RCM
RiskChainModel

Risk Assessment & Control Coordination for AI services :
Case09 Smart Appliance Optimization

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

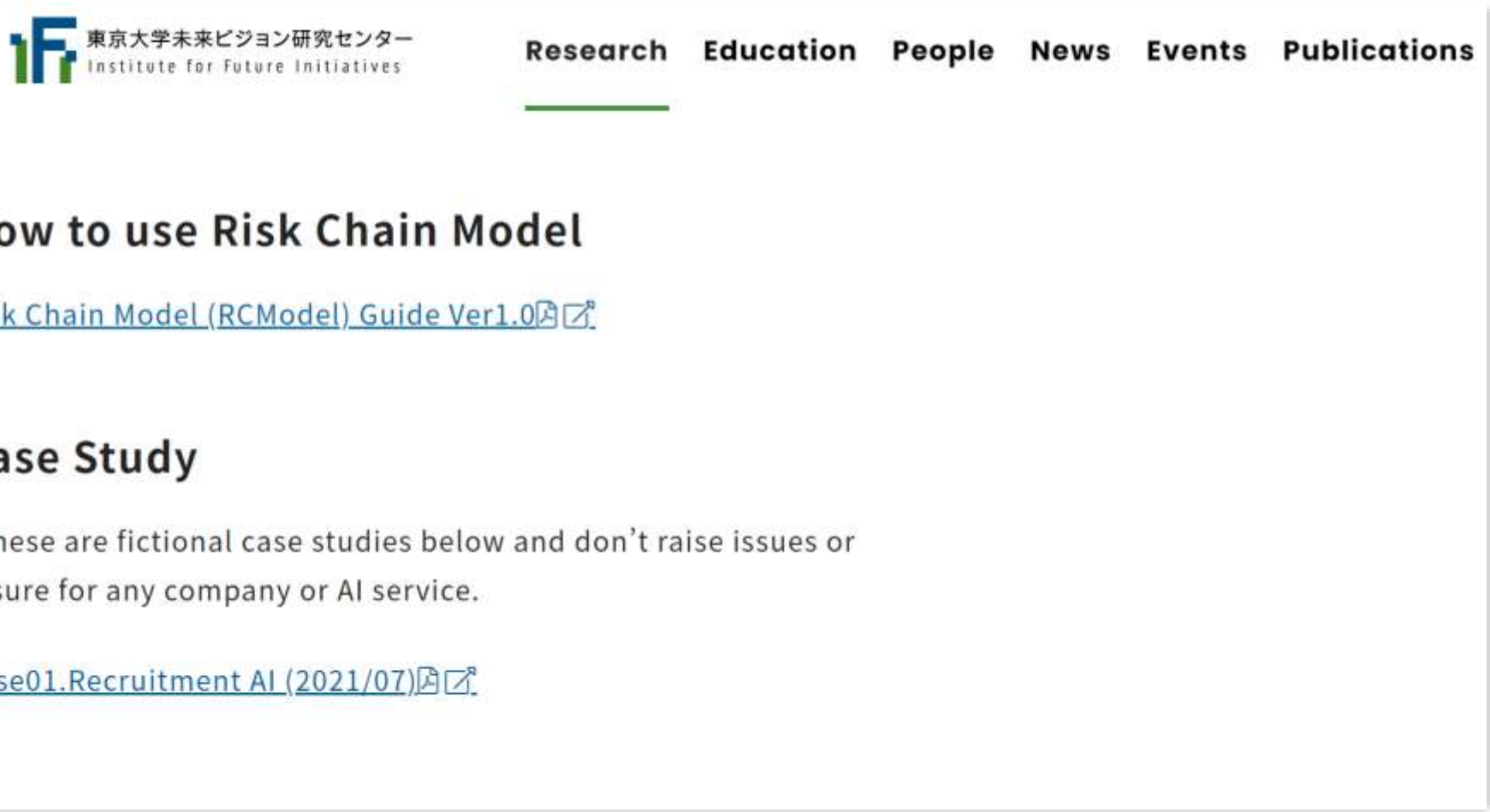
How to operate the RCModel
- Risk Assessment & Control Coordination -

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

Case09 : Smart Appliance Optimization

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

Step 1

The AI model analyzes environmental information, user behavior, etc., to optimize smart appliances. It obtains information from sensors installed in the space (user location and state, temperature, humidity, illumination, CO2 concentration), open data (weather information), and feedback from the user (opinions on stress, comfort level, etc.) and performs analysis to automatically control smart home appliances (smart refrigerators (food management, recipe suggestions, etc.), air conditioning, floor heating, air purifiers, robotic vacuum cleaners, ventilation systems, etc.).

[Values & Objectives]

- **Maintaining a comfortable space**
- **Prevention of adverse health effects**
- **Maintaining and improving economic efficiency (for both businesses and consumers)**
- **Corporate social responsibility**

[Flow of Actual Operations using AI Services]

This AI system can be installed in any given space (household, office, facility floor, etc.). Compatible home appliances are published by Company A, the company that developed this AI service. The Company sets up sensing and other settings at the time of installation, and performs periodic inspections, etc. according to the maintenance contract with the user.

Each system has its own AI model, and in principle, training data is not shared between different systems. However, training data can be shared upon request if the same user owns both systems.

