User Guide

Risk Categorization Tool

A tool for health systems (of any size) to evaluate Life and Patient Safety risks for health AI solutions



Contents

Overview

Context and scope of the Risk Categorization Tool	
Description	
Who the tool is for, when to use it, and how to interpret results	
Instructions & Example	
Step-by-step instructions for use	
Completed Example (from the CHAI Risk Work Group)	
Risk Categorization Tool	
Tool template for use	1
Authors	
Contributors and authors from the Risk Work Group	1

Overview

CHAI defines risk management for health AI solutions in four phases

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.

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

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 Life and Patient Safety risks. Future versions will include additional domains, such as Technology, Data, and Financial risks.

Risk Monitoring

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

Description

Tool Name: Risk Categorization Tool

Primary Risk Domain: Life and Patient Safety

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 life and patient safety. 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 other documentation to gather details about: Intended Use and Workflow, Primary Users, Target Patient Population, and other relevant context. In particular, leverage the CHAI Applied Model Card section "Uses and Directions" for general use case information, "Ongoing Maintenance" to support Risk Modifier #06, and "AI System Facts" to support Risk Modifier #05. Lastly, for Risk Modifiers #09-11, the team(s) responsible for predeployment risk review will need to identify how the AI solution will be used in the context of your organization.

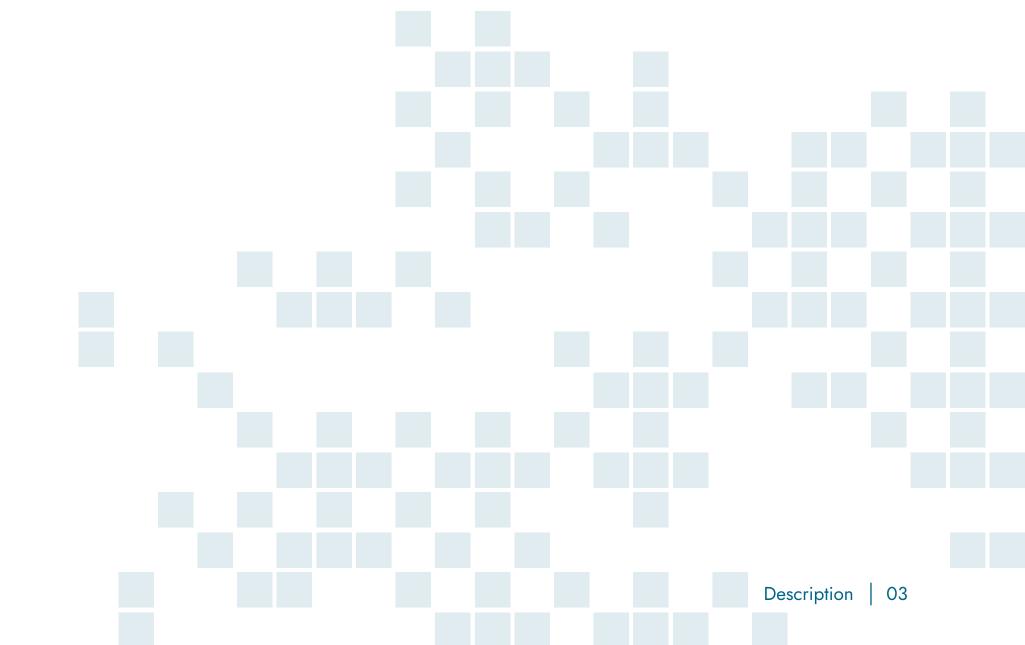
How to Score & Interpret Results

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 Domain: Primary Audience:

