Risk Framework for the Use of AI Services Driven by Citizens Themselves

Takashi Matsumoto[1,2](✉), Mika Kimura[2], Teruka Sumiya[3,4], and Tomoyo Sasao[1]

[1] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
takashi2.matsumoto@tohmatsu.co.jp
[2] Deloitte Tohmatsu Consulting LLC, 3-2-3 Marunouchi, Chiyoda-ku, Tokyo, Japan
[3] C4IR Japan, World Economic Forum, , 1-12-32 Akasaka, Minato-ku, Tokyo, Japan
[4] Pnika, 4-20-20 Onta, Higashimurayama, Tokyo, Japan

Abstract. As the issues and needs faced by citizens become more diverse and complex, there are high expectations for the utilization of machine learning AI. However, AI with complex logic is difficult for humans to interpret, and there are concerns about various risks such as the impact of environmental changes, fairness, and accountability. In addition, due to the uncertainty of the AI model, it is difficult for the AI model alone to sufficiently cope with the risks, and countermeasures in cooperation with related technologies and non-technologies are necessary. In the development of AI services for citizens' daily lives, AI developers are required to consider countermeasures by encouraging participation of citizens in order to understand various issues and concerns that citizens may have. In addition, the knowledge of various experts is also needed to examine issues related to technology, safety, legal compliance, and ethics. In this study, in the process of considering a new AI service in the living lab, we use the Risk Chain Model, which is a risk analysis framework with citizens and experts to examine whether the risk scenario and risk control for AI service can be sufficiently considered.

Keywords: Human-Centered Artificial Intelligence · AI · Smart City · Citizen · Living Lab · Risk Management

1 Introduction

As people's lifestyles diversify, the issues and needs of citizens become more diverse and complex, and there are high expectations that AI can be used to solve increasingly complex problems. Machine learning AI, which is currently widely applied, has the potential to predict solutions to complex problems with high accuracy by optimizing algorithms using large amounts of training data, which has been difficult with conventional technologies. Machine learning AI is used in many fields. The logic of AI model is difficult for humans to understand and sometimes outputs unexpected results, so it is necessary to pay attention to risks that have not been emphasized by conventional technologies [1, 2]. For example, the risk that the prediction performance of AI model is decreased due to the change in data distribution by environment changes; the risk of

N. A. Streitz and S. Konomi (Eds.): HCII 2023, LNCS 14037, pp. 97–117, 2023.
https://doi.org/10.1007/978-3-031-34609-5_7

making unfair judgments about certain groups; the risk of making significant errors in judgment due to minute information that humans cannot recognize; lack of interpretability of AI decisions, which does not convince stakeholders, etc. are concerns [3, 4]. In fact, there have been cases in which significant economic losses have been incurred as a result of the use of AI. Zillow, a real estate brokerage marketplace in the U.S., was using a service that utilizes AI to make real estate transaction decisions, but while real estate prices skyrocketed due to the pandemic, the AI continued to make real estate transaction decisions based on past transaction data, causing significant losses to the company [5]. On the other hand, risks to human-rights and ethics by AI are also becoming apparent. In the United States, COMPAS, an AI that predicts recidivism rates, became a major issue when it was pointed out that racial judgments had a strong influence on the calculation of recidivism risk [6]. In cases where AI is used to score creditworthiness for citizens, there are concerns about the lack of accountability of AI model [7] and the risks associated with the unauthorized purpose use of calculated credit scores [8]. In addition, AI model itself is fraught with uncertainty, making it difficult for AI model alone to adequately address all risks and requires coordination with measures in related technologies and non-technologies [9].

The development and utilization of AI may affect not only the users of AI, but also various stakeholders, such as the target of AI prediction, workers collaborating with AI, and providers of learning data, etc. Therefore, AI developers should communicate with multi-stakeholders and consider various impacts while developing AI [10]. Therefore, it is expected that development projects should encourage citizen participation when developing AI services that affect the lives of citizens. In recent years, open innovation has been experimented, in which citizens take the initiative in developing technologies and collaborate with various stakeholders to expand the implementation of AI service. In Barcelona, citizens have taken the initiative in building “Guifi.net”, a decentralized and managed network infrastructure, and in developing the “Smart Citizen Kit a distributed management network infrastructure [11].

While citizens are expected to participate in AI development projects, there are also issues related to technology, safety, legal compliance, and ethics in the development and utilization of AI, which require the knowledge of various experts [10]. In this study, we examined whether citizens can sufficiently consider various risks related to AI service with experts by using the Risk Chain Model, a risk analysis framework, in the case of citizen-led prototyping AI service. The Risk Chain Model is a framework in which risk scenarios affecting AI service are discussed with various stakeholders, and AI system, service providers, and users cooperate to study risk control [9]. Although case studies have been conducted on various use cases developed by enterprises, no studies have been conducted on use cases developed by citizens. In the workshop for citizen-led use case studies using AI cameras, citizens will be able to develop various risk scenarios affecting various stakeholders with experts by utilizing the Risk Chain Model, a risk analysis framework, at the living lab.

2 Related Works

Norbert organizes the following three challenges for AI technology; impossible and error-prone behavior, robustness, and lack of transparency, traceability, and accountability [12]. While the first two could be solved through technological advances, the last challenge is expected to remain [13]. Privacy considerations are essential for the use of big data in smart cities [14], but in order to enjoy the smartness of technology and develop a city that meets the needs of its citizens, citizens themselves should have the right to control what and how their data is collected and processed It has been suggested that they should have the right to control what data is collected and how it is processed [13]. In Sidewalk in Toronto, Canada, a closed forum for citizens decided how the technology would be used, and the plan was cancelled due to citizen opposition [15]. Without mechanisms for public participation, the crisis of trust will only deepen [16].

