User Guide

Risk Categorization Tool

A tool for health systems (of any size) to evaluate (1) Life & Patient Safety and (2) Technology & Data risks for health AI solutions

CHAI
Coalition for Health AI

January 2026

Contents

Overview

CHAI defines risk management for health AI solutions in four phases

1 Risk *Categorization*

Classify AI solutions as Low, Medium, or High risk during pre-deployment. This determines the level of rigor needed for later assessments and controls.

2 Risk *Assessment*

For AI solutions with (at least) High risk, conduct a detailed analysis based on your organization's risk tolerance.

3 Risk *Mitigation*

Implement and document actions (risk mitigation controls) to reduce identified risks.

4 Risk *Monitoring*

Continuously monitor the AI solution's performance and safety over time.

Current Focus

The Risk Categorization Tool helps health systems (of any size) evaluate (1) Life & Patient Safety and (2) Technology & Data risks.

Description

Tool Name: Risk Categorization Tool

Primary Risk Domains

(1) Life & Patient Safety: risks related to the life or safety of a patient or group

(2) Technology & Data: flaws, failures, vulnerabilities, or technical complexities related to AI systems, privacy or cybersecurity (confidentiality, integrity, and availability)

Who Should Use It

Health systems (any size) and teams responsible for pre-deployment risk review. Depending on team structure, one or more members may rate one or more modifiers. We recommend that all applicable modifiers are completed.

Purpose

This tool supports early AI governance by helping identify Low, Medium, or High risk levels across several risk modifiers related to (1) life & patient safety and (2) technology & data. The Tool does not produce a single risk score. The Tool highlights key dimensions of risk to guide further assessment, mitigation, and monitoring. Important to note: if your AI solution qualifies as Software as a Medical Device (SaMD), follow FDA regulations and guidance. This Tool applies to all other AI solutions.

What You Need Before Starting

Use CHAI's Applied Model Card or similar documentation to gather details about: Intended Use and Workflow, Primary Users, Target Patient Population, and other relevant context. As an example, leverage the CHAI Applied Model Card section "Uses and Directions" for general use case information, "AI System Facts" to support Risk Modifier #05 for Life & Patient Safety, and "Ongoing Maintenance" to support Risk Modifier #06 for Life & Patient Safety.

How to Score & Interpret Results

Each risk modifier is scored independently as Low, Medium, or High. If any modifier is High, a detailed risk assessment (hazard, harm, probability) is strongly recommended.

Use the results to plan mitigation strategies aligned with the risk levels of each modifier.

How to Use

Refer to the following **Example (pages 4-9)** for a step-by-step tutorial and completed illustration.

Risk Categorization Tool

Example

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

AI-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, and managing patient appointments. These products allow patients to select appointment times and providers to manage their calendars, reduce no-shows, and optimize clinic workflows. Typical features include automated reminders, real-time availability updates, and integration with EHRs. Human confirms appointment once scheduled

Step 1: Align on Use Case

Review the AI solution's use case using CHAI's Applied Model Card or equivalent documentation.

		Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
		pact on individual patient back-end functions such fice administrative tasks, alth analysis, or workflow	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care	AI solution output always reviewed by provider before any action taken	AI solution output has optional human in loop review by provider before any action taken	AI solution output is never reviewed by provider before an action is taken	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk	AI solution has no direct impact and has no affect on patient harm	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage)	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context	Used with patients who are noncomplex and stable	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians)	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally)	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	AI solution used in outpatient and non-critical settings (e.g., outpatient)	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	AI solution used in life-critical settings (e.g., emergency department)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk	There will be limited time for reaction and response planning before serious consequences of the risk	There will be very little or no time for reaction and response planning before serious consequences of the risk	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area)	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics)	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	The AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	The AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	The AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	

Risk Categorization Tool

Example

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

AI-assisted Patient Scheduling software (e.g., scheduling chatb g, and managing patient appointments. These products allow patients to select appoi ws, and optimize clinic workflows. Typical features include automated reminders, real-t intment once scheduled

Step 2: Review Risk Modifiers

Assess each risk modifier individually. Use the provided definitions for Low, Medium, and High risk.

