FUJITSU

# AI Ethics Impact Assessment Casebook

# Table of contents


- Revision history ---------- 3
- Introduction ---------- 4
- How to use it ---------- 5
- Disclaimer ---------- 6
- What is an AI Ethics Impact Assessment? ---------- 7
- Terminology, and how to interpret assessment results ---------- 11
- Examples of application ---------- 21
    1. Chatbot ---------- 22
    2. Recruitment AI ---------- 25
    3. Recidivism risk prediction ---------- 28
    4. Facial recognition by police ---------- 31
    5. Hiring decision AI for video interviews ---------- 35
    6. Loan screening AI ---------- 41
    7. Fruit grading ---------- 47
- References ---------- 50

# Revision History

| Version | Date | Summary |
| --- | --- | --- |
| v1.0 | 2022-3-31 | New |

# Introduction

- This document presents a case study of the practical application of an AI Ethics Impact Assessment, an assessment we developed as a method for assessing the ethical risks that may arise in the use of AI systems.
- In recent years, society has become aware of the ethical issues that AI raises. Examples include a facial recognition AI which produced racist results[1] and a recruitment AI that was suspended due to sexist results[2]. On the other hand, countries and organizations in Europe and elsewhere are attempting to address ethical issues by developing ethical principles and guidelines for the use of AI[3][4][5][6][7][8]. The European Commission has gone even further and proposed a draft AI regulation[9].
- Thus, while it has become essential to address ethical risks in the social implementation of AI, the fact that AI systems have multiple stakeholders and the social conditions surrounding them are changing, it is necessary to appropriately address the ethical issues which may arise from the use of AI systems. Recognizing properly how ethical issues occur is a challenge.
- With this in mind, we have developed an AI Ethics Impact Assessment to help developers and providers of AI systems assess the ethical implications of their own use cases. We applied this method to a case study built from already known cases of ethical incidents, and verified that it is possible to use this method to understand where and how ethical risks can occur in an AI system. This document presents the assessment results of these case studies.
- We hope that this document will help AI developers, AI service providers, business users, and others involved in AI to be aware of the ethical issues that can arise.
- PLEASE NOTE THAT THE AI ETHICS IMPACT ASSESSMENT DOES NOT WARRANT THAT NO ETHICAL CONCERN WILL ARISE.
- We will continue to improve this methodology through discussions and examinations with various stakeholders based on this document.

# How to use it

- Intended reader for this document
    - Sales representatives and engineers considering proposals for AI systems to their customers
    - AI algorithm developers, researchers, and data scientists
    - Companies and organizations considering applying AI systems to their own business
    - AI system users and their stakeholders
- Life cycle of the AI system covered in this document
    - This document covers the process of planning, designing, developing, and operating AI systems.
- Usage cases for this document
    - Assessing and explaining the possible ethical risks of AI by a sales representative to a customer when planning an AI system
    - Evaluating possible ethical risks of AI system components during the design phase of the AI system
    - Evaluating possible ethical risks of AI prior to the operation of an AI system
    - Evaluating possible ethical risks of AI for AI systems already in operation
    - User assessment of an AI system to evaluate the possible ethical risks of AI before implementing the AI system

# Disclaimer

- THE CASES DISCUSSED IN THIS DOCUMENT ARE FICTITIOUS, ALTHOUGH THEY ARE BASED ON REAL CASES THAT HAVE OCCURRED IN THE PAST (INFORMATION REGISTERED IN THE AI INCIDENT DATABASE[11] PUBLISHED BY THE PARTNERSHIP ON AI[10], AN INTERNATIONAL AI CONSORTIUM), AS WELL AS SIMILAR CASES AND AI USE CASES PUBLISHED BY THE ISO[12]. THE DETAILS ARE NOT THE SAME AS IN THE ACTUAL CASES, HOWEVER.
- THE CONCEPT OF ETHICS VARIES DEPENDING ON INDIVIDUAL VALUES, CULTURE, RELIGION, SOCIAL CONDITIONS, AND TECHNOLOGY IN EACH COUNTRY OR REGION, AS WELL AS THE TIMES. THE CONTENTS OF THIS DOCUMENT DO NOT REPRESENT THE OPINIONS OF FUJITSU; THEY ARE MERELY ILLUSTRATIVE EXAMPLES.
- FUJITSU DOES NOT WARRANT THAT THERE ARE NO RISKS OTHER THAN THOSE POINTED OUT IN THIS DOCUMENT; EACH ORGANIZATION OR INDIVIDUAL READER OF THIS DOCUMENT IS RESPONSIBLE FOR DETERMINING AND IMPLEMENTING MEASURES IN SPECIFIC CASES.
- FUJITSU IS NOT LIABLE FOR ANY DAMAGES INCURRED BY THE READER IN CONNECTION WITH THIS DOCUMENT OR THE METHOD DESCRIBED THEREIN.

# What is an AI Ethics Impact Assessment?

## Definition of "risk" in the present document

The present document discusses ethical risks. The terms "risk event" and "risk factor" are used in analyses. Explanations and examples of terms related to "risks" are given below.

| Terminology | Description | Example |
|---|---|---|
| Ethical risk of AI | Risks arising from ethical issues raised by AI systems, including those that have a positive impact on the AI system or its stakeholders, as well as those that have a negative impact on the same. | Injustice in a loan approval AI that biases loan approval decisions toward a particular race or gender. |
| Risk Event | Ethical risks of AI that affect stakeholders, as well as risks that stakeholders may be exposed to that result in economic loss, loss of social credibility or, conversely, in increased revenue or social credibility for the stakeholder | Using the example of a loan approval AI, having the percentage of loans deemed acceptable for a particular racial group be extremely low or unfair compared to other racial groups. |
| Risk Factors | Factors that cause ethical risks of AI, specifically risk events. Risk events can be a factor in other risk events. | When the data used to make loan decisions in a loan approval AI includes discrimination based on race or gender |

## What is an AI Ethics Impact Assessment?

- An AI Ethics Impact Assessment is a method for systematically analyzing ethical risks that may occur in AI systems based on AI ethical guidelines.
- This method identifies potential ethical issues by analyzing an AI system using checklist items that embody the AI ethical guidelines.
- The ethical guidelines for AI are based on the ethical guidelines set by the European high-level expert group on artificial intelligence (AI HLEG; Trustworthy AI)[4].
- THIS METHOD CHECKS COMPLIANCE WITH THE GUIDELINES WITHIN THE SCOPE OF THE CONTENTS OF THE AI ETHICAL GUIDELINES. THE GUIDELINES DO NOT CONTAIN ALL REAL-WORLD ETHICS. THE GUIDELINES ARE NOT INTENDED TO DETERMINE WHETHER THE AI SYSTEM IS IN ACCORDANCE WITH THE LAW.
- PLEASE ALSO SEE THE **DISCLAIMER** ON PAGE 5 FOR OTHER PRECAUTIONS REGARDING THE USE OF THE PRESENT DOCUMENT.

An overall diagram of the AI Ethics Impact Assessment is shown in Figure 1.

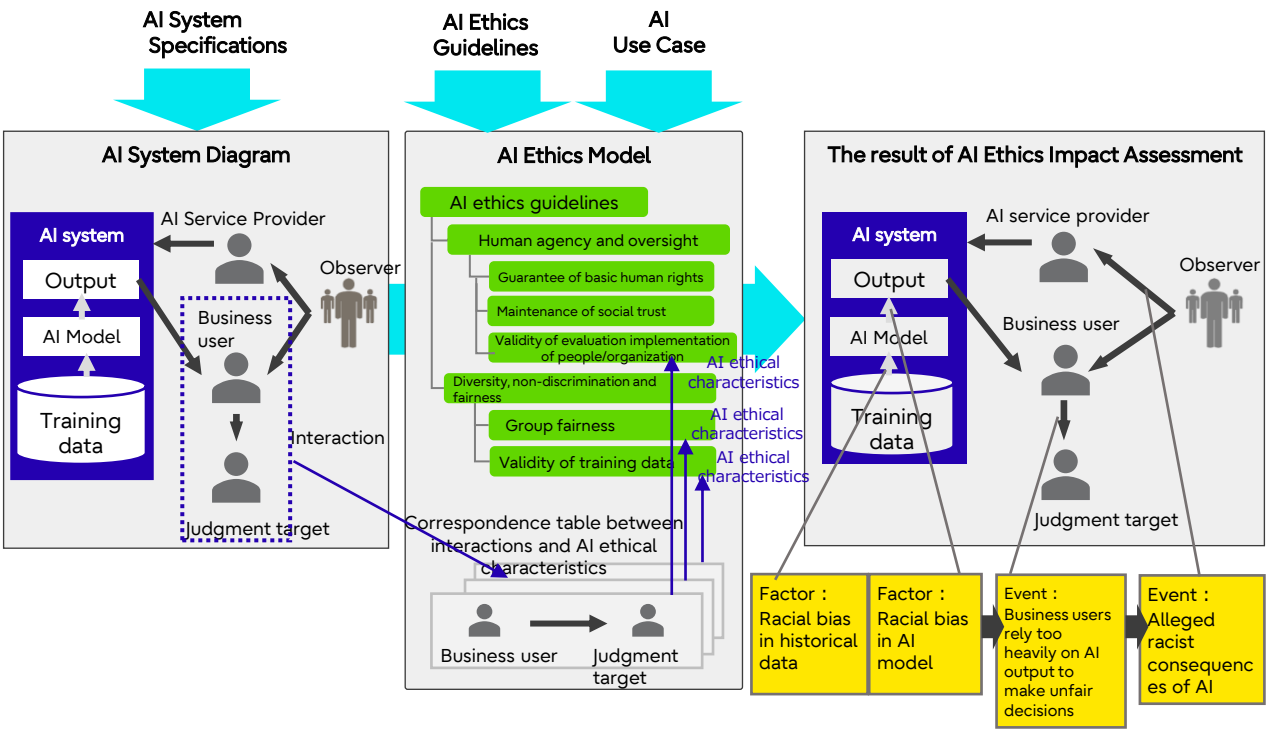



Figure 1. Overview of the AI Ethics Impact Assessment

# What is an AI Ethics Impact Assessment?

## System diagram

- The first step in an AI Ethics Impact Assessment is to create a system diagram. The system diagram identifies and describes the components and stakeholders of the AI system and their relationships with each other and with the components of the AI system. Such relationships are referred to as interactions.
- The system diagram is created by the person who evaluates the system using this method, based on the specifications and use case information about the AI system to be evaluated.