The importance of citizen involvement from the stage of city planning has been discussed in various ways [17–20]. On the other hand, the difficulty for citizens to work independently has also been pointed out [18, 21]. It is also believed that the involvement of various stakeholders allows ideas to be discussed from multiple perspectives [22] and reduces decision-making errors [16]. In order to achieve co-creation among stakeholders, it is noted that it is important to involve each stakeholder from the early stages of the project [23, 24].

3 Approach

3.1 Scope

In this study, we focus on a case in which a new AI service is examined by citizens in a workshop conducted by the living lab. In the process of studying AI services related to the daily life of citizens, we utilize the Risk Chain Model, a risk analysis framework, to examine whether it is possible to recognize various risk scenarios and to study technological and non-technological risk control by using the knowledge of both citizens and experts.

3.2 Risk Chain Model

As a risk analysis framework for AI services, we used the Risk Chain Model developed by the Institute for Future Initiatives of the University of Tokyo [9], in which the author of this paper also participates. Many research institutes [25–28] and AI development companies [29–31] have developed various tools and frameworks. However, most of them are intended to be used by developers, and require an understanding of specialized knowledge related to AI in order to use them. In addition, there is little public information on the examples of their use. The Risk Chain Model is a framework to consider risk controls for various risks associated with AI services by linking AI System/Service Provider/User [9]. It is characterized to examine risk controls in collaboration with various stakeholders, not limited to data scientists, by linking technological and non-technological components in a risk chain. Guidebooks and case studies are available so that non-AI developers can also conduct the study. As of February 2023, 12 case studies have been published, including Recruitment AI (Case01), Unstaffed Convenience Store (Case02), Verification of Recidivism Possibility AI (Case06), Driver-less Bus (Case10), etc. [32]. The Risk Chain Model is studied in the following steps.

Consider a Use Case. Before conducting the study, outline the use case of the AI service. At this stage, the values and objectives, AI model, the prediction target, the users, and the usage are clarified. For example, when considering the use case of an AI camera to be used in the city, the information shown in Table 1 should be prepared as an outline of the use case.

Table 1. Use Case Overview (Case: AI camera)

Values and objectives	AI model	Prediction target	User	Usage
• Safety of citizens • Comfortable urban environment • Reduction of security guard workload • Secondary use of data	Real-time human pose estimation	Citizen poses (cowering, violent behavior, possession of weapons, entering a restricted area)	Security guard	Notify the user when the pose of the predicted target is recognized

Risk Assessment. Identify risk scenarios that affect the use and operation of AI services. Risk scenarios are not limited to risks that may hinder business objectives, but also include ethical risks. The risk targets are not limited to AI users and forecasters, but also include operators, workers collaborating with the AI service, learning data providers, and others. Therefore, it is desirable to consider risk scenarios together with various stakeholders. For example, if risk scenarios are considered in the use case of AI cameras, risk scenarios affecting citizens, service providers, security guards, and workers are identified as shown in Table 2.

Risk Control. Risk control is examined using the map of the Risk Chain Model (See. Figure 1). There are three colored areas in the map. The green area is AI system, which includes AI model, data, system infrastructure, and other control functions. The blue area is service provider, which includes code of conduct, service operation, and communication with users. The orange region is user, which includes understanding of AI service, utilization of AI service, and usage environment. Using one map for each risk scenario, identify the relevant components (white boxes placed in each region) and draw a risk chain (red line), considering the order in which the risks are manifested. The risk chain does not necessarily consist of a single line, but may branch or aggregate, and may be a loop structure.

Table 2. Risk Scenario (Case: AI Camera)

Risk Scenarios	Affected Person
1. Some places do not achieve the expected effect	Service provider Citizens, Security guard
2. Data is not utilized	Service provider
3. Malfunction or data leakage due to cyber attacks	Service provider
4. Destruction of cameras	Service provider
5. Missed physical or violent behavior	Citizens
6. Misrecognition of normal behavior	Citizens
7. Recognition errors due to noise	Service provider Security guard
8. Recognition errors due to changes in the city environment	Service provider, Citizens, Security guard
9. Use of data for other purposes	Citizens
10. Invasion of privacy	Citizens
11. Obstruction of the landscape	Citizens, Workers
12. Deterioration of living conditions due to excessive surveillance	Citizens, Workers
13. Falling of cameras	Citizens
14. Deterioration of public safety in areas where cameras are not present	Citizens, Workers
15. Overloading of security guards due to excessive detection	Security guard
16. Missed incidents due to security guards' dependence on AI	Security guard

Risk controls are considered for each element associated in the risk chain. For each risk scenario, multiple technological and non-technological risk controls are associated with each risk scenario, so that a stepwise risk reduction can be considered. Since the contents of the study may be biased by the experience of the participants, it is desirable to conduct a role play including stakeholders who have knowledge in each area of AI System/Service Provider/User. In the use case of the AI camera, when examining the risk chain of the risk scenario "1. Some places do not achieve the expected effect", the risk control is identified as shown in Table 3.

