RCM
RiskChainModel

Risk Assessment & Control Coordination for AI services :
Case11 Guidance in Plant Operation

※ This case is detailed in Use Case 1 of the AI Business Promotion Consortium's Ethical Working Group Deliverables. It was assessed, with help from the Chiyoda Corporation, as a typical case.(Note: The Chiyoda Corporation did not assess this case directly.)
Reference(Japanese only): https://aibpc.org/?p=1523

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

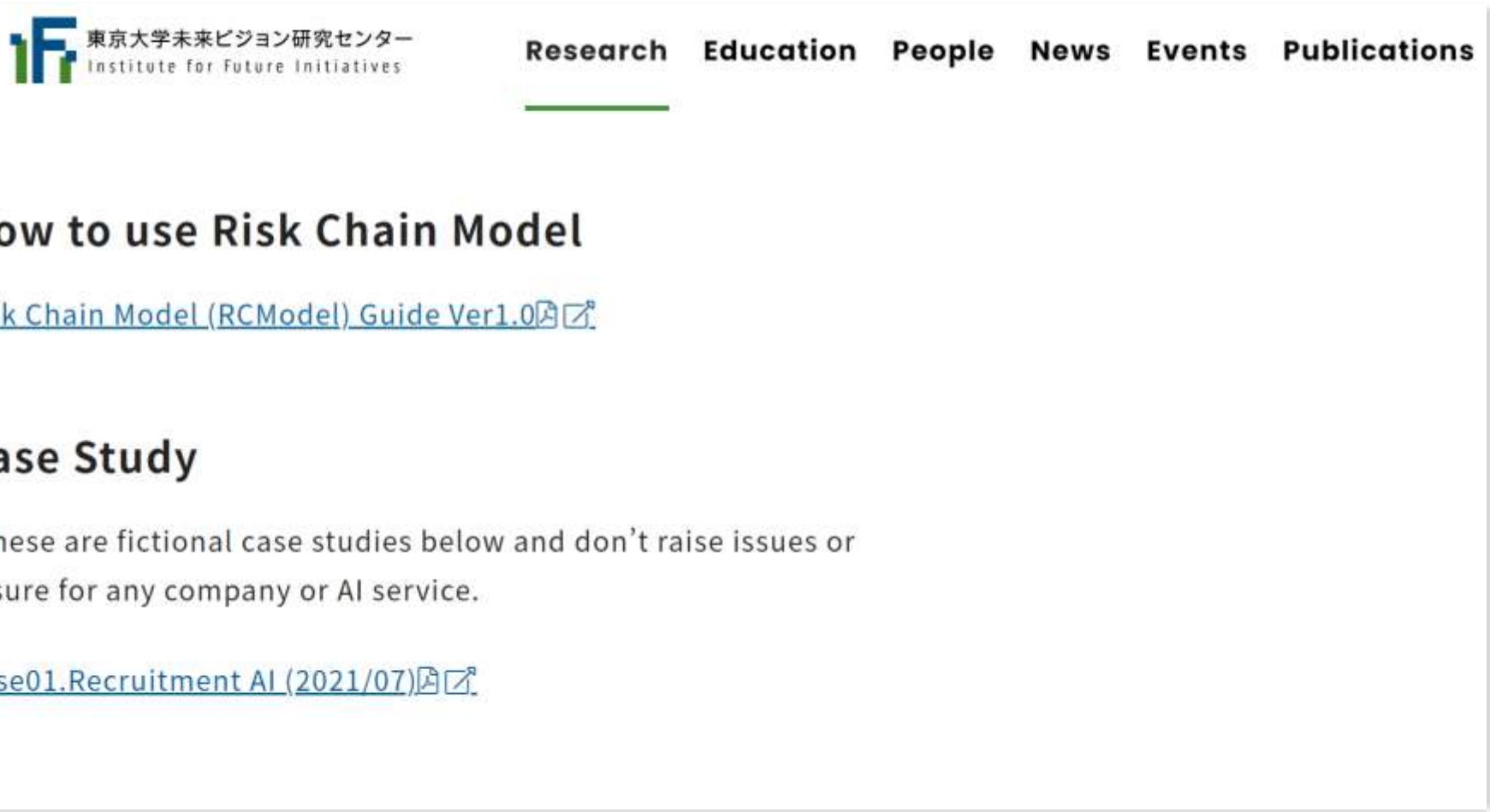
How to operate the RCModel
- Risk Assessment & Control Coordination -

Risk Assessment	Step1	Define the “values and objectives to be achieved” by AI services
	Step2	Examination of significant risk scenarios
Control Coordination	Step3	Examine the risk chain (relation of risk factors) for each important risk scenario
	Step4	Consider risk control according to the risk chain
	Step5	Organize the roles of each stakeholder based on the examination of each risk chain

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/



RCM
RiskChainModel

Case Study

Case11 : Guidance in Plant Operation

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

Step 1

Provide AI Services to Oil Company A by providing guidelines for changing raw materials (an unsteady-state operation) to the operational management division.

