

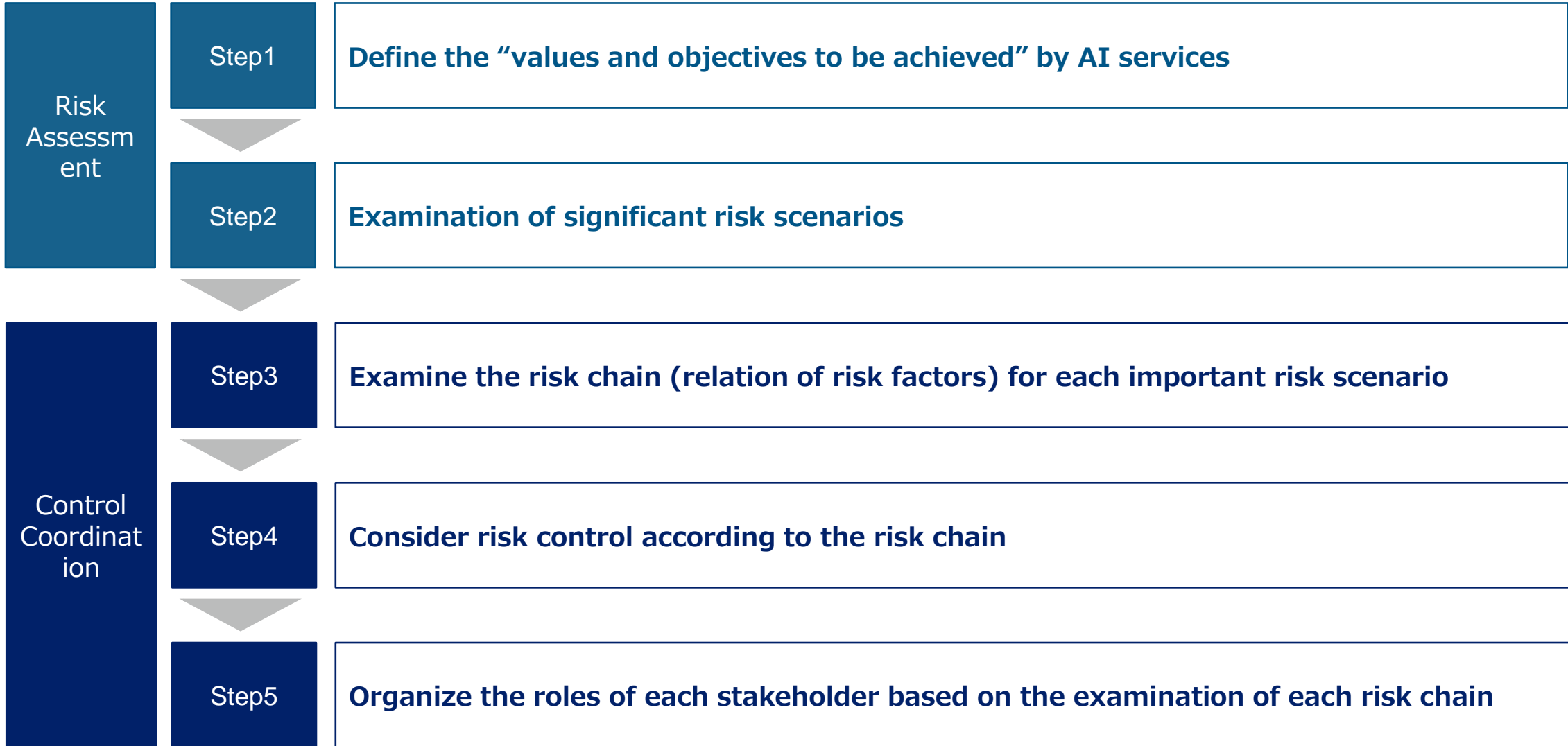
Risk Assessment & Control Coordination for AI services : Case05 Guide Robot

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/>



東京大学未来ビジョン研究センター
Institute for Future Initiatives

Research

Education

People

News

Events

Publications

How to use Risk Chain Model

[Risk Chain Model \(RCModel\) Guide Ver1.0](#)

Case Study

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

[Case01.Recruitment AI \(2021/07\)](#)

Case Study



Case05 : Guide Robot

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

This is a guide robot at the airport that covers facilities and gates. Services are realized by combining multiple special-purpose deep-learning models.

There are six robots in airport A that are built with multiple artificial intelligence (AI) models and automated processes; they provide services by interacting with specific users. The supported languages are Japanese, English, Chinese, and Korean.

The development of the deep-learning model and the body of the robot is entrusted to Co. X, but both the learning environment and the production environment are limited to the A airport. If there is a tendency that accuracy deteriorates significantly after operation, although sequential learning is not performed, Co. X will request additional learning and relearning.

[Values & Objectives]

- Improved user satisfaction (Safe guidance)
- Ensuring safety
- Reduction of burden on airport staff
- Corporate social responsibility

[Robot Status and Processing Flow]

The robot provides services while switching between the following states:

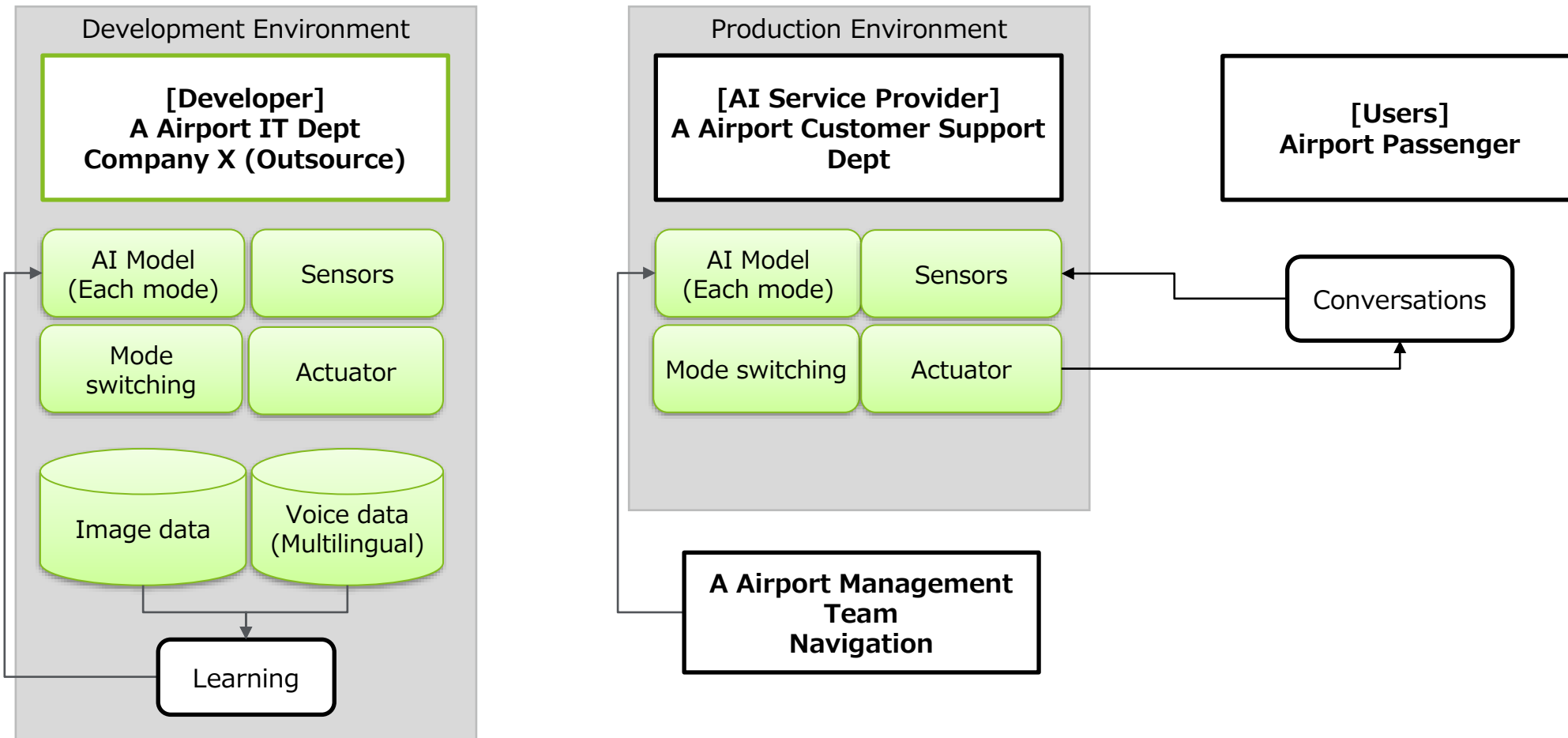
- ① [Search Mode] Recognizes spoken language (object detection/speech recognition)
- ② [Recognition Mode] Talks to airport users to guide them (object detection)
- ③ [Conversation Mode] Converses with users until the destination is identified (speech recognition/natural language processing)
- ④ [Confirmation Mode] The destination candidate to be guided is found, and the user is confirmed (speech recognition/natural language processing).
- ⑤ Guide users to the destination.



Case05 : Guide Robot

- System Overview -

AI System	A Airport IT Dept + Company X (Outsources)	Development and operation of robots and AI models
AI Service Provider	A Airport Customer Support Dept	Customer service including robot operation at A Airport
User	Airport Passenger	Use a robot to get to one's destination



Case05 : Guide Robot

- Input & Output -

[Input Data]

Data	Purpose	Collection Method	Data Manager	Including Privacy Data
Image data of the person	Learning	Always collect images from airport surveillance cameras, etc.	A Airport IT Dept (Airport Development Environment)	Yes
Voice data of the person	Learning	Always collect audio in the airport	A Airport IT Dept (Airport Development Environment)	Yes
Conversation data of the person	Learning	Always collect audio inside the airport	A Airport IT Dept (Airport Development Environment)	Yes
Image data of the person	Production	On each use	A Airport IT Dept (Airport Production Environment)	Yes
Voice data of the person	Production	On each use	A Airport IT Dept (Airport Production Environment)	Yes
Conversation data of the person	Production	On each use	A Airport IT Dept (Airport Production Environment)	Yes

[Output]

Output	Output method	Expected accuracy	User judgment	Output of evidence information	Safety risk	Connection with external system
①Choosing a language	Select the language to talk to the user (Do not output externally)	Accuracy rate (75%)	No	No	No	No
②Talk to the user	Outputs the defined voice from the robot to the user (Selects the language recognized in ①).	Accuracy rate (75%)	Yes	No	No	No
③Conversations	Outputting the conversation to the user (Select the language recognized in ①)	Accuracy rate (75%)	Yes (Conversation: Speak again)	Display Conversation	No	No
④Choosing where to guide	Select the place to guide the user	Accuracy rate (75%)	Yes (Conversation: Speak again)	Display Destination	No	No
⑤Start guide	Move toward a destination *When the robot arrives at the destination, it automatically informs the user of the end of the guidance and returns to the original arrangement.	Accuracy rate (75%)	Yes (Conversation: Speak again)	Display destination and Guidance mode	Yes (Collision with a person/Progress in a hazardous area)	Yes (Linked with location information during guidance)



