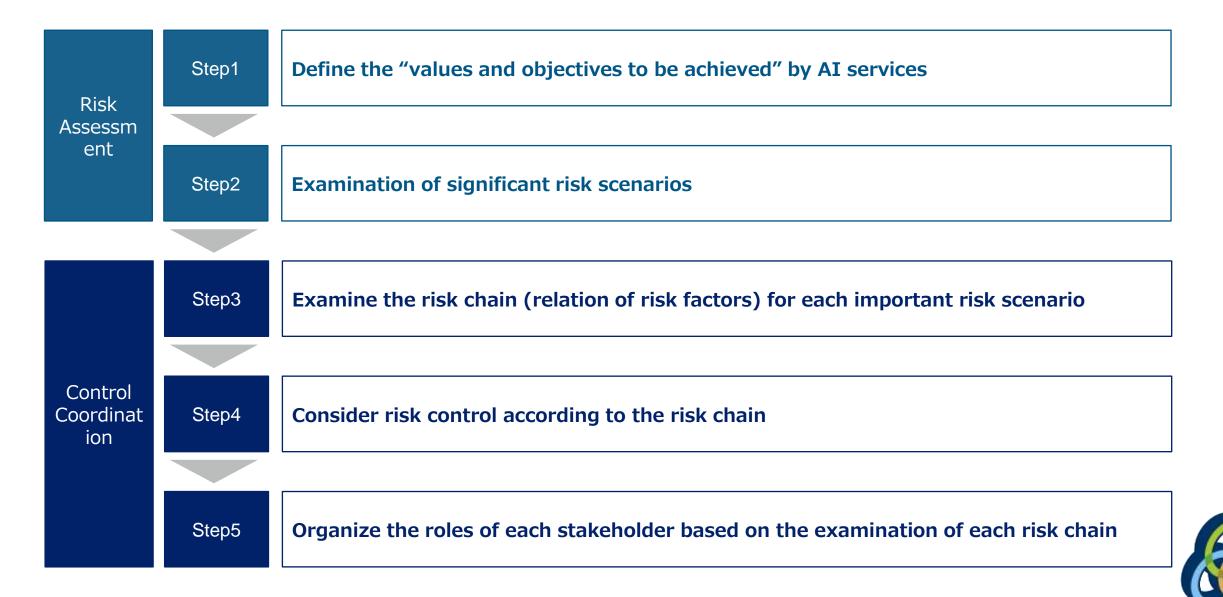


Risk Assessment & Control Coordination for AI services: Case08 Cancer Diagnostic AI

Institute for Future Initiatives, The University of Tokyo Technology Governance Research Unit AI Governance Project



How to operate the RCModel - Risk Assessment & Control Coordination -





Guide book and Case Studies of Risk Chain Model

AI Service and Risk Coordination Study Group

https://ifi.u-tokyo.ac.jp/en/projects/ai-service-and-risk-coordination/



Research Education People News Events Publications

How to use Risk Chain Model

Risk Chain Model (RCModel) Guide Ver1.0月了

Case Study

*These are fictional case studies below and don't raise issues or assure for any company or AI service.

Case01.Recruitment AI (2021/07) 译译



Case Study



Case08: Cancer Diagnostic AI

- Define the "values and objectives to be achieved" by AI services -

An early cancer detection service using deep learning.

It uses multimodal learning, incorporating patient medical history/heredity info (Data 1) and endoscopic images (Data 2) to perform real-time highlighting of places with high likelihood of cancer during endoscope medical examinations. Physicians observe output images and judge the possibility of cancer.

[Values & Objectives]

- Early cancer detection
- Cooperation with physicians
- Social responsibility as a medical institution

[Flow of Actual Operations using AI Services]

- ① The patient undergoes a medical interview prior to the endoscopic exam.
- 2 Patient medical history/heredity info (Data 1) is created from medical charts (medical history/heredity) and the medical interview.
- 3 During the endoscopic medical exam by a physician, Data 1 and Data 2 are fed into a deep learning model, and performs real-time highlighting of places on endoscopic images with high likelihood of cancer.
- 4 The physician observes the interior of the patient's body, including corresponding images, and performs appropriate judgment/medical treatment.

Uses a multimodal deep neural network (DNN), and is composed of a network that incorporates the following two types of data, and a network that outputs abnormal places in images in real time. Model development and execution environment is developed by contractor, Company A. The physician gives feedback for labels of AI model judgment result veracity. Sequential learning is not implemented, but additional learning and re-learning is implemented in cases where accuracy degrades significantly after use.

- Data 1: Table data built from patient medical chart and medical exam
- Data 2: Endoscope camera image

Case08: Cancer Diagnostic AI

- System Overview -

AI System
Company A) AI model development department
Develop AI for real-time detection of places with signs of early cancer

Real-time detection of early cancer sign locations during endoscope examination

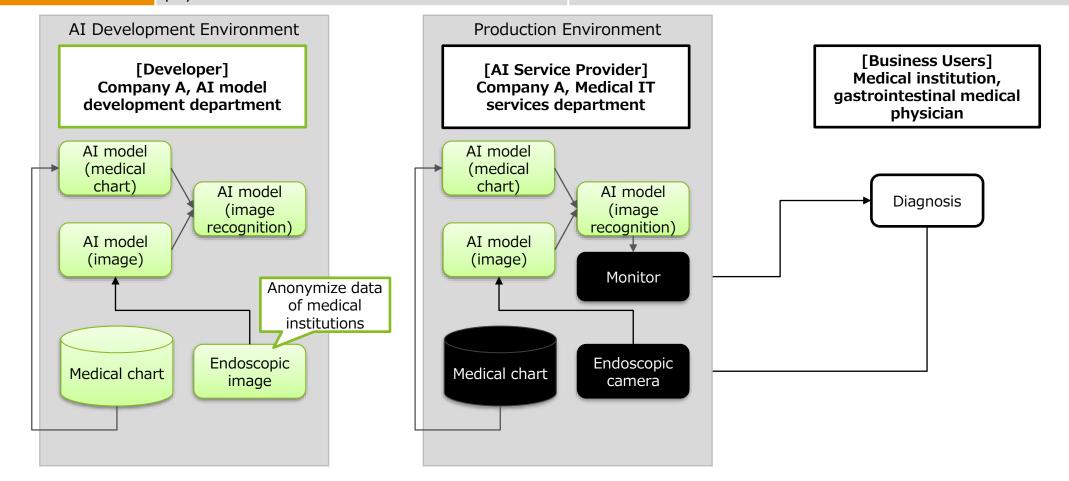
Medical institution, gastrointestinal medical physician

