specific AI tools prior to implementation and if certain AI tools may need regular training.

By educating clinicians and staff, hospitals can foster a collaborative environment where staff members are not only well versed in the advantages of AI but are also vigilant in identifying and mitigating potential biases or errors. This approach will support the safe, effective integration of AI into healthcare practices and staff workflows, promoting better outcomes and maintaining high standards of care.

Rationale: With the proliferation of AI tools in healthcare, clinicians and staff members are encountering a growing number of AI tools throughout the workplace. Providing training and education helps to ensure safe implementation and integration of AI tools into the clinical workflow.