The AI dev dept of Company B, the AI Services Provider, quantified numerous standards for ideal operation by conducting interviews with veteran plant operators. It then adjusted these standards based on past operational data and developed an evaluation function. The AI then performed theoretical operation of a dynamic simulation of the facility based on the data collected. Through trial and error, it underwent deep reinforcement learning to perform ideal operation. After being trained to consistently produce good results, the AI provided assistance to operators by monitoring live data during actual operations and outputting ideal parameters.

An AI model is now being developed with the assistance of Company C (a contractor)'s AI dev dept.

Note: While many aspects of the plant in question are automated and run by sophisticated programs, processes such as this one have a high degree of technical difficulty and must be performed by hand. Due to a decline in the number of veteran operators, there is a demand for the aid of new technology.

[Values & Objectives]

- **Reduction of nonproductive time (increased profit)** *Provision of AI model's prediction of effects after complex manual operation and multiple ideal operation parameters
- **Operator training** *Enable self-evaluation of new operators by quantifying complex operational procedures based on assessment by veteran operators, helping them transfer their skills.
- **Safely Change Raw Materials**
- **Social Responsibility**

[Flow of Actual Operations using AI Services]

Evaluation Function: Use of evaluation function to equal or exceed veterans' score (a combined expression of multiple criteria).

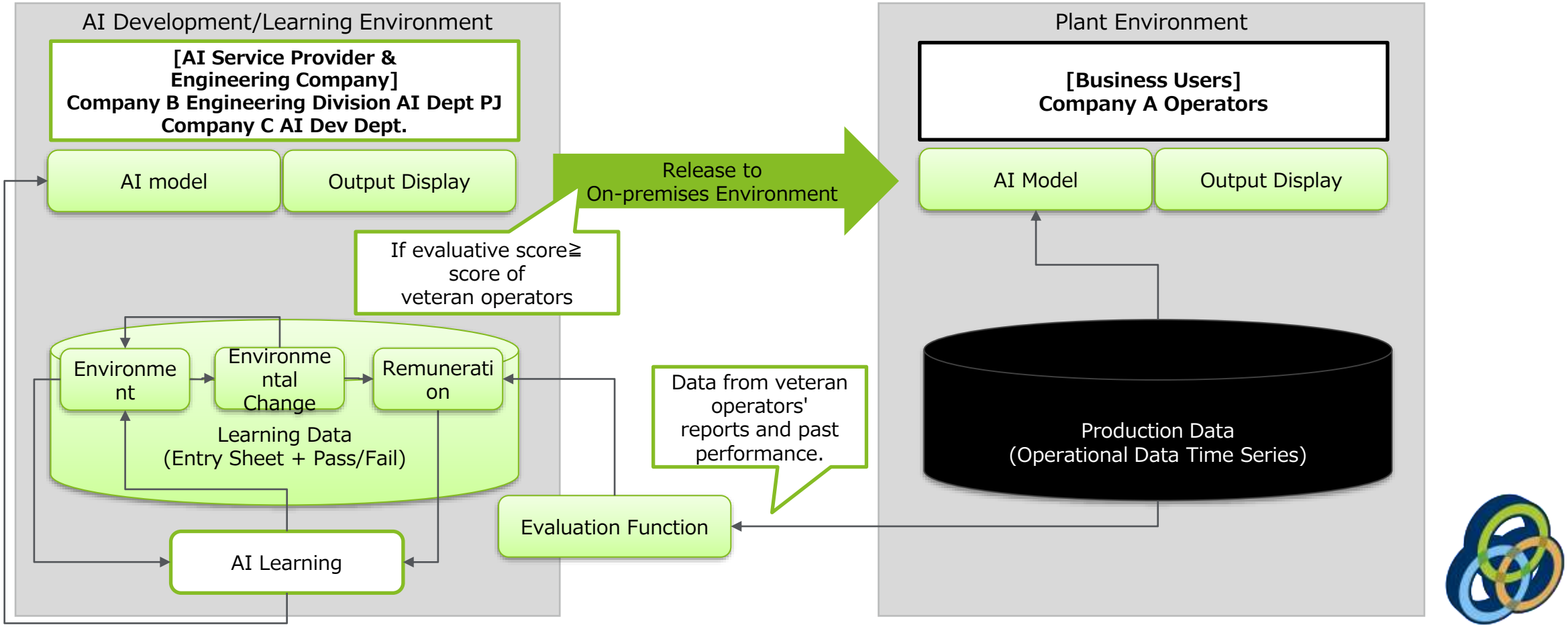
*Combines several factors, such as shortness of operating time and loss control during changeover, searching for the best overall operations from several hours to half a day ago.

*Display AI output as recommended value when it exceeds veteran operators performance based on evaluation function.