Company A) AI model development department

Real-time detection of early cancer sign locations during endoscope

examination

Conduct medical examination while referencing AI detection results





Case08: Cancer Diagnostic AI

- Input & Output -

[Input Data]

Data	Purpose	Collection Method	Data Manager	Including Privacy Data
Human medical charts (medical history, heredity, etc.)	Learning	Data collected from medical institutions (judgment by pathological diagnosis)	Medical institutions, gastrointestinal medicine	Yes
Human medical exam data	Learning	Data collected from medical institutions	Medical institutions, gastrointestinal medicine	Yes
Endoscope image data	Learning	Data collected from medical institutions	Medical institutions, gastrointestinal medicine	Yes
Human medical charts (medical history, heredity, etc.)	Production	Data collected from medical institutions	Medical institutions, gastrointestinal medicine	Yes
Human medical exam data	Production	Whenever used	Medical institutions, gastrointestinal medicine	Yes
Endoscope image data	Production	Whenever used	Medical institutions, gastrointestinal medicine	Yes



Case08: Cancer Diagnostic AI - Input & Output -

[Output]

Users	Medical institution, gastrointestinal medical physician
Output	Spots suspected of being cancer in endoscopic images
Output Method	 Endoscopic image monitor: Highlights spots with suspicion of cancer in real-time Sub-screen: Highlights features with high influence on patient medical charts
Expected Accuracy	 Photographic sensitivity (percentage of cancer images correctly judged to be cancer) over 80% Specificity (percentage of normal images correctly judged to be normal) over 90% Positive predictive value (percentage of images judged to be cancer that were actually cancer) over 85% Negative predictive value (percentage of images judged to be normal that were actually normal) over 85%
User judgment	Yes (final judgment by physician)
Output of evidence information	Highlight noteworthy sections on screen
Safety Risk	Yes (risk of misdiagnosis)
Connection with external system	No
Protocol	No





Risk Assessment



Risk Assessment

- Examination of significant risk scenarios -

	Values & Objectives		Service Requirement			Risk No. Risk Scenario				
		1-1	Ensuring prediction performance	AI prediction accuracyAI robustnessAI generalization functionality	R001	Ensuring prediction performance	Overlooked early cancer due to insufficient AI prediction accuracy, generalized functionality and robustness			
1	Early cancer detection	1-2	Real-time Diagnostics	■ Performance	R002	Limits of predictive processing ability	Cannot perform appropriate medical examination due to limits of angle of field, endoscopic camera frame, and AI examination speed			
			Response to	■ AI robustness	R003	Impact of noise	Noise interference prevents appropriate judgment			
		1-3	environmental changes	■ IoT	R004	Equipment changes	Change of image specifications (resolution, image quality, format, etc.) due to change in endoscopic camera prevents appropriate judgment			
		1-4	Protection from external attack	AI robustnessSecurity	R005	Security protection	Abnormal input of learning data or model changes caused by external attacks obstructs medical examination			
	Cooperation with Physicians	2-1	Display of decision basis	AI explainabilityUnderstandin g of user objectives	R006	Overlooking illness the AI cannot judge	Overlooked early cancer or other diseases (accidental discovery) due to lack of doubt in AI judgment			
2	(overhead reduction and appropriate	2-2	Appropriate 2 Examination Level	■ AI prediction accuracy	R007	Expanded overhead due to excessive detection	Excessive detection increases pathological diagnoses by physicians and burden to patient bodies			
	cooperation)	2-3	Contingency Plan	Switch during abnormalities	R008	Alternative operation for abnormalities	When abnormalities occur due to AI performance degradation, etc., no switch to medical examinations by physician occurs, hindering examination operation			

Risk Assessment

- Examination of significant risk scenarios -

	Values & Objectives		Service Requ	irement	Risk No.		Risk Scenario		
	■ Explainability	■ Explainability	R009	Explanation to patients	Cannot give adequate and rational explanations to patients who get cancer after getting a negative medical exam result				
	Social	3-1	Accountability	■ Verifiability	■ Verifiability	R010	Outward-facing explanation	Cannot give appropriate explanation of AI service reliability, verification from time of development, etc.	
3	Responsibilit y of Medical	3-2	Observance of standards/rules		R011	Change of standards/rules	Cannot deal with standard or guideline changes at ministries, agencies or institutions (institution evidence reflected in AI model learning data)		
	Institution	3-3	Information	■ Data	R012	Exposure of AI medical exam results	External exposure of AI medical exam results regardless of the fact that physicians perform final judgment, causing misunderstandings in patients		
			Management	management R0	R013	Privacy protection	Law violation from use of personal information (patient images, medical exam results) without receiving use objective consent		



Risk Assessment & Control Summary

- Organize the roles of each stakeholder based on the examination of each risk chain -

