Risk Assessment & Control Coordination for AI services :
Case07 Loan Examination 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 -

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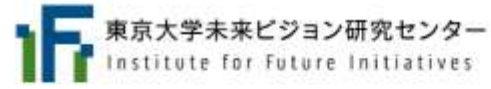
東京大学未来ビジョン研究センター
Institute for Future Initiatives

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

Case07 : Loan Examination AI

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

An AI service that offers estimated loan amounts prior to loan applications as reference info to consumers (here after "applicants") planning to apply for a loan at Bank A. The applicant enters survey the information and desired loan amount on the web screen, and obtains a score, interest rate and borrowing limit output by the AI service.

After applying the person in charge of Bank A's loan department makes a detailed loan amount estimate, and proceeds with the examination procedure with the applicant's consent. The person in charge of Bank A's loan department provides a detailed loan estimate by referring to the score, interest rate and borrowing limit output by the AI service.

The Bank A IT service department, the AI service provider, creates a learning AI model to predict estimated loan amounts with machine learning (a regression model) based on learning data collected from SNS, external credit institutions, Bank A's loan department, etc.

[Values & Objectives]

- **Sales Opportunity Expansion**

- **Assessment Quality Maintenance and Optimization**

- **Social responsibility**

Set the expected value to within 2% for bad debt rates for applicants given loans based on estimated loan amounts calculated by an AI model.
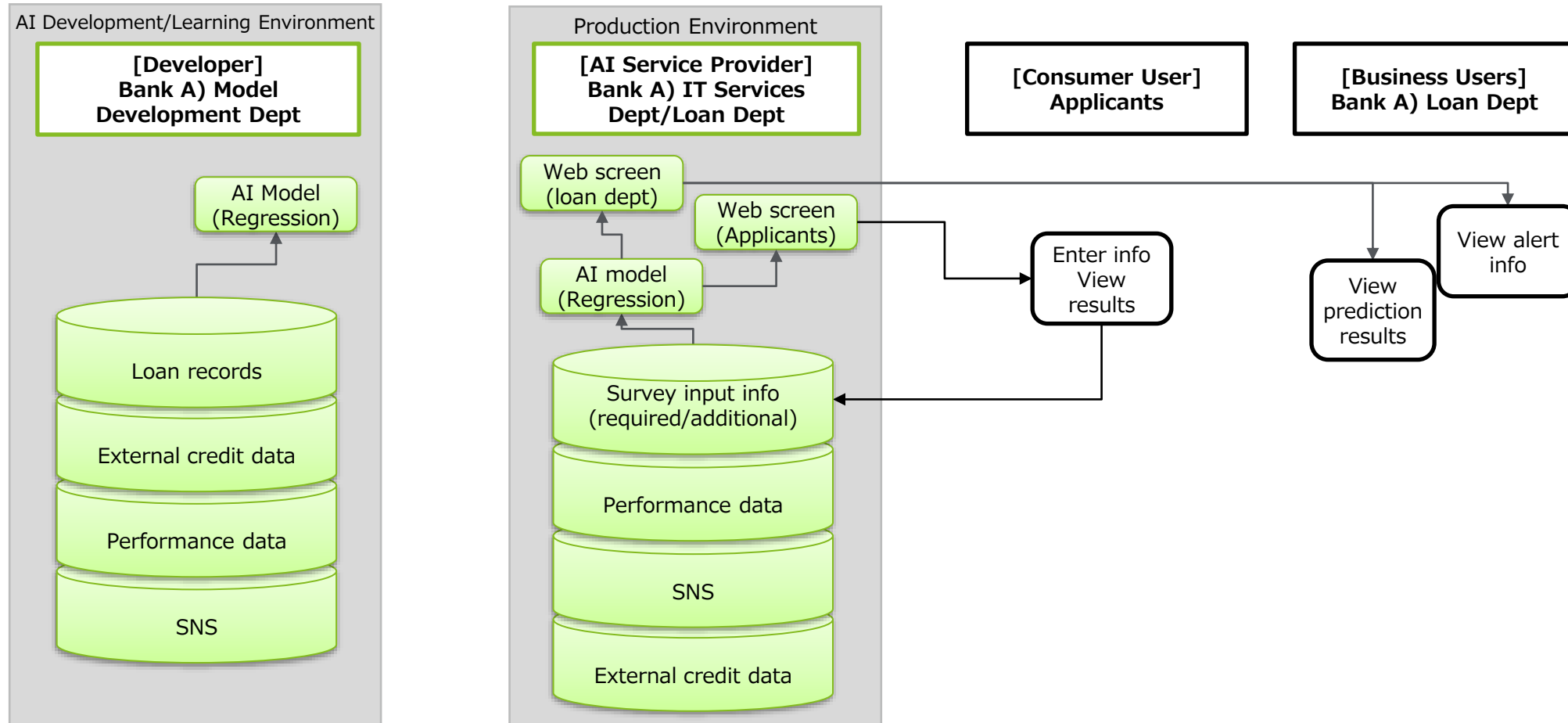
✓ The learning model applies additional learning on a periodic basis including performance data.

✓ In the event that the same applicant added additional info, predictions are executed again for score, interest rate and borrowing limit. In this case, past prediction results are recorded, which the applicant can view on the internet.

✓ When prediction results via AI fall before the applicant's desired loan amount, this is output by Bank A's loan department as alert infomation. In this case, the alert info includes a feature value contributing greatly to judgment results.

Case07 : Loan Examination AI
- System Overview -

AI System	Bank A) Model Development Department	Predict scoring and limit amounts for loan applications
AI Service Provider	Bank A) IT Services Department/Loan Department	Execute final decision for loan examination (may overturn primary AI decision)
Users	Applicant, Bank A) Loan Department Leader	Apply for loan examination

AI Development/Learning Environment

[Developer]
Bank A) Model Development Dept

AI Model (Regression)

Loan records

External credit data

Performance data

SNS

Production Environment

[AI Service Provider]
Bank A) IT Services Dept/Loan Dept

Web screen (loan dept)

Web screen (Applicants)

AI model (Regression)

Survey input info (required/additional)

Performance data

SNS

External credit data

[Consumer User]
Applicants

[Business Users]
Bank A) Loan Dept

Enter info View results

View prediction results

View alert info

Case07 : Loan Examination AI
- Input & Output -

[Input Data]

Data	Purpose	Collection Method	Data Manager	Including Privacy Data
Loan records	Learning	Anonymize and collect performance info of Bank A's loan department	Development environment of IT service department of Bank A	Yes (anonymized)
Credit data from external credit bureaus	Learning	Collect from external credit institution	Development environment of IT service department of Bank A	Yes (anonymized)
SNS	Learning	Anonymize and collect SNS info	Development environment of IT service department of Bank A	Yes (anonymized)
Survey input info (required items / additional items)	Production	Applicant enters from web service*	Public cloud environment contracted by Bank A	Yes
Information related to Bank A applicant	Production	When applicant's approved: Link with Bank A performance info	Public cloud environment contracted by Bank A	Yes (with permission)
SNS	Production	When applicant gives permission: Collect SNS info	Public cloud environment contracted by Bank A	Yes (with permission)
Credit data from external credit bureaus	Production	Collect from external credit institution	Public cloud environment contracted by Bank A	Yes

*Info entered by applicant
Required items: Name, age, gender, address, annual income, family structure, occupation, existence of other loans
Additional items: Assets, work details, educational background, lifestyle, etc.