Case11 : Guidance in Plant Operation

- System Overview -

AI System	Company B Engineering Division AI Dept PJ + Company C AI Dev. Dept.	Recommended parameters based on deep reinforcement learning used in simulation.
AI Service Provider	Company B Engineering Division AI Dept PJ	Provision of optimal parameters for changing materials in real time
Users	Company A Operators (Plant Operational Management Division)	Material Changing Operation (Parameter-setting)



Case11 : AI Guidance in Plant Operation

- Input & Output -

[Input Data]

Data	Purpose	Collection Method	Data Manager	Including Privacy Data
Past Data	Learning	Operational Data from Records	Company A Operational Supervisor (Central Control Room Server)	No
Dynamic Simulation	Learning	AI operates in dynamic simulation, improves through trial and error.	Return data generated during training to Company B.	No
Latest Data from Facility	Production	Automatic Collection by Existing System	Company A Operational Supervisor (Central Control Room Server)	No

[Output]

Users	Company A Operators
Output	Various Parameters (setpoint values of manual controllers)
Output Method	The next ideal value is displayed on terminal in the central control room of Company A's plant.
Expected Accuracy	Output when total point value of evaluation function based on deep reinforcement learning exceeds veteran operators'.
User judgment	Yes
Output of evidence information	No
Safety Risk	No
Connection with external system	No
Protocol	No

RCM
RiskChainModel

Risk Assessment

Risk Assessment

- Examination of significant risk scenarios -

Step 2

Values & Objectives		Service Requirement			Risk No.	Risk Scenario	
1	Reduction of nonproductive time (increased profit)	1-1	Maintenance of predictive performance	■ AI Prediction Accuracy ■ AI Robustness ■ AI Explainability	R001	Performance Reduction Due to Environmental Change	Performance declines when raw material composition differs from parameters initially provided and trained on.
					R002	Noise Introduction	Introduction of noise from actual environment destabilizes decision accuracy.
					R003	Change in Data Quality	Quality of data obtained from inspection/cleaning (such as catalytic alteration) varies widely, destabilizing predictive power.
		1-2	Adequate cost	–	R004	Excess costs	Operation costs are exceeded
2	Operator Training	2-1	Validity of Training Material	■ AI Explainability	R005	Unintended Application	Application of this AI model to unspecified raw materials or services will interfere with operations due to errors in judgment.
					R006	Technological Misinterpretation	A false interpretation of the AI's decision-making tendencies spread, causing operators to make mistakes.
		2-2	Troubleshooting		R007	Inability to do troubleshooting	Due to lack of emergency training, there is no way to address potential problems.
3	Safely Change Raw Materials	3-1	Ensuring safety	■ Safeguards	R008	Managing Anomalies	Inability to return to normal parameters in the event of abnormal output, interfering with operation of plant.
		3-2	Protection from External Attack	■ Robustness ■ Security	R009	Attacks on Learning Data/Models	Predictive accuracy cannot be maintained due to abnormal input of learning data or model changes caused by external attacks.
					R010	Attacks on Real-World System	Predictive accuracy cannot be maintained if sensors or controllers are damaged due to external attacks.
4	Social Responsibility	4-1	Accountability	■ Process description ■ Verifiability	R011	Response to quality audits	Inability to provide appropriate explanations when subjected to quality audits
					R012	Investigation of problems	When an external explanation is required due to the occurrence of an abnormality or a problem, the cause and preventive measures cannot be considered or explained.

Risk Assessment & Control Summary

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

Step 5

Values & Objectives		Risk No.	Risk Scenario	Uncertainly	Environmental change	Caused by user	RC	Control Summary		
								AI System	AI Service Provider	User
1	Reduction of nonproductive time (increased profit)	R001	Performance Reduction Due to Environmental Change	○			●	Examine data distribution Develop New Model Record Learning Performance	A/B Testing Understanding AI Model Change Re-learning	Understanding AI Model Change
		R002	Noise Introduction	○			●	Sensor Maintenance Data noise correction Model robustness	Examination of noise impact Understanding AI Model Change Re-learning	Understanding AI Model Change
		R003	Change in Data Quality	○	○		●	*Same as R002	*Same as R002	*Same as R002
		R004	Excess costs						Establish Appropriate Costs	Establish ROI
2	Operator Training	R005	Unintended Application			○			Explain Intended Use-Case	Understanding of AI use-case
		R006	Technological Misinterpretation	○		○	●	Judgment basis output	Make output of judgment basis/reference data easy to understand.	Final decision-making Acquisition of business expertise Identifying errors in judgment
		R007	Inability to do troubleshooting			○		Save system log	Perform systematic maintenance to address damage	Acquisition of business expertise Switch to manual operation
3	Safely Change Raw Materials	R008	Managing Anomalies	○			●	Ensure a sufficient accuracy rate Compensate for abnormal parameters Record Autocorrections	Examine rules for addressing anomalies Display autocorrection Verify model	Final decision-making
		R009	Attacks on Learning Data/Models		○			Secure underlying system	Logic Access Control	
		R010	Attacks on Real-World System		○			Monitor IoT status, etc.	Physical Access Control	