Risk Assessment



Risk Assessment

- Examination of significant risk scenarios -

Values & Objectives		Service Requirement			Risk No.	Risk Scenario	
1	Improved user satisfaction (Safe directions)	1-1	Maintaining service levels	<ul style="list-style-type: none"> Accuracy Robustness 	R001	Malfunction	Guide to the wrong destination/cause to run out of control
					R002	Insufficient service	Be late for departure
		1-2	Adopting the changes in environment	<ul style="list-style-type: none"> Robustness IoT 	R003	Impact by noise	Robots cannot make correct decisions due to noise in the airport environment (e.g., degree of congestion)
					R004	Responding to flight changes	Inability to accommodate changes in departure gates and times
		1-3	Ensuring fairness	<ul style="list-style-type: none"> Multilingual support 	R005	Unsupported languages	Inability to respond correctly when service is requested in an unsupported language, dialect, etc
2	Ensuring safety	2-1	Safety assurance	<ul style="list-style-type: none"> Contact prevention 	R006	Harm to a person	Contact with another person and injure someone while guiding
		2-2	Understanding of intrudable areas	<ul style="list-style-type: none"> Setting of prohibited area 	R007	Guidance to dangerous areas	Lead users to a prohibited area
		2-3	Responding to people with disabilities or injuries	<ul style="list-style-type: none"> User state recognition 	R008	Burden on the user's body	To guide a wheelchair user or a disabled or injured user under a burden
3	Reduction of burden on airport staff	3-1	Reducing the burden on employees	–	R009	Increased burden on airport staff	Incorrect operation of robots increases the burden on staff
4	Corporate social responsibility	4-1	Accountability	<ul style="list-style-type: none"> Process description Verifiability 	R010	Responding to trouble	When an external explanation is required due to trouble, the cause and preventive measures cannot be explained
		4-2	Data protection	<ul style="list-style-type: none"> Data management 	R011	Leakage of privacy data	Personal information (image and voice) recognized by the robot is illegally used

Risk Assessment & Control Summary

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

Values & Objectives	Risk No.	Risk Scenario	Uncertainty	Environmental change	Caused by user	RC	Control Summary		
							AI System	AI service provider	User
1 Improved user satisfaction (Safe directions)	R001	Malfunction	○			●	Prediction performance Destination modification Automatic linkage to support	Display destination Remote support	Confirmation and correction of destination
	R002	Insufficient service	○		○	●	Update of time data Speed adjustment Automatic linkage to support	Coordination with airlines Cooperation with users	Final judgment
	R003	Impact by noise	○	○		●	Microphone maintenance Robustness of the model	Model performance verification Relearning	—
	R004	Responding to flight changes	○	○		●	Automatic destination modification Automatic linkage to support	Apply schedule changes Display destination Remote support	Confirmation and correction of destination
	R005	Unsupported languages	○		○	●	Robustness of the model Test unsupported languages	Clarification of supported languages Remote support	Recognition of supported languages Contact the support team
2 Ensuring safety	R006	Harm to a person	○		○	●	Anti-collision function Sound effect while moving Collision alert	User support Investigation and disclosure of the accident Compensation for injured users	Collision precautions
	R007	Guidance to dangerous areas	○	○		●	Update of prohibited areas Direction prediction Linkage of movement states	Prohibited areas control Remote path compensation Emergency stop function	Track correction emergency stop
	R008	Burden on the user's body	○		○	●	*Same as R002	*Same as R002	*Same as R002
3 Reduction of burden on airport staff	R009	Increased burden on airport staff	○				Predictive performance of the model	Understanding the status of remote support Cost control	
4 Corporate social responsibility	R010	Responding to trouble	○				Explainability	Fault response flow	
	R011	Leakage of privacy data					Data protection	Education on professional ethics	



Organization

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

A Airport) Top Management

- Values and objectives
- Approve risk control

A Airport) Airport Management Dept

- Apply schedule changes
- Coordination with airlines

A Airport) Safety Management Dept

- Investigation of the cause of the accident

- AI Service Provider - A Airport) Customer Support Dept

- Cooperation with users
- Remote support of users
- Remote path compensation
- Emergency stop
- Clarification of supported languages
- Display destination
- Apply schedule changes
- prohibited areas control
- Model performance verification
- Microphone maintenance
- Relearning
- Fault response flow
- Investigation and disclosure of the accident
- Compensation for injured users
- Education on professional ethics

X Co) AI Dev Dept

- Predictive performance
- Destination modification
- Actuator performance
- Anti-collision function
- Sound effect while moving
- Collision alert
- Direction prediction
- Display unsupported languages
- Automatic linkage to support

A Airport) IT Dept

- Update time data
- Update prohibited areas
- Store usage logs

- User - Airport Passenger

- Confirmation and correction of destination
- Final judgment
- Collision precautions
- Track correction/Emergency stop
- Recognition of supported languages
- Contacting the Support Team