*Learning data frame and actual data frame are designed to be identical

Case07 : Loan Examination AI
- Input & Output -

[Output]

| Users	Applicants	Bank A loan department
Output	Prediction of score, interest rate and borrowing limit	Alert information: Applicant info when borrowing limit falls below the applicant's desired loan amount
Output Method	Predicted results of score, interest rate and borrowing limit are output on the web service.You can browse history info.	Lists applicant info when borrowing limit falls below the applicant's desired loan amount, and sends an email to the loan department personnel on a daily basis as alert information.
Expected Accuracy	Bad debt: within 2%	Bad debt: within 2%
User judgment	Yes	Yes
Output of evidence information	No	Yes (feature amount contributed to prediction results)
Safety Risk	No	No
Connection with external system	No	No
Protocol	No	No

Risk Assessment

Risk Assessment
- Examination of significant risk scenarios -

Values & Objectives		Service Requirement			Risk No.	Risk Scenario	
1	Sales Opportunity Expansion	1-1	Ensuring prediction performance	■ Goal setting ■ AI prediction accuracy	R001	Customer detachment due to excessive assessment	Excessive examination results in significant reduction of potential loan customers
		1-2	Real-time Diagnostics	■ Performance	R002	Insufficient processing speed	Judgment processing speed via AI is too slow, sending customers to other companies
		1-3	Protection from External Attack	■ AI robustness ■ Security	R003	Security protection	External attacks trigger changes unintended by loan performance learning data or learning models, causing mistaken learning AI.
2	Assessment Quality Maintenance and Optimization	2-1	Assessment Quality Maintenance	■ AI re-learning ■ Teacher label ■ Explainability ■ Fraud detection	R004	Loan strategy changes	Examination judgment by AI cannot handle changes to loan strategy in banks
					R005	Mistaken teacher label settings	Insolvent customers not set as teacher labels, and the same bad debt occurs frequently
					R006	Excessive AI dependence	Mistaken judgments are overlooked, caused by a lack of distrust in AI judgment, turning examination into an empty formality
					R007	False application	AI judgment results are mistaken due to an applicant inputting factually incorrect personal information, or linking the SNS of another individual
		2-2	Contingency Plan		R008	Alternative operation for abnormalities	When abnormalities occur due to AI performance degradation, etc., no switch to manual examinations occurs, hindering operations
3	Social Responsibility	3-1	Accountability	■ Explainability ■ Verifiability	R009	Explanation to applicants	Cannot give a rational explanation for examination results and loan amounts to applicants
		3-2	Observance of Ethics/Compliance	■ AI Generalization	R010	Fairness	Produce clearly unfair prediction results for specific races/sexes
		3-3	Information Management	■ Data management	R011	Score outflow, out-of-purpose use	Using/offering scoring results outside of use purpose causes disadvantages for specific clients
					R012	Privacy protection	Violation of the Personal Information Protection Act occurs through mishandling of personal information
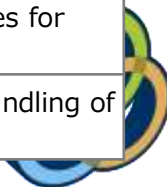
Risk Assessment&Control Summary
- Organize the roles of each stakeholder based on the examination of each risk chain -

Values & Objectives		Risk No.	Risk Scenario	Uncerta inly	Environ mental change	Caused by user	RC	Control Summary		
								AI System	AI Service Provider	User
1	Sales Opportunity Expansion	R001	Customer detachment due to excessive assessment	○			●	Classification of target data Ensure accuracy rate Judgment basis output	Validate model Re-learning	Re-learning Assessment
		R002	Insufficient processing speed	○			●	Performance assurance Alert output	Performance monitoring System environment improvement	
		R003	Security protection					Security management	Cause investigation/improvement	
2	Assessment Quality Maintenance and Optimization	R004	Loan strategy changes	○	○	○	●	*Same as R001	*Same as R001	*Same as R001
		R005	Mistaken teacher label settings	○		○	●	Automatic cooperation of external data Learning data verification	Easy-to-use UI Judgment accuracy/anomaly verification Re-learning	Accurate feedback
		R006	Excessive AI dependence	○		○	●	Judgment basis output	Examination level definition Display of decision basis	Understanding of judgment responsibility Strategic customer understanding
		R007	False application	○		○	●	Data contamination protection Judgment basis output Illegal input detection	Application content verification Assessment process restriction	Automatic control for illegal input
		R008	Alternative operation for abnormalities		○	○	●	Alert output Switch to alternative service Abnormality record	Arrangement for abnormality Cause verification System improvement	Arrangement for abnormality Manual procedure preparation
3	Social Responsibility	R009	Explanation to applicants	○		○	●	Judgment basis output Negative object processing	Easy-to-use UI Easy-to-understand display Verification/customer support	Appropriate input Petition for judgment result
		R010	Fairness	○			●	Data bias Model generalization Negative object processing	Fairness policy assessment Negative judgment disclosure	AI judgment tendency understanding Final judgment
		R011	Score outflow, out-of-purpose use					Data protection	Professional ethics education	Professional ethics education
		R012	Privacy protection					Data protection	Compliance education	Compliance education Data handling

Organization

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

- Responsible Persons -
Bank A) Management

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

Financial Services Agency

- Financial institution supervision

Bank A) Legal/Compliance

- Professional ethics education
- Compliance education

Bank A: Internal Audit

- Internal audit (business audit)

- AI Service Provider -
Bank A) IT Services Dept/ Loan Dept

- Fairness policy assessment
- Negative judgment disclosure
- Assessment process restriction
- Performance monitoring
- Validate model
- Application content verification
- Judgment accuracy/anomaly verification
- Re-learning
- Improve system environment
- Cause investigation/improvement
- Examination level definition
- Arrangement for abnormality
- Easy-to-use UI
- Display of decision basis
- Easy-to-understand display
- Customer support