	Values &	Risk	B: 1.0	Uncerta	Environ	Caused		Control Summary		
	Objectives	No.	Risk Scenario	inly	mental change	by user	RC	AI System	AI Service Provider	User
		R001	Ensuring prediction performance	0			•	Ensuring prediction performance Ensuring verifiability	Validate model AI model re-learning	Alternative operation
		R002	Limits of predictive processing ability	0			•	Performance assurance Performance record Judgment basis output	Judgment basis sub-screen output Performance verification/improvement	Understanding of judgment responsibility Understanding of technological limitations Implementation of appropriate medical exam
1	Early Cancer Detection	R003	Impact of noise	0	0		•	Noise correction of image data Model robustness Judgment basis output	Camera cleaning/storage Periodic verification of judgment basis AI model re-learning	
		R004	Equipment changes	0	0		•	Definition of image specifications Noise correction of image data Ensuring model prediction performance	PMDA approval equipment selection Function change explanation AI model re-learning	Understanding of judgment responsibility Ensuring AI knowledge Verification of re-learning necessity
		R005	Security protection					Security management	Cause investigation/improvement	
	Cooperation with	R006	Overlooking illness the AI cannot judge	0		0	•	Judgment basis output	Judgment basis sub-screen output	Understanding of judgment responsibility Understanding of AI expected accuracy Review of medical exam places
2	Physicians	R007	Expanded overhead due to excessive detection	0			•	Ensuring prediction performance Judgment basis output Ensuring verifiability	Judgment basis sub-screen output Verify causes of excessive detection Detection level definition	Sub-screen confirmation AI prediction function off Cooperation for excessive detection situations Verification of re-learning necessity
	cooperation)	R008	Alternative operation for abnormalities	0		0	•	Ensuring prediction performance Judgment basis output Temporary system stoppage	Transfer of system abnormalities Investigation of abnormality causes AI model re-learning	Alternative operation Cooperation for system abnormalities

Risk Assessment & Control Summary

- Organize the roles of each stakeholder based on the examination of each risk chain -

	Values &	Risk	Risk Scenario	Uncerta	Environ	Caused	RC	Control Summary		
	Objectives	No.	RISK SCENATIO	inly	mental change	by user	by user RC	AI System	AI Service Provider	User
		R009	Explanation to patients	0			•	Ensuring prediction performance Judgment basis output Ensuring verifiability	Essentials of explanations to patients Verification of judgment error causes Display of necessary information	Understanding of judgment responsibility Implementation of appropriate medical exam
		R010	Public explanation	0			•	Record of data understanding Record of model performance Record of verification from time of development	Access rights management Arrangement of information to disclose Public explanation	
3	Social Responsibilit y of Medical Institution	R011	Change of standards/rules	0	0			Ensuring prediction accuracy Judgment basis output Ensuring verifiability	System update Re-learning	Cooperation for standard/rule changes
		R012	Exposure of AI medical exam results			0		Data protection	Professional ethics education	Professional ethics education
		R013	Privacy protection			0		Data protection	Compliance education	Compliance education Appropriate data handling



Organization

- Organize the roles of each stakeholder based on the examination of each risk chain -

- Responsible Persons - Company A) Management

- Assessment of values and objectives to achieve
- Risk control method approval

Company A) Legal/Compliance

- Professional ethics education
- Compliance education

- AI Service Provider -Company A, Medical IT services department

- Detection level definition
- Judgment basis sub-screen output
- Camera cleaning/storage
- AI model re-learning
- Verification of prediction accuracy
- Periodic verification of judgment basis
- Performance verification/improvement
- Investigation of abnormality causes
- Verify causes of excessive detection
- Access rights settings
- PMDA approval equipment selection
- Function change explanation
- Transfer of system abnormalities

Company A) AI model development department

- Ensuring prediction performance
- Model robustness
- Judgment basis output
- Noise correction of image data
- Preservation of records from time of development

Company A) Information Systems Department

- Implementation history preservation
- Performance assurance
- Performance record
- Temporary system stoppage
- Security management
- Data protection

Medical institutions, gastrointestinal medicine

- Essentials of explanations to patients
- Display of necessary information

Ministry of Health, Labor and Welfare, PMDA Institution (Medical Exam Guidelines)

Patients

- Users -Medical institution, gastrointestinal medical physician

- Understanding of judgment responsibility
- Ensuring AI knowledge
- Understanding of technological limitations
- Implementation of appropriate medical exam
- Sub-screen confirmation
- Turn off AI prediction function
- Appropriate data handling
- Alternative operation
- Verification of re-learning necessity
- Cooperation for excessive detection situations
- Cooperation for system abnormalities
- Cooperation for standard/rule changes



