

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

Contents

Overview	
Context and scope of the Risk Categorization Tool	2
Description	
Who the tool is for, when to use it, and how to interpret results	3
Instructions & Example	
Step-by-step instructions for use	4-8
Completed Example (from the CHAI Risk Work Group)	9
Risk Categorization Tool	
Tool template for use	10-11
Authors	
Contributors and authors from the Risk Work Group	12

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.

Current Focus

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

4 *Risk Monitoring*

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

Description

Tool Name: Risk Categorization Tool (v3)

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 #04 for Life & Patient Safety, and "Ongoing Maintenance" to support Risk Modifier #05 for Life & Patient Safety.

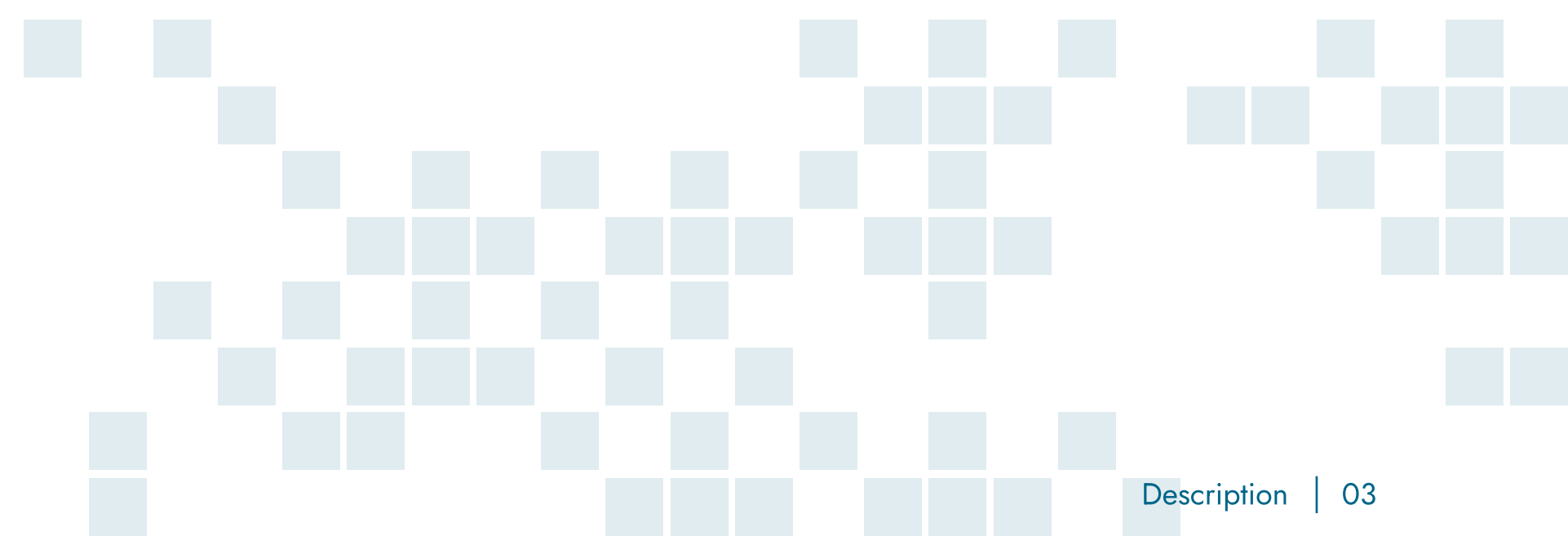
How to Score & Interpret Results

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: **Primary Audience:**
Life & Patient Safety 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
		Impact on individual patient care, such as back-end functions such as scheduling, administrative tasks, health 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 ● Medium ● High		● Low ● Medium ● High ● N/A ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	● Low ● Medium ● High		● Low ● Medium ● High ● N/A ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	Consequences Of Failure Or Error If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	● Low ● Medium ● High		● Low ● Medium ● High ● N/A ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	● Low ● Medium ● High		● Low ● Medium ● High ● N/A ● Need More Info	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 realtime 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 ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	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 ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	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 ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. "Propagation" means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 ● Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	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 ● Need More Info	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 patient safety

Use Case: AI-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient appointments. These products allow patients to select appointment times and workflows. Typical features include automated reminders, real-time availability up

Step 2: Review Risk Modifiers

Before scoring, (a) review the use case in detail; (b) read through all risk modifiers end-to-end so you understand how they relate to each other; (c) score each modifier individually using the predefined Low, Medium, and High risk levels.

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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	Low Medium High			Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	Consequences Of Failure Or Error If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	Low Medium High			Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	Low Medium High			Low Medium High N/A Need More Info	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 realtime 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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. "Propagation" means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	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 Need More Info	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 appointments. These products allow patients to select appointment times and providers to manage their calendars, reduce no-shows, and optimize 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 Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	Low Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	Consequences Of Failure Or Error If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	Low Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	Low Medium High		Low Medium High N/A Need More Info	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 realtime 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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. "Propagation" means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	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 Need More Info	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 appointments. These products allow patients to select appointment times and providers to manage their calendar and workflows. Typical features include automated reminders, real-time availability updates, and integration with EHRs. H

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. After scoring all modifiers, review your responses as a set. Did any of your responses materially change once you had the full picture? If so, update them and document the reason in Notes & Comments. This is a deliberate check on inter-modifier dependencies.