## AI Ethics Model

- An AI ethics model is a structured and concrete version of the AI Ethics Guidelines, which are proposed by the European AI HLEG as "Ethics Guidelines for Trustworthy AI" (hereafter referred to simply as "Trustworthy AI"), which AI use cases are applied to.
- The AI Ethics Model provides a reference table that shows the AI ethical characteristics (characteristics that an ethical AI system must satisfy) that correspond to each type of interaction.
- For each interaction identified in the system diagram, possible ethical risks of AI are extracted using the corresponding table. This extraction process is a task in which the analyst, based on the use case of the AI system to be analyzed, assumes a state that violates the relevant AI ethical characteristics as a specific risk scene.

## Analysis Results

- • The extracted risks are organized and described in a system diagram. The risks are shown in relation to risk events that appear in interactions with stakeholders and risk factors that cause those risk events.

# What is an AI Ethics Impact Assessment?

## AI Ethics Impact Assessment Terminology

| Terminology | Description |
|---|---|
| System Diagram | Diagram showing the components of the AI system and the stakeholders, and also describing the interactions. |
| Interaction | A relationship between any two components of an AI system (including its stakeholders). In an AI Ethics Impact Assessment, ethical risks are extracted by linking them to interactions. |
| AI Ethics Model | Principles and guidelines of AI ethics made tangible through use cases and mapped to interactions |
| AI Ethical Characteristics | Characteristics that an ethical AI system should satisfy, which are extracted by embodying AI ethical guidelines. Examples: “Guarantee fundamental human rights,” “Group fairness” |

## How to interpret assessment results

The present document is organized as follows for each case study.

1. **Overview of the AI System**

   Describes the usage cases and configuration of AI systems, as well as possible ethical issues.

2. **Use Case Overview**

   This table contains information on AI use cases required for an AI Ethics Impact Assessment.

3. **Analysis Chart**

   This chart shows the ethical risks of AI extracted from the AI Ethics Impact Assessment in relation to the interactions on the system diagram created from the use case outline.

4. **Risks Consolidation**

   This table organizes the ethical risks extracted from the AI Ethics Impact Assessment by mapping them to risk events and the risk factors that cause them.

The following pages describe the use case overview, analysis diagram, and risk consolidation.

## How to Interpret the Use Case Overview

- The Use Case Overview is a table that provides information on AI use cases required for an AI Ethics Impact Assessment.
- The items required for an AI Ethics Impact Assessment are indicated by “major items” and “medium items.” A description of each use case is provided in the “Details” column of each item.
- The table below provides a description of each item in the Use Case Overview and an example of the description.
    - Some medium items that fall under the major item "Stakeholders and their roles" are marked with an asterisk (*). The names of these items conform to those defined in the "AI Utilization Guidelines"[7] by the Ministry of Internal Affairs and Communications.

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| Industry | | | The expected type of business or industry in the AI ethics case.<br>Fill in based on the International Standard Industrial Classification[13]. | - Finance/Insurance<br>- Public service and national defense/Mandatory social security services |
| Purpose | | | Purpose of using an AI system | - To shorten loan approval time |
| Service | Service Overview | | Services provided by AI system providers using AI systems | - The AI responds to loan applicants with loan decisions |
| | Availability of customization for each customer | | Indicates whether the system requires customization for each customer | Write one of: Yes/No/Unknown |
| | Requirements | | Customer-set requirements for the service | Write one of: Yes/No/Unknown |
| Usage case | | | Indicates the characteristics of the users of the AI system and their environments, and outlines the tasks to be performed using the AI system | - Loan applicant submits application information via an app, and receives the review results via the app |

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| Stakeholders and their roles | AI Service Provider* | | A kind of stakeholder; performs operations using the developed AI system and provides various services | - Loan screening AI service developer |
| | Developer* | | A type of stakeholder that develops AI systems | - AI developer |
| | Business users* | | A type of stakeholder; people who use AI systems or AI services in the course of business | - Bank personnel who make loan decisions based on the inference results of the loan screening AI<br>- Police officers who make investigative decisions on people recognized as suspicious by facial recognition AI<br>- Doctors who make treatment decisions based on the inference results of medical image diagnosis AI |
| | Consumer-like users* | | A type of stakeholder; users of AI systems or AI services who are not business users | - Loan applicants using a loan screening AI<br>- Patients undergoing medical imaging diagnosis<br>- Chatbot users |
| | Training data provider | | A kind of stakeholder. A person who provides the original data to create training data | - Credit bureaus, banks<br>- Holders of datasets for natural language processing |
| | Source of training data | | A type of stakeholder. Persons who are directly/indirectly related to the training data provider | - Credit bureaus, banks<br>- People who provide their own face images to the face image dataset |
| | Parties involved in training data acquisition | | A type of stakeholder; people, organizations, or systems directly or indirectly involved in training data acquisition | - People who design credit scores, people to whom credit scores have been provided in the past<br>- Photographers who take facial images<br>-Doctors and technicians who take medical images |

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| Stake-holders and their roles | Inferential data providers | | A kind of stakeholder; a person who provides input data to create inferential data | - Credit bureaus, banks<br>- Hospitals that provide medical image data |
| | Inferential data source | | A type of stakeholder; people or organizations involved in some capacity with the content of inferential data | - Credit bureaus, banks<br>- Patients whose images are used in medical image data |
| | Parties involved in the acquisition of inferential data | | A kind of stakeholder; people who are directly/indirectly involved in acquiring inferential data | - Photographers who take facial images<br>- Doctors and technicians who take medical images<br>- The people in the image |
| | Observers | | A type of stakeholder.<br>A person monitoring an AI system or AI service | - Human rights organizations, media |
| | Service UI/API provider | | People, organizations, or systems that build functions for AI system users based on the inference results of the AI system | - Provider of monitors and controls for medical diagnostic imaging<br>- Provider of robot control unit for factory robots<br>- Providers of platforms such as SNS on which chatbots operate |
| | Judgment target | | Persons or organizations to be judged or evaluated by the AI system | - People captured by surveillance cameras, employees evaluated by HR AI |
| | Service authorizer | | Persons or organizations that authorize the development of AI systems or the provision of services using AI systems | - Relevant ministries, regulators |
| | Other stakeholders | | A type of stakeholder; a person or organization indirectly affected by the output from an AI system or AI service. | - Family members and business partners of users<br>- Insurance company with which the user has a contract<br>- The community to which the user belongs |

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| Presence/absence of a human-in-the-loop | | | Indicates whether or not the AI system will include the judgment of the AI system user with respect to the inference results. Risk events and risk factors to be extracted differ depending on whether or not human judgment is involved. | - The bank's loan officer makes the final decision on a loan assessment pass/fail with a human-in-the-loop.<br>No human-in-the-loop when AI inference results are communicated directly to the loan applicant |
| Presence/absence of existing methods | | | Indicates whether the task to be performed/supported by the AI system can be performed by existing methods. If existing methods can be used, efficiency and accuracy must be compared between the existing methods and the AI system. | |

# Terminology, and how to interpret assessment results

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| AI Task | Task | | An overview of the inferential data, which is the input to the AI model, and the inference results, which is the output | - Input loan applicant's application information, credit score, and transaction data, and output loan decision |
| | Problem Classification | | Types of problems handled by the AI model | Classification, recommendation, regression, natural language processing, speech recognition, image recognition, image generation |
| | Output | | AI Model Output | - Loan assessment pass/fail<br>- Conversational text of a chatbot |
| | Technology | | Source of the AI model, indicating whether it was created during the development of the AI system, by others, or a combination | Own, OSS, Own + OSS, other companies' technology |
| | Breakdown of models required to perform AI tasks | | Indicate the names of all AI models required to perform the AI task | - Speech recognition model, sentence comprehension model, facial expression recognition model |
| | Additional training with updated training data | | Indicate whether additional training is required with updated training data after the start of operation | Write one of: Yes/No/Unknown |
| | Real-time performance | | Whether real-time performance is required for processing AI tasks or not | Write one of: Yes/No/Unknown |
| | Verification of AI Tasks<br>(Consistency with decisions made by existing methods) | | If similar tasks have been performed by existing means, an indication of the accuracy of the AI task or differences in individual results compared to existing means | - Comparison with decisions made by loan reviewers |

# Terminology, and how to interpret assessment results

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| Training Data | Breakdown, provider and source | | Breakdown of training data and respective data providers and data sources | - Transaction data of past loan applicants (banks) |
| | Teacher labels | | Details of teacher labels to be set on training data | - Loan availability/non-availability |
| | Protective attributes, attributes used for fairness verification and mitigation | | Protective attributes included in training data<br>(gender, age, nationality, race) | - Gender, age |
| | Presence/absence of personal information | | Indicates whether personal information is included in the training data | Indicate whether yes/no/unknown |
| | Notes on data acquisition | | Things to keep in mind when acquiring training data | Obtain permission from those involved with the training data |
| | Data Storage | | Whether training data can be stored or not and precautions to be taken when storing training data | Training data may not be stored |