Risk Assessment & Control Summary

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

Step 5

Values & Objectives		Risk No.	Risk Scenario	Uncertainly	Environmental change	Caused by user	RC	Control Summary		
								AI System	AI Service Provider	User
4	Social Responsibility	R011	Response to quality audits	○			●	Record data understanding Record model performance Record control log	Perform internal audits Respond to external audits Access Control	
		R012	Investigation of problems	○				Preserve log data	System operation monitoring Fault handling	Switch to manual operation

Organization

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

Step 5

- Responsible Persons -
Co. B Management

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

- Responsible persons -
Co. A Management

Co. B Quality Control

- Performing internal audits
- Responding to external audits

Co. A Internal Auditing

- Implementation of internal audits

External Auditor

Service Provider
Co. B Engineering Division (AI Development)

- Explanation of use case/materials
- Clear expression of grounds for judgment
- Shared awareness of rules for addressing abnormal parameters.
- Establishment of adequate cost
- Performance of internal audits

- Examine data distribution
- Verification through A/B testing
- Examination of noise impact
- Determination of need for change in AI model
- Understanding AI Model Change
- Changing of AI model

AI Model
Co. C Model Dev. Dept.

- Develop New Model
- Model robustness
- Output of model judgment basis
- Autocorrection of abnormal parameters

AI System
Co. B AI Development PJ

- Sensor maintenance
- Data noise correction
- Saving of usage logs
- System operation monitoring

- Users -
Co. A Operational Management Division at plant.

- Final decision-making
- Understanding of AI use-case
- Acquisition of business expertise
- Identifying errors in judgment
- Shared awareness of rules for addressing abnormal parameters.
- Understanding AI Model Change
- Establish ROI
- Alternative operation

RCM
RiskChainModel

Control Coordination

Control Coordination

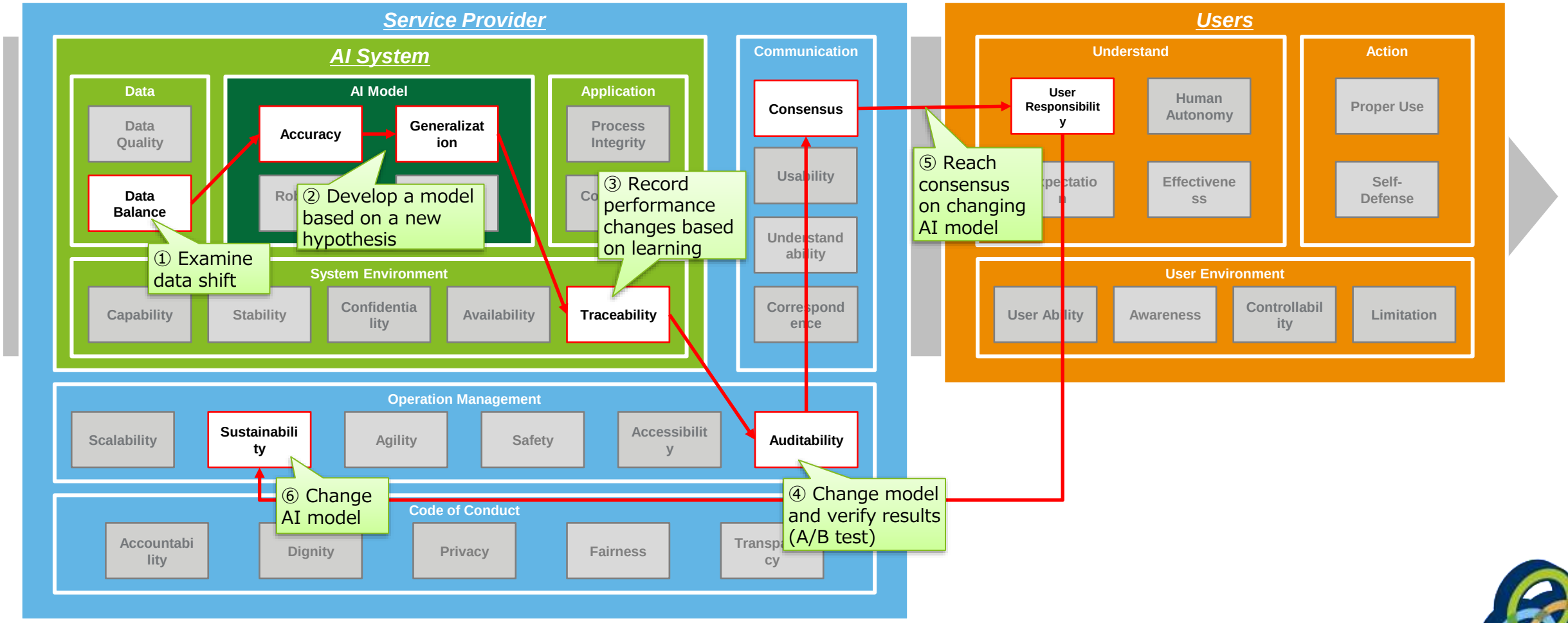
- Examine the risk chain (relation of risk factors) for each important risk scenario -

Step 3

R001

Performance Reduction Due to Environmental Change

Performance declines when raw material composition differs from parameters initially provided and trained on.