Other Users

- Collision precautions



Control Coordination



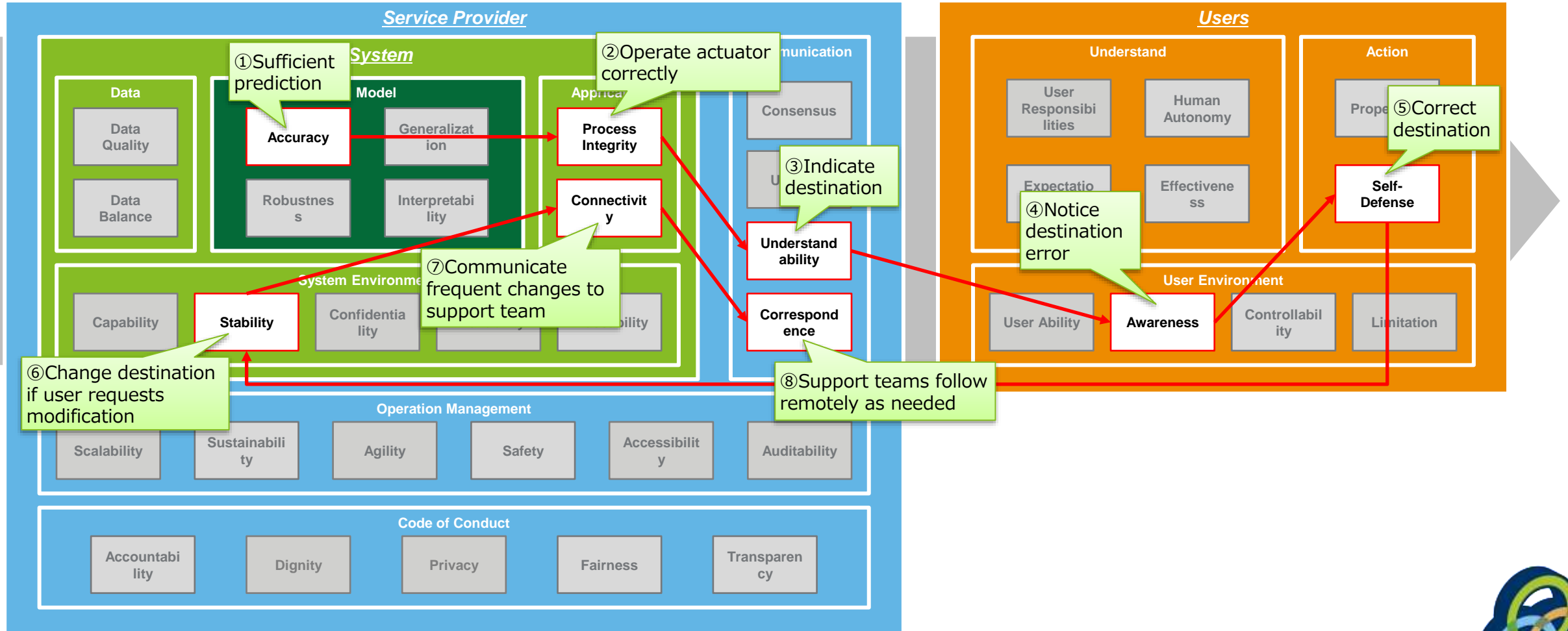
Control Coordination

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

R001

Malfunction

Guide to the wrong destination/cause to run out of control



Risk Control

- Consider risk control according to the risk chain -

R001

Malfunction

Guide to the wrong destination/cause to run out of control

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
<p>①[Accuracy] Develop models with sufficient prediction (AI Dev Dept, Co. X)</p> <p>②[Process Integrity] Operate the actuator correctly (A Airport IT Dept/AI Dev Dept, Co. X)</p> <p>⑥[Stability] Change destination if user requests modification (Robot)</p> <p>⑦[Connectivity] Frequent changes may not be recognized correctly and will be automatically communicated to the support team (Robot)</p>	<p>③[Understandability] Display the destination on the robot to the user (A Airport Customer Support Dept)</p> <p>⑧[Correspondence] Support teams follow remotely as needed (A Airport Customer Support Dept)</p>	<p>④[Awareness] Notice a robot's destination error (Airport Passenger)</p> <p>⑤[Self-Defense] Correct destination (Airport Passenger)</p>