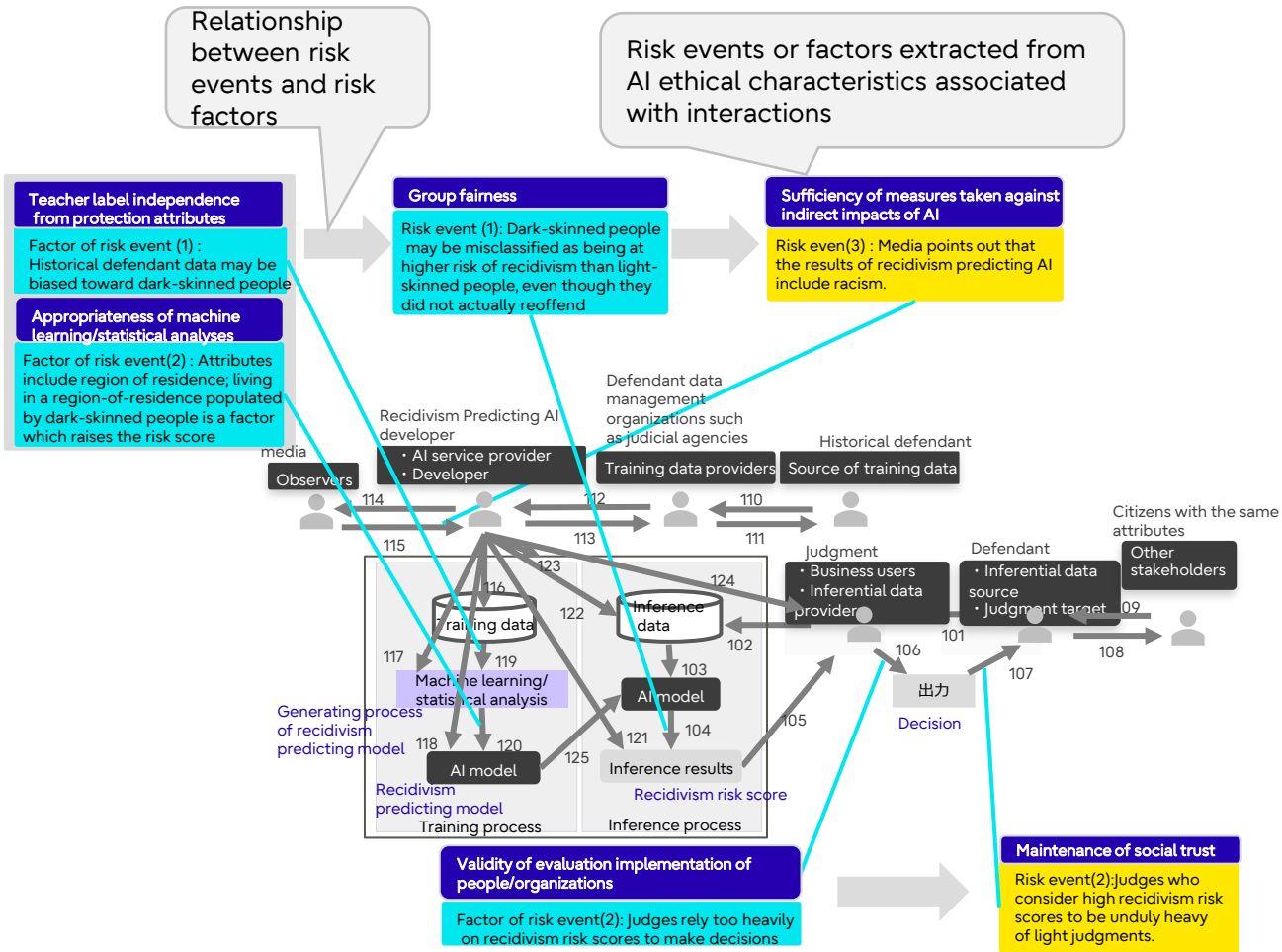
# Terminology, and how to interpret

| Major Item | Medium Item | Details | Item Description | Example |
|---|---|---|---|---|
| Inferential data | Breakdown, provider and source | | Breakdown of inferential data and respective data providers and data sources | - Credit scores of loan applicants (credit bureaus) |
| | Protective attributes, attributes used for fairness verification and mitigation | | Protected attributes included in inferential data<br>(gender, age, nationality, race) | - Gender, age |
| | Presence/absence of personal information | | Indicates whether personal information is included in the inferential data | Indicate whether yes/no/unknown |
| | Notes on data acquisition | | Things to keep in mind when acquiring inferential data | Is it necessary to obtain permission from the data source? What are the terms of permission, etc. |
| AI System Output | Things to keep in mind during output | | Things to keep in mind during AI system output | - Provide a point of contact to discuss the results of AI |
| Citations, reference articles | | | Incident Database Registration<br>Link to URL or explanatory article | |

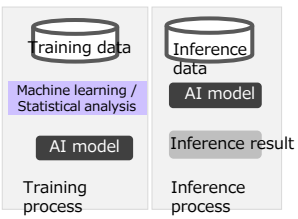
# Terminology, and how to interpret assessment results

## How to Interpret the Analysis Chart

- System diagram showing risk events and risk factors extracted by the AI Ethics Impact Assessment in relation to interactions.




Description of symbols in analysis chart

| | |
|---|---|
| | Stakeholder |
|   | AI system |
| | Interaction (Number describes Interaction ID) |

| | |
|---|---|
|  | Risk event |
| | Risk factor |
| | AI ethical characteristics |
|  | Correspondence between an interaction and a risk event or risk factor |

## How to Interpret Risk Consolidation

- Table that organizes risk events extracted by AI Ethics Impact Assessment and the risk factors that caused them by mapping them to each other.

Describes risk events and the risk factors that cause them side by side

Rows with the Risk Factor column left blank describe the risk event and the corresponding AI ethical characteristics and interaction ID.

The other rows describe the risk factors listed and their corresponding AI ethical characteristics and interaction ID.

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| (1) Dark-skinned people may be misclassified as being at higher risk of re-offending than light-skinned people, even though they did not actually re-offend | | Group fairness | 104 |
| | Attributes include region of residence; living in a region-of-residence populated by dark-skinned people is a factor which raises the risk score | Appropriateness of machine learning/statistical analyses | 120 |
| | Historical defendant data may be biased toward dark-skinned people | Independence of teacher labels from protective attributes | 119 |
| (2) Judges who consider high re-offending risk scores to be unduly heavy or light judgments. | | Maintenance of social trust | 107 |
| | Judges rely too heavily on re-offending risk scores to make decisions | Validity of evaluation implementation of people/organizations | 106 |
| (3) Media points out that the results of re-offense-predicting AI include racism. | | Sufficiency of measures taken against indirect impacts of AI | 114 |
| | Risk Event (1) | Group fairness | 104 |

Numbers the risk event if it is a factor in another risk event, and lists the number in the risk factor.

# Examples of application

## List of applications

| No. | Name | Industry (International Standard Industrial Classification[13]) | AI Task | Data Type | Possible ethical issues |
|---|---|---|---|---|---|
| 1 | Chatbots | Arts, entertainment and recreation | Sentence comprehension and generation | Natural language | Transmits discriminatory expressions. |
| 2 | Recruitment AI | Management and support services industry | Classification | Table Data | Discrimination against women in hiring results. |
| 3 | Re-offending Risk Prediction | Public service/National defense and mandatory social security services | Classification | Table Data | AI's prediction results are unfair to dark-skinned people. |
| 4 | Facial Recognition by Police | Public service/National defense and mandatory social security services | Classification | Image | An AI matches a suspect's facial image captured by a surveillance camera with a citizen's facial database, resulting in the false arrest of an unrelated citizen. |
| 5 | Hiring Decision AI for Video Interviews | Management and support services industry | Classification | Image | The judgment result by the facial expression recognition AI may not be fair. |
| 6 | Loan Screening AI | Finance and insurance industry | Classification | Table Data | Women and young business owners are less likely to be screened than older male business owners. |
| 7 | Fruit Grading | Agriculture, forestry and fisheries | Classification | Image | Used to evaluate the farmer, which is not the intended purpose. |

# 1. Chatbots

## Overview of the AI System

- AI System Usage Cases
  - A chatbot to talk with unspecified users on SNS
- AI System Organization
  - AI responds to questions from users on social media
  - Sentence comprehension/generation model
  - Learning how to converse with an unspecified user
- Typical ethical issues
  - Chatbot learns how to converse with malicious users, and sends racist, sexist, or violent messages.

## Use Case Overview

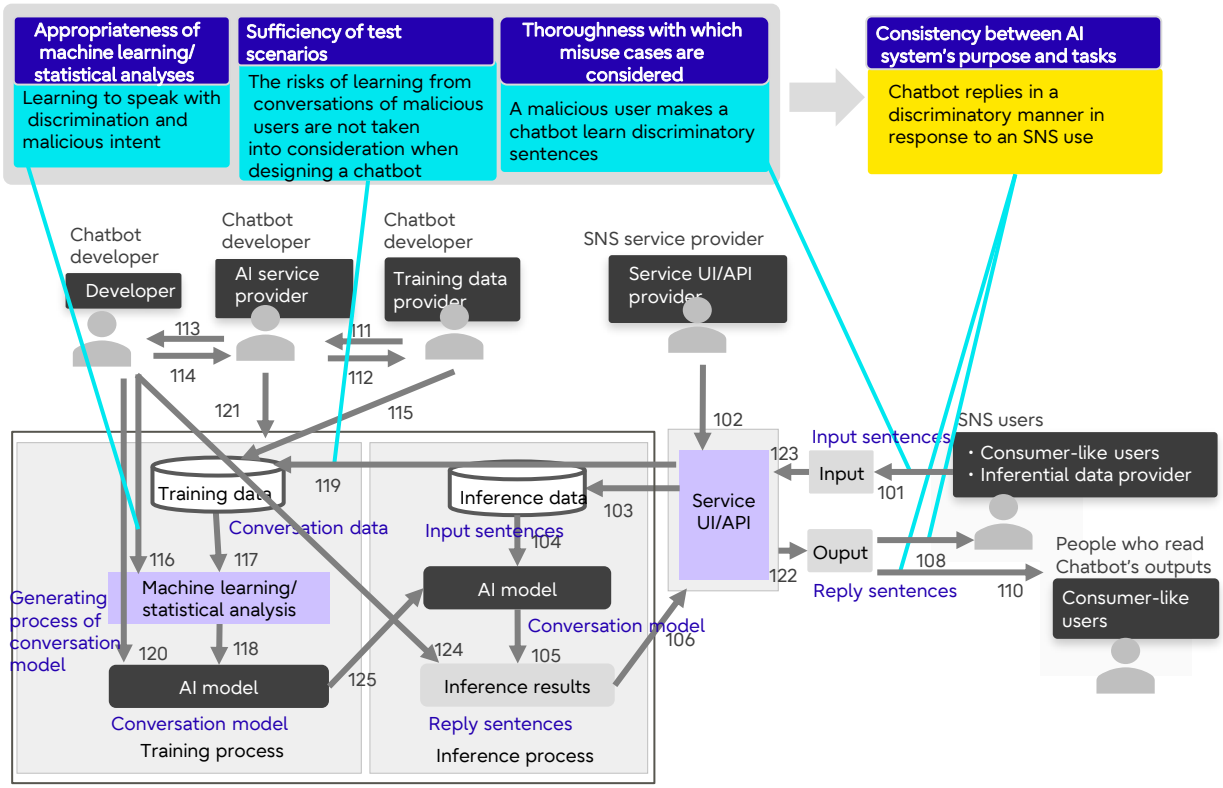
| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Entertainment |
| Purpose | | Chatbot that learns to converse freely with users |
| Service | Service overview | Chatbot that responds meaningfully to what users on social media send to it |
| | Availability of customization for each customer | N/A |
| | Requirements | N/A |
| Usage case | | Conversations between the user and the chatbot |
| Stakeholders and their roles | AI service provider | Chatbot development vendor |
| | Developer | Chatbot development vendor |
| | Business users | N/A |
| | Training data provider | Chatbot development vendor |
| | Source of training data | Chatbot development vendor |
| | Parties involved in training data acquisition | Unknown |
| | Inferential data providers | Social media users |
| | Inferential data source | Social media users |
| | Parties involved in the acquisition of inferential data | Unknown |
| | Consumer-like users | Social media users |
| | Observers | N/A |
| | Service UI/API provider | Social media service provider |
| | Judgment target | N/A |
| | Service Authorizer | N/A |