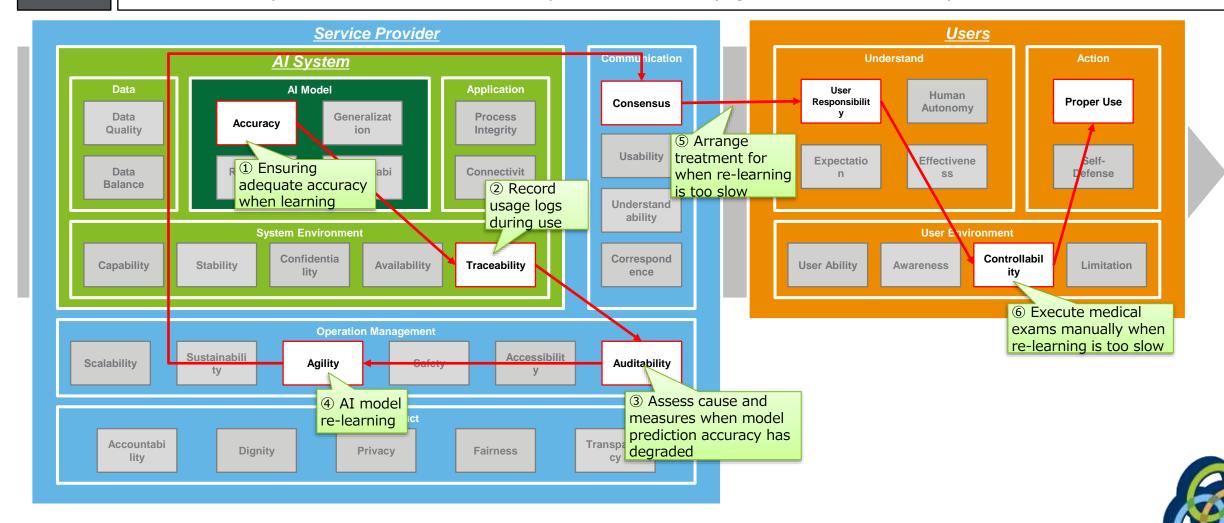


- Examine the risk chain (relation of risk factors) for each important risk scenario -

R001

Ensuring prediction performance

Overlooked early cancer due to insufficient AI prediction accuracy, generalized functionality and robustness



- Consider risk control according to the risk chain -

R001

Ensuring prediction performance

Overlooked early cancer due to insufficient AI prediction accuracy, generalized functionality and robustness

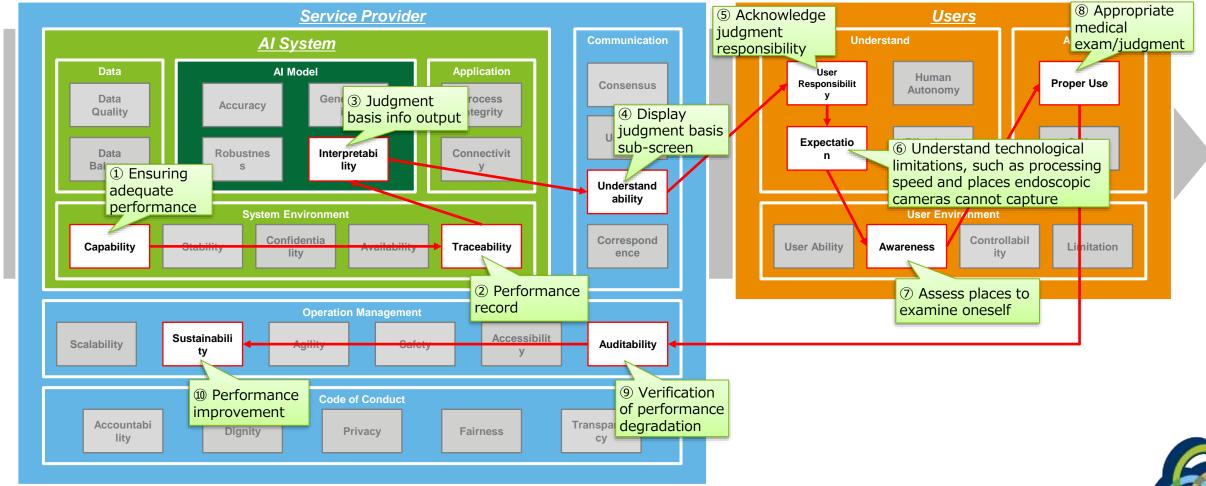
	Risk Control	
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)
[Accuracy] Ensure adequate accuracy rate during learning (Company A, AI development team) [Traceability] Ensure log info during use (Company A information systems department)	 ③ [Auditability] Assess causes and countermeasures of model prediction accuracy degradation (Company A medical IT services department) ④ [Agility] Request AI model re-learning to preserve adequate precision (Company A medical IT services department) ⑤ [Consensus] Make arrangements for alternate manual operations when re-learning takes too much time (Company A medical IT services department) 	(§) [User Responsibility] Make arrangements for alternate manual operations (medical exams) when re-learning takes too much time (Physician (§) [Controllability/Proper Use] Use alternate manual operations (medical exams) when relearning takes too much time (Physician)

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R002

Limits of predictive processing ability

Cannot perform appropriate medical examination due to limits of angle of field, endoscopic camera frame, and AI examination speed





- Consider risk control according to the risk chain -

R002

Limits of predictive processing ability
Cannot perform appropriate medical examination due to limits of angle of field, endoscopic camera frame, and AI examination speed