Risk Control

- Consider risk control according to the risk chain -

Step 4

R001	**Performance Reduction Due to Environmental Change** Performance declines when raw material composition differs from parameters initially provided and trained on.

Risk Control		
AI System **(AI Dev. Team, Co. B + AI Dev. Dept., Co. C)**	**Service Provider** **(AI Dev. Team, Co. B)**	**Users** **(Plant Operators, Co. A)**
① [Data Balance] Examining data shift (AI Dev Dept., Co. B) ② [Accuracy/Generalization] Retrain/Redevelop model based on new hypothesis (AI Dev. Team, Co. B/AI Dev Dept., Co. C) ③ [Traceability] Record performance while learning with new model (AI Dev. Team, Co. B)	④ [Auditability] Perform A/B testing to verify performance of new & old models (AI Dev Team, Co. B) ⑤ [Consensus] Reach consensus on changing AI model (AI Dev. Team, Co. B) ⑥ [Sustainability] Change AI model (AI Dev. Team, Co. B)	⑤ [User Responsibility] Reach consensus on changing AI model (Plant Operators, Co. A)

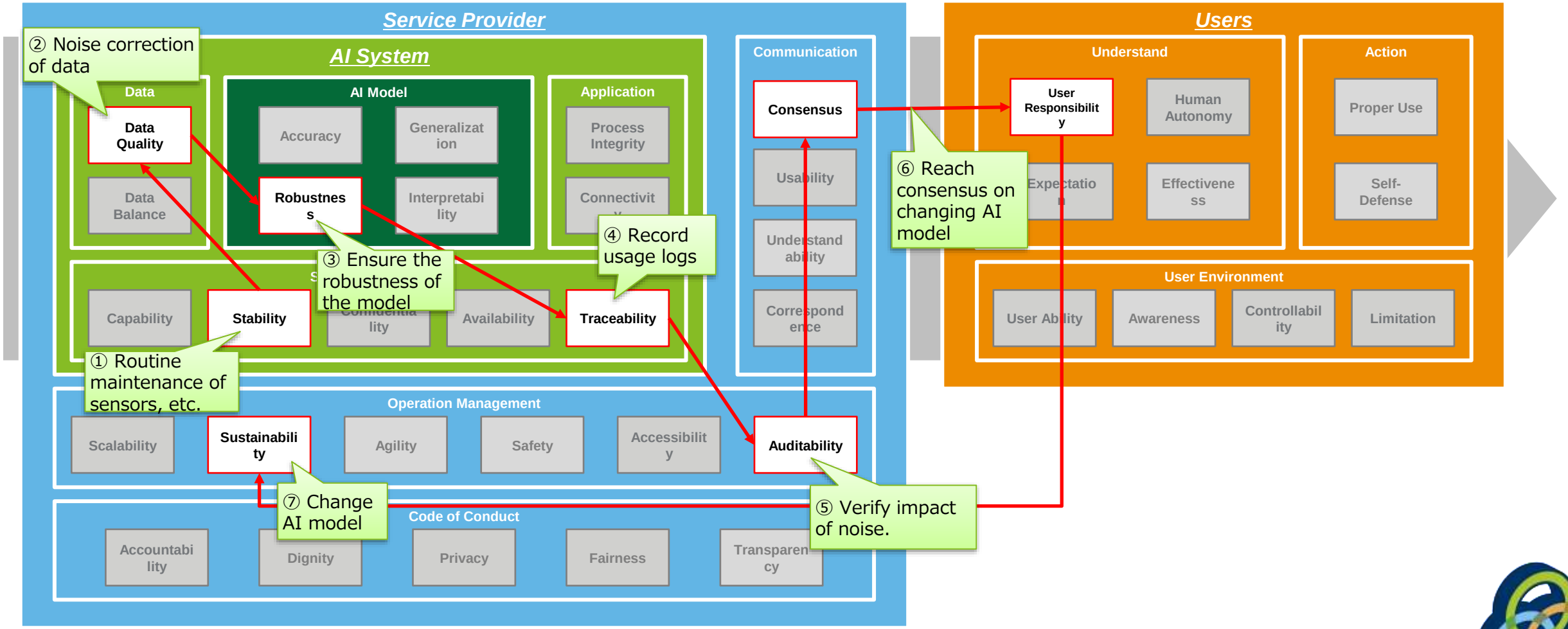
Control Coordination

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

Step 3

R002

Impact of Noise
Introduction of noise destabilizes decision accuracy.



Risk Control

- Consider risk control according to the risk chain -

Step 4

R002	**Impact of noise** Introduction of noise destabilizes decision accuracy.