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Distance From Patient** How physically or operationally close the AI solution is to the patient	No direct impact on individual patient care, support back-end functions such as back office administrative tasks, population health analysis, or workflow optimization	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care	AI solution output always reviewed by provider before any action taken	AI solution output has optional human in loop review by provider before any action taken	AI solution output is never reviewed by provider before an action is taken	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk	AI solution has no direct impact and has no affect on patient harm	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage)	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context	Used with patients who are noncomplex and stable	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians)	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally)	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	AI solution used in outpatient and non-critical settings (e.g., outpatient)	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	AI solution used in life-critical settings (e.g., emergency department)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk	There will be limited time for reaction and response planning before serious consequences of the risk	There will be very little or no time for reaction and response planning before serious consequences of the risk	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area)	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics)	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	The AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	The AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	The AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	

Risk Categorization Tool

Example

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

AI-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, and manag
appointments. These products allow patients to select appointment times and providers to manage their calendars, reduce no-shows, and opt
workflows. Typical features include automated reminders, real-time availability updates, and integration with EHRs. Human confirms appointment once

Step 3: Determine Risk Level

Record the team's ratings, including rationale and supporting evidence, as able.

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Distance From Patient** How physically or operationally close the AI solution is to the patient	No direct impact on individual patient care, support back-end functions such as back office administrative tasks, population health analysis, or workflow optimization	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction	Low 5 Medium 0 High 3	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care	AI solution output always reviewed by provider before any action taken	AI solution output has optional human in loop review by provider before any action taken	AI solution output is never reviewed by provider before an action is taken	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk	AI solution has no direct impact and has no affect on patient harm	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage)	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context	Used with patients who are noncomplex and stable	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians)	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally)	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	AI solution used in outpatient and non-critical settings (e.g., outpatient)	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	AI solution used in life-critical settings (e.g., emergency department)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk	There will be limited time for reaction and response planning before serious consequences of the risk	There will be very little or no time for reaction and response planning before serious consequences of the risk	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area)	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics)	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	The AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	The AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	The AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	

Risk Categorization Tool

Example

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

AI-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, ar
appointments. These products allow patients to select appointment times and providers to manage their calendars, reduce no-shows,
workflows. Typical features include automated reminders, real-time availability updates, and integration with EHRs. Human confirms appointm

Step 4: Document Risk Level

Based on the most frequently selected risk level from team ratings, assign an overall risk level (Low, Medium, High) for each risk modifier in the "Response" column.

*Based on the responses to each modifier (if you have more than one team member responding), consider toggling the "Response" field using the majority response. If there is a tie, consider discussing further to gain majority consensus or defaulting to the higher risk category.

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Distance From Patient** How physically or operationally close the AI solution is to the patient	No direct impact on individual patient care, support back-end functions such as back office administrative tasks, population health analysis, or workflow optimization	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction	Low 5 Medium 0 High 3	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care	AI solution output always reviewed by provider before any action taken	AI solution output has optional human in loop review by provider before any action taken	AI solution output is never reviewed by provider before an action is taken	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk	AI solution has no direct impact and has no affect on patient harm	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage)	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context	Used with patients who are noncomplex and stable	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians)	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally)	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	AI solution used in outpatient and non-critical settings (e.g., outpatient)	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	AI solution used in life-critical settings (e.g., emergency department)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk	There will be limited time for reaction and response planning before serious consequences of the risk	There will be very little or no time for reaction and response planning before serious consequences of the risk	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area)	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics)	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	The AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	The AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	The AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	

Risk Categorization Tool

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

AI-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, and managing patient appointments. These products allow patients to select appointment times and providers to manage their calendars, reduce no-shows, and optimize clinic workflows. Typical features include automated reminders, real-time availability updates, and integration with EHRs. Human confirms appointment once scheduled

Step 5: Address Risk Modifiers

Apply organizational risk mitigation controls for each risk modifier

Note: CHAI has not yet developed a Risk Assessment – use your organization's processes.

If risk modifier #01 is Low risk, apply your organization's Low risk mitigation controls; if any risk modifier is rated High, conduct a rigorous risk assessment for that modifier.