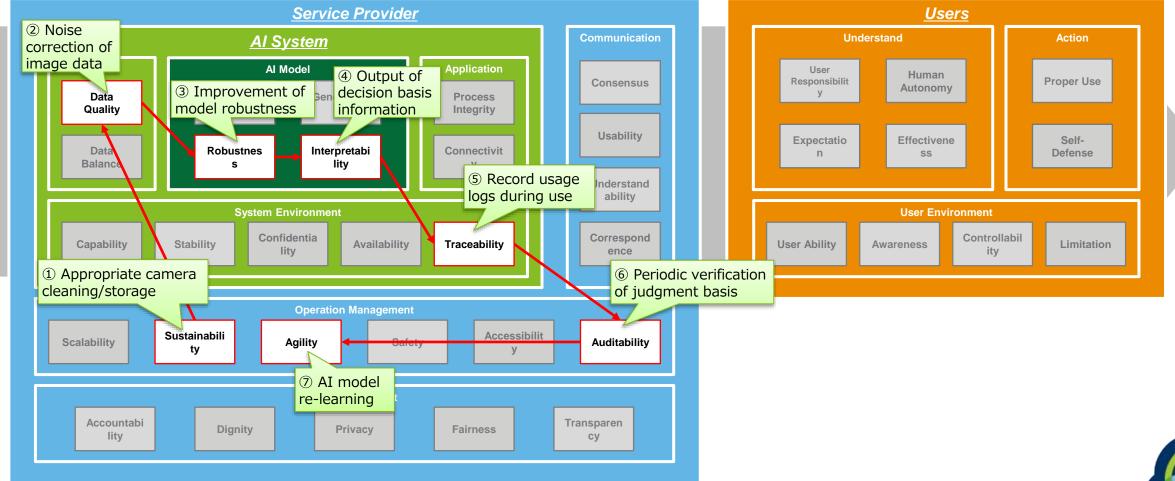
	Risk Control	
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)
① [Capability] Assure suitable processing speed with the entire AI system (Company A, AI development team)	④ [Understandability] Output judgment basis info (places of interest in image) to sub-screen (Company A medical IT services department)	⑤ [User Responsibility] Understand that responsibility of final judgment lies with the physician (Physician)
② [Traceability] Record performance burden situation (Company A information systems department)	[Auditability] Confirm if performance degradation occurs frequently (Company A medical IT services department)	6 [Expectation] Understand AI predictive process speed and spots that cannot be captured with endoscopic cameras (Physician)
③ [Interpretability] Output judgment basis (places of interest in image) (Company A, AI development team)	(I) [Sustainability] Improve AI system processing function to assure adequate performance (Company A medical IT services department)	② [Awareness] Check for discrepancies between highlighted parts displayed in real time and still images output onto the sub-screen (Physician)
		® [Proper Use] Appropriately examine spots marked by AI while covering for spots the endoscopic camera cannot capture with other medical examination methods (Physician)

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R003

Impact of noise

Noise interference prevents appropriate judgment





- Consider risk control according to the risk chain -

R003

Impact of noise

Noise interference prevents appropriate judgment

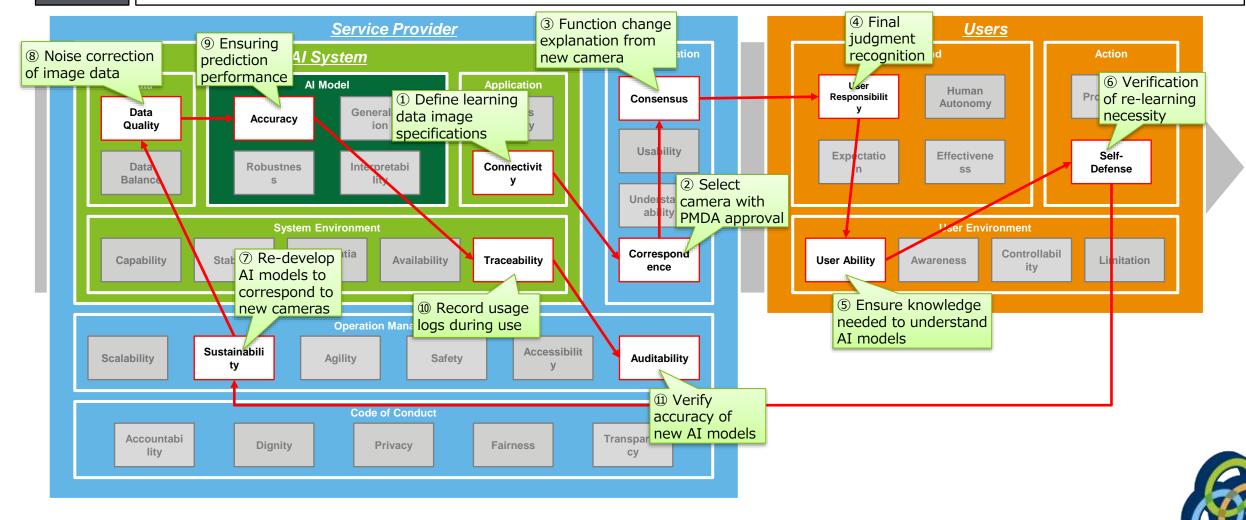
	Risk Control	
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)
② [Data Quality] Address degradation of image data by noise correction (Company A AI development team) ③ [Robustness] Use learning to enhance model robustness (Company A AI development team) ④ [Interpretability] Output model judgment basis (places of interest in image) (Company A, AI development team) ⑤ [Traceability] Ensure AI judgment result info during use (Company A information systems department)	① [Sustainability] Clean camera with appropriate methods (Company A medical IT services department) ⑥ [Auditability] Periodically check samples to see if clear errors occur for places of interest in images (Company A medical IT services department + physician) ⑦ [Agility] Request AI model re-learning to preserve adequate precision (Company A medical IT services department)	physician)

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R004

Equipment changes

Change of image specifications (resolution, image quality, format, etc.) due to change in endoscopic camera prevents appropriate judgment



- Consider risk control according to the risk chain -

R004

Equipment changesChange of image specifications (resolution, image quality, format, etc.) due to change in endoscopic camera prevents appropriate judgment