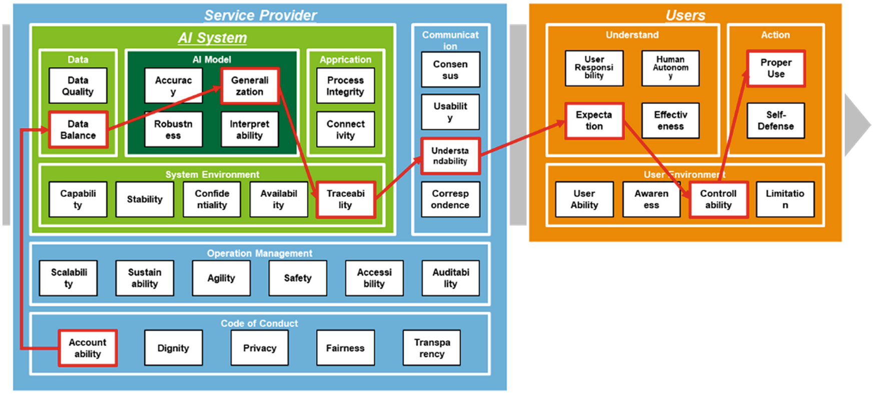

Fig. 1. Risk Chain Model

Table 3. Risk Controls (Risk Scenario: 1. Some places do not achieve the expected effect)

Risk Scenario	Risk Controls (AI System)	Risk Controls (Service Provider)	Risk Controls (Users)
1. Some places do not achieve the expected effect	(3) Ensure the quality of training data (Data Quality) (4) Ensure generalization of AI model for location (Generalization) (5) Recording of AI decision results (Traceability)	(1) Ensure that benefits are not biased by locations (fairness) (2) Clarify expectations for AI service (Accountability) (6) Monitoring AI performance (Auditability) (7) Explanation of expected effects to citizens (Consensus) (11) Improvement of AI model (Sustainability) (12) Development of individual AI models for locations (Capability)	(8) Explanation of the expected effects of AI cameras (Expectation) (9) Preparation of means to provide opinions to the management (Controllability) (10) Citizens can submit their opinions to the operator (Proper-Use)

3.3 Living Lab - Urban Design Studio for Everyone of Urban Design Center Kashiwa-No-Ha (UDCK)

Urban Design Center Kashiwa-no-ha (UDCK), based in Kashiwa City, Chiba Prefecture, Japan, operates a living lab called "Urban Design Studio for everyone" [33]. Based on this living lab, we conducted a workshop to propose three AI services using AI cameras by citizens for the purpose of solving problems in the city. A group of citizens will discuss a use case. Prototyping and user surveys were conducted to the extent technically feasible to improve the feasibility of the use case, and the use case was finally presented in a public forum including online distribution.

3.4 Method of This Project

The Risk Chain Model was utilized in the prototyping of each use case conducted in the living lab workshop. The results of each use case were qualitatively evaluated in terms of “Various risk scenarios” and “Technical and non-technical risk control”.

Various Risk Scenarios. For each use case, risk assessment (3.2) was conducted to qualitatively evaluate the risk scenarios including safety, economic efficiency, human rights (fairness and privacy), etc. Comprehensiveness of risk scenarios (1) and Diversity of stakeholders exposed to risk (2) were qualitatively evaluated. In addition, we evaluated the existence of risk scenarios recognized through the knowledge of citizens or experts (3) who participated in the discussion.

Technical and Non-technical Risk Control. The risk controls were identified in each use case using a risk chain (3.2). We evaluated whether risk control is considered in each domain of AI System (4)/Service Provider (5)/Users (6) without relying on a single technological element. In addition, we evaluated the existence of risk control which does not exist in the existing case and is considered by the citizens (7).

4 Experiment

4.1 Workshop

The workshop was conducted from May to September 2022 (Table 4) at UDCK.

Lecture (May 14 2022 to Jun 04 2022). Before examining use cases, a three-day lecture was held for workshop participants. The participants examined the problems of the city (Day1) and the expectations of the AI camera (Day2) using LEGO Serious Play. In Day3, we conducted a study using the Risk Chain Model for AI cameras in order to learn how to use the Risk Chain Model.

Group Work: Use Case Ideation (Jun 04 2022 to Jul 22 2022). After three days of lectures, the participants were organized into groups based on the similarities in the issues and expectations of AI cameras presented by each participant on Day 1 and Day 2. Each group started to consider use cases led by the participants. Three use cases, “Recommendation of walking routes for pets,” “Visualization of the excitement of events in the city,” and “Signage that collects and visualizes Signage that collects and visualizes the good and bad points of the city” were discussed.

Interim Presentation (Jul 23 2022). The three groups presented the challenges and proposed solutions for the use case. Experts in the field of smart city and technology governance (promoters of initiatives in other local governments, project promoters at the World Economic Forum, consultants, university researchers, Kashiwa City officials, and area management companies) provided feedback on each group's presentation. The feedback included ideas on the problem setting, personas, and realization methods, as well as reference cases. General citizens participated as audience members, and the presentations were streamed on online.

Group Work: Prototyping Use Case (Jul 24 2022 to Sep 16 2022). Each group improved the use case based on the comments made at the interim presentation, and conducted prototyping. The prototypes were created as web applications, etc., and improvements were made based on questionnaires to test users. In this prototyping phase, risk scenario and risk control were examined using the Risk Chain Model.

Final Presentation (Jul 23 2022). The three groups presented the final contents of their use cases. The proposals included the target problem, solution, contents of prototyping (including survey results), important risk scenarios considered in the Risk Chain Model, and risk control (functions to be incorporated into service operation). As with the interim presentations, feedback from experts was provided, and the presentations were streamed online with the participation of the general audience.