Bank A) Model Development Dept

- Ensure a sufficient accuracy rate
- Model generalization
- Judgment basis output
- Classification of target data
- Learning data verification
- Data bias adjustments

Bank A) IT Services Dept

- Performance assurance
- Abnormality record
- Security management
- Switch to alternative service
- Automatic data linkage
- Alert output
- Illegal input detection

- Users -
Bank A: Loan Department Leader

- Re-learning Assessment
- Accurate feedback
- Understanding responsibility
- Strategic customer understanding
- Understanding AI judgment tendency
- Arrangement for abnormality
- Manual procedure preparation

- Users -
Applicants

- Appropriate input
- Petition for judgment result
- Automatic control for illegal input
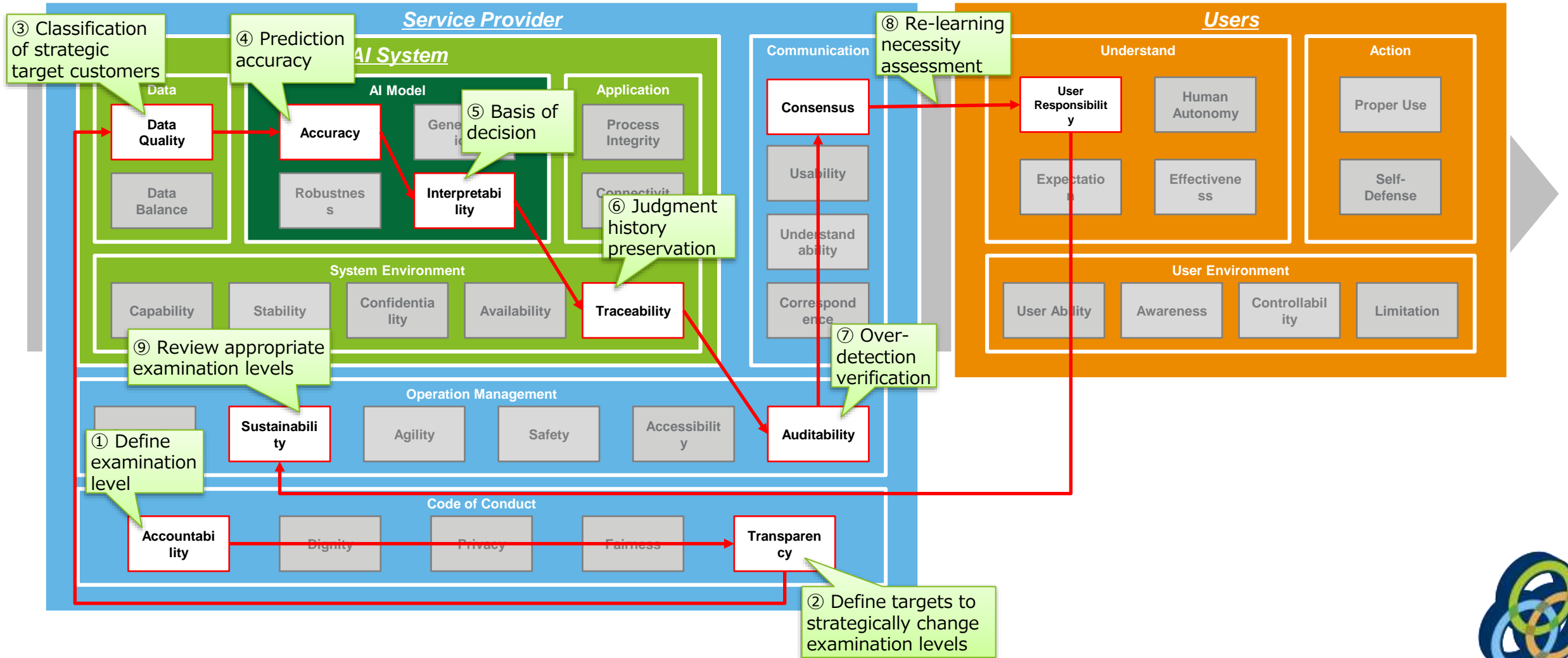
Control Coordination

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

R001	**Customer detachment due to excessive assessment** Excessive examination results in significant reduction of potential loan customers

Service Provider

AI System

③ Classification of strategic target customers

④ Prediction accuracy

⑤ Basis of decision

Data
- Data Quality
- Data Balance

AI Model
- Generic
- Accuracy
- Interpretability
- Robustness

Application
- Process Integrity
- Connectivity

⑥ Judgment history preservation

System Environment
- Capability
- Stability
- Confidentiality
- Availability
- Traceability

⑨ Review appropriate examination levels

① Define examination level

Operation Management
- Sustainability
- Agility
- Safety
- Accessibility
- Auditability

⑦ Over-detection verification

Code of Conduct
- Accountability
- Dignity
- Privacy
- Fairness
- Transparency

② Define targets to strategically change examination levels

Communication
- Consensus
- Usability
- Understandability
- Correspondence

⑧ Re-learning necessity assessment

Users

Understand
- User Responsibility
- Human Autonomy
- Expectation
- Effectiveness

Action
- Proper Use
- Self-Defense

User Environment
- User Ability
- Awareness
- Controllability
- Limitation

Risk Control
- Consider risk control according to the risk chain -

R001	**Customer detachment due to excessive assessment** Excessive examination results in significant reduction of potential loan customers