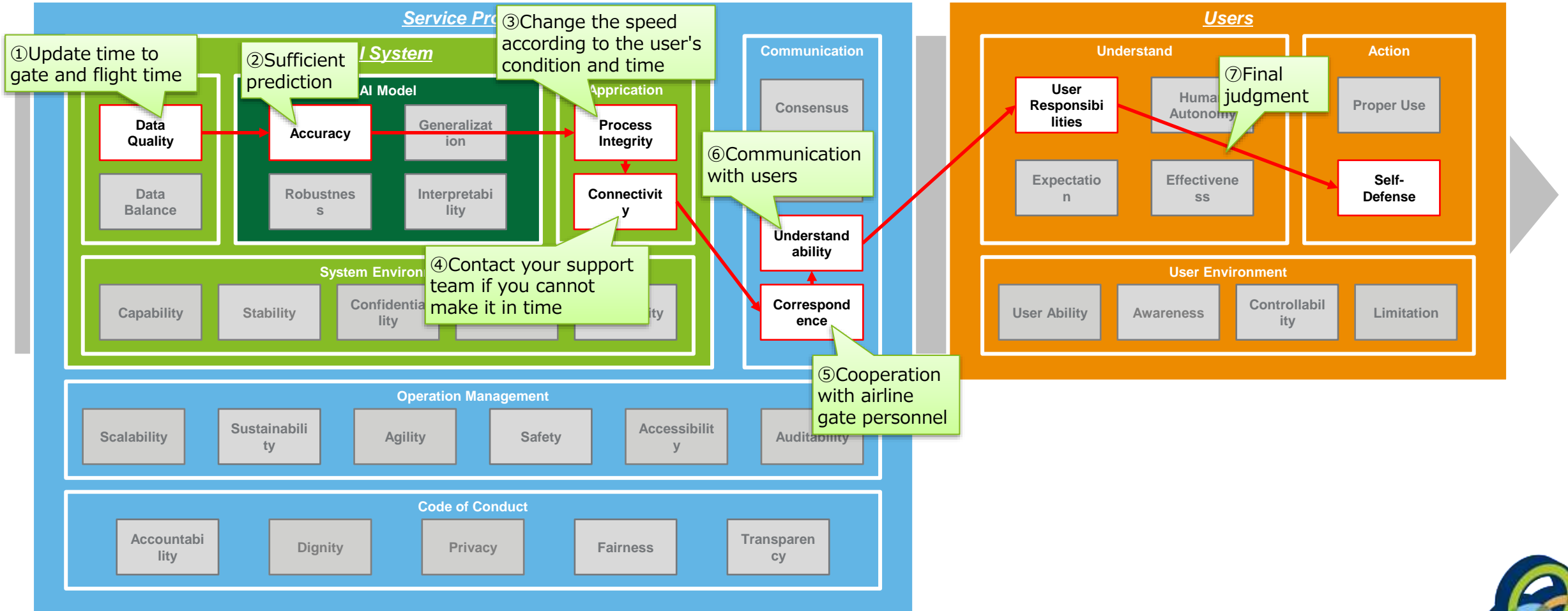


Control Coordination

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

R002

Insufficient service
Be late for departure



Risk Control

- Consider risk control according to the risk chain -

R002

Insufficient service
Be late for departure

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
<p>①[Data Quality] Update the required time to the gate and flight time information (A Airport IT Dept)</p> <p>②[Accuracy] Develop models with sufficient prediction (AI Dev Dept, Co. X)</p> <p>③[Process Integrity] Change the speed according to the user's condition and time (A Airport IT Dept/AI Dev Dept, Co. X)</p> <p>④[Connectivity] Automatically communicate to support team if not in time (Robot)</p>	<p>⑤[Correspondence] Check with the gate staff of the airline to see if they can respond (A Airport Customer Support Dept)</p> <p>⑥[Understandability] (If you can't make it in time for boarding) Contact the user and suggest a flight change, etc (A Airport Customer Support Dept/Airline)</p>	<p>⑦[User Responsibility/Self-Defense] Make final decisions such as processing changed flights (Airport Passenger)</p>



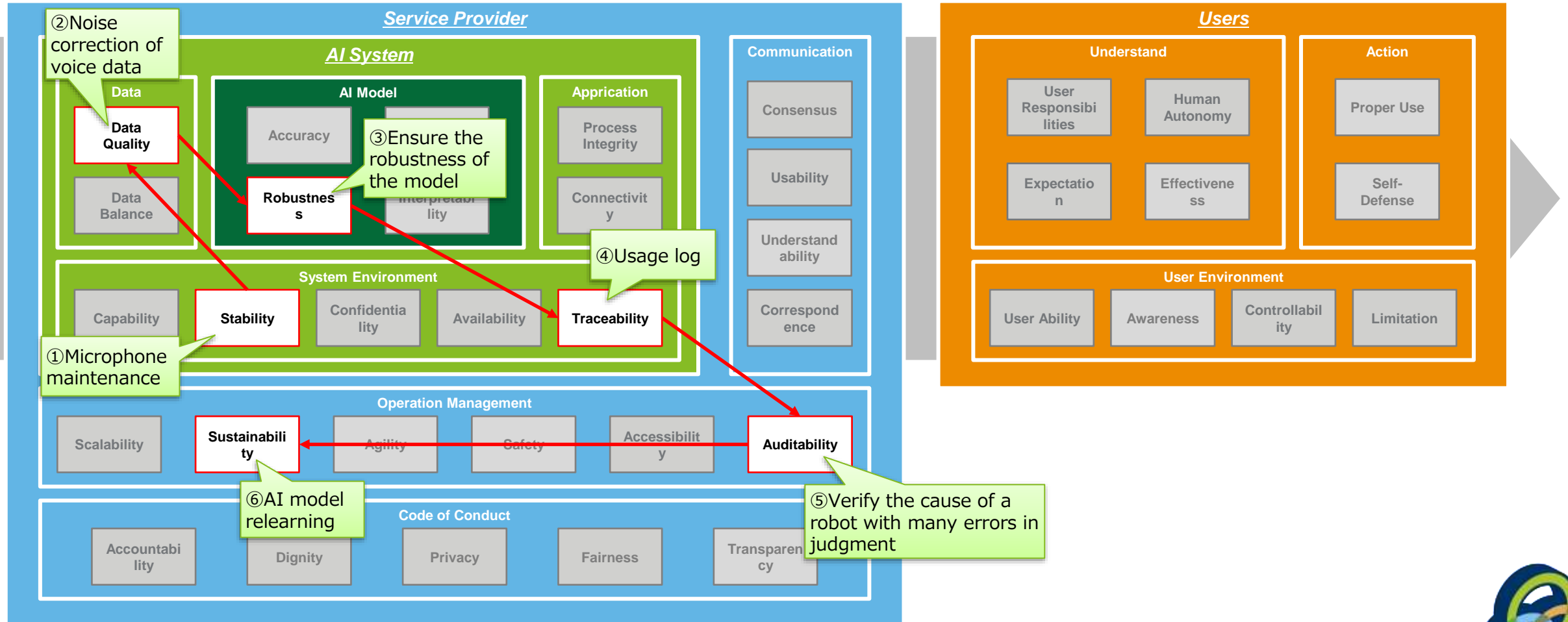
Control Coordination

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

R003

Impact by noise

Robots cannot make correct decisions due to noise in the airport environment (degree of congestion, etc)



Risk Control

- Consider risk control according to the risk chain -

R003

Impact by noise

Robots cannot make correct decisions due to noise in the airport environment (degree of congestion, etc)

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
①[Stability] Properly maintain sound collection microphones (A Airport IT Dept/AI Dev Dept, Co. X)	⑤ [Auditability] Verify the cause of a robot with many errors in judgment (A Airport Customer Support Dept)	
②[Data Quality] Degradation of voice data by noise correction, etc (A Airport IT Dept/AI Dev Dept, Co. X)	⑥[Sustainability] Relearn AI models of robots with many errors in judgment (A Airport Customer Support Dept/A Airport IT Dept/AI Dev Dept, Co. X)	
③[Robustness] Learning to enhance the robustness of the model (A Airport IT Dept)		
④[Traceability] Store AI judgment and user interaction history (A Airport IT Dept)		