Risk Control		
AI System **(AI Dev. Team, Co. B + AI Dev. Dept., Co. C)**	**Service Provider** **(AI Dev. Team, Co. B)**	**Users** **(Plant Operators, Co. A)**
① [Stability] Perform regular maintenance on sensors (AI Dev. Team, Co. B) ② [Data Quality] Address degradation of data by noise correction (AI Dev. Team, Co. B) ③ [Robustness] Use learning to enhance model robustness (AI Dev Team, Co. B/AI Dev Dept., Co. C) ④ [Traceability] Store information on AI judgment results when in use (AI Dev Team, Co. B)	⑤ [Auditability] Perform A/B testing to confirm performance of new & old models (AI Dev. Team, Co. B) ⑥ [Consensus] Reach consensus on changing AI model (AI Dev. Team, Co. B) ⑦ [Sustainability] Change AI model (AI Dev. Team, Co. B)	⑥ [User Responsibility] Reach consensus on changing AI model (Plant Operators, Co. A)

Control Coordination

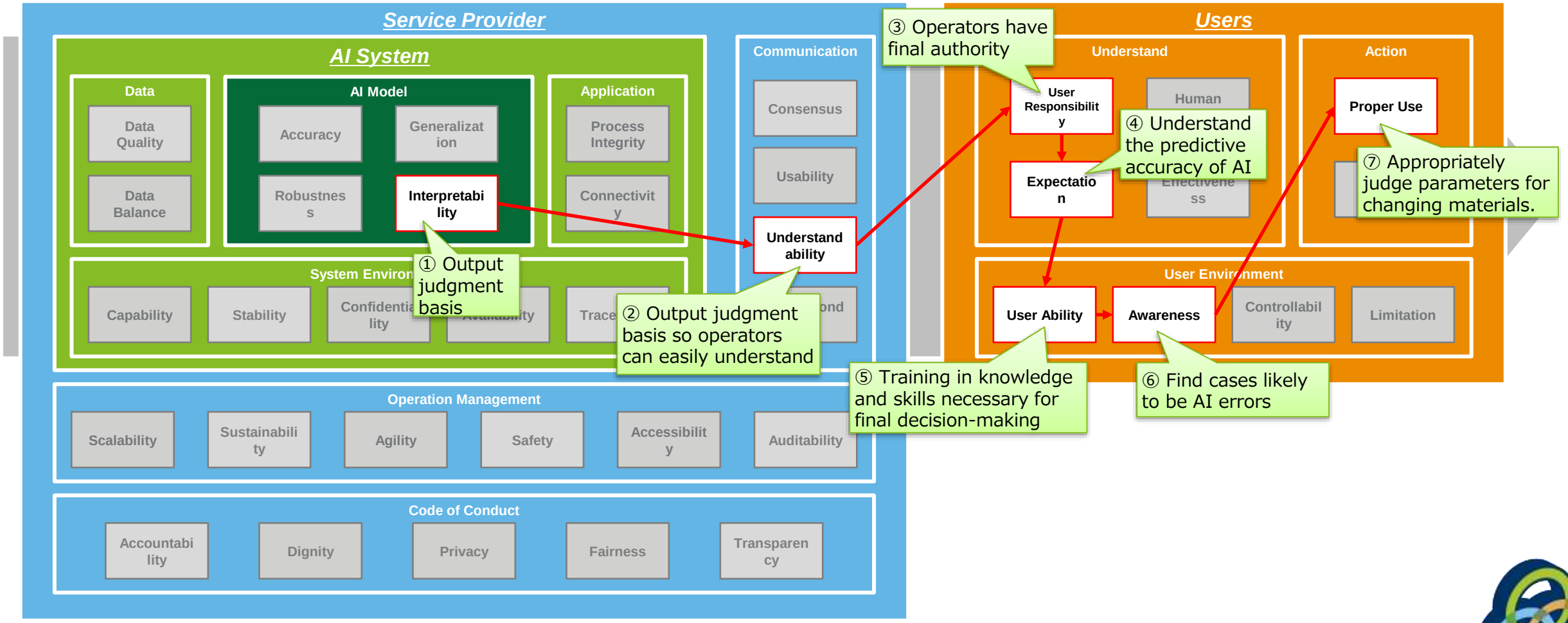
- Examine the risk chain (relation of risk factors) for each important risk scenario -

Step 3

R006

Technological Misinterpretation

If the basis for AI decision-making is not understood, operators will become reliant on the AI and be unable to correct errors.



Risk Control

- Consider risk control according to the risk chain -

Step 4

R006	**Technological Misinterpretation** If the basis for AI decision-making is not understood, operators will become reliant on the AI and be unable to correct errors.