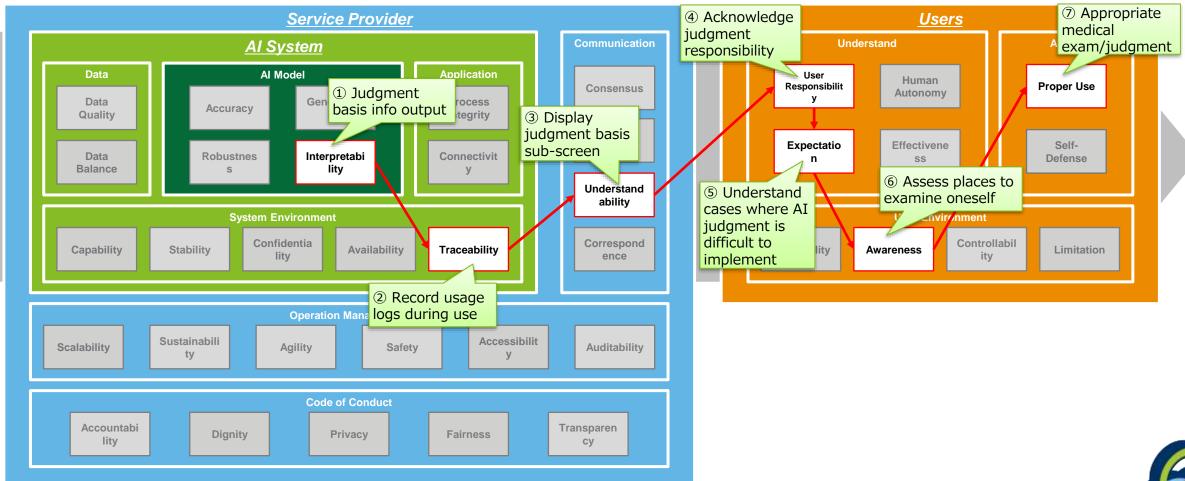
	Risk Control	
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)
① [Connectivity] Define learning data image specifications (resolution, brightness, etc.) (Company A AI development team) ⑧ [Data Quality] Address degradation of image data by noise correction (Company A AI development team) ⑨ [Accuracy] Ensure adequate model prediction functionality (Company A, AI development team) ⑩ [Traceability] Ensure AI judgment result info during use (Company A information systems department)	 [2] [Correspondence] Select new cameras from those with PMDA approval (Company A medical IT services department) [3] [Consensus] Explain function changes from new cameras to physicians and ask for re-learning judgment (Company A medical IT services department) [7] [Sustainability] Re-develop AI models corresponding to new camera (Company A medical IT services department) [8] [Auditability] Verify prediction accuracy of new AI model (Company A medical IT services department) 	 ④ [User Responsibility] Understand that final judgment lies with the physician (Physician) ⑤ [User Ability] Ensure required knowledge for understanding of AI model functions (Physician) ⑥ [Self-Defense] Judge the necessity of relearning (Physician)

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R006

Overlooking illness the AI cannot judge

Overlooked early cancer or other diseases (accidental discovery) due to lack of doubt in AI judgment





- Consider risk control according to the risk chain -

R006

Overlooking illness the AI cannot judge

Overlooked early cancer or other diseases (accidental discovery) due to lack of doubt in AI judgment

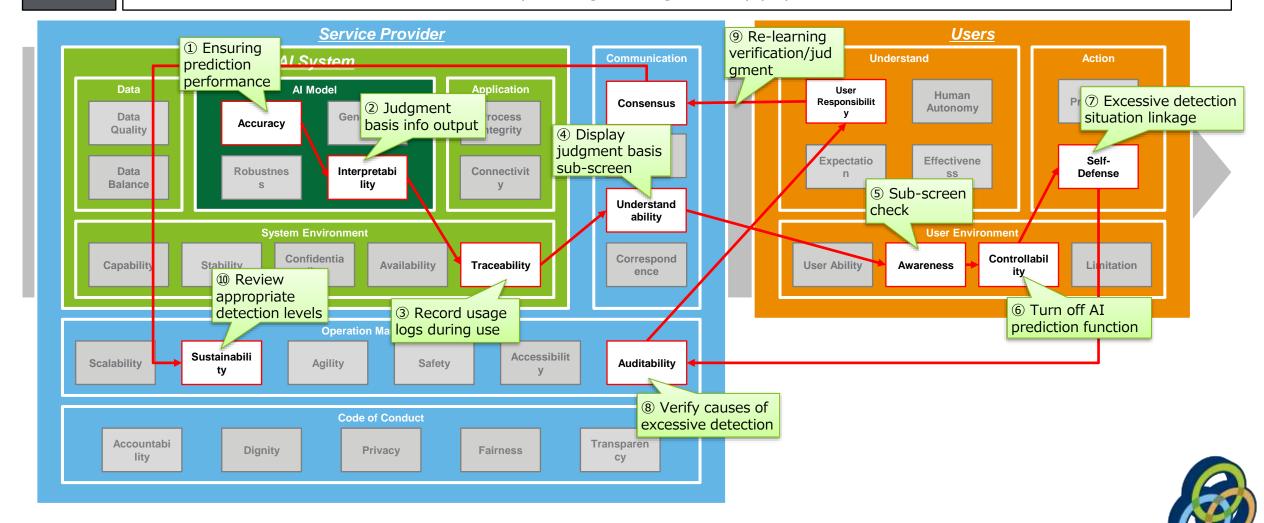
	Risk Control	
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)
[Interpretability] Output model judgment basis (places of interest in image) (Company A, AI development team) [Traceability] Ensure AI judgment result info during use (Company A information systems department)	③ [Understandability] Output judgment basis info (places of interest in image) to sub-screen (Company A medical IT services department)	 ④ [User Responsibility] Understand that final judgment lies with the physician (Physician) ⑤ [Expectation] Understand patient/illness cases where AI prediction is difficult to implement (Physician) ⑥ [Awareness] Judge places that should be examined oneself according to the state of AI detection generation, including places that cannot be captured by endoscopic cameras (Physician) ⑦ [Proper Use] Appropriately examine spots marked by AI while covering for spots the endoscopic camera cannot capture with other medical examination methods (Physician)