The model is based on multimodal machine learning using various types of information. The expected accuracy is initially set using a general comfort level as a benchmark, and each system's AI model learns independently as it receives user feedback after use. User feedback is as follows, obtained via a smartphone application, and tuning of the AI model is performed every 1-2 weeks.

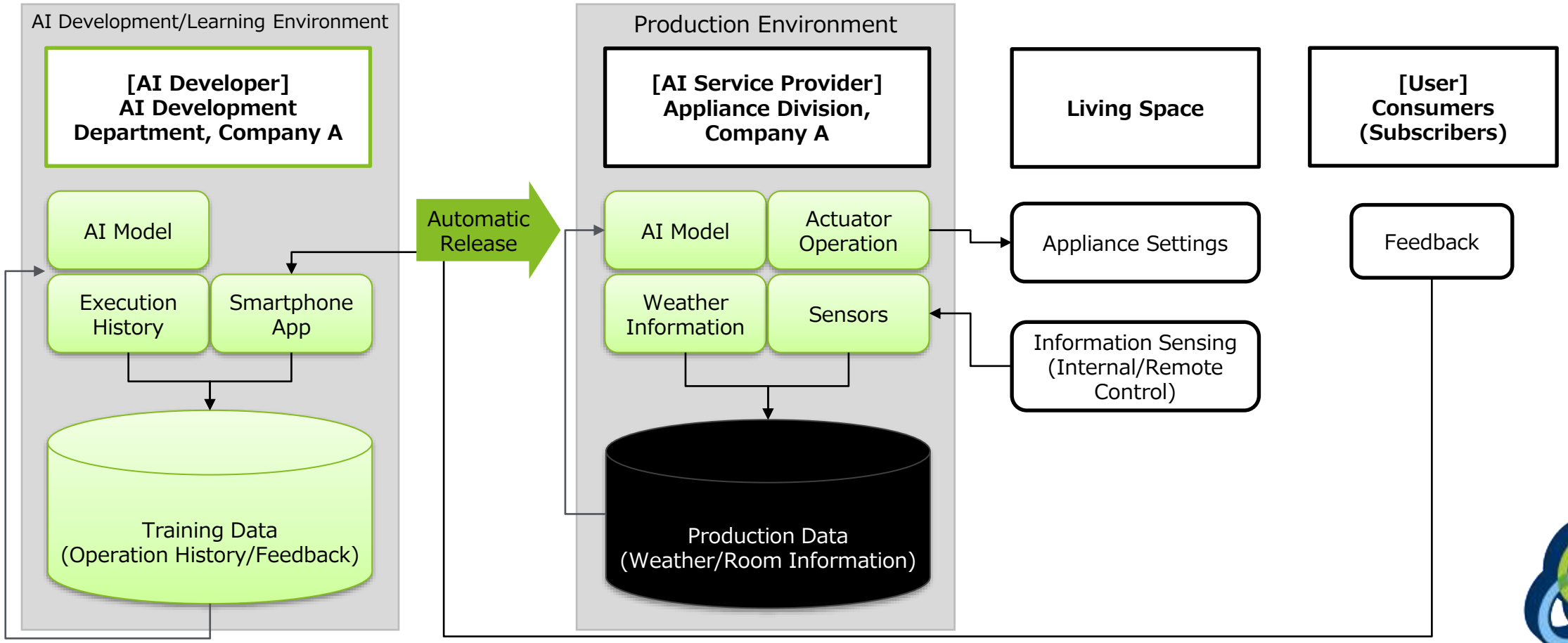
Stress: automatically collected when the app is left running near the user, such as when sleeping

Opinions on comfort level: the user can answer from his/her smartphone at any time.

Case09 : Smart Appliance Optimization

- System Overview -

AI System	Company A) AI Development Department	Predicts optimal settings for smart appliances
AI Service Provider	Company A) Appliance Division	Provides services to users in conjunction with smart appliances
Users	Consumers (Subscribers)	Use this AI service in their own living space



Case09 : Smart Appliance Optimization

- Input & Output -

[Input Data]

Data	Purpose	Collection Method	Data Manager	Including Privacy Data
Room information (size, temperature, humidity, illumination, CO2 concentration, etc.)	Learning	Automatically collected by sensors installed in the room (usually inside the appliances)	Company A's cloud environment (includes all users)	Yes (room information)
User feedback (stress, comfort level)	Learning	Collected from users via a smartphone app (stress: automatic collection, comfort level: user input)	Company A's cloud environment (includes all users)	Yes
Weather information	Learning	Collected from external sources	Respective open data information	No
Room information (size, temperature, humidity, illumination, CO2 concentration, etc.)	Production	Current information automatically collected by sensors installed in the room	Company A's cloud environment	Yes (room information)
Weather information	Production	Collected from external sources	Respective open data information	No

Case09 : Smart Appliance Optimization

- Input & Output -

[Output]

Users	Occupants of the spaces in which this system operates
Output	Control of home appliances (temperature and humidity control, ventilation)
Output Method	Commands to smart appliances
Expected Accuracy	No strict settings, but minimize negative feedback from users.
User judgment	Users can turn off automatic control by AI
Output of evidence information	No
Safety Risk	Adverse health effects
Connection with external system	AI systems in multiple spaces belonging to the same user may share training data
Protocol	Possibility of sharing training data