Table 4. Workshop Schedule

Date	Activities	Location
May 14 2022	Kick off Lecture Day1: LEGO Serious Play	UDCK
May 28 2022	Lecture Day2: LEGO Serious Play	UDCK
Jun 4 2022	Lecture Day3: How to use Risk Chain Model	UDCK
From Jun 4 2022 to Jul 22 2022	Group work: Use case ideation	Online and UDCK
Jul 23 2022	Interim presentation	UDCK
From Jul 24 2022 to Sep 16 2022	Group work: Prototyping (*)	Online and UDCK
Sep 17 2022	Final Presentation	UDCK

* Conducted analysis using Risk Chain Model

4.2 Use Case

The workshop participants were divided into groups to examine use cases. In the study using the Risk Chain Model, the study is conducted as a role play among AI System/Service Provider/User, which requires at least four participants per use case, including one facilitator. Although the number of use cases has not been strictly considered, we decided to consider multiple use cases. A total of 13 people including citizens, workers, and students of Kashiwa-no-ha participated in the workshop. Therefore, the workshop participants were organized into three groups of at least four persons in each group (Table 5). The workshop management team supported the promotion of each group as needed.

Group A: Recommendation of Walking Routes for Pets. This use case was examined under the theme of making a friendly town where people can live without worrying about rules and manners. To avoid stressful encounters between pet owners who walk their dogs in the Kashiwa-no-ha area and citizens who are not good with dogs, AI service estimated comfortable dog walking routes for both citizens based on information obtained by AI cameras, and displayed on a smartphone application, which provides a recommendation of walking routes for pet owners. The AI camera acquires data on location, time, temperature, humidity, surface temperature, brightness, and human flow, and estimates appropriate walking courses and times.

Group B: Visualization of the Excitement of Events in the City. This use case was examined under the theme of enabling citizens to feel the excitement of the city by using AI cameras installed in Kashiwa-no-ha area to recognize the emotions of people gathering in the area and displaying icons and event information on a map of Kashiwa-no-ha area on the Internet. AI cameras at each location recognize people, strollers, bicycles, etc., and analyze the emotions of the people based on their facial expressions, etc., to express the fun of each location (many people in high spirits, parents and children with children gathered, active exercise, etc.).

Group C: Signage that Collects and Visualizes the Good and Bad Points of the City. This use case was examined with the theme of enabling citizens to recognize the good and bad points of the city and to work together to improve the city. AI classification model classified into the positive information (e.g., seasonal flowers have bloomed) and negative information (e.g., roads and facilities are broken) collected in the city recognized by AI cameras installed in the Kashiwa-no-ha area recognize and posted by citizens that submit photos and free text via smartphones, etc. and visualized on signage and websites posted in the city.

4.3 Risk Chain Model in This Workshop

Each group gathered at UDCK to discuss the Risk Chain Model. The research team of the Risk Chain Model at the University of Tokyo, participated in the discussion and gave advice on how to use the model in response to the participants' questions.

The risk assessment was conducted with researchers on AI governance, researchers on smart cities, and personnel from area management companies in each group. Specifically, participants raised risk scenarios on post-it notes, and risk scenarios to be identified were organized by grouping them (See. Figure 2).

When discussing risk control, it is desirable to include participants who have knowledge of AI System/Service Provider/User. However, it is assumed that it may be difficult to cover all the participants with knowledge in some groups, so we created our own 47 cards as candidates for risk control, referring to the case study [32] of the Risk Chain Model published by the University of Tokyo. When conducting the study using the Risk Chain Model, participants at each table are divided into the roles of AI System/Service

Table 5. Group Overview

Group	Theme	Number of people	Use Case	Role of AI camera
A	Making a friendly town where people can live without worrying about rules and manners	4	Recommendation of walking routes for pets	Understanding Dog Behavior in the City
B	Enabling citizens to feel the excitement of the city	5	Visualization of the excitement of events in the city	Emotional analysis through images
C	Enabling citizens to recognize the good and bad points of the city and to work together to improve the city	4	Signage that collects and visualizes the good and bad points of the city	Recognize images corresponding to good and bad points of a city

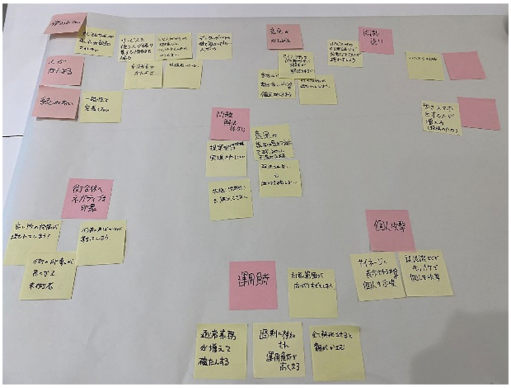

Fig. 2. Risk Assessment

Provider/User and role-play. Participants discussed risk control, referring to the cards as necessary. After that, the risk control cards were placed on the mat (See. Figure 3) on which the map of the Risk Chain Model was outputted, and the risk control of technical and non-technical risks were related by drawing the risk chain (See. Figure 4). When a risk control that did not exist on the prepared cards was considered, we placed it on a post-it and linked it to other risk controls by drawing a risk chain.

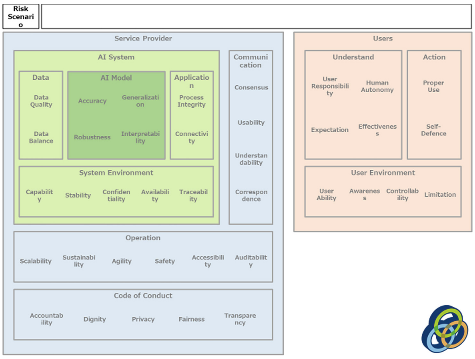

Fig. 3. Risk Chain Model map for the workshop

Fig. 4. Workshop

5 Results

5.1 Risk Scenarios Identified in Use Cases

Finally, for each use case, the risk scenarios listed in Table 6 were recognized. In order to examine the comprehensiveness of risk scenarios, we classified each risk scenario based on its contents. The classification items are based on the Dimensions of Trustworthy AI described in the book "Trustworthy AI" by Beena Ammanath [3]: fair and impartial (Fair), robust and reliable (Robust), respectful of privacy (Privacy), safe and secure (Safe), responsible and accountable (Responsible), and transparent and explainable (Transparent). Risks related to economic loss and satisfaction of citizens are classified as (Responsible).