Risk Control		
AI System (AI Dev. Team, Co. B + AI Dev. Dept., Co. C)	**Service Provider (AI Dev. Team, Co. B)**	**Users (Plant Operators, Co. A)**
① [Interpretability] Output the judgment basis of the model (AI Dev. Team, Co. B/AI Dev. Dept., Co. C)	② [Understandability] Output judgment basis, past records, etc. in easily understood form (AI Dev. Team, Co. B)	③ [User Responsibility] It must be understood that operators have final authority (Plant Operators, Co. A) ④ [Expectation] Users must understand the accuracy of model judgments (Plant Operators, Co. A) ⑤ [User Ability] Provide training in knowledge and skills necessary for operation (Plant Operators, Co. A) ⑥ [Awareness] Find places where AI seems to have made mistakes (Plant Operators, Co. A) ⑦ [Proper Use] Accurately judge parameters for changing materials (Plant Operators, Co. A)

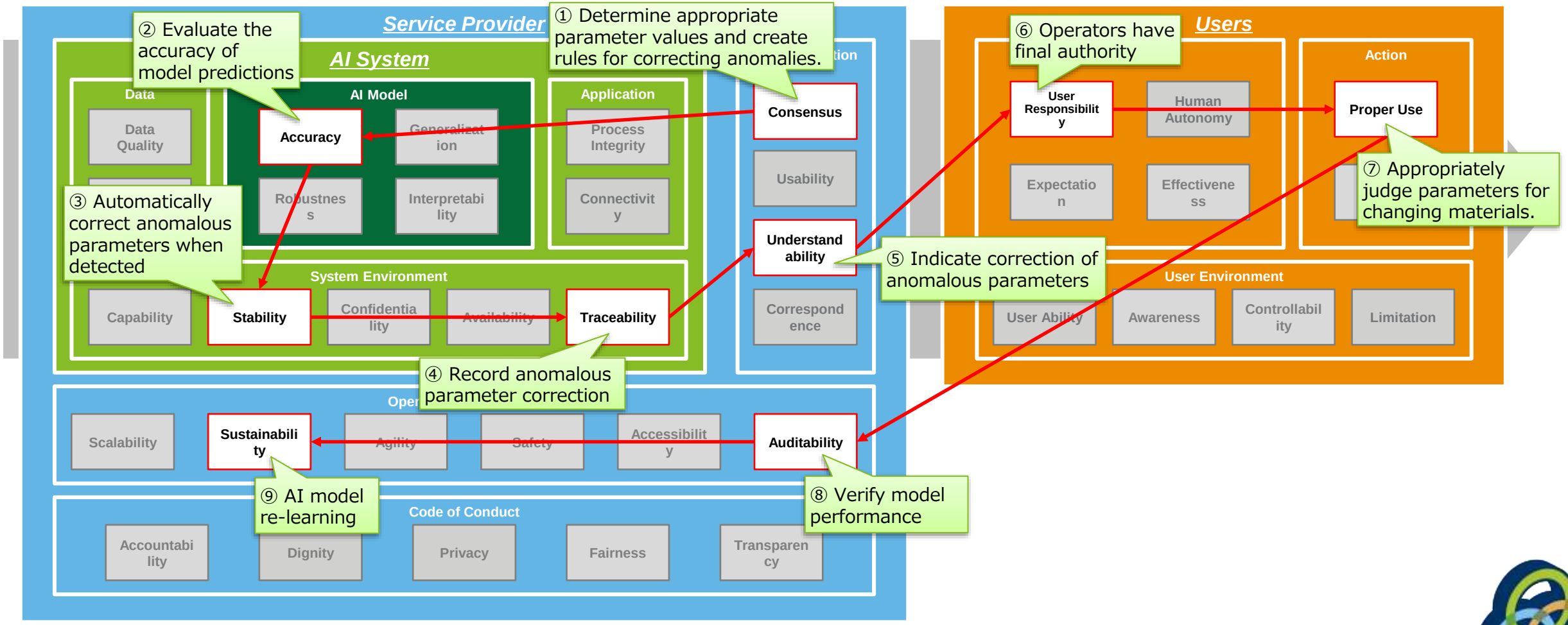
Control Coordination

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

Step 3

R008

Managing Anomalies
Inability to return to normal parameters in the event of abnormal output, interfering with operation of plant.



Risk Control

- Consider risk control according to the risk chain -

Step 4

R008

Managing Anomalies
Inability to return to normal parameters in the event of abnormal output, interfering with operation of plant.

Risk Control		
AI System (AI Dev. Team, Co. B + AI Dev. Dept., Co. C)	**Service Provider (AI Dev. Team, Co. B)**	**Users (Plant Operators, Co. A)**
② [Accuracy] Implement learning to ensure model prediction accuracy (AI Dev. Team, Co. B/AI Dev. Dept., Co. C) ③ [Stability] Automatically correct anomalous parameters when detected (AI Dev. Team, Co. B) ④ [Traceability] Record anomalous parameter correction (AI Dev. Team, Co. B)	① [Consensus] Determine appropriate parameter values and create rules for correcting anomalies. (Plant Operators, Co. A + AI Dev. Team, Co. B) ⑤ [Understandability] Indicate correction of anomalous parameters (AI Dev. Team, Co. B) ⑧ [Auditability] Test model performance and determine need for re-learning (AI Dev. Team, Co. B) ⑨ [Sustainability] Change AI model (AI Dev. Team, Co. B)	⑥ [User Responsibility] It must be understood that operators have final authority (Plant Operators, Co. A) ⑦ [Proper Use] Accurately judge parameters for changing materials (Plant Operators, Co. A)