Example

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Distance From Patient** How physically or operationally close the AI solution is to the patient	No direct impact on individual patient care, support back-end functions such as back office administrative tasks, population health analysis, or workflow optimization	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction	Low 5 Medium 0 High 3	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care	AI solution output always reviewed by provider before any action taken	AI solution output has optional human in loop review by provider before any action taken	AI solution output is never reviewed by provider before an action is taken	Low 6 Medium 2 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk	AI solution has no direct impact and has no affect on patient harm	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage)	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis)	Low 3 Medium 5 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context	Used with patients who are noncomplex and stable	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians)	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	Low 0 Medium 2 High 6		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally)	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	Low 0 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	AI solution used in outpatient and non-critical settings (e.g., outpatient)	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	AI solution used in life-critical settings (e.g., emergency department)	Low 1 Medium 4 High 2		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk	There will be limited time for reaction and response planning before serious consequences of the risk	There will be very little or no time for reaction and response planning before serious consequences of the risk	Low 4 Medium 0 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area)	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics)	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	Low 0 Medium 3 High 1		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	The AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	The AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	The AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low 0 Medium 3 High 1		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low 0 Medium 4 High 0		Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	

Risk Categorization Tool

Example

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

AI-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, and managing patient appointments. These products allow patients to select appointment times and providers to manage their calendars, reduce no-shows, and optimize clinic workflows. Typical features include automated reminders, real-time availability updates, and integration with EHRs. Human confirms appointment once scheduled

Step 6: Complete All Risk Modifiers

Ensure all risk modifiers are reviewed, categorized, and documented. Complete all risk modifiers for (1) Life & Patient Safety and 2) Technology & Data.

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Distance From Patient** How physically or operationally close the AI solution is to the patient	No direct impact on individual patient care, support back-end functions such as back office administrative tasks, population health analysis, or workflow optimization	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction	Low 5 Medium 0 High 3	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care	AI solution output always reviewed by provider before any action taken	AI solution output has optional human in loop review by provider before any action taken	AI solution output is never reviewed by provider before an action is taken	Low 6 Medium 2 High 0	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk	AI solution has no direct impact and has no affect on patient harm	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage)	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis)	Low 4 Medium 4 High 0	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context	Used with patients who are noncomplex and stable	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians)	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	Low 0 Medium 2 High 6	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities	Low 0 Medium 0 High 0	*Abstain b/c don't have this use case information*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally)	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	Low 0 Medium 0 High 0	*Abstain b/c don't have this use case information*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	AI solution used in outpatient and non-critical settings (e.g., outpatient)	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	AI solution used in life-critical settings (e.g., emergency department)	Low 1 Medium 4 High 2	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk	There will be limited time for reaction and response planning before serious consequences of the risk	There will be very little or no time for reaction and response planning before serious consequences of the risk	Low 4 Medium 0 High 0	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area)	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics)	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	Low 0 Medium 3 High 1	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low 0 Medium 3 High 1	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low 0 Medium 4 High 0	*Detailed rationale, artifacts, and supporting evidence, as able.*	Low Medium High N/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	*Additional notes and comments*

Risk Categorization Tool

Risk Domain:

Life & Patient Safety

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Distance From Patient** How physically or operationally close the AI solution is to the patient.	No direct impact on individual patient care, support back-end functions such as back office administrative tasks, population health analysis, or workflow optimization.	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots.	AI solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; AI solution is directly involved in patient care/patient interaction.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
02	**Human In The Loop** The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before they affect patient care.	AI solution output always reviewed by provider before any action taken.	AI solution output has optional human in loop review by provider before any action taken.	AI solution output is never reviewed by provider before an action is taken.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
03	**Consequences Of Failure Or Error** The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk.	AI solution has no direct impact and has no affect on patient harm.	Errors may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but does not cause long-term damage).	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., AI misinterprets critical diagnostic imaging or fails to detect sepsis).	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
04	**Patient Population Vulnerability** The degree to which the patient population affected is vulnerable (e.g., pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context.	Used with patients who are noncomplex and stable.	Used with patients who are medically complex but stable (e.g., patients with heart failure but on stable medication, being seen by primary care physicians).	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state).	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
05	**Level Of Difficulty Monitoring AI Solution Output** How robust is the AI solution's monitoring capabilities? How resource intensive will the AI solution be to monitor output and performance?; Depends on both the AI solution provider and health system capabilities.	Embedded real-time monitoring and/or capability of real-time monitoring.	AI solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports.	Monitoring needs to be developed before implementation of solution; and/or manual monitoring that requires resource intensive activities.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
06	**Data Transparency** The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution.	Health system has complete access to training data of the underlying model(s) for the AI solution; lowest level of detail for the data/datasets are shared and available (e.g., AI solution developed internally).	Health system has partial access to training data of the underlying model(s) for the AI solution; some level of detail for the data/datasets are shared and available.	Health system has no access to training data of the underlying model(s) for the AI solution; no components of the data/datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable).	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
07	**Clinical Level Of Care** Does the AI solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)?	AI solution used in outpatient and non-critical settings (e.g., outpatient).	AI solution used in inpatient or urgent, but non-critical settings (e.g., inpatient).	AI solution used in life-critical settings (e.g., emergency department).	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
08	**React Time** Assuming the AI output is incorrect, how quickly a decision or intervention can be made	There will be time for reaction and response planning before serious consequences of the risk.	There will be limited time for reaction and response planning before serious consequences of the risk.	There will be very little or no time for reaction and response planning before serious consequences of the risk.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
09	**Breadth Of Potential Harm** The breadth of potential harm the AI solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across locations or institutions.	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot, one site, one clinic, or limited geographical area).	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes prediction across outpatient clinics).	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization).	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
10	**Integrated Error Propagation Risk** The degree to which the AI solution's integration within the broader health IT environment increases the potential for errors to cascade across systems, workflows, and clinical decisions. This includes both the breadth of technical integration and the depth of inter-dependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. Errors are unlikely to spread beyond the immediate user or use case.	AI solution is integrated into specific modules or workflows but has limited cross-functional connections. Errors could impact related components but are unlikely to cause widespread disruptions.	AI solution is deeply embedded across multiple systems and workflows. Its outputs are widely relied upon and shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
11	**Population Sensitivity Or Disparity Risk** The risk that the AI solution could exacerbate health disparities or biases affecting sensitive populations based on race, gender, SES, etc.	There is minimal to no risk of the AI solution's output contributing to health disparities.	There is risk of the AI solution contributing to health disparities, especially if mitigation strategies are not implemented effectively or continuously evaluated.	There could be significant risk of contributing to health disparities, such as high potential to cause harm through unequal diagnosis, treatment, or outcomes; the system could reinforce or worsen existing healthcare inequities, especially for vulnerable groups.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	

Risk Categorization Tool

Risk Domain:

Technology & Data

Primary Audience:

Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments
01	**Use Of Sensitive Data** The degree of risk depends on whether AI solutions rely on synthetic or Expert-Determined data (low risk), HIPAA Safe Harbor de-identified data with residual re-identification potential (medium risk), or identified PII/PHI directly exposed during use (high risk).	Only synthetic, nonconfidential, or nonproprietary data – or data de-identified under Expert Determination (per HIPAA) – are used. No individually identifiable elements are present, and privacy or re-identification risks are minimal (residual risk as determined by expert statistical analyses), making residual confidentiality risk low.	Data are de-identified following HIPAA Safe Harbor standards, with direct identifiers removed but some residual re-identification risk remaining (e.g., through data linkage), or falls under a HIPAA exception. While confidentiality risks are reduced, technical and organizational safeguards are still needed to maintain integrity and prevent unintended disclosures. Unlike Expert Determination, Safe Harbor does not provide statistical proof of minimal risk.	Data contain personally identifiable information (PII), protected health information (PHI), or other confidential or proprietary content and does not fall under a HIPAA exception. Identifiers or sensitive elements are directly available to systems or vendors, creating elevated confidentiality and privacy risks if safeguards fail.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
02	**Accuracy, Completeness, And Veracity Of Data Used For AI Model Training And Operation** How accurate, complete, and veracious are the data sources and pipelines feeding the AI model? Consider validation, freshness of updates, and conformance to quality and interoperability standards.	Training data come from highly reputable and trusted sources (e.g., well-established health systems, certified registries, national datasets). Sources are subject to automated integrity checks (e.g., missingness, duplication, range or semantic validation). When appropriate, training data are standardized to common formats or ontologies (e.g., ICD, SNOMED, LOINC, FHIR). Update frequency keeps data current, with little risk of corruption or degradation.	Training data originate from sources that are generally reputable but may have known limitations or variability (e.g., vendor-managed datasets, multi-site extracts with inconsistent coding practices). Gaps in data quality exist – such as lagged refresh cycles, incomplete fields, or partial mapping to standards – but mitigation measures (e.g., cross-checks, imputation, reconciliation against reference datasets) are applied. While usable, confidence in long-term veracity and timeliness is reduced compared to highly reputable sources.	Training data are drawn from unreliable, poorly governed, or opaque sources (e.g., siloed local systems with little oversight, proprietary vendor datasets with unclear provenance, or ad-hoc data pulls). Data are stale or outdated, and quality defects (e.g., high missingness, unvalidated values, inconsistent identifiers) frequently enter the pipeline. Training data are fragmented or stored in proprietary formats with little or no harmonization, even when appropriate, posing significant risks to the reliability of downstream AI outputs.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
03	**Sufficiency And Representativeness Of Data Used For AI Model Training And Operation** How representative and sufficient are the training and/or testing datasets – whether originating from the organization, health system or solely from the vendor – for the intended use case, considering breadth, depth, and alignment with the populations served? Note: Bias is often unavoidable, however transparency is important to help organizations determine appropriate mitigation steps (e.g., tuning with more representative data, training staff for limited/cautious use, assistive vs. autonomous implementation preferences, etc.)	Data are large, longitudinal, and use case representative across key variables, with adequate samples for relevant subpopulations. There is high confidence in the data's ability to support reliable and bias mitigated model performance.	Data are adequate for the primary population but have limitations in depth (e.g., no historical patient data captured) or breadth (e.g., few sites, missing subgroups). These limitations are documented, understood, and mitigated through supplemental data, external model validation, or adjusted model use. Risk depends on whether the organization owns the data or relies on vendor sources.	Data are small, narrow in scope, or lack use case representativeness. There is a significant risk of poor generalizability or unintended harm to one or more subgroups. Risks are elevated when vendor-provided data provenance is not shared, data cannot be supplemented for intended use, or model performance cannot be externally validated across relevant datasets.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
04	**AI Model Security Vulnerabilities For Technology** The extent to which an AI model is exposed to security risks based on its technical deployment surface (application-bound, internal, or internet-facing).	The model operates only within the application boundary, with no external interfaces or integrations, minimizing exposure to external threats. The isolated, self-contained design provides high resiliency and assured business continuity (e.g., the application continues to function even if external networks are disrupted).	The model is internal-facing (accessible within the organization's network), where exposure to insider threats or lateral movement attacks is possible if security controls are weak. The deployment has moderate resiliency and business continuity, as operations depend on the stability of the internal network and safeguards (e.g., if the hospital intranet fails, AI services are disrupted but external systems remain unaffected).	The model is internet-facing, exposed to external networks, making it a target for adversarial inputs, denial-of-service attempts, or unauthorized access. The deployment has low resiliency and business continuity, as disruptions or compromises could cascade widely across users and systems (e.g., an outage of the internet-facing service could halt critical clinical workflows across multiple sites).	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
05	**AI Model Security Vulnerabilities For Data Handling** The extent to which an AI model is exposed to security risks based on its data handling practices (static training, periodic updates, or continuous unsupervised learning in production).	Training occurs entirely outside of the production environment and the model is static (not updated post-deployment), reducing the risk of data poisoning or integrity loss.	Training is performed outside of production but the model is periodically updated with new data, creating potential vulnerabilities if update pipelines are not rigorously secured or validated.	The model continuously learns in production through unsupervised ingestion of live data, raising risks of data poisoning, drift, and unmonitored integrity breaches that directly affect outputs.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
06	**AI Model Lifecycle Management And Updates** How effectively are AI models planned, developed, updated, validated, and configured throughout their lifecycle, separate from real-time monitoring? This includes retraining cadence, validation of updates, configuration and version control, and controls for adaptive vs. locked algorithms.	AI models follow a documented lifecycle for design, retraining, validation, and controlled deployment of updates. Configuration management and version control are fully automated and auditable, with clear rollback and approval steps. For example, model updates are versioned, tested, staged (e.g., canary deployment), validated, and have rollback capability.	AI models have lifecycle processes, but they are applied inconsistently. Updates and validations may occur only for major releases or rely on partial automation, and configuration management has gaps in traceability or rollback. For example, model updates are controlled but lack rollback or partial validation.	AI models lack a structured lifecycle or update process. Retraining and validation occur ad hoc, and configuration or version records are incomplete or absent, leaving models at risk of remaining outdated or improperly tuned. For example, model updates may have uncontrolled retraining, no version tracking, or updates are pushed without validation.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
07	**AI Monitoring, Incident Detection And Response** How well can the organization (vendor, implementing organization, or combination) appropriately monitor AI models and detect, classify, and respond to incidents – including output anomalies, emergent bias or drift, hallucinations, security breaches, or AI impacts on dependent systems? Clarifies integration with enterprise IT incident management vs. AI safety incidents.	Automated, timely (periodic or continuous based on relevance to use case) monitoring tracks performance, security, and emergent bias or drift. Anomaly alerts have predefined severity levels and integrate with both AI safety and enterprise IT incident workflows, and clear response protocols guide investigation and mitigation.	Some monitoring exists, often in batch or aggregate form. Detection of anomalies, emergent bias, or drift may rely on manual review, and response processes are only partially defined or linked to enterprise incident management.	Monitoring is minimal or absent, allowing emergent bias, drift, or output errors to persist undetected. There is no clear separation between IT operational incidents and AI safety incidents, and escalation or response procedures are missing.	Low Medium High		○ Low ○ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ○ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	
08	**AI Detection And Traceability** How well can end users (providers, patients, staff, as appropriate) detect when and how AI is influencing outputs, decisions, or workflows (e.g., transparency into the AI solutions use and integration into downstream tasks)?	All AI-generated outputs are clearly labeled and traceable (for both human-facing outputs and machine-to-machine interactions); users are informed when AI is involved in recommendations or automation. Full audit trails exist.	Some AI outputs are identifiable, but others may blend with non-AI content; labeling or auditability is partially implemented (though, partial audit trails can make investigations after an incident more difficult).	AI involvement is not visible to end users; decisions or recommendations may be made or influenced by AI without detection, leading to risks such as automation bias, silent failures, or unsafe overrides. Invisible AI influence can conflict with regulatory labeling or patient consent requirements.	Low Medium High		○ Low ◉ Medium ○ High ○ N/A	○ Apply organizational low risk mitigation controls ◉ Apply organizational medium risk mitigation controls ○ Apply organizational high risk mitigation controls	