# 1. Chatbots

| Major Item | Medium Item | Details |
|---|---|---|
| Stakeholders and their roles | Other Stakeholders | People, media, and other such organizations that see chatbot conversations |
| Presence/absence of a human-in-the-loop | | N/A |
| Presence/absence of existing methods | | N/A |
| AI Task | Task | Sentence comprehension and generation |
| | Problem Classification | Natural language processing |
| | Output label | Sentences |
| | Technology | Unknown |
| | Breakdown of models required to perform AI tasks | Sentence comprehension/generation model |
| | Update/additional training of training data | Yes |
| | Real-time performance | Yes |
| | Verification of AI tasks (consistency with decisions made by existing means) | N/A |
| Training Data | Breakdown, provider and source | Social media user input text (provider and source: development vendor, social media user) |
| | Teacher labels | N/A |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | N/A |
| | Notes on data acquisition | Unknown |
| | Data Storage | N/A |
| Inferential data | Breakdown and data owner(s) | Social media user input text (provider and source: social media user) |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | N/A |
| | Notes on data acquisition | N/A |
| AI System Output | Things to keep in mind during output | N/A |
| Citations, reference articles | | PAI AI incident database: incident ID #6<br>https://incidentdatabase.ai/cite/6 #undefined |

# 1. Chatbots

## Analysis Chart




## Risks Consolidation

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| Chatbot replies in a discriminatory manner in response to an SNS user | | Consistency between AI system's purpose and tasks | 108, 110 |
| | Learning to speak with discrimination and malicious intent | Appropriateness of machine learning/statistical analyses | 116 |
| | The risks of learning from conversations of malicious users are not taken into consideration when designing a chatbot | Sufficiency of test scenarios | 119 |
| | A malicious user makes a chatbot learn discriminatory sentences | Thoroughness with which misuse cases are considered | 101 |

# 2. Recruitment AI

## Overview of the AI System

- AI System Usage Cases
  - In recruiting: screening candidates for employment interviews based on their resumes
- AI System Organization
  - Training is done on past job applicant resumes and hiring results to generate a hiring decision AI. Race and sex are not included in the resume information.
- Typical ethical issues
  - AI's candidate recruitment is extremely biased in favor of men.

## Use Case Overview

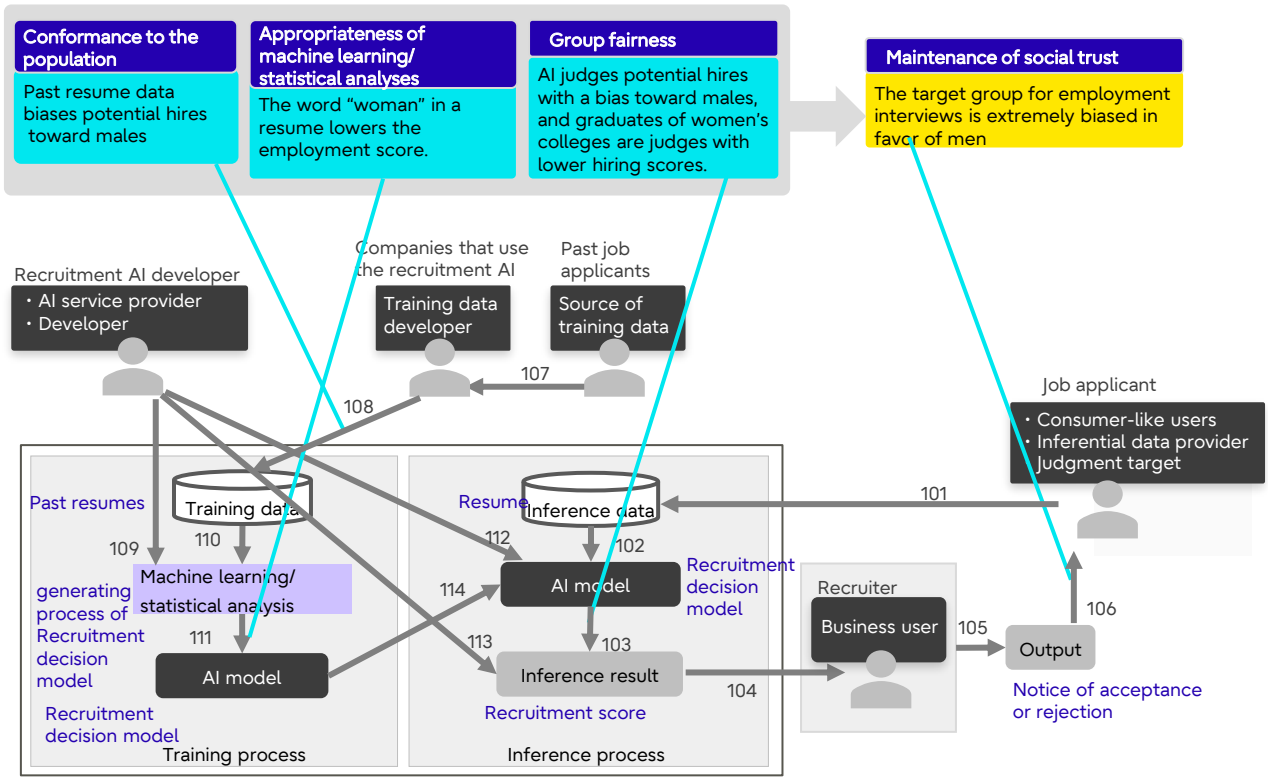
| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Management and support services industry |
| Purpose | | Narrowing down the number of candidates for employment interviews |
| Service | Service overview | AI tool that automatically categorizes the resumes of your job candidates and selects the most promising candidates |
| | Availability of customization for each customer | N/A |
| | Requirements | N/A |
| Usage case | | Automatically categorizes the resumes of job candidates and selects the most promising candidates |
| Stakeholders and their roles | AI service provider | Development vendors for a recruitment AI |
| | Developer | Development vendors for a recruitment AI |
| | Business users | Recruiter for a company that uses the recruitment AI |
| | Training data provider | Companies that use the recruitment AI |
| | Source of training data | Companies that use the recruitment AI |
| | Parties involved in training data acquisition | Past Job applicants |
| | Inferential data providers | Job applicants |
| | Inferential data source | Job applicants |
| | Parties involved in the acquisition of inferential data | Unknown |
| | Consumer-like users | Job applicants |
| | Observers | Unknown |
| | Service UI/API provider | Job applicants |
| | Judgment target | Job applicants |
| | Service authorizer | Unknown |
| | Other stakeholders | Unknown |

# 2. Recruitment AI

| Major Item | Medium Item | Details |
|---|---|---|
| Presence/absence of a human-in-the-loop | | N/A |
| Presence/absence of existing methods | | Scoring job applicants based on resume content |
| AI Task | Task | Determining the level of a potential candidacy of an applicant based on resume data |
| | Problem Classification | Classification issues |
| | Output label | Score on a 5-point scale |
| | Technology | N/A |
| | Breakdown of models required to perform AI tasks | N/A |
| | Update/additional training of training data | N/A |
| | Real-time performance | N/A |
| | Verification of AI tasks (consistency with decisions made by existing means) | Decision by the hiring manager |
| Training Data | Breakdown, provider and source | Past resumes (Provider: company using the recruitment AI, Source: Past job applicants) |
| | Teacher labels | Score on a 5-point scale |
| | Protective attributes, attributes used for fairness verification and mitigation | Sex |
| | Presence/absence of personal information | Unknown |
| | Notes on data acquisition | Unknown |
| | Data Storage | Unknown |
| Inferential data | Breakdown, provider and source | Resume (Provider/source: job applicant) |
| | Protective attributes, attributes used for fairness verification and mitigation | Sex |
| | Presence/absence of personal information | Yes |
| | Notes on data acquisition | Unknown |
| AI System Output | Things to keep in mind during output | Unknown |
| Citations, reference articles | | PAI incident database #37 https://incidentdatabase.ai/ cite/37 |

# 2. Recruitment AI

## Analysis Chart




## Risks Consolidation

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| The target group for employment interviews is extremely biased in favor of men. | | Maintenance of social trust | 106 |
| | Past resume data biases potential hires toward males | Conformance to the population | 108 |
| | The word "woman" in a resume lowers the employment score. | Appropriateness of machine learning/statistical analyses | 111 |
| | AI judges potential hires with a bias toward males, and graduates of women's colleges are judged with lower hiring scores. | Group fairness | 103 |

# 3. Recidivism risk prediction

## Overview of the AI System

- AI System Usage Cases
  - Recidivism risk prediction AI: Using information about a defendant(*) as input data, the AI determines the defendant's risk of recidivism on a 10-point scale. Judges use the recidivism risk score as a reference when deciding whether a defendant should be released on parole or sentenced.

    (*) Includes the defendant's criminal history, drug use, education and employment level. Race is not included.
- Typical ethical issues
  - Recidivism risk prediction AI leads to racist outcomes

*THE LEGAL SYSTEMS, SUCH AS THE CRIMINAL AND JUDICIAL SYSTEMS, OF EACH JURISDICTION ARE NOT INCLUDED IN THE SCOPE OF EVALUATION IN THIS DOCUMENT.