RCM
RiskChainModel

Risk Assessment

Risk Assessment

- Examination of significant risk scenarios -

Step 2

Values & Objectives		Service Requirement			Risk No.	Risk Scenario	
1	Maintaining a comfortable space	1-1	Ensuring proper operation and performance	■ AI prediction accuracy ■ AI robustness ■ Easy-to-use UI	R001	Inadequate usability	Inability to properly operate AI services results in a less comfortable space, which drives customers away
					R002	Impact of noise	Noise in the sensors degrades the accuracy of AI decisions
		1-2	Adapting to changes	■ Management of compatible devices ■ Collection of teacher data	R003	Change in cohabitants	Inability to cope with changes in residents (childbirth, nursing care, etc.)
					R004	Response to environmental changes	Inability to cope with another residence or seasonal changes
		1-3	Proper use by users	■ User aids	R005	Connection of unsupported devices	Abnormal behavior occurs when unsupported device are forcibly connected
2	Prevention of adverse health effects	2-1	Managing Anomalies	■ Ensuring safety	R006	Health deterioration due to abnormal behavior	Abnormal settings are sent to the devices, resulting in health hazards for users
		2-2	Ensuring fairness	■ Control of abnormal instructions	R007	Malicious feedback	Providing malicious feedback worsens the health of certain cohabitants (aids domestic violence)
3	Maintaining and improving economic efficiency	3-1	Business-side economic management		R008	Excess costs	Operation costs are exceeded
		3-2	Consumer-side economic management		R009	Increased utility costs	Continuous execution of fine behaviors increases utility costs
4	Corporate social responsibility	4-1	Accountability	■ Process description ■ Verifiability	R010	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.
		4-2	Information management	■ Data management	R011	Protection of personal information	Training data and execution history information are leaked to outside parties and misused
		4-3	Environmental impact	■ Proper execution	R012	Increased greenhouse gas emissions	Continuous execution of fine behaviors increases greenhouse gas emissions

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	Maintaining a comfortable space (prevention of adverse effects on health, etc.)	R001	Inadequate usability	○			●	Saving feedback Model updates	Validating feedback User surveys	Feedback
		R002	Impact of noise	○	○		●	Equipment maintenance Data noise correction Model robustness	Examination of maintenance and cleaning methods Collaboration with compatible device manufacturers Notes to users	Remote control placement and cleaning Alerts to the app
		R003	Change in cohabitants	○	○	○	●	Initialization of training data Model updates	Automated development environment	Understanding of dietary and environmental constraints Model initialization function
		R004	Response to environmental changes	○	○		●	Automatic selection of training data Model updates	Automated development environment	Training data sharing function
		R005	Connection of unsupported devices			○		Restrictions on unsupported devices	Clarification of compatible devices	Selecting the right equipment
2	Prevention of adverse health effects	R006	Health deterioration due to abnormal behavior	○			●	Basis of decision Verifiability Automatic correction of abnormal values	Understanding of safety thresholds UI for switching to manual operation Collaboration with compatible device manufacturers Confirmation of abnormality causes	Consultations for abnormal behavior
		R007	Malicious feedback			○	●	Restrictions on abnormal feedback Automatic correction of abnormal values Verifiability	Understanding of safety thresholds Verification of malicious use Consideration of responses with the legal department, etc.	Alerts for abnormal feedback

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
3	Maintaining and improving economic efficiency (for both businesses and consumers)	R008	Excess costs						Cost management	
		R009	Increased utility costs					Adoption of energy-saving technologies		Cost management
4	Corporate social responsibility	R010	Investigation of problems	○				Preserve log data	System operation monitoring Fault handling	
		R011	Protection of personal information					Data protection	Security management	
		R012	Increased greenhouse gas emissions					Prevention of excessive operation Adoption of energy-saving technologies		

Organization

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

Step 5

- Responsible Persons -
Company A) Management

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

Consumer Affairs Agency

Company A)
Legal/Compliance

- Dealing with malicious users

Company A)
Quality Control

- Implementation of internal audits
- Responding to external audits

- AI Service Provider -
Company A) Appliance Division

- Preparation of an automated development environment
- Understanding of safety thresholds
- UI for switching to manual operation
- Validation of feedback
- Clarification of compatible devices
- Examination of inspection and maintenance methods
- Clarification of abnormal conditions
- System operation monitoring
- Security management

Company A)
AI Development Department

- Predictive performance of the model
- Model robustness
- Securing training data
- Automatic selection of training data
- Automatic correction of abnormal values
- Automatic alert processing
- Sensor specifications and protocols

Company A)
Customer Support

- User surveys
- Handling maintenance
- Dealing with malicious users

- Users -
Consumers (Subscribers)

- Proper use

(App)

- Comfort level feedback
- Model initialization function
- Training data sharing function
- Alert function