*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 level.

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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	Low Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	Consequences Of Failure Or Error If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	Low Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	Low Medium High		Low Medium High N/A Need More Info	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 realtime 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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. “Propagation” means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	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 Need More Info	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.

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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	Low Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
03	Consequences Of Failure Or Error If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	Low Medium High		Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	Low Medium High		Low Medium High N/A Need More Info	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 realtime 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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
06	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
07	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. “Propagation” means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	
09	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 Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	

Example

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 Medium High 503		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	Low Medium High 620		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	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 If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	Low Medium High 440		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	Low Medium High 026		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	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 Medium High 000		Abstain b/c don't have this use case information	Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments
06	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 400		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments
07	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 031		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. "Propagation" means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 031		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments
09	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 040		Detailed rationale, artifacts, and supporting evidence, as able.	Low Medium High N/A Need More Info	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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
02	Decision Autonomy The extent to which the AI holds independent decision-making authority relative to a human, and therefore how much human judgment stands between the AI's output and a real-world action.	Assistive: Human decides. The AI informs or recommends; a human reviews and makes the decision. Risk is entirely on the human reviewer.	Augmentative: Human supervises. The AI executes within a defined scope; a human is on the loop with the ability to intervene or override. Risk depends on whether oversight is meaningful (clinician has time, context, and authority to override) or nominal (rubber-stamping).	Autonomous: AI decides. The AI makes and acts on the decision without case-by-case human review. Risk shifts almost entirely to the AI system, its validation, and its ongoing monitoring.	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
03	Consequences Of Failure Or Error If AI Output Is Incorrect Assuming the AI produces an incorrect output that contributes to a negative outcome, what is the worst-case severity of patient harm that could result? Clinical outcomes are higher risk.	AI solution has no direct impact. An incorrect output cannot meaningfully harm a patient.	An incorrect output may lead to temporary discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care). An incorrect output may result in temporary or reversible harm that requires medical intervention (e.g., prescribing the wrong medication dose that requires monitoring but no lasting injury).	An incorrect output may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., critical diagnostic imaging misinterpreted, missed sepsis).	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
04	Use Context And Complexity The risk introduced by the setting in which the AI is used (e.g., administrative/back office, outpatient, inpatient, ED) or the complexity of the primary patient population impacted. Note that both the criticality of the use context and the complexity of the patient population can impact a user's time and cognitive resources to review outputs, catch errors, etc.	AI solution is used in non-critical settings (e.g., outpatient or back-office support) and/or primarily on patients who are non-complex and stable (e.g. patients coming for routine annual visit).	AI solution used in inpatient or urgent (non-life-critical) settings, and/or used primarily with patients who are medically complex, but stable (e.g. patients with multiple chronic conditions, but who are stable on medication, and being seen by primary care physician).	AI solution used in life-critical or emergency settings and/or primarily used with patients who are medically complex or unstable (e.g. patient with heart failure that is not stable on medications).	Low Medium High		Low Medium High N/A Need More Info	☐ 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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
06	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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
07	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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
08	Cross-System Propagation Risk The degree to which an incorrect AI output (or unavailability of the AI service) can propagate across integrated systems and workflows, magnifying its downstream impact. “Propagation” means technical and operational cascade across systems.	AI solution is functionally isolated, with minimal integration into other digital systems or workflows. An incorrect output is 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. An incorrect output could affect related components but is unlikely to cause widespread disruption.	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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
09	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 Need More Info	☐ 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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
02	Accuracy Of Data How correct, valid, and conformant are the values in the training and operational data? This modifier addresses semantic validation, identifier consistency, deduplication, and conformance to coding standards and ontologies (e.g., ICD, SNOMED, LOINC, FHIR).	Values are subject to automated validation (e.g., range or semantic checks, deduplication). When appropriate, data are standardized to common formats or ontologies (e.g., ICD, SNOMED, LOINC, FHIR), ensuring consistent, correctly coded values.	Sources are generally reliable, but coding practices may be inconsistent across sites, with only partial mapping to standards. Mitigation measures (e.g., cross-checks, reconciliation against reference datasets) are applied, though confidence in value-level correctness is reduced.	Quality defects such as unvalidated values and inconsistent identifiers frequently enter the pipeline. Data are 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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