Group A: Recommendation of Walking Routes for Pets. Five risk scenarios were identified in this group. The participants in this group were pet owners, and the risk "1. Dangerous induction" was raised due to dog allergies, contact with humans (especially with children), and so on. In addition, a researcher of AI Governance who participated in the discussion commented on the need to modernize appropriate walking routes if there are changes in streets and facilities, and "3. Inability to cope with changes in the urban environment" was recognized. The "4. Division of citizens" was discussed as a risk to pet owners and dog-phobic citizens who are fixed from the walking course. Other risks to the operator were identified as “2. Inaccurate information leading users away” and “5. Maintenance burden such as updating equipment and information”.

Group B: Visualization of the Excitement in the City. Nine risk scenarios were identified in this group. “1. Failure to recognize citizens and damage the reputation of the event” was raised as a risk for AI recognition performance. Fairness in terms of appearance, location, organizers, etc. was also a concern, and “3. Smiles are not recognized due to differences in appearance”, “4. Unfair places”, “5. Unfair representation of excitement by organizers” were raised. A data scientist who participated in the discussion commented on a hostile case [4] in which AI makes different judgments based on minute information that humans cannot see, and “2. Adding noise to images to make certain events appear less exciting” was discussed. “7. Crowding into events that limit the number of people” was identified as a risk to event participants. “6. Negative impact on neighborhoods due to excessive crowding”, “8. People are forced to participate in the event” and “9. Visualization of citizens’ private life in the city” were also identified.

Group C: Signage that Collects and Visualizes the Good and Bad Points of the City. Six risk scenarios were identified in this group. In this case, citizens submit issues, which are then classified by AI and visualized by citizens. Therefore, the risks associated with building a relationship with citizens are “2. Issues are not resolved and citizens’ dissatisfaction grows” and “5. Important issues are not recognized due to misclassification of positive/negative”. “3. Issues of certain groups are not recognized due to the bias of users” and “4. Issues posted become personal attacks” were identified as risks related to the fairness and uncertainty of AI. “1. Negative impression of the city” and “6. Excessive recognition of issues increases the load on the local government” were also identified as city management.

5.2 Risk Controls Identified in Use Cases

Each group examined the risk control corresponding to the risk scenario using the Risk Chain Model. Table 7 shows the results of selecting the most important functions in the operation of AI service from the identified risk controls.

Group A: Recommendation of Walking Routes for Pets. In the AI System, it is considered to perform “Adjustment of bias of training data (dog breed, etc.)” and “Output of decision basis (number of dogs, breeds, etc.)” so that various breeds can be judged fairly. And “Correction of output results (walking course)” when there is an accident on the walking course recommended by AI. In the Service Provider, a team for “Coordination of event information in town and at facilities” should be organized, and “Consideration of individual models for each location” should be performed for locations where the prediction accuracy is clearly different. For users (citizens), the smartphone application will provide “Registration of pet information (breed, allergies, etc.)” and “Visualization of hazard information (construction, congestion, allergic reaction, etc.) on walking courses” in the smartphone application.

Group B: Visualization of the Excitement in the City. In the AI System to be able to appropriately identify various persons, including those wearing or not wearing masks,

Table 6. Risk Scenario (Use cases)

Gr	Use Case	Risk Scenarios	Affected Person	Classification
A	Recommendation of walking routes for pets	1. Dangerous induction (dog allergy outbreak, contact with humans or other animals in close quarters)	Citizens, Pet	Safe
		2. Inaccurate information leading users away	Service provider	Responsible
		3. Inability to cope with changes in the urban environment	Citizens, Pet	Robust
		4. Division of citizens (pet owners or citizens who do not like dogs)*	Citizens	Fair
		5. Maintenance burden such as updating equipment and information	Service provider	Robust
B	Visualization of the excitement in the city	1. Failure to recognize citizens and damage the reputation of the event	Event Organizers	Robust
		2. Adding noise to images to make certain events appear less exciting*	Event Organizers	Robust
		3. Smiles are not recognized due to differences in appearance	Event participants	Fair
		4. Unpopular places	Citizens	Fair
		5. Unfair representation of excitement by organizers	Event Organizers	Fair
		6. Negative impact on neighborhoods due to excessive crowding	Citizens	Responsible
		7. Crowding into events that limit the number of people	Event participants	Safe

(*continued*)

Table 6. (*continued*)

Gr	Use Case	Risk Scenarios	Affected Person	Classification
		8. People are forced to participate in the event	Citizens	Responsible
		9. Visualization of citizens' private life in the city	Citizens	Privacy
C	Signage that collects and visualizes the good and bad points of the city	1. Negative impression of the city as a whole	Government	Responsible
		2. Issues are not resolved and citizens' dissatisfaction grows	Service providers	Responsible
		3. Issues of certain groups are not recognized due to the bias of users	Citizens	Fair
		4. Issues posted become personal attacks	Citizens	Privacy
		5. Important issues are not recognized due to misclassification of positive/negative	Citizens	Robust
		6. Excessive recognition of issues increases the load on the local government	Service providers	Responsible

* Risk scenarios considered by incorporating the opinions of non-citizens experts

in various event environments, it was considered to have functions such as "Camera noise cancellation", "Sufficient training data (presence/absence of masks, etc.)", and "Ensure generalization performance of AI model" as well as "Detection of crowding.". For the Service Provider, "Privacy and fairness policies" and "Setting a maximum number of dense people per location" were considered in the operation, and "Description of AI camera locations, detection details" was discussed for event organizers and participants. For users (citizens, event organizers), "Web-based publication of event excitement and crowding" and "Complaint window from Neighbors" and "Notification to event organizers and participants" were considered.