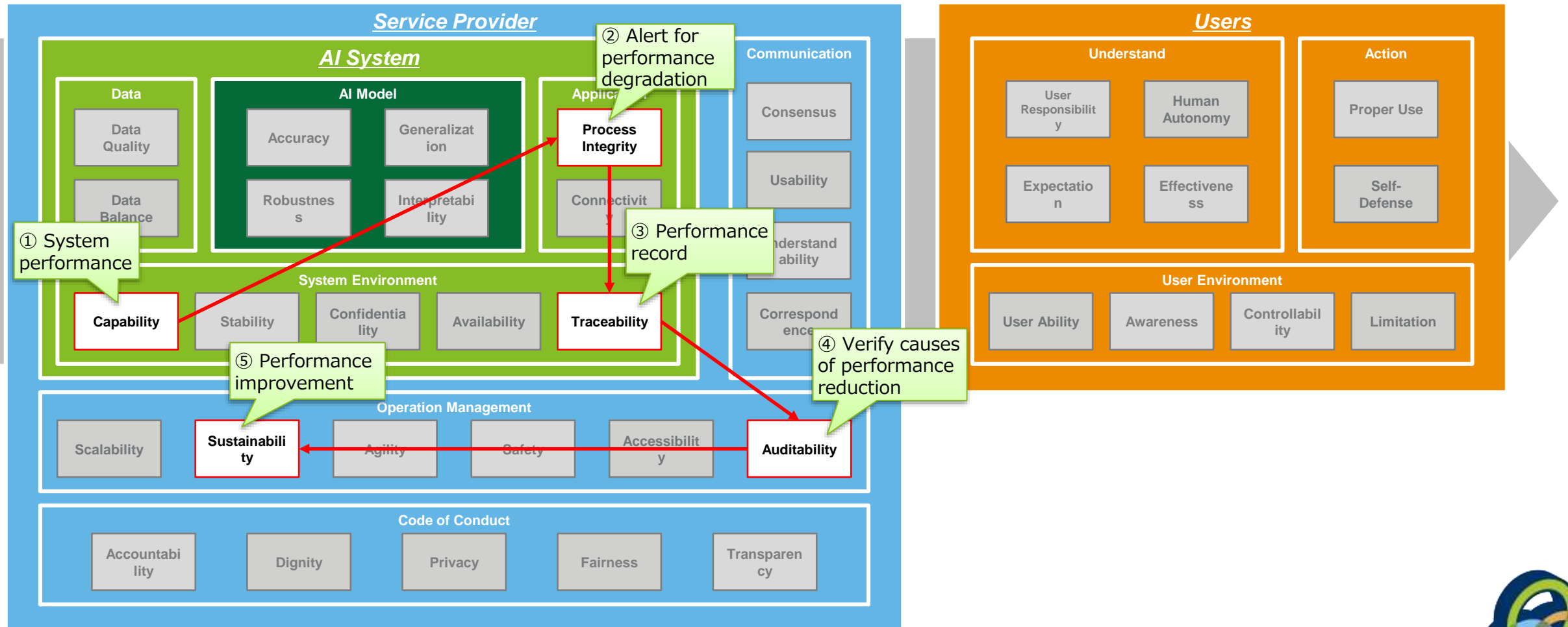| Risk Control		
AI system **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
③ [Data Quality] Clearly divide data of strategic target customers (Bank A model development department) ④ [Accuracy] Implement learning to ensure model prediction accuracy (Bank A model development department) ⑤ [Interpretability] Output model judgment basis (Bank A model development department) ⑥ [Traceability] Store AI judgment history (Bank A IT services department)	① [Accountability] Define examination level to achieve by AI (Bank A loan department) ② [Transparency] Clarify target to strategically alter examination level (Bank A loan department) ⑦ [Auditability] Periodically verify model function, and verify if excessive denial takes place (Bank A IT services department/loan department) ⑧ [Consensus] Assess whether the state of judgment accuracy is explained and relearned (Bank A IT services department/loan department) ⑨ [Sustainability] Review appropriate examination levels (model function) based on maintenance costs (Bank A IT services department/loan department)	⑧ [User Responsibility] Assess whether the state of judgment accuracy is explained and relearned (Bank A IT services department/loan department)

Risk Control

- Consider risk control according to the risk chain -

R002	**Insufficient processing speed** Judgment processing speed via AI is too slow, sending customers to other companies

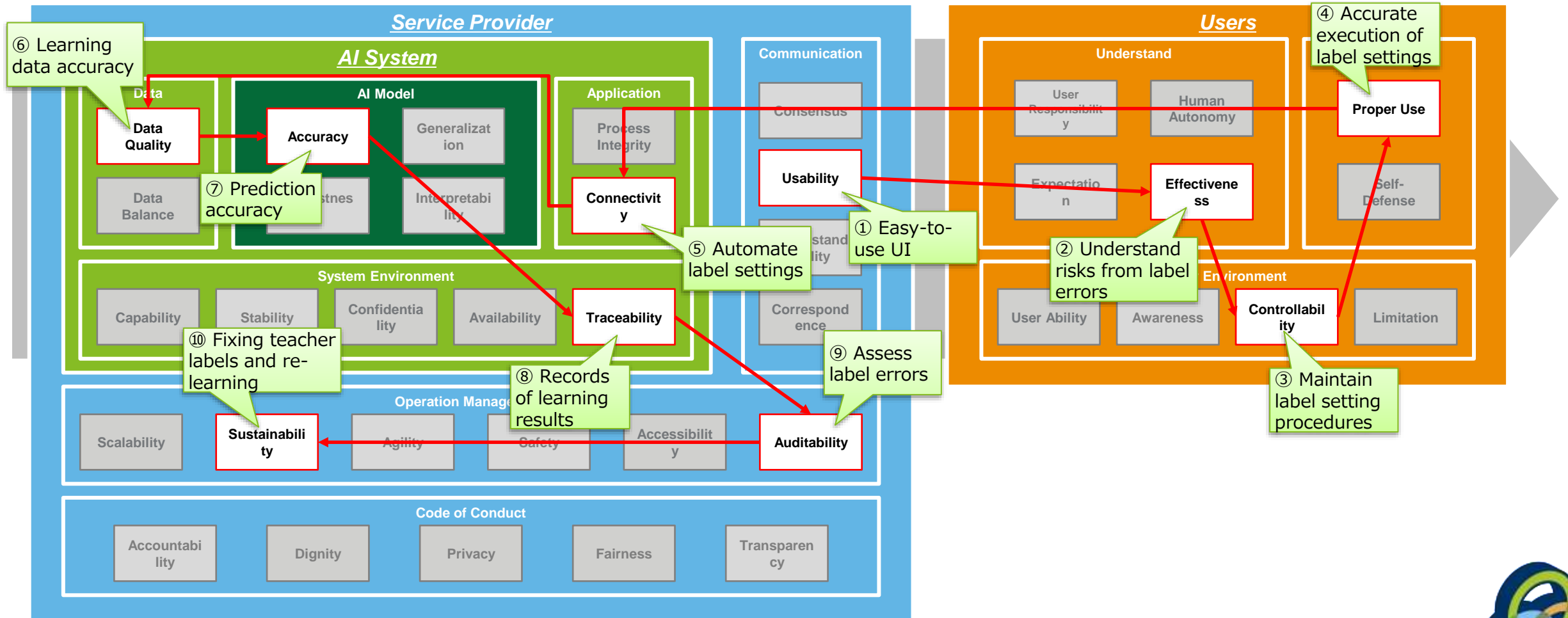| Risk Control		
AI System **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
① [Capability] Assure suitable processing speed with the entire AI system (Bank A model development department) ② [Process Integrity] Output alerts when prediction performance degradation is observed (Bank A model development department) ③ [Traceability] Record performance burden situation (Bank A IT services department)	④ [Auditability] Assess causes of performance burden (Bank A IT services department) ⑤ [Sustainability] Improve AI system processing function to assure adequate performance (Bank A IT services department)	