Life & Patient Safety Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

Al-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

tep 1: Align on Use Case	Definition	Medium Risk Definition	High Risk Definition	Team Ratings	Rationale/Evidence	Response	Action(s)	Notes & Comments				
Review the Al solution's use case using		Indirect impact on patient care, access to care, or informational use, such as		• Low								
HAI's Applied Model Card or uivalent documentation.		scheduling, transportation, non-clinical		• Medium 0								
			directly involved in patient care/patient interaction	High								
Human In The Loop Al solution	output always reviewed by	Al solution output has optional human	Al solution output is never reviewed by	• Low								
	fore any action taken	in loop review by provider before any action taken		Medium								
overriding the AI solution outputs before they affect patient care				High								
The severity and likelihood of negative no affect on	nas no direct impact and has no patient harm		permanent damage to body structure, disability, or death (e.g., Al	Low								
outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect			misinterprets critical diagnostic imaging	Medium								
information; clinical consequences are higher risk		intervention (e.g., prescribing the wrong medication dose that requires	and the state of t	High								
		monitoring but does not cause long- term damage)										
Patient Population Vulnerability Used with p	patients who are noncomplex	Used with patients who are medically	Used with patients who are medically	Low								
The degree to which the patient population affected is vulnerable (e.g.,		complex but stable (e.g., patients with	complex but unstable (e.g., patients with heart failure and in unstable state)	• Medium 0								
pediatrics, elderly, low health literacy, marginalized groups); Depends on		being seen by primary care physicians)		High								
clinical setting and presentation context												
	real-time monitoring and/or freal-time monitoring		before implementation of solution; and/									
How robust is the AI solution's monitoring capabilities? How resource		monitoring capabilities; periodic	or manual monitoring that requires resource intensive activities	Medium 0								
intensive will the AI solution be to monitor output and performance?;		reports		High								
Depends on both the AI solution provider and health system capabilities												
		Health system has partial access to			• Low							
accessibility of the data sources and for the Al so	olution; lowest level of detail	for the AI solution; some level of detail	data of the underlying model(s) for the Al solution; no components of the data/	• Medium •								
the Al solution available (e	ta/datasets are shared and e.g., Al solution developed				datasets are shared or available (e.g., data provenance and data catalog/dictionary unavailable)	High						
internally)			Cicional y unavanable)									
Clinical Level Of Care Al solution	used in outpatient and non-	At solution used in inpatient or urgent	At solution used in life critical settings									
Does the AI solution operate in a critical setting	ngs (e.g., outpatient)	Al solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)		·		but non-critical settings (e.g., inpatient)		LowMedium				
clinically sensitive or high-risk setting that requires a higher level of care (e.g.,												
inpatient, outpatient, emergency department, etc.)				• High								
		There will be limited time for reaction and response planning before serious		Low								
quickly a decision or intervention can be made	ces of the risk	consequences of the risk	serious consequences of the risk	Medium								
				• High								
n una la anna f		Affects a moderate number of patients		• Low								
solution could cause to patients if it of setting	gs (e.g., rare disease	(e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes	systems (e.g., enterprise-wide triage	Medium								
I DCHOTHS HIGGHGGHY, 7 (33C33 HOW DIGGGHY I	one clinic, or limited	prediction across outpatient clinics)		• High								
iocanons or maniunons												
Integrated Error Propagation Risk The AI solu	ution is functionally isolated,	The AI solution is integrated into	The AI solution is deeply embedded	Low								
The degree to which the AI solution's integration within the broader health IT with minim	nal integration into other ems or workflows. Errors are	specific modules or workflows but has limited cross-functional connections.	across multiple systems and workflows. Its outputs are widely relied upon and	Medium								
environment increases the potential for unlikely to		Errors could impact related components but are unlikely to cause widespread	shared, increasing the chance that a single point of failure could cascade	• High								
workflows, and clinical decisions. This includes both the breadth of technical		disruptions.	across care settings, decisions, or resource allocations.									
integration and the depth of inter- dependence, reflecting how embedded												
the AI solution is and how errors in one part could propagate to others.												
Population Sensitivity Or Disparity Risk There is m	inimal to no risk of the AI	There is risk of the AI solution	There could be significant risk of	• Low								
		contributing to health disparities,		• Medium 0								
affecting sensitive populations based on race, gender, SES, etc.			unequal diagnosis, treatment, or outcomes; the system could reinforce or	• High								
			worsen existing healthcare inequities, especially for vulnerable groups.	J								



Risk Domain:

Life & Patient Safety

Primary Audience:

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

Al-assisted Patient Scheduling software (e.g., scheduling chatb appointments. These products allow patients to select appointments workflows. Typical features include automated reminders, real

Assess each risk modifier individually. Use the ws, and optimize clinic provided definitions for Low, Medium, and High risk. pintment once scheduled

	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	Al solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; Al solution is directly involved in patient care/patient interaction	Low Medium High				
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	Al solution output always reviewed by provider before any action taken	Al solution output has optional human in loop review by provider before any action taken	Al solution output is never reviewed by provider before an action is taken	LowMediumHigh		LowMediumHighN/A		
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	Al solution has no direct impact and has no affect on patient harm	lasting health effects (e.g., minor delays	permanent damage to body structure, disability, or death (e.g., Al misinterprets critical diagnostic imaging	Low Medium High		LowMediumHighN/A		
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	complex but stable (e.g., patients with	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	LowMediumHigh		LowMediumHighN/A		
05	Level Of Difficulty Monitoring Al Solution Output How robust is the Al solution's monitoring capabilities? How resource intensive will the Al solution be to monitor output and performance?; Depends on both the Al solution provider and health system capabilities	capability of real-time monitoring	Al solution includes partial real-time monitoring capabilities; health system still requires partial development of monitoring capabilities; periodic reports	before implementation of solution; and/ or manual monitoring that requires					
06	Data Transparency The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate the AI solution	training data of the underlying model(s) for the AI solution; lowest level of detail	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	data of the underlying model(s) for the AI solution; no components of the data/	• Medium •		LowMediumHighN/A		
07	Clinical Level Of Care Does the Al solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	Al solution used in outpatient and non-critical settings (e.g., outpatient)	Al solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)		LowMediumHigh				
08	React Time Assuming the AI output is incorrect, how quickly a decision or intervention can be made		There will be limited time for reaction and response planning before serious consequences of the risk		LowMediumHigh		LowMediumHighN/A		
09	Breadth Of Potential Harm The breadth of potential harm the Al solution could cause to patients if it performs incorrectly; Assess how broadly the Al solution is deployed across locations or institutions	number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot,	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)	facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma			LowMediumHighN/A		
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 interdependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	with minimal integration into other digital systems or workflows. Errors are	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.	across multiple systems and workflows. Its outputs are widely relied upon and	• Medium •		LowMediumHighN/A		
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.				• Medium 0				

Example

Risk Domain: Primary Audience:

Life & Patient Safety Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

Al-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, and manage 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.