## Use Case Overview

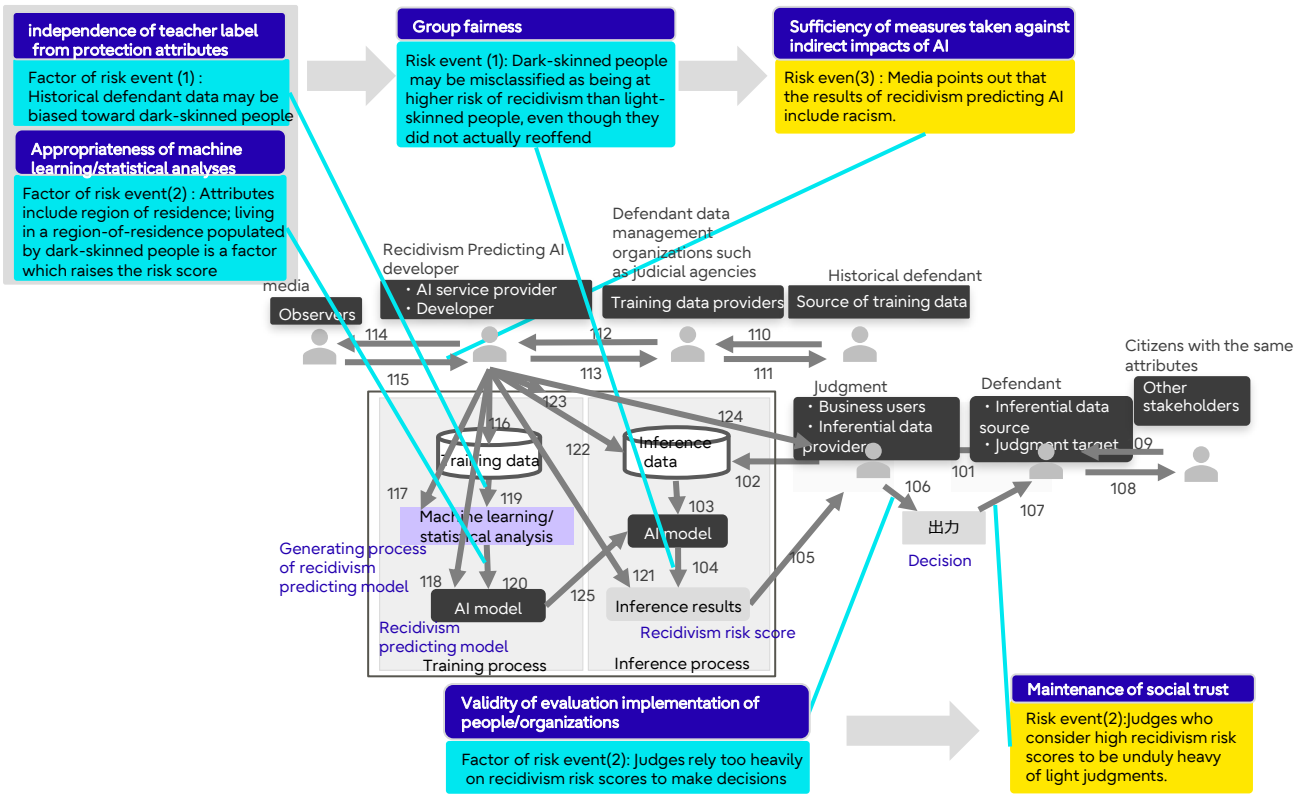
| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Public service/National defense and mandatory social security services |
| Purpose | | Reference for use by judges to make decisions regarding parole and sentencing |
| Service | Service Overview | Determine the defendant's risk of recidivism on a 10-point scale. |
| | Availability of customization for each customer | N/A |
| | Requirements | Unknown |
| Usage case | | Judges will refer to the recidivism risk score for decisions regarding pre-trial parole and determining whether a defendant will reoffend . |
| Stakeholders and their roles | AI service provider | Recidivism risk prediction AI developer vendor |
| | Developer | Recidivism risk prediction AI developer vendor |
| | Business users | Court judges |
| | Training data provider | Organizations that manage defendant data |
| | Source of training data | Past defendants |
| | Parties involved in training data acquisition | Unknown |
| | Inferential data providers | Court |
| | Inferential data source | Defendant |
| | Parties involved in the acquisition of inferential data | Unknown |
| | Consumer-like users | N/A |
| | Observers | Media, general public (e.g., people who read news in the media, and people who are interested in analyzing the data) |

# 3. Recidivism risk prediction

| Major Item | Medium Item | Details |
|---|---|---|
| Stakeholders and their roles | Service UI/API provider | N/A |
| | Judgment target | Defendant |
| | Service authorizer | N/A |
| | Other stakeholders | People with the same attributes as the defendant |
| Presence/absence of a human-in-the-loop | | N/A |
| Presence/absence of existing methods | | Judgment passed down by the judge |
| AI Task | Task | Estimate the defendant's risk of recidivism on a 10-point scale |
| | Problem Classification | Classification issues |
| | Output | 10-point recidivism risk score |
| | Technology | Unknown |
| | Breakdown of models required to perform AI tasks | Recidivism risk prediction model |
| | Update/additional training of training data | Unknown |
| | Real-time performance | Unknown |
| | Verification of AI tasks (consistency with decisions made by existing means) | Prior cases |
| Training Data | Breakdown, provider and source | Information about the defendant's past (includes criminal history, drug use, education and employment level; race is not included) (Provider and source: the authorities) |
| | Teacher labels | 10-point recidivism risk score |
| | Protective attributes, attributes used for fairness verification and mitigation | Age, sex |
| | Presence/absence of personal information | Unknown |
| | Notes on data acquisition | Unknown |
| | Data Storage | Unknown |
| Inferential data | Breakdown, provider and source | Information about the defendant (not including race)<br>(Provider: the court, source: the defendant) |
| | Protective attributes, attributes used for fairness verification and mitigation | Age, sex |
| | Presence/absence of personal information | Yes |
| | Notes on data acquisition | Unknown |
| AI System Output | Things to keep in mind during output | Unknown |
| Citations, reference articles | | PAI Artificial Intelligence Incident Database #11<br>https://incidentdatabase.ai/cite/11/ |

# 3. Recidivism risk prediction

## Analysis Chart




## Risks Consolidation

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| (1) Dark-skinned people may be misclassified as being at higher risk of recidivism than light-skinned people, even though they did not actually reoffend | | Group fairness | 104 |
| | Attributes include region of residence; living in a region-of-residence populated by dark-skinned people is a factor which raises the risk score | Appropriateness of machine learning/statistical analyses | 120 |
| | Historical defendant data may be biased toward dark-skinned people | Independence of teacher labels from protective attributes | 119 |
| (2) Judges who consider high recidivism risk scores to be unduly heavy or light judgments. | | Maintenance of social trust | 107 |
| | Judges rely too heavily on recidivism risk scores to make decisions | Validity of evaluation implementation of people/organizations | 106 |
| (3) Media points out that the results of recidivism predicting AI include racism. | | Sufficiency of measures taken against indirect impacts of AI | 114 |
| | Risk Event (1) | Group fairness | 104 |

# 4. Facial recognition by police

## Overview of the AI System

■ AI System Usage Cases

- Match the images of people captured by surveillance cameras with a facial recognition database, and extract similar people from the database.

■ AI System Organization

- Using images of people captured by surveillance cameras as input, the facial recognition AI matches them against a database of faces to extract people who look like them.

■ Typical ethical issues

- • Facial recognition AI incorrectly identifies an unrelated person as a suspect, leading to a false arrest.

※THE LEGAL SYSTEMS, SUCH AS THE CRIMINAL AND JUDICIAL SYSTEMS, OF EACH JURISDICTION ARE NOT INCLUDED IN THE SCOPE OF EVALUATION IN THIS DOCUMENT.

## Use Case Overview

| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Public service/National defense and mandatory social security services |
| Purpose | | Identify criminals |
| Service | Service Overview | Determining if the input facial image matches the person in the surveillance camera image |
| | Availability of customization for each customer | N/A |
| | Requirements | N/A |
| Usage case | | Match the images of people captured by surveillance cameras with a facial recognition database and extract similar people from the database. |

# 4. Facial recognition by police

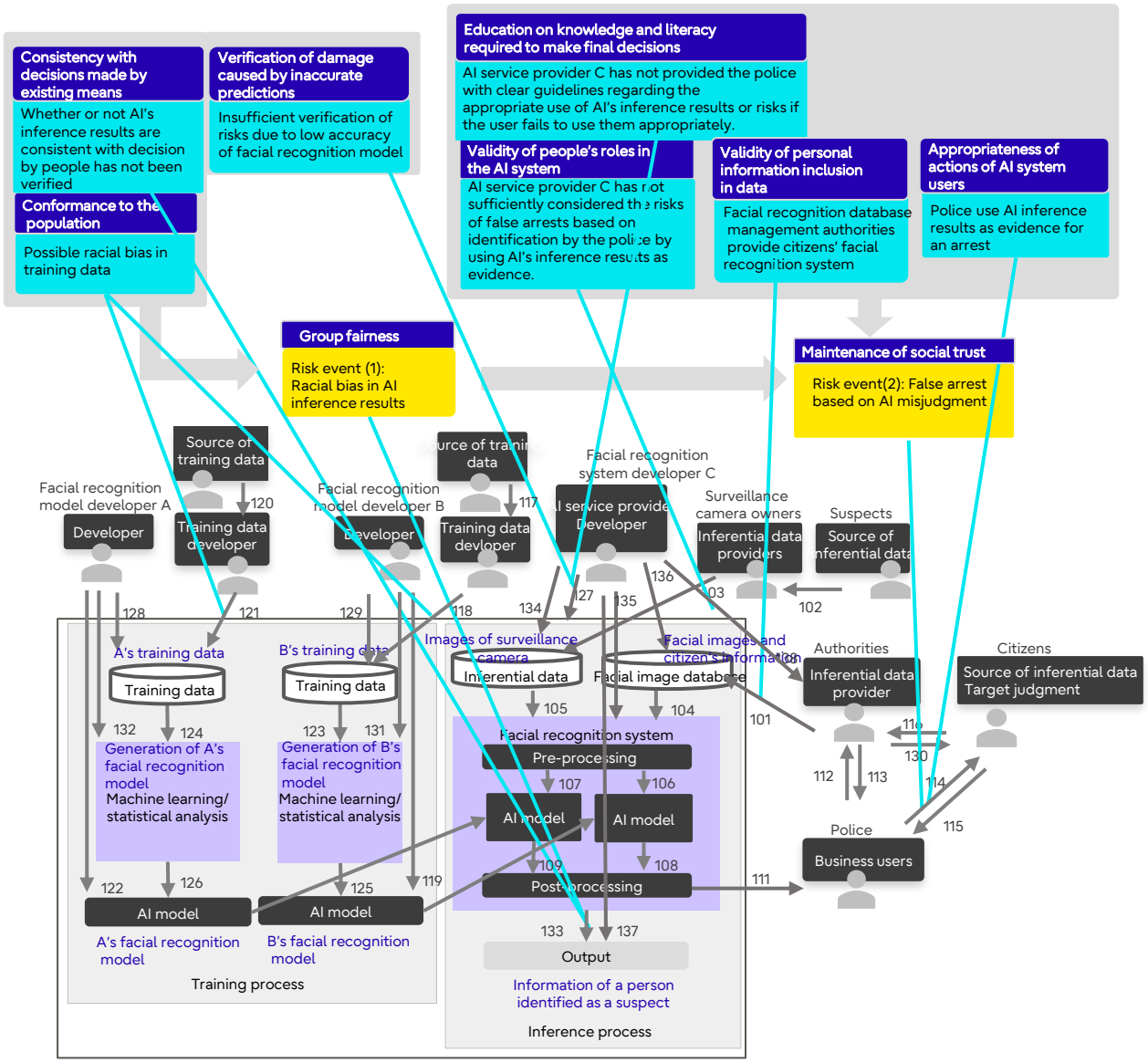
| Major Item | Medium Item | Details |
|---|---|---|
| Stakeholders and their roles | AI service provider | Facial recognition system vendor: “C Inc.” |
| | Developer | Facial recognition system vendor C Inc., facial recognition model vendor A Inc., facial recognition model vendor B Inc. |
| | Business users | Police |
| | Training data provider | Unknown |
| | Source of training data | Unknown |
| | Parties involved in training data acquisition | Unknown |
| | Inferential data providers | (1) Surveillance camera owners, (2) Facial recognition database management authorities |
| | Inferential data source | (1) Suspects caught on surveillance cameras, (2) Citizens |
| | Parties involved in the acquisition of inferential data | Unknown |
| | Consumer-like users | N/A |
| | Observers | Unknown |
| | Service UI/API provider | N/A |
| | Judgment target | Citizens |
| | Service authorizer | N/A |
| | Other stakeholders | Citizens |