Risk Control
- Consider risk control according to the risk chain -

R005	**Mistaken teacher label settings** Insolvent customers not set as teacher labels, and the same bad debt occurs frequently

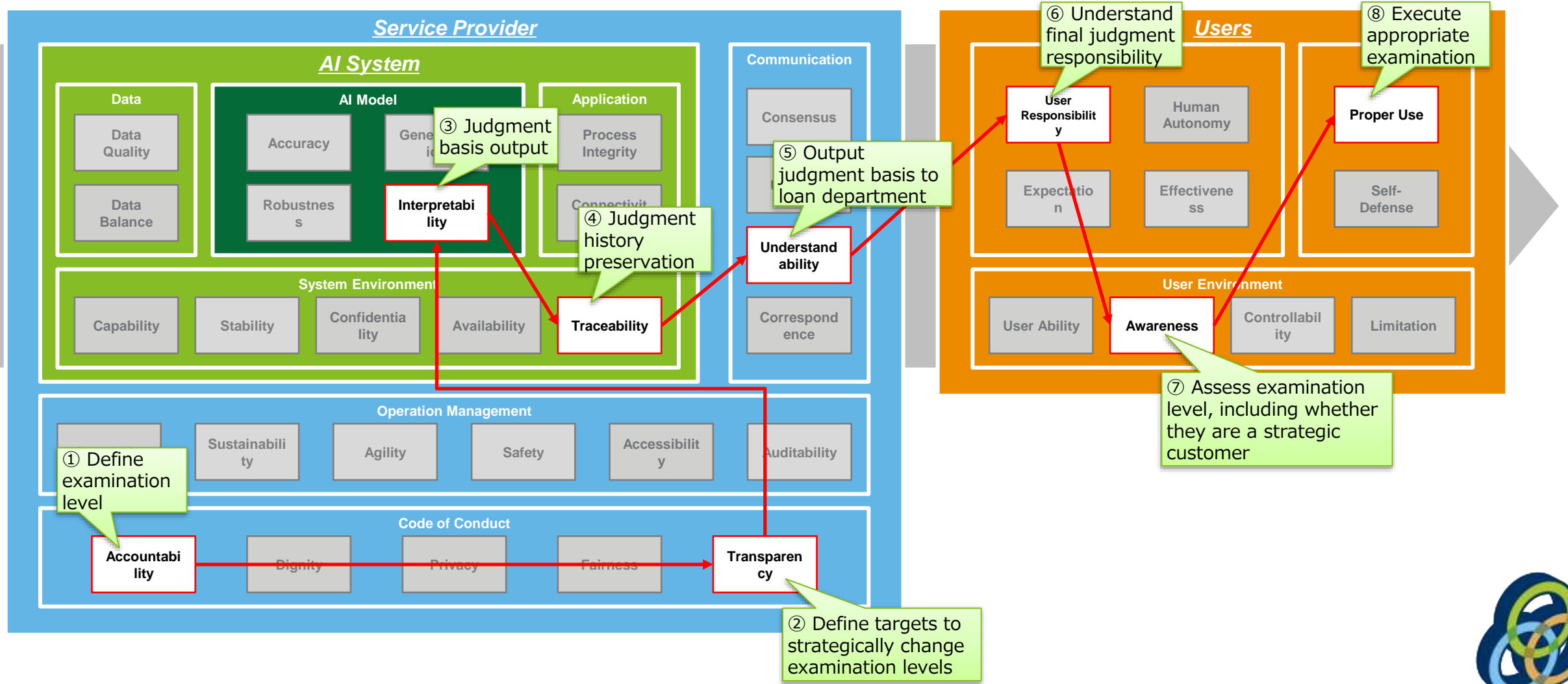| Risk Control		
AI System **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
⑤ [Connectivity] Automatically link teacher label settings for customers from internal systems (Bank A model development department) ⑥ [Data Quality] Assess learning data accuracy (Bank A model development department) ⑦ [Accuracy] Implement learning to ensure model prediction accuracy (Bank A model development department) ⑧ [Traceability] Preserve prediction results at each learning level and anomaly content (Bank A IT services department)	① [Usability] Prepare an easy-to-use UI that makes operation mistakes difficult (Bank A IT services department) ⑨ [Auditability] Assess prediction performance changes and anomalies in learning data (data thought to be label errors) (Bank A, IT services department) ⑩ [Sustainability] Correct data teacher labels and relearn AI models as necessary (Bank A IT services department)	② [Effectiveness] Recognize effect of feedback errors on AI model performance degradation (Loan department Bent personnel) ③ [Controllability] Set up procedures to provide accurate feedback (Loan department personnel) ④ [Proper Use] Execute accurate feedback (Loan department personnel)

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

R006

Excessive AI dependence
Mistaken judgments are overlooked, caused by a lack of distrust in AI judgment, turning examination into an empty formality

Risk Control
- Consider risk control according to the risk chain -

R006	**Excessive AI dependence** Mistaken judgments are overlooked, caused by a lack of distrust in AI judgment, turning examination into an empty formality