Compatible device manufacturers

- Cooperation in inspection and maintenance

Cloud Services Department, Company A

- Recording of user feedback
- Recording of decision basis

Cohabitants

RCM
RiskChainModel

Control Coordination

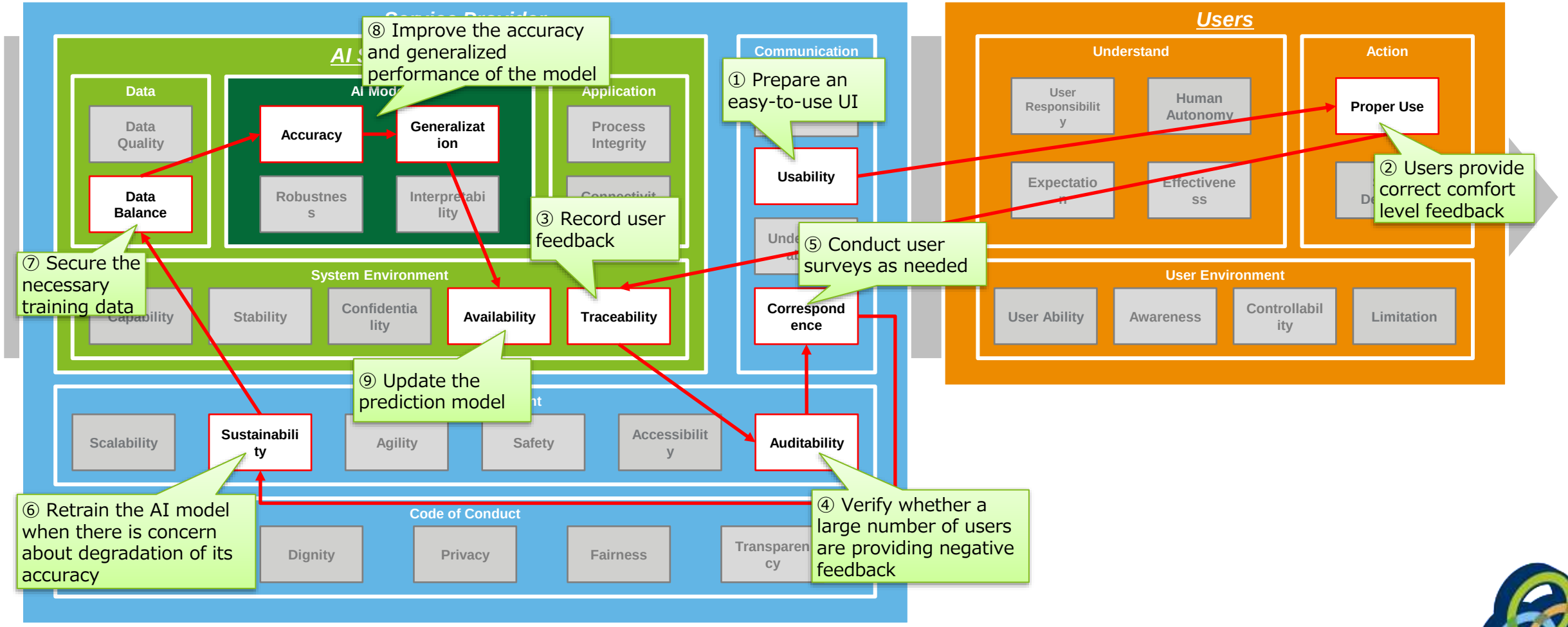
Control Coordination

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

Step 3

R001

Inadequate usability
Inability to properly operate AI services results in a less comfortable space, which drives customers away



Risk Control

- Consider risk control according to the risk chain -

Step 4

R001	**Inadequate usability** Inability to properly operate AI services results in a less comfortable space, which drives customers away

Risk Control		
AI System (AI Development Department, Company A)	**Service Provider (Appliance Division, Company A)**	**Users (Consumers (Subscribers))**
③ [Traceability] Record user feedback (Company A, Cloud Services Department) ⑦ [Data Balance] Secure sufficient training data (AI Development Department, Company A) ⑧ [Accuracy] [Generalization] Ensure sufficient model accuracy and generalized performance during training (AI Development Department, Company A) ⑨ [Availability] Update the prediction model (AI Development Department, Company A)	① [Usability] Design an easy-to-use UI (Appliance Division, Company A) ④ [Auditability] Verify whether a large number of users are providing negative feedback (Appliance Division, Company A) ⑤ [Correspondence] Conduct user surveys as necessary (Appliance Division, Company A/Customer Service, Company A) ⑥ [Sustainability] Request re-training of AI models to ensure sufficient accuracy (Appliance Division, Company A)	② [Proper Use] Provide feedback on comfort level via the smartphone app (Consumers)