Group C: Signage that Collects and Visualizes the Good and Bad Points of the City. In the AI System, the function to perform "Ensure generalization performance of AI model", "Record decision results and posting frequency", and "Output high frequency was considered. In the Service Provider, "Positive/negative and non-biased output expressions" are performed, and "Verification of contributor bias/high frequency keywords/retention issues" in signage are performed. In addition, "Publicity at facilities/events" was considered to reduce the bias of contributors. For users (citizens), "Collection of opinions from citizens" including the design of SNS hashtags was considered necessary to recognize issues that should be prioritized.

Table 7. Service Management Functions

Gr	Use Case	AI System	Service Provider	Users
A	Recommendation of walking routes for pets	- Adjustment of bias of training data (dog breed, etc.) - Output of decision basis (number of dogs, breeds, etc.) - Correction of output results (walking course)	- Coordination of event information in town and at facilities - Monitoring of AI decision results - Consideration of individual models for each location	- Registration of pet information (breed, allergies, etc.)* - Visualization of hazard information (construction, congestion, allergies, etc.) on walking courses
B	Visualization of the excitement in the city	- Camera noise cancellation - Sufficient training data (presence/absence of masks, etc.) - Ensure generalization performance of AI model - Detection of crowding	- Privacy and fairness policies - Setting a maximum number of dense people per location* - Description of AI camera locations, detection details, and expression methods	- Web-based publication of event excitement and crowding - Complaint window from neighbors - Notification to event organizers and participants
C	Signage that collects and visualizes the good and bad points of the city	- Ensure generalization performance of AI model - Record decision results and posting frequency - Output high frequency keywords and stagnant issues	- Positive/negative and non-biased output expressions* Positive/negative and non-biased output expressions - Verification of contributor bias/high frequency keywords/retention issues - Publicity at facilities/events*	- Collection of opinions from citizens (dedicated SNS hashtag)

* Added risk control that does not exist on the cards prepared

Each group identified the necessary risk control functions through the above risk studies, and presented their final proposals for each use case at the final presentation meeting. The proposals included important risk scenarios related to the use case and functions that should be incorporated into the service operation as risk control.

6 Consideration

6.1 Various Risk Scenarios

Based on the risk scenarios identified in each use case, we qualitatively evaluated the comprehensiveness of risk scenarios (1) and diversity of stakeholders exposed to risk (2). In addition, we evaluated the existence of risk scenarios (3) that could be recognized from the viewpoints of both citizens and experts in the discussion.

Comprehensiveness of Risk Scenarios (1). In Group A, two "Robust" risks such as changes in the urban environment, one "Safe" risk related to citizens and pets, one "Responsible" risk that is user-independent, and one "Fair" risk as a division of citizens were identified. In Group B, three risks of "Fair" related to event organizers and citizens, two risks of "Robust" related to the prediction performance of AI, two risks of "Responsible" related to citizens' satisfaction, one risk of "Safe" related to participants was identified, and one "Privacy" risk related to citizens was identified. In Group C, three "Responsible" risks related to the reputation of the city as a whole and its operation, and the operational burden of the letters themselves, one risk each for "Fair" and "Privacy" of citizens, and one "Robust" risk related to the accuracy of the classification by AI were identified.

No risk scenario classified as "Transparent" was identified in all groups. However, referring to the results of the risk control study (Table 7), some of the risk scenarios were related to "Transparent," such as the basis of AI decisions and visualization of information to stakeholders. In Group A, "Output of decision basis (number of dogs, breeds, etc.)" and "Visualization of hazard information (construction, congestion, allergies, etc.) on walking courses" are identified as risk control. In Group B, "Description of AI camera locations, detection details, and expression methods" and "Notification to event organizers and participants" are identified as risk control. In Group C, "Verification of contributor bias/high frequency keywords/retention issues" are identified. It was confirmed that risk scenarios were comprehensively examined in each group (Table 8).

Diversity of Stakeholders Exposed to Risk (2). In Group A, three risks affecting citizens who are users of AI services and their pets are considered. Among them, risks specific to animals (allergies) and risks to citizens who are not pet owners are also considered. Two risks were also identified for the operator of the AI service. In Group B, four risks affecting citizens regardless of whether they participate in the event or not, three risks affecting the organizer of the event, and two risks affecting the participants were identified. In Group C, three risks affecting citizens, two risks affecting AI service operators, and one risk affecting the government were identified. It was confirmed that each group was able to consider risks to various stakeholders.

Risk Scenarios Recognized Through the Knowledge of Citizens or Experts (3). The risk scenarios that could be identified from the perspective specific to citizens were identified. In Group A, "1. Dangerous induction" was raised against the background of many children living in the Kashiwa-no-ha area. "3. Inability to cope with changes in

Table 8. Classification of Risk Scenarios

Gr	Use Case	Fair	Robust	Privacy	Safe	Responsible	Transparent
A	Recommendation of walking routes for pets	1	2		1	1	*
B	Visualization of the excitement in the city	3	2	1	1	2	*
C	Signage that collects and visualizes the good and bad points of the city	1	1	1		3	*

* Considered in risk controls (Table 7)

the urban environment" was also raised as a reason for the boredom of a standardized life. In Group B, the risks affecting the lives of citizens were "6. Negative impact on neighborhoods due to excessive crowding", "8. People are forced to participate in the event", "9. Visualization of citizens' private life in the city. In Group C, "1. Negative impression of the city", "2. Issues are not resolved and citizens' dissatisfaction grows", and "5. Issues are not recognized due to misclassification of positive/negative" were raised.