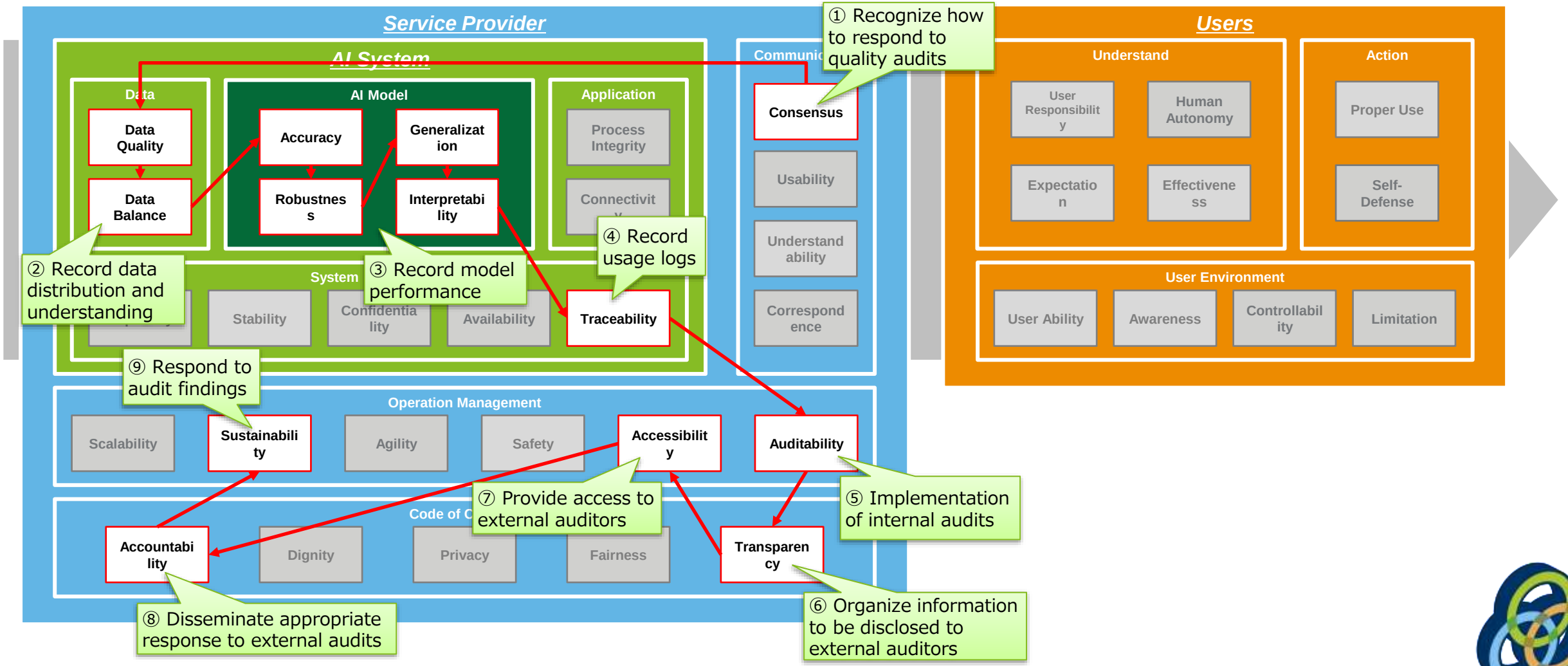
Control Coordination

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

Step 3

R011

Response to Quality Audits
Inability to provide appropriate explanations when subjected to quality audits



Risk Control

- Consider risk control according to the risk chain -

Step 4

R011	**Response to Quality Audits** Inability to provide appropriate explanations when subjected to quality audits

Risk Control		
AI System **(AI Dev. Team, Co. B + AI Dev. Dept., Co. C)**	**Service Provider** **(AI Dev. Team, Co. B)**	**Users** **(Plant Operators, Co. A)**
② [Data Quality/Data Balance] Record data distribution and understanding (AI Dev. Team, Co. B) ③ [Accuracy/Robustness/Generalization/Interpretability] Record model performance (AI Dev. Team, Co. B) ④ [Traceability] Store information on AI judgment results when in use (AI Dev. Team, Co. B)	① [Consensus] Determine appropriate parameter values and create rules for correcting anomalies. (Plant Operators, Co. A + AI Dev. Team, Co. B) ⑤ [Auditability] Conduct internal audits and respond in advance (Internal Audit Dept., Co. A) ⑥ [Transparency] Prepare information to be provided to external auditors (Plant Operational Management, Co. A + AI Dev. Team, Co. B) ⑦ [Accessibility] Establish necessary access rights for external auditors (AI Dev. Team, Co. B) ⑧ [Accountability] Disseminate appropriate response to external audits (Plant Operational Management, Co. A + AI Dev. Team, Co. B) ⑨ [Sustainability] Address issues discovered during audits (AI Dev. Team, Co. B)	