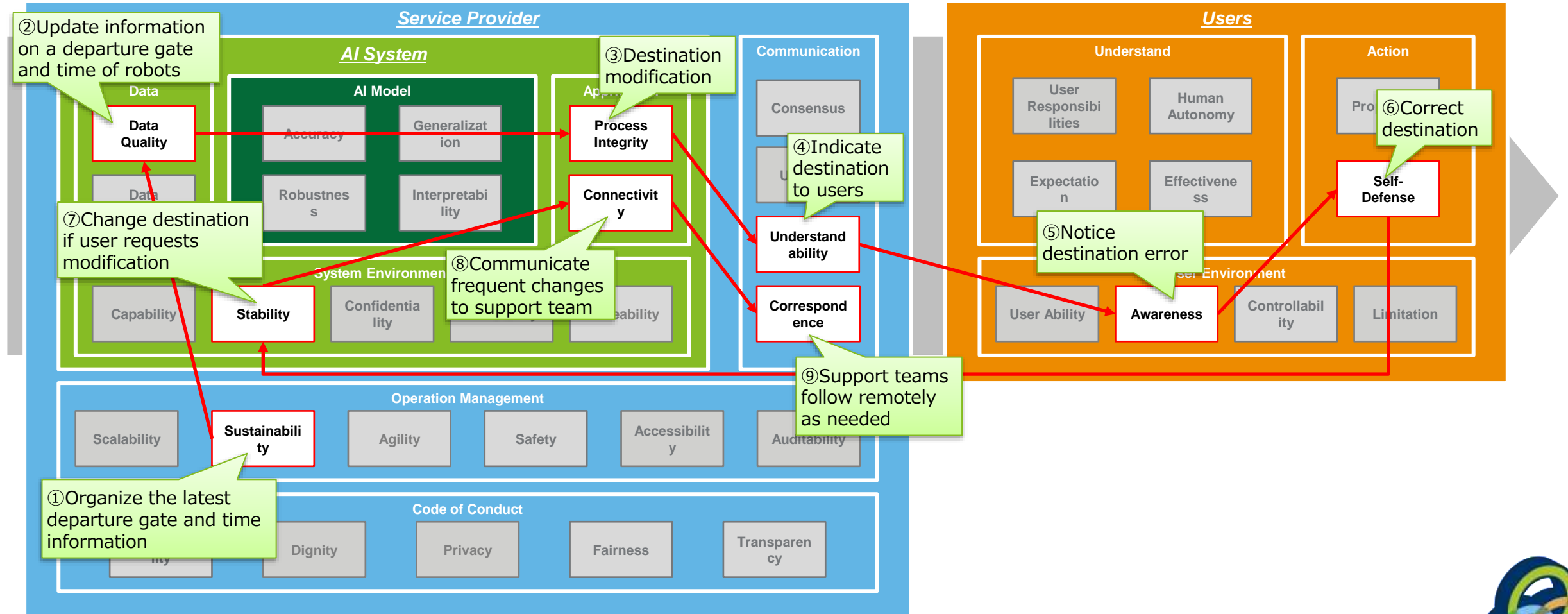
Control Coordination

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

R004

Responding to Flight Changes

Inability to accommodate changes in departure gates and times



Risk Control

- Consider risk control according to the risk chain -

R004

Responding to Flight Changes

Inability to accommodate changes in departure gates and times

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
<p>②[Data Quality] Update information on a departure gate and a departure time of an unmanned robot (A Airport IT Dept)</p> <p>③[Process Integrity] Automatically modifying destination (Robot)</p> <p>⑦[Stability] Change destination if user requests modification (Robot)</p> <p>⑧[Connectivity] Frequent changes may not be recognized correctly and will be automatically communicated to the support team (robot)</p>	<p>①[Sustainability] Organize the latest departure gate and departure time information (A Airport Customer Support Dept)</p> <p>④[Understandability] Display the destination on the robot to the user (A Airport Customer Support Dept)</p> <p>⑨[Correspondence] Support teams follow remotely as needed (A Airport Customer Support Dept)</p>	<p>⑤[Awareness] Notice a robot's destination error (Airport Passenger)</p> <p>⑥[Self-Defense] Correct destination (Airport Passenger)</p>