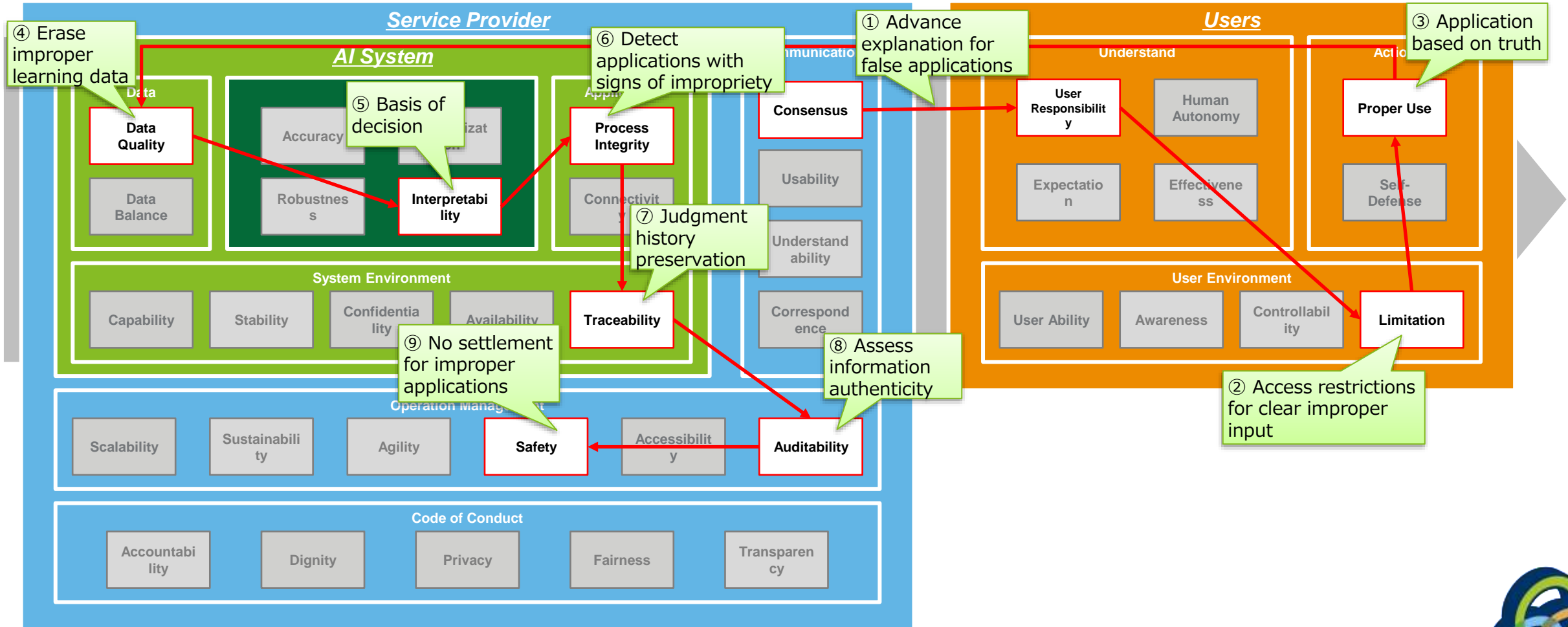| Risk Control		
AI System (Bank A: Model Development Department)	**Service Provider** (Bank A: IT Services Department/Loan Department)	**Users** (Applicant/Loan Department Leader)
③ [Interpretability] Output model judgment basis (Bank A model development department) ④ [Traceability] Store AI judgment result info during use (Bank A IT services department)	① [Accountability] Define examination level to achieve by AI (Bank A loan department) ② [Transparency] Clarify target to strategically alter examination level (Bank A loan department) ⑤ [Understandability] Output judgment basis info to loan department (Bank A, IT services department/loan department)	⑥ [User Responsibility] Understand that final judgment lies with loan department personnel (Loan department personnel) ⑦ [Awareness] Assess examination level, including whether they are a strategic customer (Loan department personnel) ⑧ [Proper Use] Execute appropriate examination (Loan department personnel)

Control Coordination

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

R007	**False application** AI judgment results are mistaken due to an applicant inputting factually incorrect personal information, or linking the SNS of another individual

Service Provider

④ Erase improper learning data

⑥ Detect applications with signs of impropriety

① Advance explanation for false applications

Users

③ Application based on truth

AI System

Communication

Understand

Action

⑤ Basis of decision

Data

Consensus

User Responsibility

Human Autonomy

Proper Use

Data Quality

Accuracy

Generalization

Application

Process Integrity

Usability

Expectation

Effectiveness

Self-Defense

Data Balance

Robustness

Interpretability

Connectivity

⑦ Judgment history preservation

Understandability

System Environment

Capability

Stability

Confidentiality

Availability

Traceability

Correspondence

User Environment

User Ability

Awareness

Controllability

Limitation

⑨ No settlement for improper applications

⑧ Assess information authenticity

Operation Management

Scalability

Sustainability

Agility

Safety

Accessibility

Auditability

② Access restrictions for clear improper input

Code of Conduct

Accountability

Dignity

Privacy

Fairness

Transparency

Risk Control

- Consider risk control according to the risk chain -

R007	**False application** AI judgment results are mistaken due to an applicant inputting factually incorrect personal information, or linking the SNS of another individual

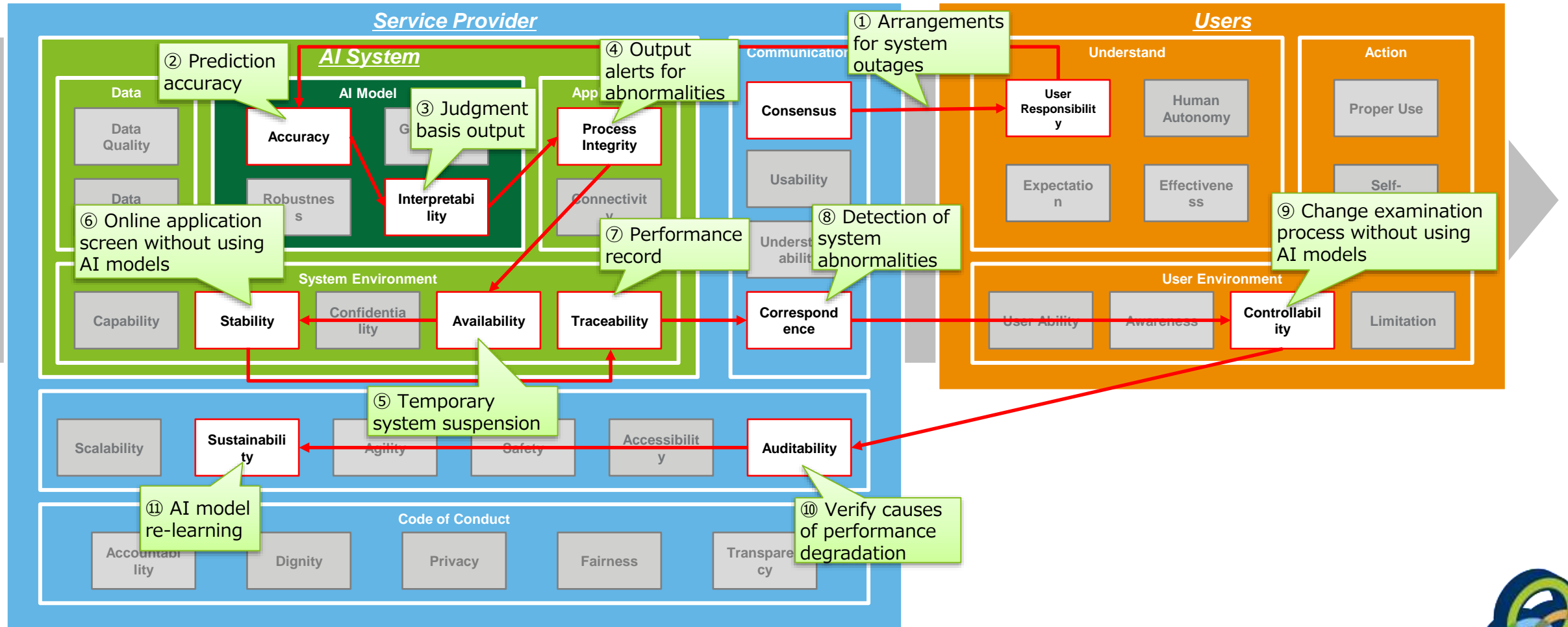| Risk Control		
AI System **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
④ [Data Quality] Revision and removal when evident falsehoods exist within input data, avoiding data examination triggering improper prediction results (Bank A model development department) ⑤ [Interpretability] Output model judgment basis (Bank A model development department) ⑥ [Process Integrity] Output alerts when improper input such as contradictory input info is observed (Bank A model development department) ⑦ [Traceability] Store AI judgment result info during use (Bank A IT services department)	① [Consensus] Explain responsibility/penalties in the event that falsehoods are included in applicant info (Bank A loan department/applicant) ⑧ [Auditability] Verify info with applicant when input info authenticity is suspicious (Bank A loan department) ⑨ [Safety] No examination process for improper input (Bank A loan department)	① [User Responsibility] Explain responsibility/penalties in the event that falsehoods are included in applicant info (Bank A loan department/applicant) ② [Limitation] Implement access restrictions when clear improper input is identified (based on illegal SNS accounts, etc.) (Applicant) ③ [Proper Use] Execute applications based on truth (Applicant)