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R007

Expanded overhead due to excessive detection

Excessive detection increases overhead of pathological diagnoses by physicians



- Consider risk control according to the risk chain -

R007

Expanded overhead due to excessive detection

Excessive detection increases overhead of pathological diagnoses by physicians

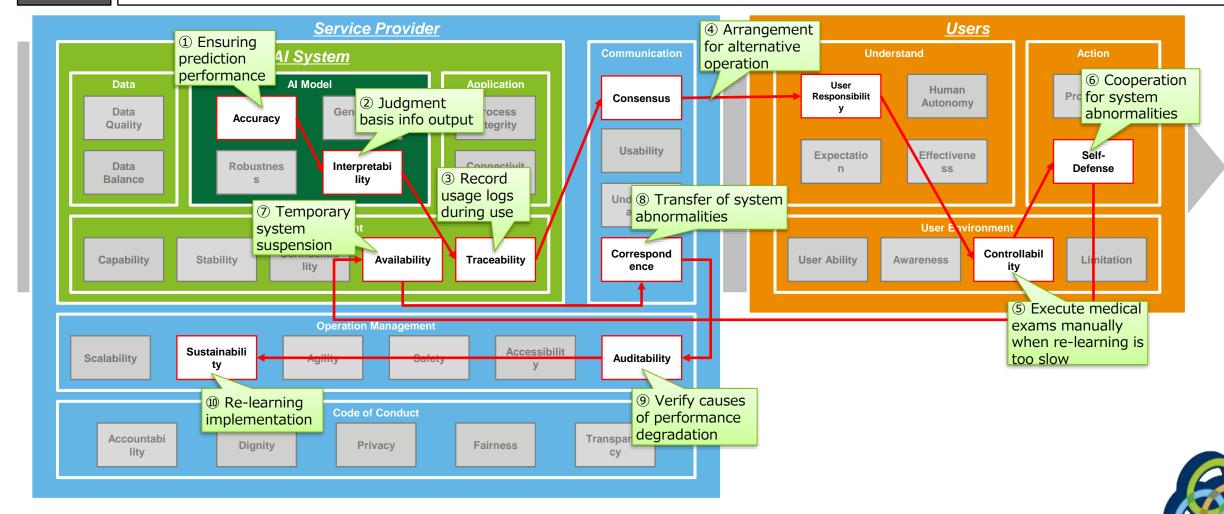
	Risk Control	
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)
① [Accuracy] Implement learning to ensure model prediction accuracy (Company A AI development team) ② [Interpretability] Output model judgment basis (places of interest in image) (Company A, AI development team) ③ [Traceability] Ensure AI judgment history (Company A information systems department)	 ④ [Understandability] Output judgment basis info (places of interest in image) to sub-screen (Company A medical IT services department) ⑧ [Auditability] Verify model functionality and confirm reasons for excessive detection (Company A medical IT services department) ⑨ [Consensus] Assess whether the state of judgment accuracy is explained and relearned (Company A medical IT services department) ⑩ [Sustainability] Review appropriate detection levels (model function) based on maintenance costs (Company A medical IT services department/Company A AI development team) 	 ⑤ [Awareness] Check highlighted parts displayed in real time (Physician) ⑥ [Controllability] Turn off display of AI prediction results at the physician's judgment if it seems to interfere with medical examination (Physician) ⑦ [Self-Defense] Cooperate with the development is deemed in the event that excessive detection is evident and re-learning assessment is deemed necessary (Physician) ⑨ [User Responsibility] Understand the appropriate detection level, and assess whether re-learning occurs (Physician)

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R008

Alternative operation for abnormalities

When abnormalities occur due to AI performance degradation, etc., no switch to medical examinations by physician occurs, hindering examination operation



- Consider risk control according to the risk chain -

R008

Alternative operation for abnormalitiesWhen abnormalities occur due to AI performance degradation, etc., no switch to medical examinations by physician occurs, hindering examination operation

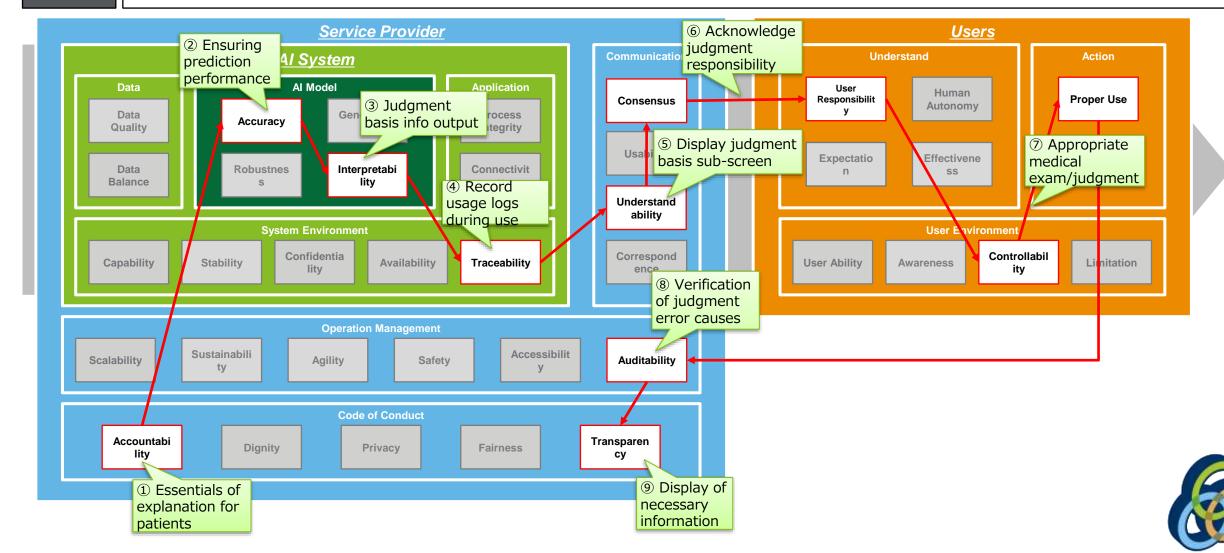
Risk Control			
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)	
① [Accuracy] Implement learning to ensure model prediction accuracy (Company A AI development team)	④ [Consensus] Make arrangements for alternate manual operations when re-learning takes too much time (Company A medical IT services department)	④ [User Responsibility] Make arrangements for alternate manual operations (medical exams) when re-learning takes too much time (Physician)	
② [Interpretability] Output model judgment basis (places of interest in image) (Company A, AI development team)	 (a) [Correspondence] Cooperate with other physicians when system abnormalities occur (Company A medical IT services department) 	⑤ [Controllability] Use alternate manual operations (medical exams) when re-learning takes too much time (Physician)	
 [3] [Traceability] Ensure AI judgment history (Company A information systems department) [7] [Availability] Render system temporarily 	(9) [Auditability] Assess causes and countermeasures of model performance degradation (Company A medical IT services	6 [Self-Defense] Cooperate with the development side in the event that abnormalities are evident in the system (Physician)	
unusable (Company A medical IT services department/Company A information systems department)	department) ① [Sustainability] Request AI model re-learning to preserve adequate precision (Company A IT services department)		
	Services departments		