					rationale												
	Risk Modifier	Low Risk Definition	Medium Risk Definition	High Risk Definition	Team Rat	ings	Rationale/Evidence	Response	Action(s)	Notes & Comments							
)1	Distance From Patient How physically or operationally close the	No direct impact on individual patient care, support back-end functions such	Indirect impact on patient care, access to care, or informational use, such as	Al solution has semi-direct involvement in patient care; such as, used by a	Low		Detailed rationale, artifacts, and supporting evidence, as able.										
	Al solution is to the patient	as back office administrative tasks, population health analysis, or workflow	scheduling, transportation, non-clinical informational chatbots	healthcare professional as part of a broader clinical judgment; Al solution is	Medium												
		optimization		directly involved in patient care/patient interaction	High	3											
2	Human In The Loop	Al solution output always reviewed by		Al solution output is never reviewed by	Low												
	The extent to which human oversight is involved in reviewing, verifying, or	provider before any action taken	in loop review by provider before any action taken	provider before an action is taken	Medium												
	overriding the AI solution outputs before they affect patient care				High												
)3	Consequences Of Failure Or Error	Al solution has no direct impact and has	Errors may lead to temporary	Errors may lead to permanent harm.	Low												
J	The severity and likelihood of negative outcomes (e.g., morbidity, mortality) if	no affect on patient harm	discomfort or inconvenience, with no	permanent damage to body structure, disability, or death (e.g., Al	Medium												
	the Al solution fails or provides incorrect information; clinical consequences are		in care); Errors may result in temporary or reversible harm that requires medical	misinterprets critical diagnostic imaging or fails to detect sepsis)	High												
	higher risk		intervention (e.g., prescribing the wrong medication dose that requires		- Tilgii												
			monitoring but does not cause long- term damage)														
4	Patient Population Vulnerability	Used with patients who are noncomplex	Used with patients who are medically	Used with patients who are medically	Low												
	The degree to which the patient population affected is vulnerable (e.g.,	and stable	heart failure but on stable medication,	complex but unstable (e.g., patients with heart failure and in unstable state)	Medium												
	pediatrics, elderly, low health literacy, marginalized groups); Depends on		being seen by primary care physicians)		High												
	clinical setting and presentation context																
)5	Level Of Difficulty Monitoring AI	Embedded real-time monitoring and/or	Al solution includes partial real-time	Monitoring needs to be developed	• Low												
))	Solution Output How robust is the Al solution's	capability of real-time monitoring	Al solution includes partial real-time bility of real-time monitoring still requires partial development of monitoring capabilities; periodic reports	before implementation of solution; and/													
	monitoring capabilities? How resource intensive will the Al solution be to				High												
	monitor output and performance?; Depends on both the AI solution	;			Tilgii												
	provider and health system capabilities																
6	Data Transparency		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		Low												
	The clarity, completeness, and accessibility of the data sources and datasets used to train, test, and validate	for the AI solution; lowest level of detail		Al solution; no components of the data/	Medium												
	the Al solution				High												
7	Clinical Level Of Care	Al solution used in outpatient and non-									Al solution used in life-critical settings	Low					
	Does the AI solution operate in a clinically sensitive or high-risk setting that	critical settings (e.g., outpatient)	but non-critical settings (e.g., inpatient)	but non-critical settings (e.g., inpatient)	but non-critical settings (e.g., inpatient)	but non-critical settings (e.g., inpatient)	(e.g., emergency department)	Medium									
	requires a higher level of care (e.g., inpatient, outpatient, emergency				High												
	department, etc.)																
8	React Time	There will be time for reaction and	There will be limited time for reaction	There will be very little or no time for	Low												
0	Assuming the AI output is incorrect, how quickly a decision or intervention can be		and response planning before serious consequences of the risk		Medium												
	made		consequences of the risk		High												
	Proposition Of Potential House	Affords a single individual on a small	Affects a made mate mumber of mations	Detential for wideenward house conse													
9	The breadth of potential harm the AI	number of patients in a limited number	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple	facilities, populations, or entire health													
	solution could cause to patients if it performs incorrectly; Assess how broadly the AI solution is deployed across		units or clinics (e.g., diabetes		Medium												
	locations or institutions	geographical area)			High												
0	Integrated Error Propagation Risk The degree to which the AI solution's	with minimal integration into other	The AI solution is integrated into specific modules or workflows but has	across multiple systems and workflows.													
	integration within the broader health IT environment increases the potential for	unlikely to spread beyond the	limited cross-functional connections. Errors could impact related components	shared, increasing the chance that a	Medium												
	errors to cascade across systems, workflows, and clinical decisions. This includes both the broadth of technical	immediate user or use case.	disruptions.	single point of failure could cascade across care settings, decisions, or resource allocations.	High												
	includes both the breadth of technical integration and the depth of interdependence, reflecting how embedded			. 5555. 55 GIIOGGIIOIIO.													
	the Al solution is and how errors in one part could propagate to others.																
		The second secon		The second of th													
1	Population Sensitivity Or Disparity Risk The risk that the Al solution could	solution's output contributing to health	There is risk of the AI solution contributing to health disparities,														
	exacerbate health disparities or biases affecting sensitive populations based on	disparities.		unequal diagnosis, treatment, or outcomes; the system could reinforce or	Medium												
	race, gender, SES, etc.			worsen existing healthcare inequities, especially for vulnerable groups.	High												

Risk Domain:

Primary Audience:

Life & Patient Safety

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

Use Case:

Al-assisted Patient Scheduling software (e.g., scheduling chatbot) - for outpatient, primary care clinics - simplifies booking, rescheduling, a 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 appointing the confirmation of the confir

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

Risk Domain:

Primary Audience:

Life & Patient Safety

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

Use Case:

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



Example

Risk Domain: Primary Audience:

Life & Patient Safety Health systems (any size) and teams responsible for pre-deployment risk review

Use Case:

Al-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 Modifiers Ensure all risk modifiers (1–11) are reviewed, categorized, and documented.

									ed, calegorized, and documented.				
	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 Al solution is to the patient	the care, support back-end functions such as back office administrative tasks, scheduling, trans	Indirect impact on patient care, access to care, or informational use, such as scheduling, transportation, non-clinical informational chatbots	Al solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; Al solution is directly involved in patient care/patient interaction	LowMediumHigh3	Detailed rationale, artifacts, and supporting evidence, as able.	LowMediumHigh	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments				
				merachon			● N/A						
02	Human In The Loop The extent to which human oversight is involved in reviewing, verifying, or overriding the AI solution outputs before	Al solution output always reviewed by provider before any action taken	Al solution output has optional human in loop review by provider before any action taken	Al solution output is never reviewed by provider before an action is taken	Low 6 Medium 2	Detailed rationale, artifacts, and supporting evidence, as able.	Low	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	they affect patient care				• High 0		HighN/A	Apply organizational high risk mitigation controls					
03	Consequences Of Failure Or Error The severity and likelihood of negative	Al 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	Errors may lead to permanent harm, permanent damage to body structure, disability, or death (e.g., Al	Low 4 Medium 4	Detailed rationale, artifacts, and supporting evidence, as able.	LowMedium	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	outcomes (e.g., morbidity, mortality) if the AI solution fails or provides incorrect information; clinical consequences are higher risk		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)	misinterprets critical diagnostic imaging or fails to detect sepsis)	High 0		High N/A	Apply organizational high risk mitigation controls					
04	Patient Population Vulnerability The degree to which the patient population affected is vulnerable (e.g.,	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,	Used with patients who are medically complex but unstable (e.g., patients with heart failure and in unstable state)	LowMedium2	Detailed rationale, artifacts, and supporting evidence, as able.	LowMedium	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	pediatrics, elderly, low health literacy, marginalized groups); Depends on clinical setting and presentation context		being seen by primary care physicians)		High 6		High N/A	Apply organizational high risk mitigation controls					
05	Level Of Difficulty Monitoring Al Solution Output How robust is the Al solution's	capability of real-time monitoring	Embedded real-time monitoring and/or capability of real-time monitoring				Al solution includes partial real-time monitoring capabilities; health system still requires partial development of	before implementation of solution; and/ or manual monitoring that requires	Low 0	Abstain b/c don't have this use case information	Low	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	
	monitoring capabilities? How resource intensive will the Al solution be to monitor output and performance?; Depends on both the Al solution provider and health system capabilities		monitoring capabilities; periodic reports	resource intensive activities	Medium 0 High 0		High N/A	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	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	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	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.,	Low 0 Medium 0	Abstain b/c don't have this use case information	LowMedium	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls					
	the AI solution	available (e.g., AI solution developed internally)	available	a rando.		data provenance and data catalog/ dictionary unavailable)	• High		High N/A	Apply organizational high risk mitigation controls			
07	Clinical Level Of Care Does the Al solution operate in a clinically sensitive or high-risk setting that	Al solution used in outpatient and non- critical settings (e.g., outpatient)	Al solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	Al solution used in life-critical settings (e.g., emergency department)	Low 1 Medium 4	Detailed rationale, artifacts, and supporting evidence, as able.	Low	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)				• High 2		HighN/A	Apply organizational high risk mitigation controls					
08	React Time Assuming the AI output is incorrect, how quickly a decision or intervention can be	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	Detailed rationale, artifacts, and supporting evidence, as able.	Low	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	made				• High		HighN/A	Apply organizational high risk mitigation controls					
09	Breadth Of Potential Harm The breadth of potential harm the Al solution could cause to patients if it performs incorrectly; Assess how broadly	Affects a single individual or a small number of patients in a limited number of settings (e.g., rare disease diagnostics, single-department pilot,	Affects a moderate number of patients (e.g., roughly half of patient population), possibly across multiple units or clinics (e.g., diabetes	Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma	Low 0 Medium 3	Detailed rationale, artifacts, and supporting evidence, as able.	LowMedium	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	the AI solution is deployed across locations or institutions	one site, one clinic, or limited geographical area)	prediction across outpatient clinics)	prioritization)	• High		HighN/A	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	with minimal integration into other digital systems or workflows. Errors are	The AI solution is integrated into specific modules or workflows but has limited cross-functional connections.	across multiple systems and workflows. Its outputs are widely relied upon and	Low 0 Medium 3	Detailed rationale, artifacts, and supporting evidence, as able.	LowMedium	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls	Additional notes and comments				
	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 interdependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	unlikely to spread beyond the immediate user or use case.	Errors could impact related components but are unlikely to cause widespread disruptions.	shared, increasing the chance that a single point of failure could cascade across care settings, decisions, or resource allocations.	• High 1		HighN/A	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 Al 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.	LowMediumHighO	Detailed rationale, artifacts, and supporting evidence, as able.	LowMediumHighN/A	Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls	Additional notes and comments				