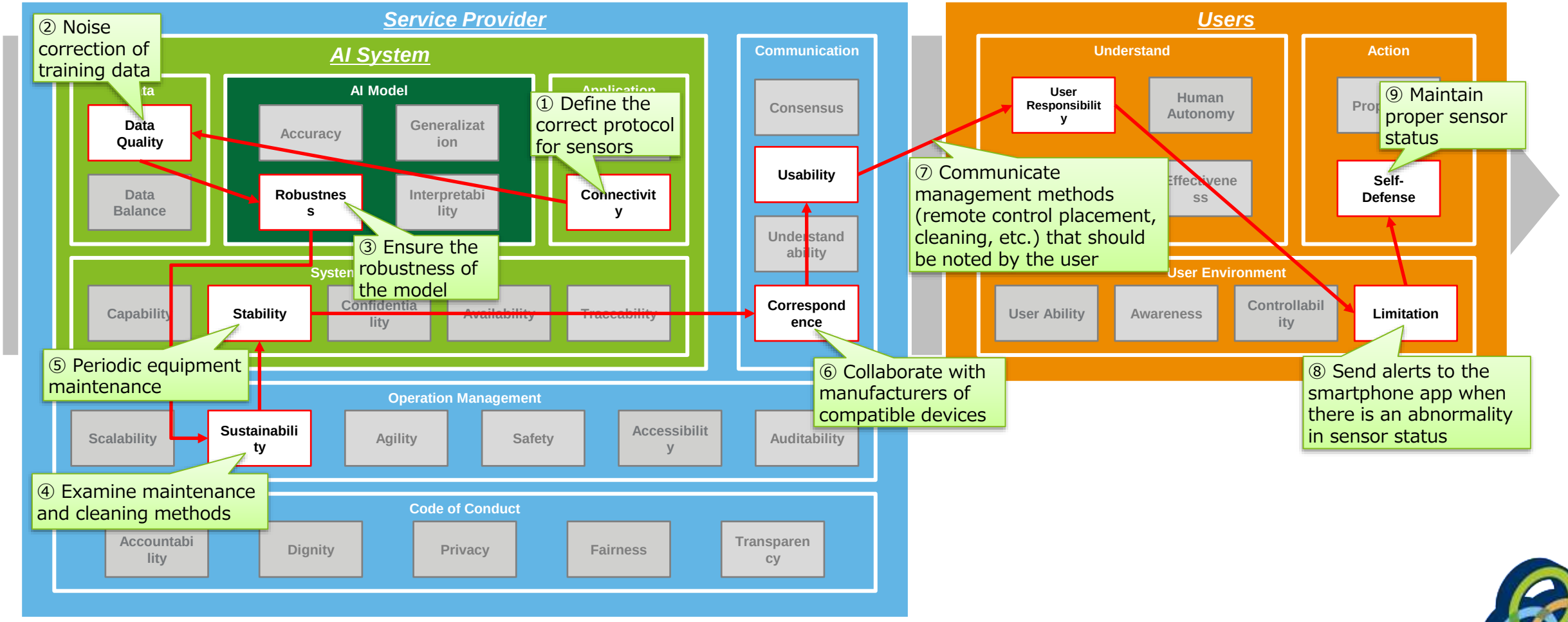
Control Coordination

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

Step 3

R002

Impact of noise
Noise in the sensors degrades the accuracy of AI decisions



Risk Control

- Consider risk control according to the risk chain -

Step 4

R002	**Impact of noise** Noise in the sensors degrades the accuracy of AI decisions

Risk Control		
AI System (AI Development Department, Company A)	**Service Provider (Appliance Division, Company A)**	**Users (Consumers (Subscribers))**
① [Connectivity] Implement sensors according to the correct specifications and protocols (AI Development Department, Company A) ② [Data Quality] Address degradation of training data through noise correction, etc. (AI Development Department, Company A) ③ [Robustness] Train the model to increase its robustness (AI Development Department, Company A) ⑤ [Stability] Perform proper maintenance of sensors and other components of smart appliances (AI Development Department, Company A)	④ [Sustainability] Examine methods of maintaining and cleaning smart appliances (Appliance Division, Company A) ⑥ [Correspondence] Collaborate with manufacturers of compatible devices on maintenance (Appliance Division, Company A + Compatible Device Manufacturers) ⑦ [Usability] Communicate management methods (remote control placement, cleaning, etc.) that should be noted by the user (Customer Support, Company A)	⑦ [User Responsibility] Understand the management methods (remote control placement, cleaning, etc.) that should be noted by the user (Consumers) ⑧ [Limitation] Send alerts to the smartphone app when there is an abnormality in sensor status (Appliance Division, Company A) ⑨ [Proper Use] Properly maintain the condition of the sensors (Consumers)