Control Coordination
- 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 manual examinations occurs, hindering operations

Risk Control

- Consider risk control according to the risk chain -

R008	**Alternative operation for abnormalities** When abnormalities occur due to AI performance degradation, etc., no switch to manual examinations occurs, hindering operations

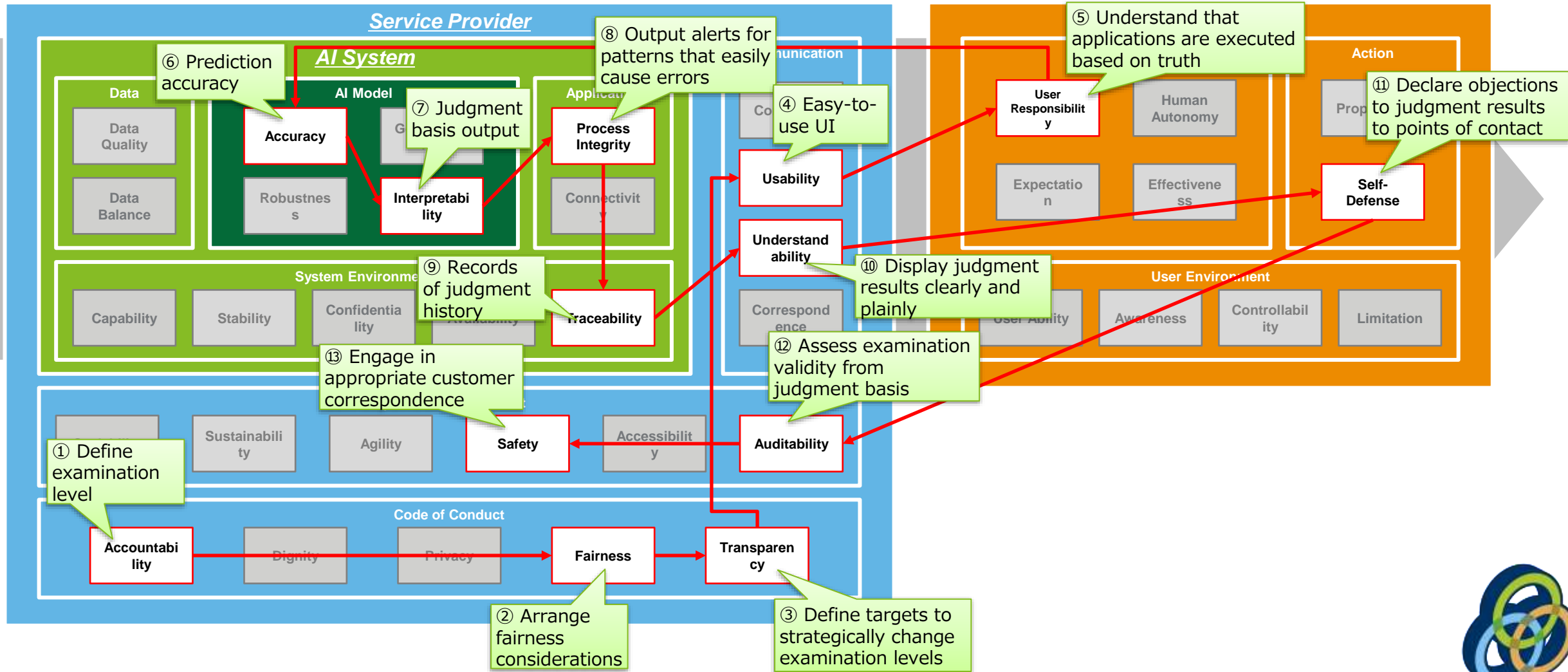| Risk Control		
AI System **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
② [Accuracy] Implement learning to ensure model prediction accuracy (Bank A model development department) ③ [Interpretability] Output model judgment basis (Bank A model development department) ④ [Process Integrity] Output alerts when system abnormalities (including marked prediction rate degradation) is observed (Bank A model development department) ⑤ [Availability] Render system temporarily unusable (Bank A IT services department) ⑥ [Stability] Boot online application screen without using AI models (Bank A IT services department) ⑦ [Traceability] Preserve performance results, etc. (Bank A, IT services department)	① [Consensus] Make arrangements for alternate operations during system outages (Bank A IT services department/loan department personnel) ⑧ [Correspondence] Link with loan department personnel when system abnormalities occur (Bank A IT services department/loan department personnel) ⑩ [Auditability] Assess causes and countermeasures of model performance degradation (Bank A IT services department/loan department) ⑪ [Sustainability] Request AI model re-learning to preserve adequate precision (Bank A IT services department/loan department)	① [User Responsibility] Make arrangements for alternate manual operations when re-learning takes too much time (Bank A IT services department/loan department personnel) ⑨ [Controllability] Change to operations without using AI models (Bank A loan department personnel)