Risk scenarios identified by the experts' findings were also identified. In Group A, a risk scenario "3. Inability to cope with changes in the urban environment" was identified by an AI governance expert who commented that AI cannot make appropriate predictions unless the data is updated. In Group B, the data scientist who participated in the discussion provided knowledge on the robustness of AI, and identified "2. Adding noise to images to make certain events appear less exciting. It was confirmed that a new risk scenario was identified by the knowledge provided by both citizens and experts.

6.2 Technical and Non-technical Risk Control

Based on the risk control examined using the risk chain in each use case, we qualitatively evaluated whether a risk control that relates AI System (4)/Service Provider (5)/User (6) without relying on a single technological element is considered. We also evaluated whether there exists a risk control (7) considered by citizens, which is not included in the existing risk chain models.

Risk Control of AI System (4). As risk controls common to each group, risk controls related to the collection of training data (Group A "Adjustment of bias of training data (dog breed, etc.)", Group B "Sufficient training data (presence/absence of masks, etc.)") and fairness of AI model (Groups B and C "Ensure generalization performance of AI model") were examined. In addition, information output for the purpose of abnormality monitoring (Group A: "Output of decision basis (number of dogs, breeds, etc.)", Group B: "Detection of crowding", and Group C: "Output high frequency keywords and stagnant issues") were examined. As a risk control specific to each group, Group A examined

"Correction of output results (walking course)" as a response to changes in the urban environment (especially when accidents, construction, or other hazards occur). In Group B, "Camera noise cancellation" was identified to maintain the quality of image information recognized by AI. In Group C, "Record decision results and posting frequency" was also considered to monitor the bias of posted assignments.

Risk Control of Service Provider (5). Monitoring (Group A "Monitoring of AI decision results" and Group C "Verification of contributor bias/high frequency keywords/retention issues") was considered as risk controls common to all groups. In addition, individual measures for each location in the city (Group A "Consideration of individual models for each location" and Group B "Setting a maximum number of dense people per location") were considered. As a risk control specific to each group, Group A identified "Coordination of event information in town and at facilities" as a response to changes in the urban environment. In Group B, it was considered to establish "Privacy and fairness policies" for citizens, event organizers, and participants, and to provide "Description of AI camera locations, detection details, and expression methods" for citizens, event organizers, and participants. In Group C, it was considered that "Positive/negative and non-biased output expressions" for city information displayed on signage and "Publicity at facilities/events" to reduce user bias.

Risk Control of Users (6). As risk controls common to each group, risk controls related to the collection of opinions from citizens (Group B: "Complaint window from neighbors" and Group C: "Collection of opinions from citizens (dedicated SNS hashtag)") were considered as risk controls common to each group. As risk controls specific to each group, in Group A, the "Registration of pet information (breed, allergies, etc.)" and "Visualization of hazard information (construction)" were considered to prevent pets from being led to places where allergies may occur. In Group B, "Web-based publication of event excitement and crowding" and "Notification to event organizers and participants" were considered to manage the crowding situation at event sites. In each group, the Risk Chain Model was used.

In each group, it was confirmed that the risk control in each of AI System/Service Provider/User can be sufficiently studied by using the Risk Chain Model.

Risk Control Added by Citizens (7). In this study of risk control using the Risk Chain Model, we created 47 cards as candidates of risk control by referring to the case study of the Risk Chain Model published by the University of Tokyo [32], for the purpose of supplementing our expertise. In the actual study, risk controls that did not exist in the risk control cards were examined by citizens. In Group A, "Registration of pet information (breed, allergies, etc.)" in the user domain, "Setting a maximum number of dense people per location" in Group B, and "Positive/negative and non-biased output expressions" in Group C in the service provider domain. In all groups, it was confirmed that there existed risk controls that were not included in the existing risk chain model studies, but were newly considered by the citizens.

6.3 Results

The results of this study confirm that "various risk scenarios" and "technical and non-technical risk control" can be considered by using the Risk Chain Model, which is a risk analysis framework, in cooperation with citizens and experts. The stakeholders affected by the utilization of AI are wide-ranging, and it is difficult to determine the extent to which they should be involved in the process. In this study, the target population was the citizens who applied for the living lab. Citizens who participate in the living lab often have high literacy and motivation, so a method to collect the opinions of the silent majority is required. In addition, in the study of risk control, it is necessary to secure the human and resource resources needed to realize the service, and if the costs incurred by the service are not controlled within an appropriate budget range, it will be difficult to sustain the AI service.

7 Conclusion

In the process of prototyping a new AI service in the living lab workshop, the risk was comprehensively identified by the collaboration between citizens and experts using a risk analysis framework. The risk control measures were reflected in the AI service proposals from the citizens. It is expected that the number of cases like Barcelona, where citizens take the initiative in technology implementation, will increase in the future. We expect that various technologies including AI will be socially implemented as trusted services by considering sufficient risk control by multi-stakeholders as in this study.