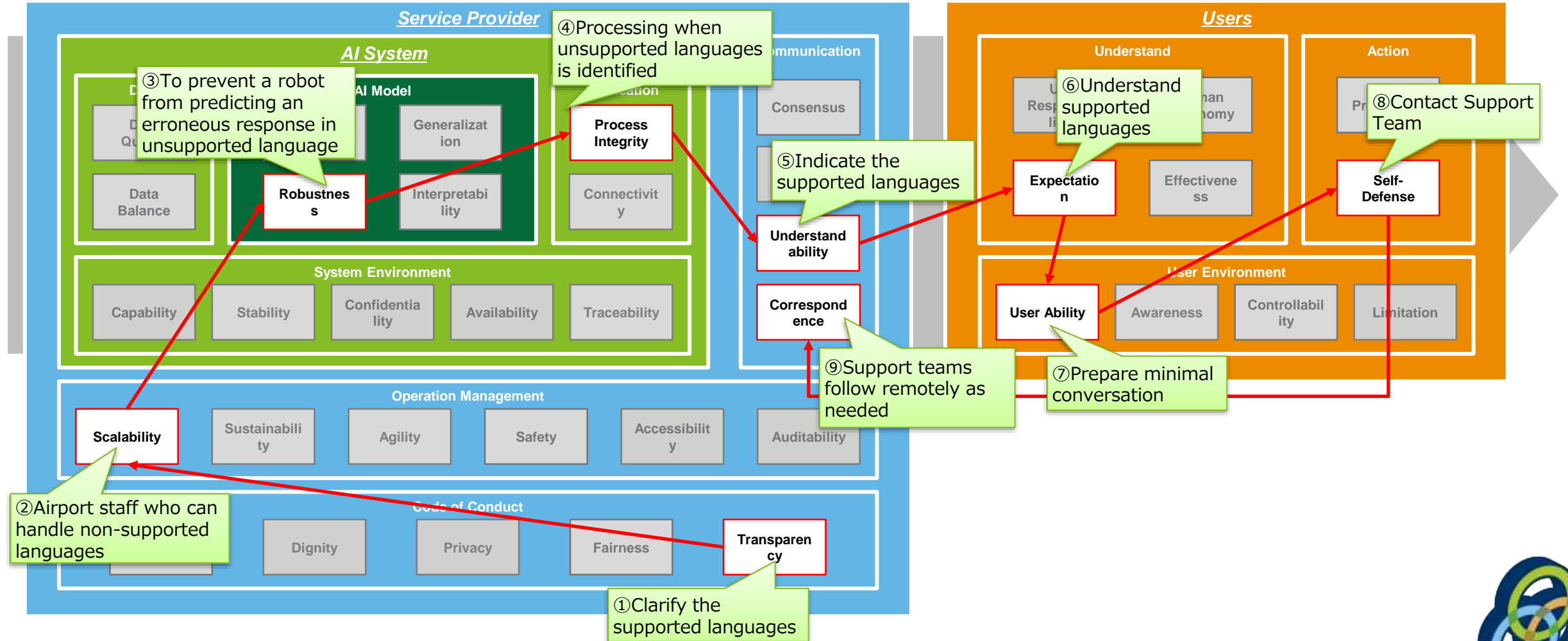
Control Coordination

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

R005

Unsupported languages

Inability to respond correctly when service is requested in an unsupported language, dialect, etc



Risk Control

- Consider risk control according to the risk chain -

R005

Unsupported languages

Inability to respond correctly when service is requested in an unsupported language, dialect, etc

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
<p>③[Robustness] To prevent the robot from predicting an erroneous response in another language to an unsupported language (A Airport IT Dept/AI Dev Dept, Co. X)</p> <p>④[Process Integrity] When the robot identifies an unsupported language, it displays to the user that the language is not supported (Robot)</p>	<p>①[Transparency] Clearly organize the supported languages by robots (A Airport Customer Support Dept)</p> <p>②[Scalability] Provide staff who can support various languages (A Airport Customer Support Dept)</p> <p>⑤[Understandability] Indicate the supported languages for users (A Airport Customer Support Dept)</p> <p>⑨[Correspondence] Support teams follow remotely as needed (A Airport Customer Support Dept)</p>	<p>⑥[Expectation] Understand supported languages (Airport Passenger)</p> <p>⑦[User Ability] Helps users find the minimum conversation they need to ask for direction (Airport Passenger)</p> <p>⑧[Self-Defense] Contact airport staff for help if you don't know your departure gate (Airport Passenger)</p>