Risk Domain:

Primary Audience:

Life & Patient Safety

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	Al solution has semi-direct involvement in patient care; such as, used by a healthcare professional as part of a broader clinical judgment; Al solution is directly involved in patient care/patient interaction	LowMediumHigh		LowMediumHighN/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	Al solution output always reviewed by provider before any action taken	Al solution output has optional human in loop review by provider before any action taken	Al solution output is never reviewed by provider before an action is taken	LowMediumHigh		LowMediumHighN/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	Al solution has no direct impact and has no affect on patient harm	discomfort or inconvenience, with no lasting health effects (e.g., minor delays in care); Errors may result in temporary	misinterprets critical diagnostic imaging	LowMediumHigh		LowMediumHighN/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)	LowMediumHigh		LowMediumHighN/A	 Apply organizational low risk mitigation controls Apply organizational medium risk mitigation controls Apply organizational high risk mitigation controls 	
05	Level Of Difficulty Monitoring Al Solution Output How robust is the Al solution's monitoring capabilities? How resource intensive will the Al solution be to monitor output and performance?; Depends on both the Al solution provider and health system capabilities	Embedded real-time monitoring and/or capability of real-time monitoring	monitoring capabilities; health system still requires partial development of	Monitoring needs to be developed before implementation of solution; and/ or manual monitoring that requires resource intensive activities	LowMediumHigh		LowMediumHighN/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)	LowMediumHigh		LowMediumHighN/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 Al solution operate in a clinically sensitive or high-risk setting that requires a higher level of care (e.g., inpatient, outpatient, emergency department, etc.)	Al solution used in outpatient and non-critical settings (e.g., outpatient)	Al solution used in inpatient or urgent, but non-critical settings (e.g., inpatient)	Al solution used in life-critical settings (e.g., emergency department)	LowMediumHigh		LowMediumHighN/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	LowMediumHigh		LowMediumHighN/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 Al solution could cause to patients if it performs incorrectly; Assess how broadly the Al 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)		Potential for widespread harm—across facilities, populations, or entire health systems (e.g., enterprise-wide triage algorithm, regional EMS AI for trauma prioritization)	LowMediumHigh		LowMediumHighN/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 interdependence, reflecting how embedded the AI solution is and how errors in one part could propagate to others.	with minimal integration into other	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.	across multiple systems and workflows.	LowMediumHigh		LowMediumHighN/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 Al 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.	LowMediumHigh		LowMediumHighN/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

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 Sam Pinson, Nixon Law

Taylor Rhoades, Mercy

Brenna Loufek, Mayo

Ben Kaplan, Mount Sinai

Christine Palermo, Encore Health

January Choy, Memorial Sloan Kettering

Brenton Hill, CHAI

Merage Ghane, CHAI

Anthony DiDonato, CHAI

Greg Shemancik, CHAI



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 <u>creativecommons.org/</u> <u>licenses/by-nc-nd/4.0/</u>.