References

1. OECD: OECD Multilingual Summaries Artificial Intelligence in Society, June 11 2019
2. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**, 389–399 (2019)
3. Ammanath, B.: Trustworthy AI: A Business Guide for Navigating Trust and Ethics in AI, 1st edn. Wiley, USA (2022)
4. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and Harnessing Adversarial Examples, 3rd version (2015). https://arxiv.org/abs/1412.6572, last accessed February 10 2023
5. Datta, A.: The $500mm+ Debacle at Zillow Offers – What Went Wrong with the AI Models?. insideBIGDATA, December 19 2022. https://insidebigdata.com/2021/12/13/the-500mm-debacle-at-zillow-offers-what-went-wrong-with-the-ai-models/, last accessed February 10 2023
6. ProPUBLICA Homepage. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, last accessed February 10 2023
7. Kiviat, B.: The moral limits of predictive practices: the case of credit-based insurance scores. Am. Sociol. Rev. **84**(6), 1134–1158 (2019)
8. epic.org Homepage. https://epic.org/epic-files-complaint-with-ftc-about-airbnbs-secret-trustworthiness-scores/, last accessed February 10 2023
9. Matsumoto, T., Ema, A.: RCModel, a Risk Chain Model for Risk Reduction in AI Service. The University of Tokyo (2020). https://ifi.u-tokyo.ac.jp/en/news/4815/, last accessed February 10 2023
10. Ema, A., et al.: Future Relations between Humans and Artificial Intelligence: A Stakeholder Opinion Survey in Japan. IEEE (2016). https://ieeexplore.ieee.org/document/7790979, last accessed February 10 2023
11. Capdevila, I., Zarlenga, M.I.: Smart city or smart citizens? The Barcelona case. J. Strateg. Manag. **8**(3), 266–282 (2015)

12. Streitz, N.: Beyond ‘smart-only’ cities: redefining the ‘smart-everything’ paradigm. J. Ambient. Intell. Humaniz. Comput. **10**(2), 791–812 (2019)
13. Streitz, N.A., Riedmann-Streitz, C.: Rethinking ‘smart’ islands: toward humane, self-aware, and cooperative hybrid islands. Interactions **29**(3), 54–60 (2022)
14. Lim, C., Kim, K.J., Maglio, P.P.: Smart cities with big data: reference models, challenges, and considerations. Cities **82**, 86–99 (2018)
15. Goodman, E.P., Powles, J.: Urbanism under google: lessons from sidewalk Toronto. Fordham L. Rev **88**, 457 (2019)
16. Fiorino, D.J.: Citizen participation and environmental risk: a survey of institutional mechanisms. Sci. Technol. Human Values **15**(2), 226–243 (1990)
17. Arnstein, S.R.: A ladder of citizen participation. J. Am. Inst. Plann. **35**(4), 216–224 (1969)
18. Cardullo, P., Kitchin, R.: Being a ‘citizen’ in the smart city: up and down the scaffold of smart citizen participation in Dublin, Ireland. Geo Journal **84**(1), 1–13 (2019). https://doi.org/10.1007/s10708-018-9845-8
19. Goodman, N., Zwick, A., Spicer, Z., Carlsen, N.: Public engagement in smart city development: lessons from communities in Canada’s smart city challenge. The Canadian Geographer/Le Géographe canadien **64**(3), 416–432 (2020)
20. de Hoop, E., Moss, T., Smith, A., Löffler, E.: Knowing and governing smart cities: Four cases of citizen engagement with digital urbanism. Urban Governance **1**(2), 61–71 (2021)
21. IEEE/ACM 41st International Conference on Software Engineering 2019, pp. 41–50. IEEE (2019)
22. Salvia, G., Morello, E.: Sharing cities and citizens sharing: perceptions and practices in Milan. Cities **98**, 102592 (2020)
23. Kalinauskaite, I., Brankaert, R., Lu, Y., Bekker, T., Brombacher, A., Vos, S.: Facing societal challenges in living labs: towards a conceptual framework to facilitate transdisciplinary collaborations. Sustainability **13**(2), 614 (2021)
24. Mahmoud, I., Morello, E.: Co-creation pathway for urban nature-based solutions: testing a shared-governance approach in three cities and nine action labs. In: Smart and Sustainable Planning for Cities and Regions: Results of SSPCR 2019—Open Access Contributions, 3, pp. 259–276. Springer International Publishing (2021). https://doi.org/10.1007/978-3-030-57764-3_17
25. AI Risk Management Framework Second Draft, pp. 10–16, NIST, USA, August 18, 2022. https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf, last accessed February 10 2023
26. Machine Learning Quality Management Guideline 2nd Edition, AIST, Japan, May 16 2022. https://www.digiarc.aist.go.jp/en/publication/aiqm/aiqm-guideline-en-2.1.1.0057-e26-signed.pdf, last accessed February 10 2023
27. Model AI Governance Framework second edition, IMDA and PDPC, Singapore, January 21 2020. http://go.gov.sg/ai-gov-mf-2, last accessed February 10 2023
28. Ethics, Transparency and Accountability Framework for Automated Decision-Making, Central Digital and Data Office, Cabinet Office, and Office for Artificial Intelligence, UK, May 13 2021. https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making, last accessed February 10 2023
29. Google: Responsible AI practices,. https://ai.google/responsibilities/responsible-ai-practices/, last accessed February 10 2023
30. Microsoft: Microsoft Responsible AI Standard, v2, June 2022. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf, last accessed February 10 2023
31. AI Ethics Impact Assessment White Paper, Casebook, and Practice Guide, Fujitsu, February 21 2021. https://www.fujitsu.com/global/about/research/technology/aiethics/, last accessed February 10 2023

32. AI Service and Risk Coordination Study Group. The University of Tokyo. https://ifi.u-tokyo.ac.jp/en/projects/ai-service-and-risk-coordination, last accessed February 10 2023
33. Urban Design Center Kashiwa-no-ha (UDCK). https://www.udck.jp/, last accessed February 10 2023