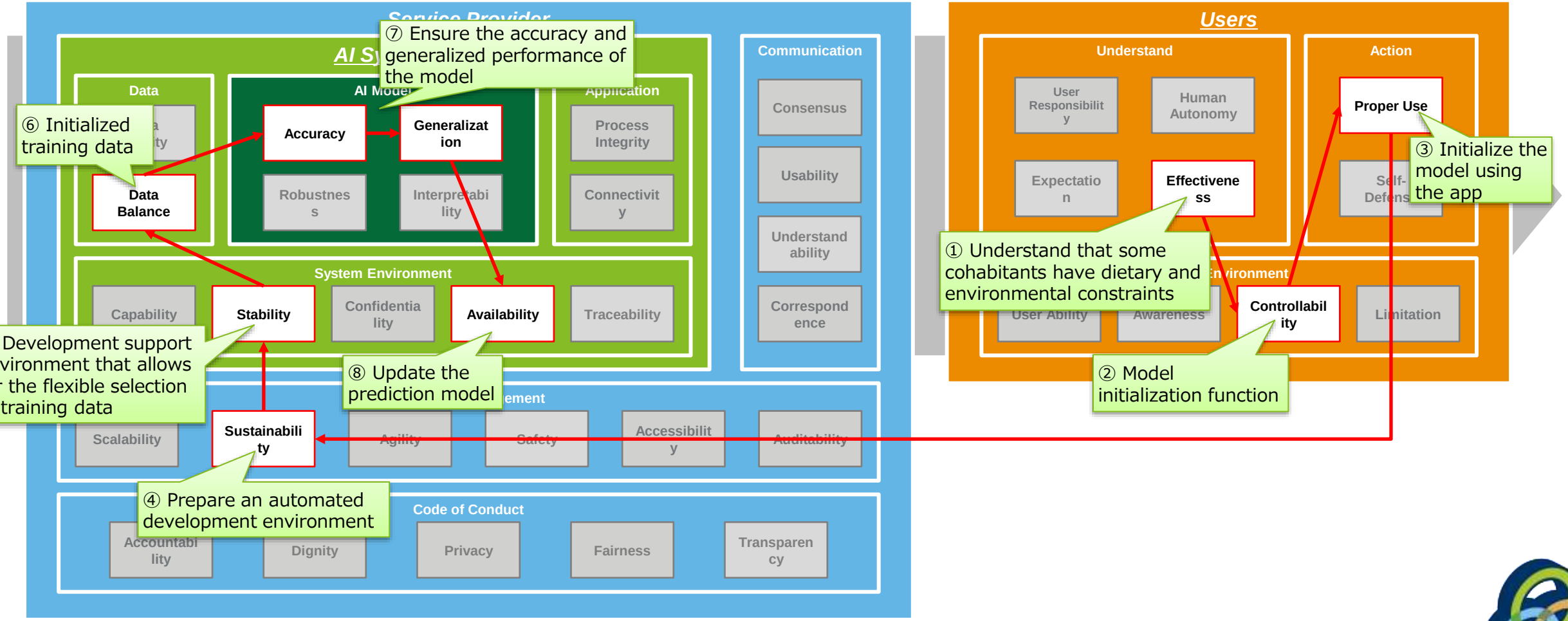
Control Coordination

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

Step 3

R003

Change in cohabitants
Inability to cope with changes in residents (childbirth, nursing care, etc.)



Risk Control

- Consider risk control according to the risk chain -

Step 4

R003	**Change in cohabitants** Inability to cope with changes in residents (childbirth, nursing care, etc.)

Risk Control		
AI System (AI Development Department, Company A)	**Service Provider (Appliance Division, Company A)**	**Users (Consumers (Subscribers))**
⑤ [Stability] Provide a development support environment that allows for the flexible selection of training data (AI Development Department, Company A) ⑥ [Data Balance] Prepare initialized training data (AI Development Department, Company A) ⑦ [Accuracy] [Generalization] Ensure sufficient model accuracy and generalized performance during training (AI Development Department, Company A) ⑧ [Availability] Update the prediction model (AI Development Department, Company A)	④ [Sustainability] Prepare an automated development environment (Appliance Division, Company A)	① [Effective] Understand that some cohabitants may have dietary, environmental, or other constraints (Consumers) ② [Controllability] Include a model initialization function in the application (Appliance Division, Company A) ③ [Proper Use] Initialize the model (Consumers)

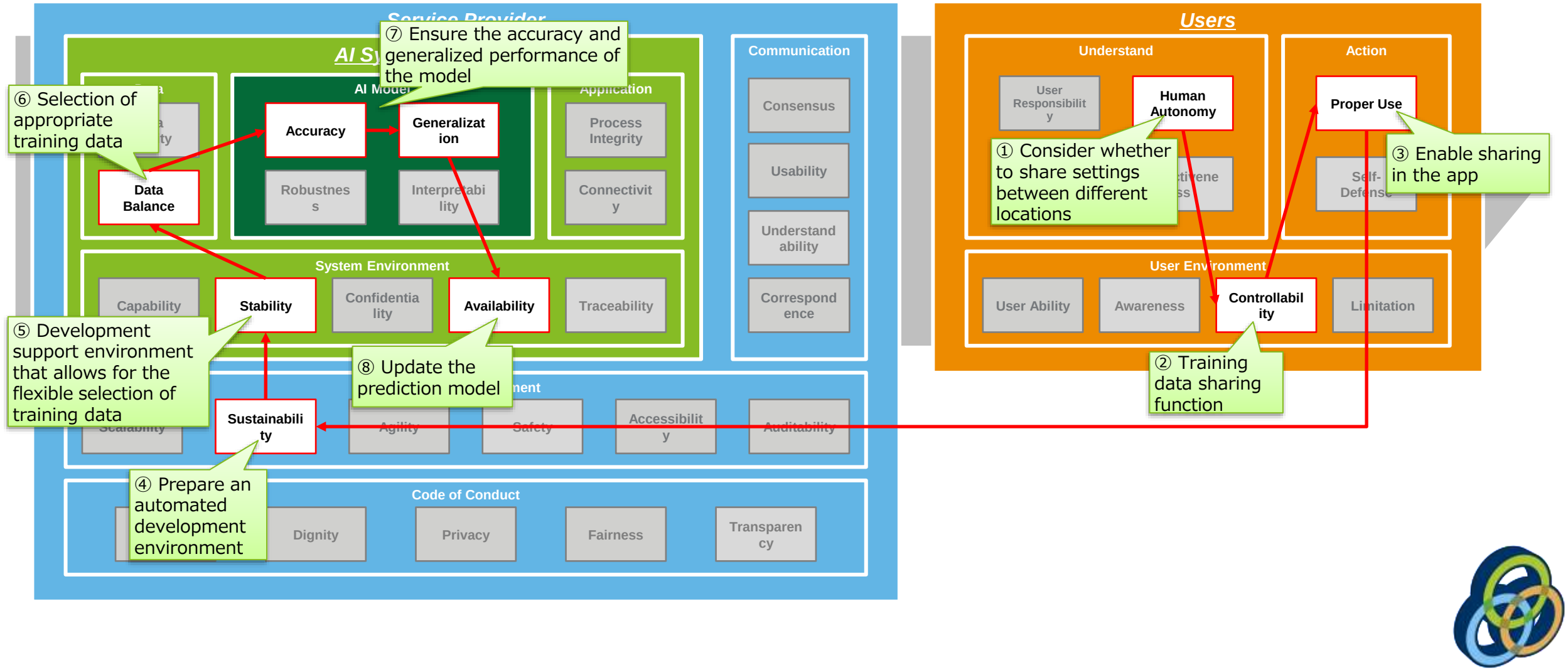
Control Coordination

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

Step 3

R004

Response to environmental changes
Inability to cope with another residence or seasonal changes