Authors

Special thank you to the contributing members of the CHAI Risk Work Group, including:

Eric Henry, Brooke & Associates

Ashley Beecy, Sutter Health

Shubham Goel, Sutter Health

Britt Anderson, UnitedHealth Group

Nicoleta Economou, Duke

Christina Silcox, Duke

Lindsay Mico, Providence

Vivek Tomer, Providence

Taylor Anderson, Stanford

Howard Strasberg, Wolters Kluwer

Noelle Vidal, University of California

Joshua Miller, University of Rochester Medical Center

Larry Vernaglia, Foley & Lardner LLP

Teresa Luke, HealthPartners

Selvi Ramalingam, Emory Healthcare

Monica Kedzierski, Claritev

Gary Herrington, CareSouth

Sam Pinson, Nixon Law

Kathleen Snyder, Lumeris

Alaap Shah, Epstein Becker & Green, P.C.

Taylor Rhoades, Mercy

Ruby Chen, Mercy

Brenna Loufek, Mayo

Ben Kaplan, Mount Sinai

Arash Kia, Mount Sinai

Holly Meidl, Ascension

Aslam Merchant, Ascension

Jennifer Sloane, San Joaquin General Hospital

Joseph Izzo, San Joaquin General Hospital

Christine Palermo, Encore Health

January Choy, Memorial Sloan Kettering

Ivan Pan, Memorial Sloan Kettering

Brenton Hill, CHAI

Merage Ghane, CHAI

Anthony DiDonato, CHAI

Greg Shemancik, CHAI

Get in touch at admin@chai.org

Visit our website

Connect on LinkedIn

The previous content is work in-progress and in draft form available for public comment. The mention or sharing of any examples, products, organizations, or individuals does not indicate any endorsement of those examples, products, organizations, or individuals by the Coalition for Health AI (CHAI). Any examples provided here are still under review for alignment with existing standards and instructions. We welcome feedback and stress-testing of the tool in draft form.

The information provided in this document is for general informational purposes only and does not constitute legal advice. It is not intended to create, and receipt or review of it does not establish, an attorney-client relationship.

This document should not be relied upon as a substitute for consulting with qualified legal or compliance professionals. Organizations and individuals are encouraged to seek advice specific to their unique circumstances to ensure adherence to applicable laws, regulations, and standards.

CHAI

Coalition for Health AI