03	Completeness Of Data How complete is the data? This modifier addresses missingness, field population, duplication and fragmentation, and coverage relative to the intended population.	Sources are subject to automated integrity checks for missingness and duplication. Fields are well-populated and coverage is sufficient for the intended use, with little risk of gaps or fragmentation.	Gaps in completeness exist, such as incomplete fields or partial coverage, but mitigation measures (e.g., imputation, reconciliation against reference datasets) are applied. While usable, confidence is reduced compared to highly complete sources.	Quality defects such as high missingness frequently enter the pipeline, and data are fragmented across sources, posing significant risks to the reliability of downstream AI outputs.	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
04	Veracity Of Data How truthful, trustworthy, and current are the data sources and pipelines? This modifier addresses source reputability and provenance, governance and oversight, and freshness/refresh cadence.	Training data come from highly reputable and trusted sources (e.g., well-established health systems, certified registries, national datasets). Provenance is clear, and 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), including lagged refresh cycles. 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, posing significant risks to the reliability of downstream AI outputs.	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
05	Data Transparency The clarity, completeness, and accessibility of information about the data sources and datasets used to train, test, and validate the AI solution.	Health system has complete information about the data sources and datasets used to train, test, and validate the underlying model(s); detailed documentation (e.g., data provenance and a data catalog/dictionary) is shared and available (e.g., AI solution developed internally).	Health system has partial information about the data sources and datasets used to train, test, and validate the underlying model(s); some documentation about the data/datasets is shared and available.	Health system has no information about the data sources and datasets used to train, test, and validate the underlying model(s); no documentation about the data/datasets is shared or available (e.g., data provenance and data catalog/dictionary unavailable).	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
06	Sufficiency And Representativeness Of Data Used For AI Model Training And Operation How well do the training and operational data cover the populations, settings, and subgroups the AI will encounter in deployment? Note: Bias is often unavoidable; transparency is important to help organizations determine appropriate mitigation steps (e.g., tuning with more representative data, training staff for limited/cautious use).	Data are large, longitudinal, and representative of the population in which it will be used across key variables and subgroups.	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).	Data are small, narrow in scope, or lack population representativeness. There is a significant risk of poor generalizability.	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
07	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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
08	AI Model Lifecycle Management And Updates How effectively the AI model is designed, developed, retrained, validated, version-controlled, and deployed across its lifecycle. This includes (a) the model lifecycle governance and (b) the data handling regime that drives lifecycle risk.	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. Or, model training is static or performed in a tightly controlled, validated pipeline outside production, reducing the risk of data poisoning or integrity loss.	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. Or, model training is performed outside production, but the model is periodically updated with new data, creating potential vulnerabilities if update pipelines are not rigorously secured or validated.	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 pushed without validation. Or, the model continuously learns in production through unsupervised ingestion of live data, raising risks of data poisoning, drift, and unmonitored integrity breaches.	Low Medium High		Low Medium High N/A Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
09	AI Monitoring, Incident Detection And Response At runtime (post-deployment), how effectively can the organization (vendor, implementing organization, or combination) monitor the AI solution and underlying model(s) to detect, classify, and respond to incidents (e.g., output anomalies, emergent bias or drift, hallucinations, security breaches, downstream impacts on dependent systems, or data integrity issues such as poisoning, drift, or unmonitored integrity breaches at runtime)?	Automated, timely (periodic or continuous based on relevance to use case) monitoring tracks performance, security, and emergent bias or drift. Runtime data integrity safeguards detect poisoning attempts, anomalous data ingest, and integrity breaches. Anomaly alerts have predefined severity levels and integrate with both AI safety and enterprise IT incident management workflows, and clear response protocols guide investigation and mitigation.	Some monitoring exists, often in batch or aggregate form. Detection of anomalies, emergent bias, drift, or data integrity issues 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, output errors, or data integrity breaches (e.g., poisoning, unmonitored integrity loss) 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 Need More Info	☐ Apply organizational low risk mitigation controls ☐ Apply organizational medium risk mitigation controls ☐ Apply organizational high risk mitigation controls	
10	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 Need More Info	☐ 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

Todd F. Dardas, Wolters Kluwer;
UpToDate

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

Scott Ross, Opala

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

Jennifer Foster, San Joaquin General Hospital

Joseph A. Izzo, San Joaquin General Hospital

Christine Palermo, Encore Health

January Choy, Memorial Sloan Kettering

Ivan Pan, Memorial Sloan Kettering

Michael Blumental, HealthLeap

Mariana de Oliveira, Complear

Bitu Behrouzi, MaineHealth Maine Medical
Center; Tufts University School of Medicine

Joseph Seliski, Individual Contributor

Merage Ghane, CHAI

Anthony DiDonato, CHAI

Brenton Hill, 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.

This document is licensed under a **Creative Commons Attribution-Non-Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0)**.

You are free to share this material (copy and redistribute it in any medium or format) under the following terms:

Attribution: You must give appropriate credit, provide a link to the license, and indicate if changes were made.

Noncommercial: You may not use the material for commercial purposes.

No Derivatives: If you remix, transform, or build upon the material, you may not distribute the modified material.

For more information about this license, visit creativecommons.org/licenses/by-nc-nd/4.0/.