Risk Control

- Consider risk control according to the risk chain -

Step 4

R004	**Response to environmental changes** Inability to cope with another residence or seasonal changes

Risk Control		
AI System (AI Development Department, Company A)	**Service Provider (Appliance Division, Company A)**	**Users (Consumers (Subscribers))**
⑤ [Stability] Provide a development support environment that allows for the flexible selection of training data (AI Development Department, Company A) ⑥ [Data Balance] Secure appropriate training data (AI Development Department, Company A) ⑦ [Accuracy] [Generalization] Ensure sufficient model accuracy and generalized performance during training (AI Development Department, Company A) ⑧ [Availability] Update the prediction model (AI Development Department, Company A)	④ [Sustainability] Prepare an automated development environment (Appliance Division, Company A)	① [Human Autonomy] Consider whether to share settings between different locations (Consumers) ② [Controllability] Include a function to share training data between different locations in the application (Appliance Division, Company A) ③ [Proper Use] Use the application to set up the sharing of training data between different locations (Consumers)

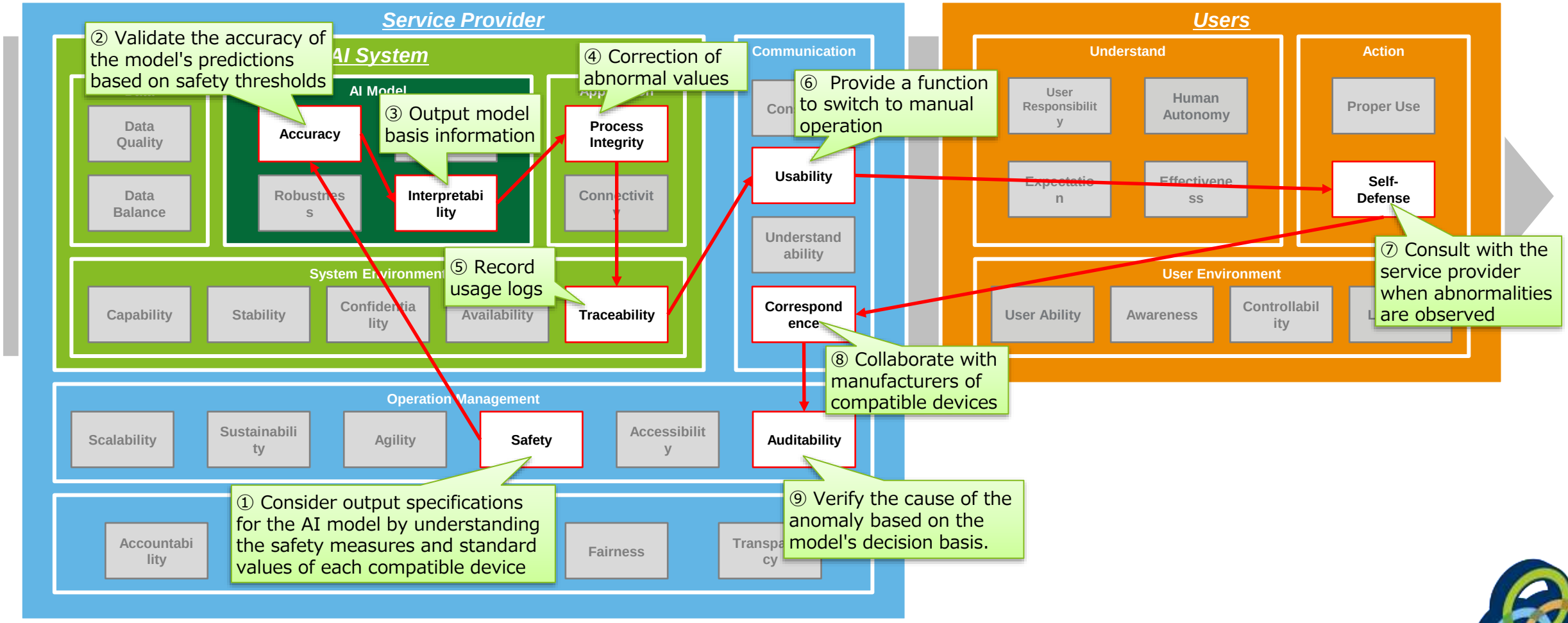
Control Coordination

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

Step 3

R006

Health deterioration due to abnormal behavior
Abnormal settings are sent to the devices, resulting in health hazards for users



Risk Control

- Consider risk control according to the risk chain -

Step 4

R006	**Health deterioration due to abnormal behavior** Abnormal settings are sent to the devices, resulting in health hazards for users