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R009

Explanation to patients

Cannot give rational explanations to patients who get cancer after getting a negative medical exam result



- Consider risk control according to the risk chain -

R009

Explanation to patients

Cannot give rational explanations to patients who get cancer after getting a negative medical exam result

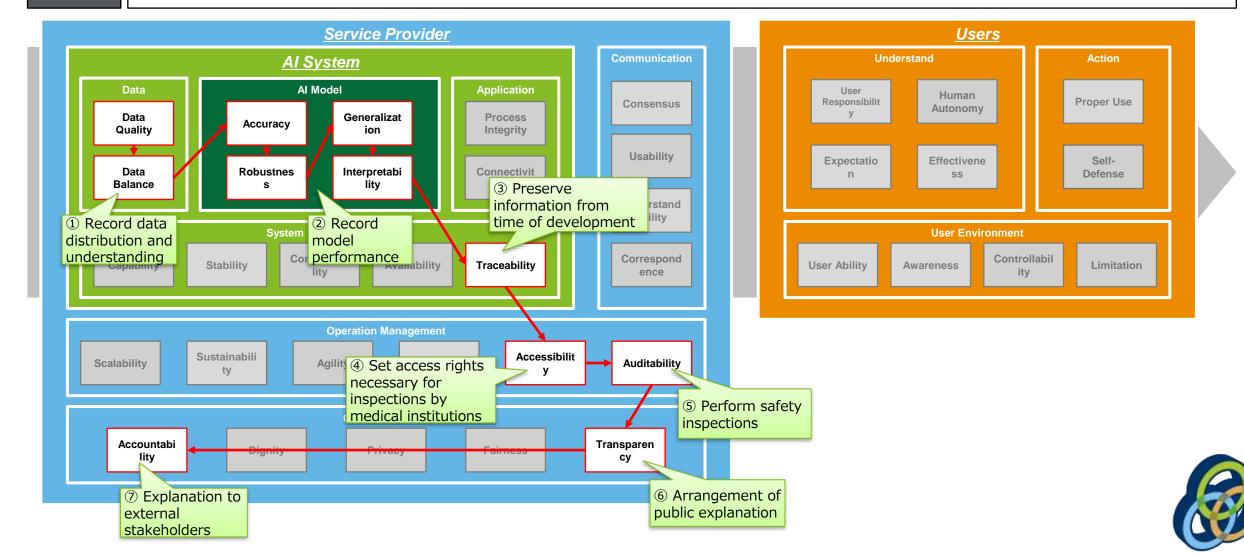
Risk Control			
Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)		
① [Accountability] Assess essentials/content of required explanations for patients (Hospital T)	6 [User Responsibility] Reach consensus on the fact that final judgment lies with the physician (Physician)		
⑤ [Understandability] Output judgment basis info (places of interest in image) to sub-screen (Company A medical IT services department)	② [Controllability/Proper Use] Perform medical examinations based on final judgment (Physician)		
6 [Consensus] Reach consensus on the fact that final judgment lies with the physician (Company A medical IT services department)			
[Auditability] Assess causes of cases where errors occurred in model judgment (Company A medical IT services department)			
			
	Service Provider (Company A, Medical IT services department) ① [Accountability] Assess essentials/content of required explanations for patients (Hospital T) ⑤ [Understandability] Output judgment basis info (places of interest in image) to sub-screen (Company A medical IT services department) ⑥ [Consensus] Reach consensus on the fact that final judgment lies with the physician (Company A medical IT services department) ⑧ [Auditability] Assess causes of cases where errors occurred in model judgment (Company A medical IT services department) ⑨ [Transparency] Disclosure necessary information (Company A medical IT services		

- Examine the risk chain (relation of risk factors) for each important risk scenario -

R010

Outward-facing explanation

Cannot give public explanation of AI service reliability



- Consider risk control according to the risk chain -

R010

Outward-facing explanation

Cannot give public explanation of AI service reliability

Risk Control			
AI System (Company A, AI model development department)	Service Provider (Company A, Medical IT services department)	Users (Medical institution, gastrointestinal medical physician)	
① [Data Quality/Data Balance] Record data distribution and understanding (Company A, AI development team)	④ [Accessibility] Establish access rights necessary for safety inspections by medical institutions (Company A, medical IT services department)		
② [Accuracy/Robustness/Generalization/Interpretability] Record model performance (Company A, AI development team)	⑤ [Auditability] Periodic implementation of safety inspections (Hospital T & Company A medical IT services department)		
③ [Traceability] Record verification results from time of development (Company A, AI development team)	 [6] [Transparency] Arrange information on the assumption of public explanation (Hospital T & Company A medical IT services department) [7] [Accountability] Explain to external stakeholders as necessary (Hospital T & Company A medical IT services department) 		