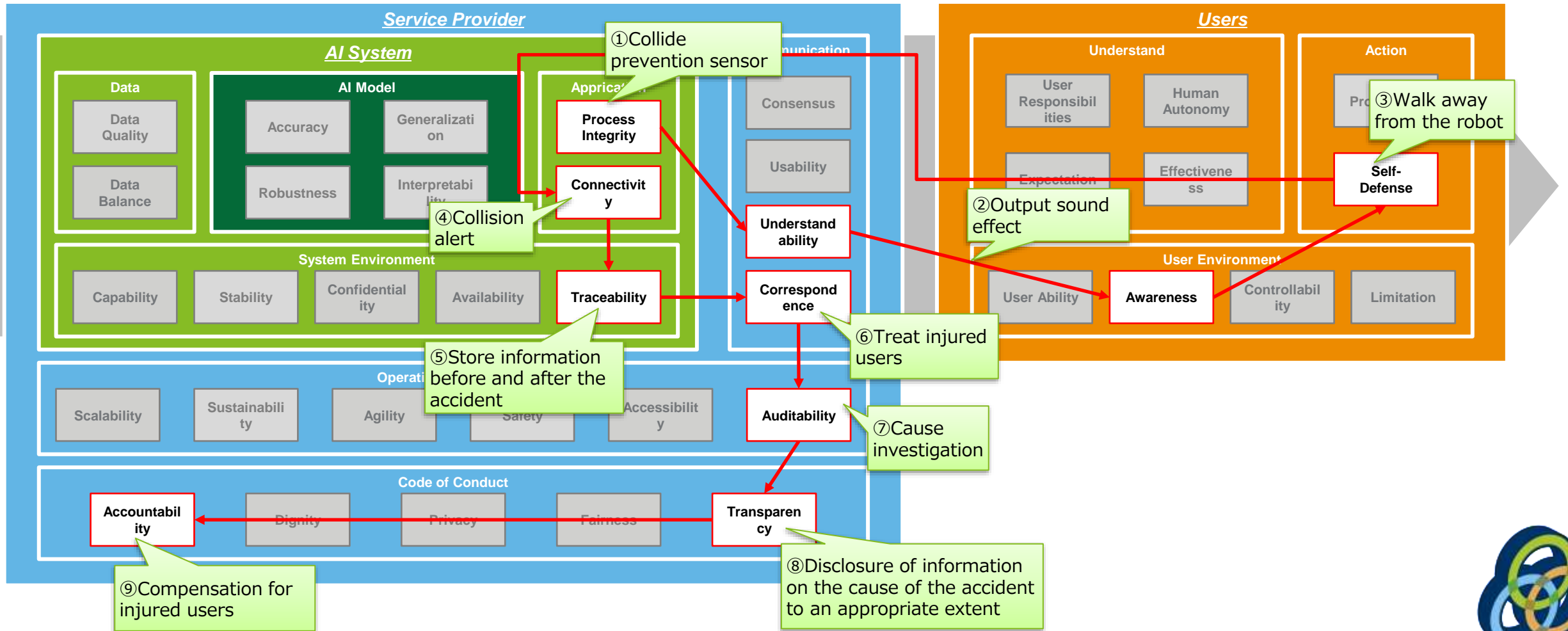
Control Coordination

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

R006

Harm to a person

Collide with another person and injure someone while guiding



Risk Control

- Consider risk control according to the risk chain -

R006

Harm to a person

Collide with another person and injure someone while guiding

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
<p>①[Process Integrity] To prevent collisions with people or objects, the robot stops suddenly when the sensor reacts (A Airport IT Dept)</p> <p>④[Connectivity] When a robot collide with person or object, it sends an alert including the location of the accident to the management department (A Airport IT Dept)</p> <p>⑤[Traceability] Store information before and after the accident (A Airport IT Dept)</p>	<p>②[Understandability] Outputs sound effects when moving so that the robot can be recognized (A Airport Customer Support Dept)</p> <p>⑥[Correspondence] Treat injured users (A Airport Customer Support Dept)</p> <p>⑦[Auditability] Verifying the cause of the accident (A Airport Safety Management Dept)</p> <p>⑧[Transparency] Disclose information on the cause of the accident to an appropriate extent (A Airport Top Management/Airport Management Dept)</p> <p>⑨[Accountability] Compensation for injured users (A Airport Top Management/Airport Management Dept)</p>	<p>②[Awareness] Outputs sound effects when moving so that the robot can be recognized (Airport Passenger)</p> <p>③[Self-Defense] Walk away from the robot (Airport Passenger)</p>



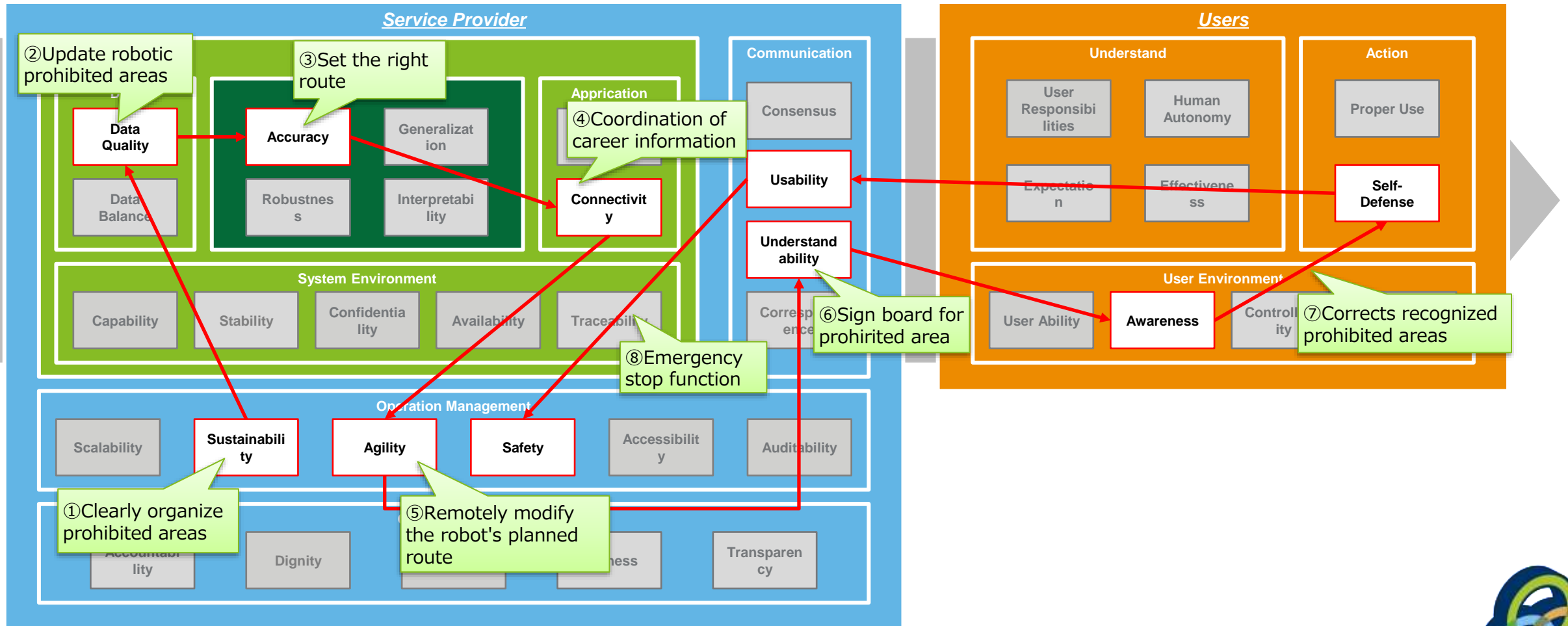
Control Coordination

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

R007

Guidance to dangerous areas

Lead users to a prohibited area



Risk Control

- Consider risk control according to the risk chain -

R007

Guidance to dangerous areas

Lead users to a prohibited area

Risk Control		
AI system (A Airport IT Dept/AI Dev Dept, Co. X)	Service Provider (A Airport Customer Support Dept)	User (Airport Passenger)
<p>②[Data Quality] Properly set prohibited areas for robots (A Airport Information Systems Department)</p> <p>③[Accuracy] Set the route so that the robot does not go to prohibited areas (Robot)</p> <p>④[Connectivity] Link robot route information to support team (Robot)</p>	<p>①[Sustainability] Clearly organize prohibited areas (A Airport Customer Support Dept)</p> <p>⑤[Agility] Remotely modify the robot's planned route (A Airport Customer Support Dept)</p> <p>⑥[Understandability] Set up signboards, etc., to make it easier to understand prohibited areas (A Airport Customer Support Dept)</p> <p>⑧[Usability/Safety] Emergency stop or modification is performed when the robot is moving into the prohibited area (A Airport Customer Support Dept)</p>	<p>⑦[Awareness/Self-Defense] When the user notices that the robot is going to the restricted area, the user corrects it (Airport Passenger)</p>