Risk Control		
AI System (AI Development Department, Company A)	**Service Provider (Appliance Division, Company A)**	**Users (Consumers (Subscribers))**
② [Accuracy] Validate the accuracy of the model's predictions based on safety thresholds (AI Development Department, Company A) ③ [Interpretability] Output the model's decision basis (AI Development Department, Company A) ④ [Process Integrity] Automatically compensate for abnormal values (AI Development Department, Company A) ⑤ [Traceability] Save usage logs along with the decision basis (AI Development Department, Company A)	① [Safety] Consider output specifications for the AI model by understanding the safety measures and standard values of each compatible device (Appliance Division, Company A) ⑥ [Usability] Provide a function/UI to switch to manual operation (Appliance Division, Company A) ⑧ [Correspondence] Perform equipment maintenance in cooperation with manufacturers of compatible devices (Appliance Division, Company A + Compatible Device Manufacturers) ⑨ [Auditability] Verify the cause of the anomaly based on the model's decision basis (Appliance Division, Company A)	⑦ [Self-Defense] Consult with the service provider when abnormalities are observed (Consumers)

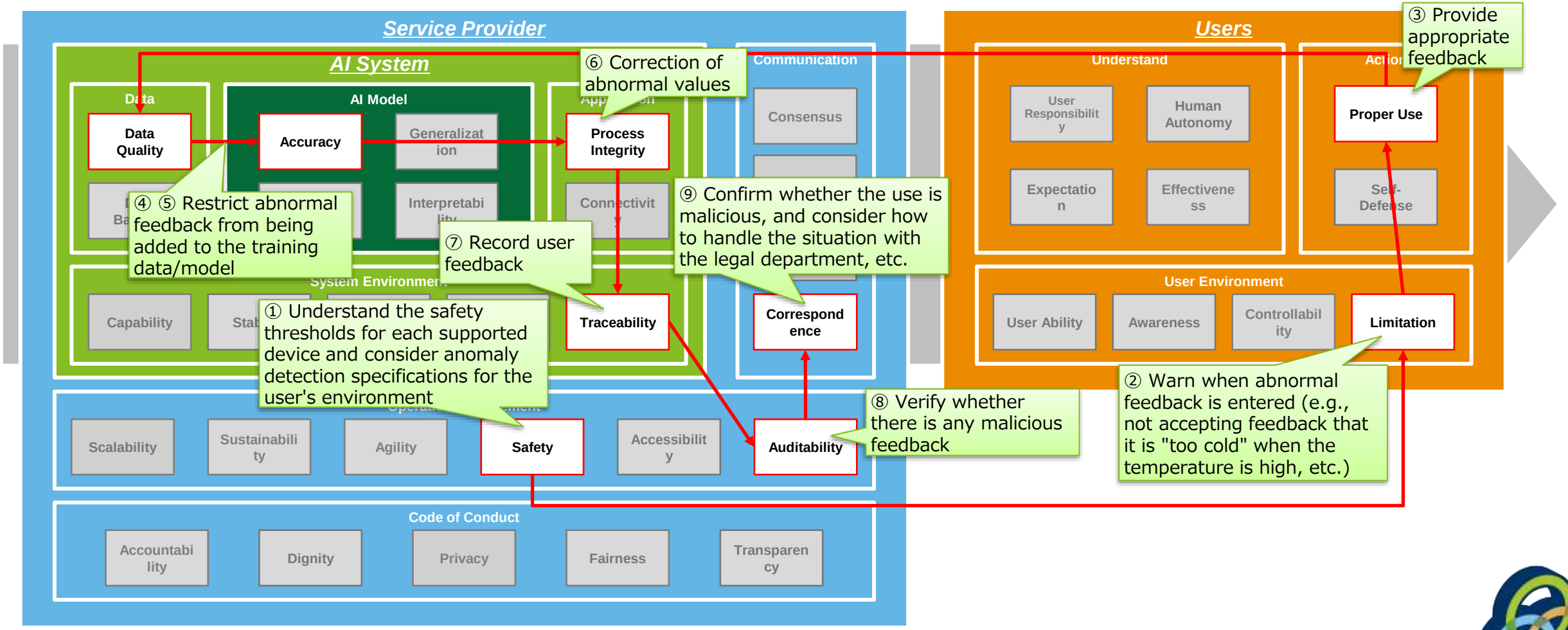
Control Coordination

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

Step 3

R007

Malicious feedback
Providing incorrect feedback worsens the health of certain cohabitants (aids domestic violence)



Risk Control

- Consider risk control according to the risk chain -

Step 4

R007	**Malicious feedback** Providing incorrect feedback worsens the health of certain cohabitants (aids domestic violence)

Risk Control		
AI System (AI Development Department, Company A)	**Service Provider (Appliance Division, Company A)**	**Users (Consumers (Subscribers))**
④ [Data Quality] Restrict abnormal feedback from being added to training data (AI Development Department, Company A) ⑤[Accuracy] Restrict abnormal feedback from being added to the predictive model (AI Development Department, Company A) ⑤ [Process Integrity] Automatically compensate for abnormal values (AI Development Department, Company A) ⑥ [Traceability] Record user feedback (AI Development Department, Company A)	① [Safety] Understand the safety thresholds for each supported device and consider anomaly detection specifications for the user's environment (Appliance Division, Company A) ⑦ [Auditability] Verify whether there is any malicious feedback (Appliance Division, Company A) ⑧ [Sustainability] Verify the facts when a malicious user is recognized and consider necessary actions (Customer Service, Company A/Legal and Compliance Department, Company A)	② [Limitation] Warn when abnormal feedback is entered (e.g., not accepting feedback that it is "too cold" when the temperature is high, etc.) (Appliance Division, Company A) ③ [Proper Use] Provide appropriate feedback (Consumers)