# 4. Facial recognition by police

| Major Item | Medium Item | Details |
|---|---|---|
| Presence/absence of a human-in-the-loop | | Yes (police) |
| Presence/absence of existing methods | | Human judgment |
| AI Task | Task | Detects face images similar to the face of the person to be identified from the input face image data. |
| | Problem Classification | Facial recognition |
| | Output | Is there a match for the face? |
| | Technology | Facial recognition technology |
| | Breakdown of models required to perform AI tasks | Facial recognition model |
| | Update/additional training of training data | Unknown |
| | Real-time performance | Unknown |
| | Verification of AI tasks (consistency with decisions made by existing means) | Unknown |
| Training Data | Breakdown, provider and source | Unknown |
| | Teacher labels | Unknown |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | N/A |
| | Notes on data acquisition | Unknown |
| | Notes on data storage | N/A |
| Inferential data | Breakdown, provider and source | Surveillance camera images (provider: surveillance camera owner, source: person in the image), citizen facial recognition database (provider: facial recognition management authority, source: citizen) |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | Yes |
| | Notes on data acquisition | Unknown |
| AI System Output | Things to keep in mind during output | N/A |
| Citations, reference articles | | https://incidentdatabase.ai/cite/74 |

# 4. Facial recognition by police

## Analysis Chart



# 4. Facial recognition by police

## Risks Consolidation

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| (1) Racial bias in AI inference results. | | Group fairness | 133 |
| | Possible racial bias in training data | Conformance to the population | 118, 121 |
| | Whether or not AI's inference results are consistent with decisions by people has not been verified | Consistency with decisions made by existing means | 133 |
| | Insufficient verification of risks due to low accuracy of facial recognition model | Verification of damage caused by inaccurate predictions | 127 |
| (2) False arrest based on AI misjudgment | | Maintenance of social trust | 114 |
| | Risk Event (1) | Group fairness | 133 |
| | AI service provider, C Inc., has not sufficiently considered the risks of false arrests based on identification by the police by using AI's inference results as evidence | Validity of people's roles in the AI system | 127 |
| | AI service provider, C Inc., does not provide the police with clear guidelines regarding the appropriate use of AI's inference results or risks if the user fails to use them appropriately | Education on knowledge and literacy required to make final decisions | 138 |
| | Facial recognition database management authorities provide citizens' facial recognition databases to the facial recognition system. | Validity of personal information inclusion in data | 101 |
| | Police use AI inference results as evidence for an arrest | Appropriateness of actions of AI system users | 114 |

# 5. Hiring decision AI for video interviews

## Overview of the AI System

- AI System Usage Cases
  - Hiring decision AI for video interviews: An AI system for evaluating job applicants based on interview videos of job applicants
  - Job applicants sit in front of a webcam and answer questions, and the AI evaluates their characteristics based on their words, voice, and facial expressions.
- AI System Organization
  - The AI system has a speech recognition model and a facial expression recognition model. Each model outputs a job applicant evaluation result using the characteristics of the job applicant determined from the video interview.
- Typical ethical issues
  - Racial bias in the facial expression recognition model leads to racist results in the AI's assessment of job applicants.

## Use Case Overview

| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Management and support services industry |
| Purpose | | Automated screening of video interviews of job applicants |
| Service | Service Overview | The AI evaluates job candidates as they answer interview questions on video, and evaluates their characteristics based on their words, voice, and facial expressions |
| | Availability of customization for each customer | Unknown |
| | Requirements | Unknown |
| Usage case | | Job applicants sit in front of a webcam and answer questions, and the AI evaluates their characteristics based on their words, voice, and facial expressions |

# 5. Hiring decision AI for video interviews

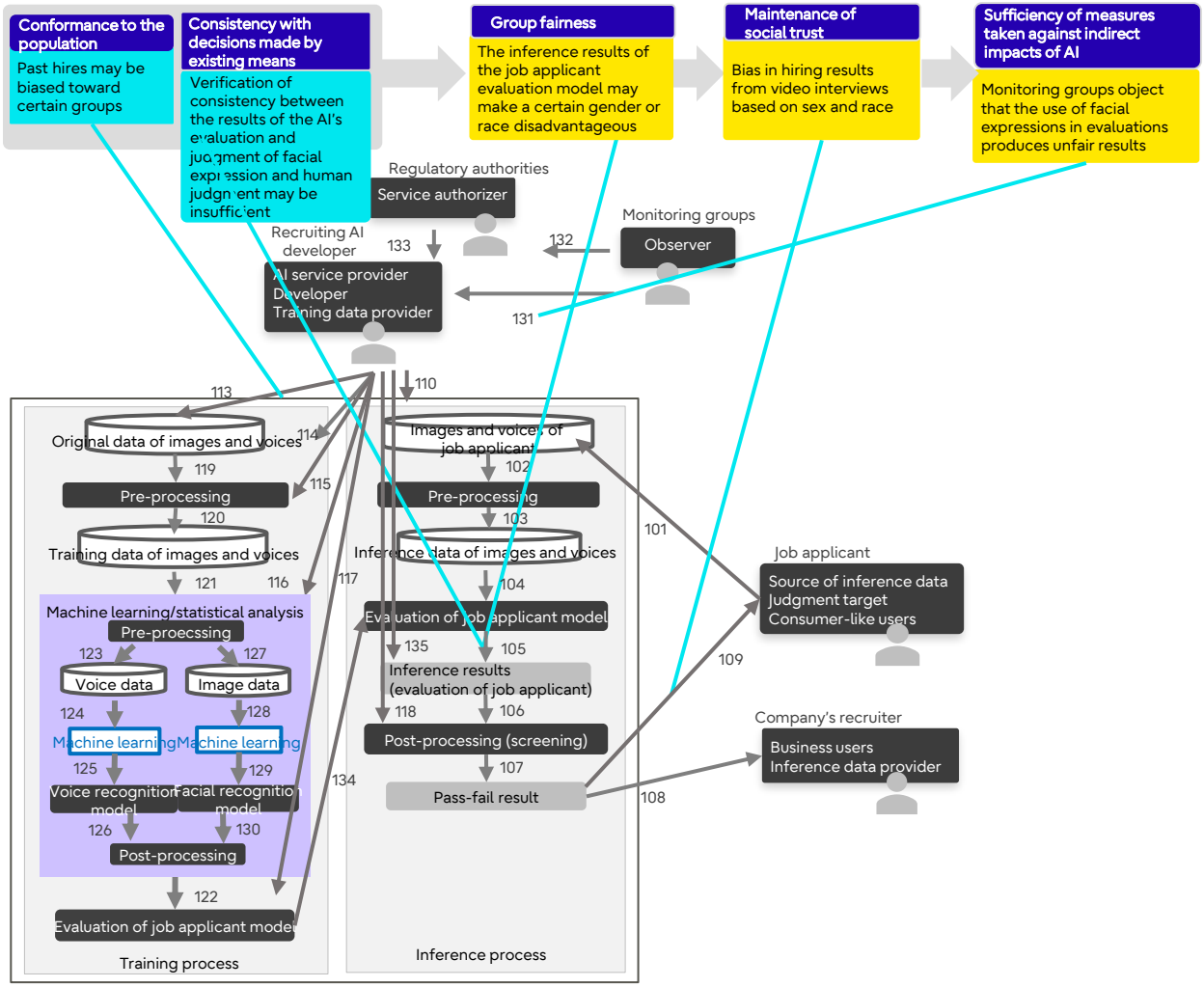
| Major Item | Medium Item | Details |
|---|---|---|
| Stakeholders and their roles | AI service provider | Recruitment AI service providers |
| | Developer | Recruitment AI service providers |
| | Business users | Corporate human resources departments |
| | Training data provider | Recruitment AI service providers |
| | Source of training data | Unknown |
| | Parties involved in training data acquisition | Unknown |
| | Inferential data providers | Job applicants |
| | Inferential data source | Job applicants |
| | Parties involved in the acquisition of inferential data | Unknown |
| | Consumer-like users | Job applicants |
| | Service UI/API provider | Recruitment AI service providers |
| | Judgment target | Job applicants |
| | Service authorizer | N/A |
| | Other stakeholders | Family members of job applicants, community |

# 5. Hiring decision AI for video interviews

| Major Item | Medium Item | Details |
|---|---|---|
| Presence/absence of a human-in-the-loop | | N/A |
| Presence/absence of existing methods | | Human judgment |
| AI Task | Task | Evaluating job applicant ` based on interview videos of job applicants. |
| | Problem Classification | Classification issues |
| | Output | Job applicant evaluation score |
| | Technology | Unknown |
| | Breakdown of models required to perform AI tasks | Facial expression analysis, acoustic analysis, and natural language processing |
| | Update/additional training of training data | Unknown |
| | Real-time performance | Yes |
| | Verification of AI tasks (consistency with decisions made by existing means) | Unknown |
| Training Data | Breakdown, provider and source | Facial expression data, voice data, conversation data (provider and source: Recruitment AI service provider) |
| | Teacher labels | Personality characteristics assessment |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | N/A |
| | Notes on data acquisition | Unknown |
| | Notes on data storage | Unknown |
| Inferential data | Breakdown, provider and source | Interview video (provider: business user's company, source: job applicant) |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | Yes |
| | Notes on data acquisition | Comply with privacy laws and regulations |
| AI System Output | Things to keep in mind during output | Unknown |
| Citations, reference articles | | https://incidentdatabase.ai/cite/95 |

# 5. Hiring decision AI for video interviews

## Analysis Chart



# 5. Hiring decision AI for video interviews

## Risks Consolidation

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| (1) The inference results of the job applicant evaluation model may make a certain gender or race disadvantageous | | Group fairness | 105 |
| | Verification of consistency between the results of the AI's evaluation and judgment of facial expressions and human judgment may be insufficient. | Consistency with decisions made by existing means | 105 |
| | Past hires may be biased toward certain groups | Conformance to the population | 113 |
| (2) Bias in hiring results from video interviews based on sex and race | | Maintenance of social trust | 109 |
| | Risk Event (1) | Group fairness | 105 |
| (3) Monitoring groups object that the use of facial expressions in evaluations produces unfair results. | | Sufficiency of measures taken against indirect impacts of AI | 131 |
| | Risk event (2) | Maintenance of social trust | 109 |

# 6. Loan screening AI

## Overview of the AI System

### Case 1

- AI System Usage Cases
  - In the loan screening process of banks, AI makes loan decisions that were previously made by loan officers.
- AI System Organization
  - Loan applicants enter the applicant information, loan amount, and repayment period into the loan screening AI service application.
  - Based on the input information, the loan screening AI obtains the applicant's transaction data and credit score at the bank, determines whether the loan is acceptable or not, and notifies the loan applicant of the result.
- Typical ethical issues
  - Racial and gender bias in AI financing results.