Control Coordination

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

Step 3

R009

Explanation to applicants
Cannot give a rational explanation for examination results and loan amounts to applicants

Risk Control
- Consider risk control according to the risk chain -

R009	**Explanation to applicants** Cannot give a rational explanation for examination results and loan amounts to applicants

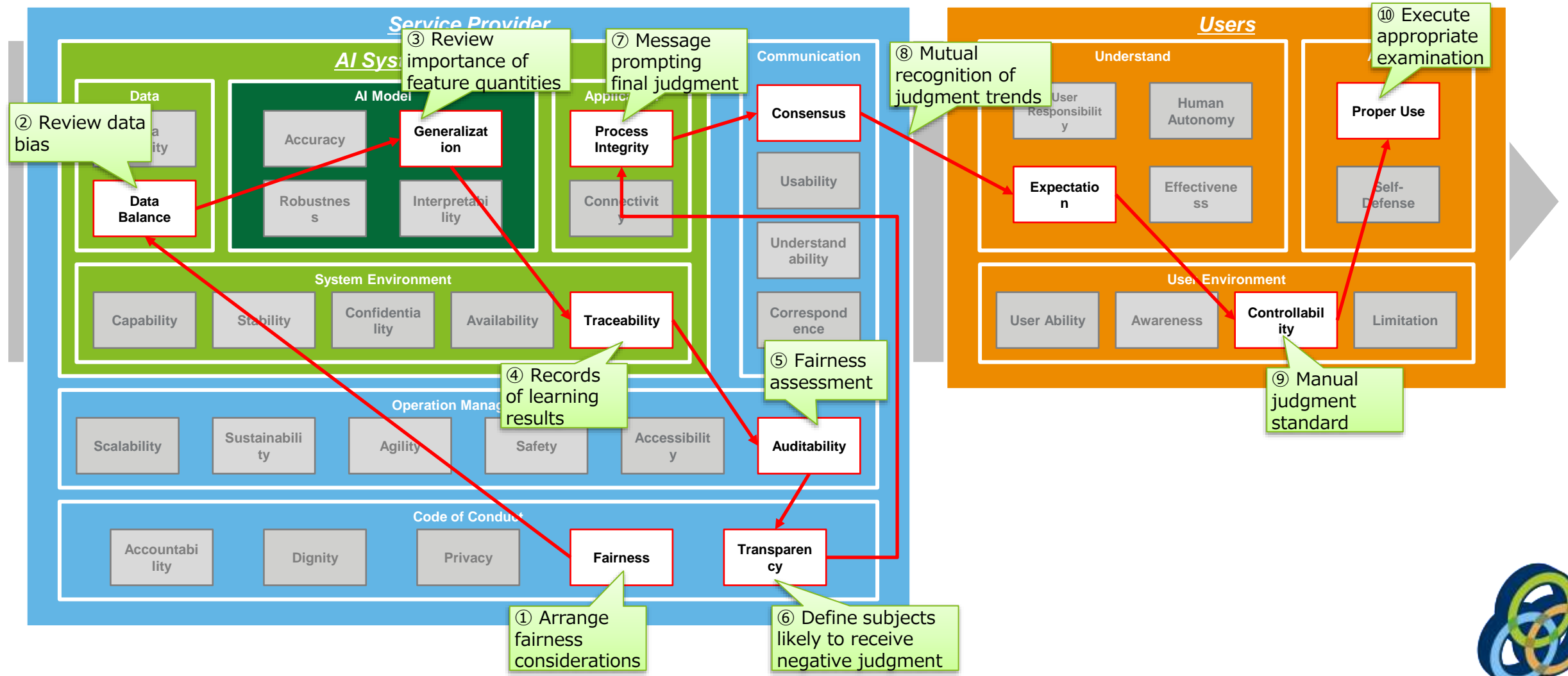| Risk Control		
AI System **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
⑥ [Accuracy] Implement learning to ensure model prediction accuracy (Bank A model development department) ⑦ [Interpretability] Output model judgment basis (Bank A model development department) ⑧ [Process Integrity] Output automatic processing for final judgment in the examination division for subjects without stable prediction results (Bank A model development department) ⑨ [Traceability] Store AI judgment history (Bank A IT services department)	① [Accountability] Define examination level to achieve by AI (Bank A loan department) ② [Fairness] Assess fairness considerations (Bank A loan department) ③ [Transparency] Clarify subjects of strategic examination level lowering/subjects where negative judgment isn't excluded (Bank A IT services department/loan department) ④ [Usability] Prepare UI to prevent input errors (Bank A IT services department/loan department) ⑩ [Understandability] Display judgment basis clearly and plainly (Bank A, IT services department/loan department) ⑫ [Auditability] Assess excessive judgment/validity of examination from judgment basis info, etc. (Bank A IT services department/loan department) ⑬ [Safety] Engage in appropriate customer correspondence (Bank A loan department)	⑤ [User Responsibility] Understand that applications are executed based on truth (Applicant) ⑪ [Self-Defense] Declare events of unsatisfactory judgment results to bank points of contact (Applicant)

Risk Control
- Consider risk control according to the risk chain -

R010	**Fairness** Produce clearly unfair prediction results for specific races/sexes

Risk Control		
AI System **(Bank A: Model Development Department)**	**Service Provider** **(Bank A: IT Services Department/Loan Department)**	**Users** **(Applicant/Loan Department Leader)**
② [Data Balance] Review data bias (Bank A model development department) ③ [Generalization] Review importance of feature quantities (Bank A model development department) ④ [Traceability] Record assessment results from learning(Bank A model development department) ⑦ [Process Integrity] Output automatic processing for final judgment in the examination division for subjects without stable prediction results (Bank A model development department)	① [Fairness] Assess fairness considerations (Bank A loan department) ⑤ [Auditability] Assess model fairness (Bank A IT services department/loan department) ⑥ [Transparency] Clarify targets that cannot be excluded from negative judgment due to reasons such as insufficient data (Bank A IT services department/loan department) ⑧ [Consensus] Mutually recognize prediction performance and negative judgment trends (Bank A IT services department/loan department)	⑧ [Expectation] Mutually recognize prediction performance and negative judgment trends (Bank A IT services department/loan department) ⑨ [Controllability] Arrange manual operation methods (Bank A loan department personnel) ⑩ [Proper Use] Execute appropriate examination (Bank A loan department personnel)