### Case 2

- AI System Usage Cases
  - In the loan screening process of banks, AI makes support decisions that were previously made by loan officers.
- AI System Organization
  - Loan applicants enter the applicant information, loan amount, and repayment period into the loan screening AI service application.
  - Based on the input information, the loan screening AI obtains the applicant's transaction data and credit score at the bank, determines whether the loan is acceptable or not, and notifies the loan officer of the result.
  - The loan officer makes a final decision on the loan based on the AI results and responds to the loan applicant.
- Typical ethical issues
  - Racial and gender bias in AI financing results.

# 6. Loan screening AI

## Use Case Overview (for both Case 1 and Case 2)

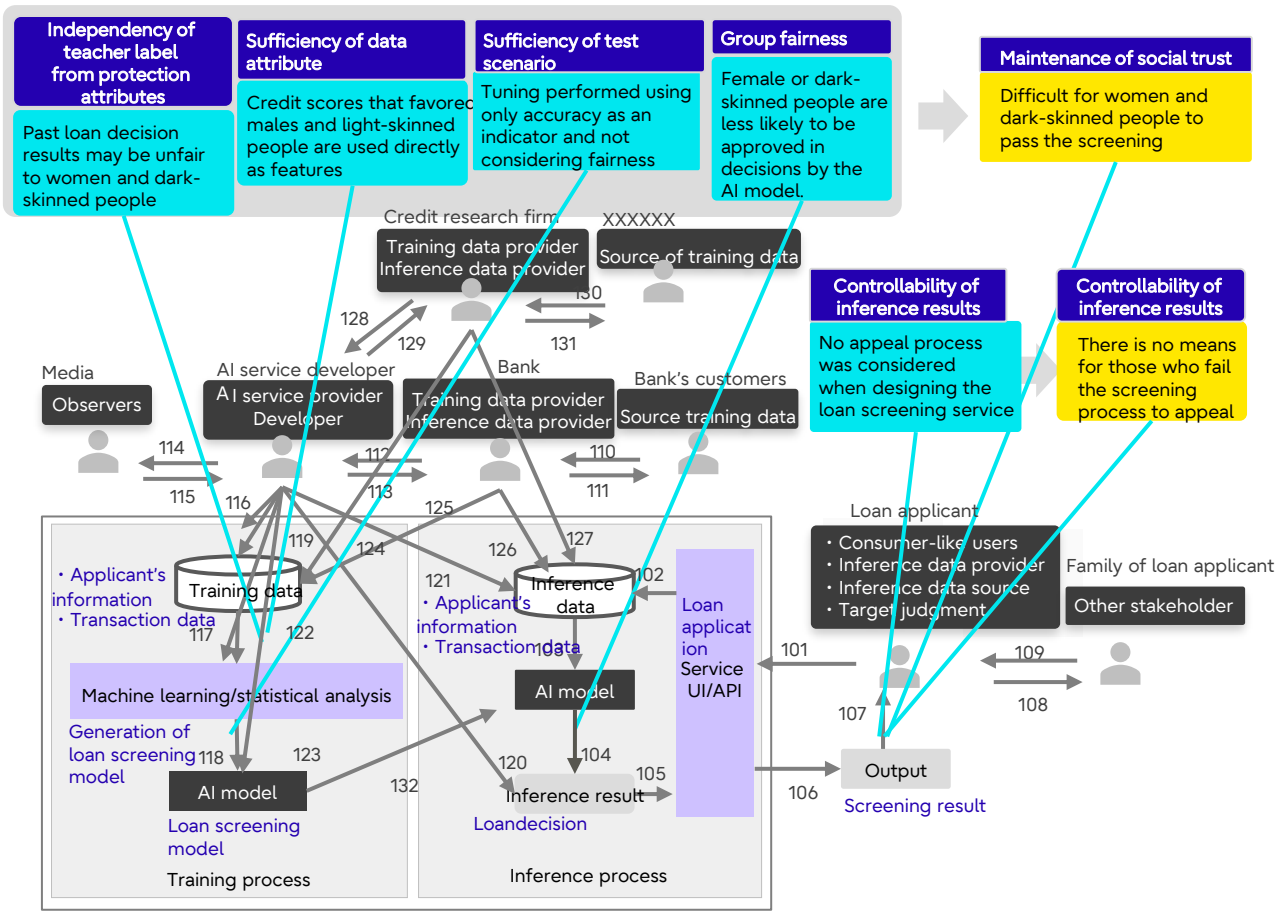
| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Finance and insurance industry |
| Purpose | | Reduced loan approval time |
| Service | Service Overview | Short Loan Review Service |
| | Availability of customization for each customer | N/A |
| | Requirements | N/A |
| Usage case | | Response to applicants with the screening results within a short period of time after the loan screening application is submitted. |
| Stakeholders and their roles | AI service provider | Banks |
| | Developer | AI service vendor |
| | Business Users | Banks |
| | Training data provider | Banks |
| | Source of training data | Banks |
| | Parties involved in training data acquisition | Bank customers |
| | Inferential data providers | Loan applicants |
| | Inferential data source | Loan applicants |
| | Parties involved in the acquisition of inferential data | Unknown |
| | Consumer-like users | Loan applicants |
| | Observers | Media, financial authorities |
| | Service UI/API provider | AI service vendor |
| | Judgment target | Loan applicants |
| | Service Authorizer | Financial authorities |
| | Other Stakeholders | Family members of loan applicants |
| Presence/absence of a human-in-the-loop | | Case 1: None (AI decision results are relayed directly to the loan applicant)<br>Case 2: Yes (Loan officers make a final decision based on the AI judgment results and reply to the loan applicant) |
| Presence/absence of existing methods | | Yes (to be determined by the loan officer) |

# 6. Loan screening AI

| Major Item | Medium Item | Details |
|---|---|---|
| AI Task | Task | Using machine learning and statistical analysis to determine whether a loan applicant is eligible for a loan or not |
| | Problem Classification | Classification issues |
| | Output | Loan availability/non-availability |
| | Technology | Independent + open source software |
| | Breakdown of models required to perform AI tasks | Loan screening model |
| | Update/additional training of training data | N/A |
| | Real-time performance | Yes |
| | Verification of AI tasks (consistency with decisions made by existing means) | Yes |
| Training Data | Breakdown, provider and source | Details of financing, transaction data (provider: bank, acquisition element: customer), credit score (provider: credit investigation agency, source: unknown) |
| | Teacher labels | Repayment record |
| | Protective attributes, attributes used for fairness verification and mitigation | Gender, age, race |
| | Presence/absence of personal information | N/A |
| | Notes on data acquisition | Unknown |
| | Data Storage | Unknown |
| Inferential data | Breakdown, provider and source | Application details, transaction data, credit score (provider and source: loan applicant) |
| | Protective attributes, attributes used for fairness verification and mitigation | Gender, age, race |
| | Presence/absence of personal information | Yes |
| | Notes on data acquisition | Unknown |
| AI System Output | Things to keep in mind during output | Unknown |
| Citations, reference articles | | ISO IEC TR 24030:2021 Artificial Intelligence(AI) - Use cases, https://www.iso.org/standard/77610.html |

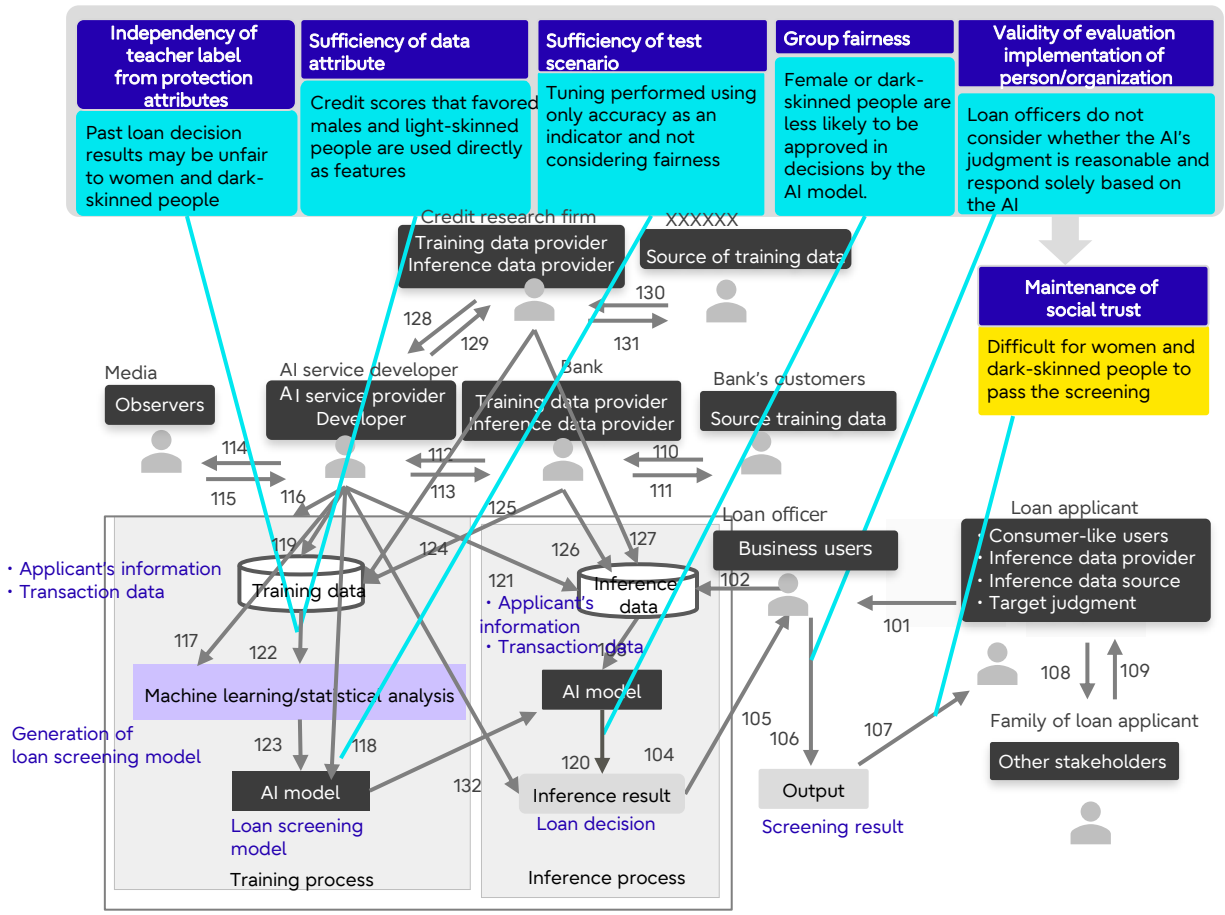
# 6. Loan screening AI

## Analysis Chart Case1



# 6. Loan screening AI

## Analysis Chart Case2



# 6. Loan screening AI

## Risks Consolidation for Case1

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| Difficult for women and dark-skinned people to pass the interview stage | | Maintenance of social trust | 107 |
| | Female or dark-skinned people are less likely to be approved in decisions by the AI model | Group fairness | 104 |
| | Tuning is performed using only accuracy as an indicator and not considering fairness | Sufficiency of test scenarios | 118 |
| | Credit scores that favored males and light-skinned people are used directly as features | Sufficiency of data attribute | 122 |
| | Past loan decision results may be unfair to women and dark-skinned people | Independency from teacher label protection attributes | 122 |
| There is no means for those who fail the screening process to appeal | | Controllability of inference results | 107 |
| | No appeal process was considered when designing the loan review service | Controllability of inference results | 107 |

## Risks Consolidation for Case2

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| Difficult for women and dark-skinned people to pass the interview stage | | Maintenance of social trust | 107 |
| | Female or dark-skinned people are less likely to be approved in decisions by the AI model | Group fairness | 104 |
| | Tuning is performed using only accuracy as an indicator and not considering fairness | Sufficiency of test scenarios | 118 |
| | Credit scores that favored males and light-skinned people are used directly as features | Sufficiency of data attribute | 122 |
| | Past loan decision results may be unfair to women and dark-skinned people | Independency from teacher label protection attributes | 122 |
| | Loan officers do not consider whether the AI's judgment is reasonable and respond solely based on the AI | Validity of evaluation implementation of person/organizations | 106 |

# 7. Fruit grading

- AI System Usage Cases
  - Using an AI that determines the grade of fruit based on an image of it: an agricultural organization determines the grade of fruit collected from farmers and pays the farmers according to the grade
- AI System Organization
  - Training section: The grading AI model is generated by fruit images for training.
  - Prediction section: Fruits collected from farmers are transported by conveyor belts, and an image is taken of each fruit, and the AI determines its grade. The judgment results are aggregated for each farmer in post-processing to determine the payment for each farmer.
- Typical ethical issues
  - Possibility for agricultural organizations to analyze the data of grading results and infer the ranking of farmers without the agreement of farmers

## Use Case Overview

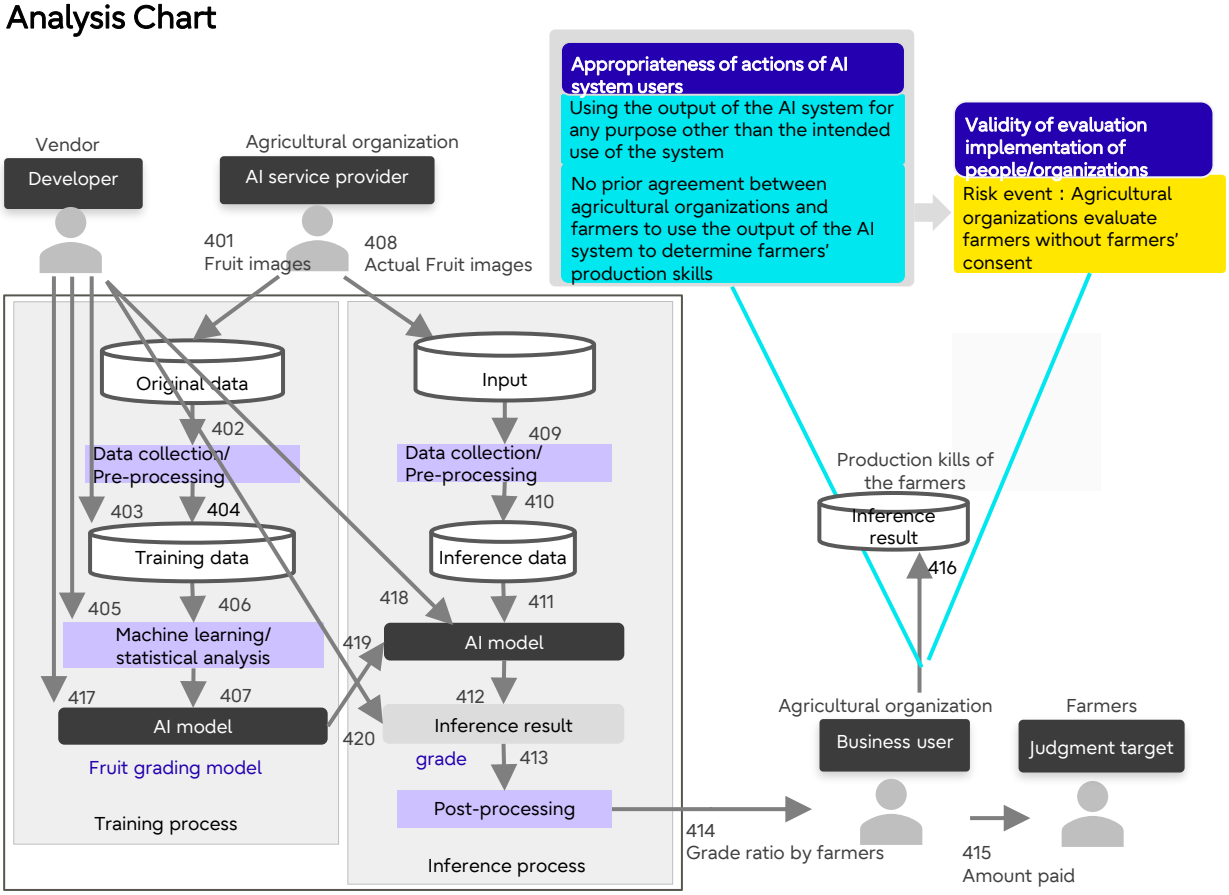
| Major Item | Medium Item | Details |
|---|---|---|
| Industry | | Agriculture |
| Purpose | | Determination of the amount to be paid to farmers |
| Service | Service Overview | Agricultural organizations use fruit grading AI to determine the amount of payment for each farmer. |
| | Availability of customization for each customer | N/A |
| | Requirements | N/A |
| Usage case | | An AI system determines the grade of fruit collected from farmers.The amount paid to farmers is determined from the proportion of each grade in the farmer's harvest. |
| Stakeholders and their roles | AI service provider | AI development vendor |
| | Developer | AI development vendor |
| | Business users | Agricultural organizations |
| | Training data provider | Agricultural organizations |
| | Source of training data | Farmers |
| | Parties involved in training data acquisition | N/A |
| | Inferential data providers | Agricultural organizations |
| | Inferential data source | Farmers |

# 7. Fruit grading

| Major Item | Medium Item | Details |
|---|---|---|
| Stakeholders and their roles | Parties involved in the acquisition of inferential data | N/A |
| | Consumer-like users | Farmers |
| | Observers | N/A |
| | Service UI/API provider | N/A |
| | Judgment target | Farmers |
| | Service authorizer | N/A |
| | Other stakeholders | N/A |
| Presence/absence of a human-in-the-loop | | N/A |
| Presence/absence of existing methods | | Yes |
| AI Task | Task | Determining the grade from an image of a fruit |
| | Problem Classification | Classification |
| | Output | Grade |
| | Technology | Independent + open source software |
| | Breakdown of models required to perform AI tasks | Fruit grading model |
| | Update/additional training of training data | N/A |
| | Real-time performance | Yes |
| | Verification of AI tasks (consistency with decisions made by existing means) | Grade identification by experienced workers from agricultural organizations |
| Training Data | Breakdown, provider and source | Fruit image (provider and source: agricultural organizations) |
| | Teacher labels | Fruit grade |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | Farmers who shipped fruit |
| | Notes on data acquisition | N/A |
| | Notes on data storage | N/A |
| Inferential data | Breakdown, provider and source | Fruit image (provider: agricultural organizations source: farmers) |
| | Protective attributes, attributes used for fairness verification and mitigation | N/A |
| | Presence/absence of personal information | Farmers who shipped fruit |
| | Notes on data acquisition | N/A |
| AI System Output | Things to keep in mind during output | N/A |
| Citations, reference articles | | N/A |

# 7. Fruit grading

## Analysis Chart




## Risks Consolidation

| Risk Event | Risk Factors | AI Ethical Characteristics | Interaction ID |
|---|---|---|---|
| Agricultural organizations evaluate farmers without farmers' consent | | Validity of evaluation implementation of people/organizations | 416 |
| | Using the output of the AI system for any purpose other than the intended use of the system | Appropriateness of actions of AI system users | 416 |
| | No prior agreement between agricultural organizations and farmers to use the output of the AI system to determine farmers' production skills | Appropriateness of actions of AI system users | 416 |

# References


[1] https://incidentdatabase.ai/cite/16

[2] https://incidentdatabase.ai/cite/37

[3] https://www.eismd.eu/wpcontent/uploads/2019/11/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society_compressed.pdf

[4] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[5] https://oecd.ai/en/ai-principles

[6] https://ethicsinaction.ieee.org/

[7] https://www.soumu.go.jp/main_content/000637097.pdf

[8] https://www8.cao.go.jp/cstp/aigensoku.pdf

[9] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN

[10] https://partnershiponai.org/

[11] https://incidentdatabase.ai/

[12] https://www.iso.org/standard/77610.html

[13] https://www.jil.go.jp/kokunai/statistics/databook/2019/03/d2019_T3-01A.pdf

© FUJITSU LIMITED 2022
This material is licensed by Creative Commons under the following terms:.
No Modification. 4.0 International (CC BY-ND 4.0)
For more information about licensing terms, see the following site:.
http://creativecommons.org/licenses/by-nd/4.0/