

AI Value Platform Accelerates Data Valorization by Consolidating SDSC's Elemental Technologies

Takahiro Hata, Midori Kodama, Miho Fujishima, Yu Adachi, Kenichi Fukuda, and Yuki Yokohata

Abstract

We at NTT Smart Data Science Center (SDSC) are developing an artificial intelligence (AI) value platform to accelerate data valorization by consolidating the results from various industry domains that we have been working on. By cataloging analytical case studies, we aim to support problem design and accelerate valorization by automatically experimenting with processing techniques for aggregated real data, thus shortening the initial trial-and-error phase of data analysis. In this article, we introduce our efforts toward implementing the AI value platform.

Keywords: AI value platform, data analysis, mathematical optimization

1. Introduction

At NTT Smart Data Science Center (SDSC), we have been addressing customer challenges in various areas such as lifestyle, consumption, mobility, logistics, and resources, particularly to enable urban development through Digital Twin Computing (DTC) (Fig. 1).

Looking back at these achievements, it is apparent that the source of value in the real world (physical space) is based on real data acquired under the constraints imposed by reality. We have come to understand that this can be explained by the mechanism of a cyber physical system, where optimal states are calculated in the digital twin (DT) in cyber space and the results are returned as controls or recommendations for the real world (Fig. 2). Optimization, playing the most crucial role here, has been studied extensively, even with complex and vast models. However, with advancements in artificial intelligence (AI) and machine-learning technologies, it has become feasible to solve optimization problems by learning from rich and well-structured data within reasonable computation time.

To achieve valorization in the real world, however, various constraints specific to each field such as measurement conditions, operational conditions, and costs must be met. Often, raw data as measured cannot be directly used without preprocessing.

In light of these circumstances, SDSC has been developing an AI value platform that consolidates core technologies and insights from various domains of urban DTC, which SDSC has been addressing, to accelerate data valorization to bridge the gap between the physical and cyber realms and return value to the real world on the basis of the data.

The AI value platform defines and embodies new value in actual industry domains through AI, making each value tangible as services for business operations and customer value within those domains. It encompasses common analysis functions such as real-data processing, prediction, and optimization, as well as functions required for actual data integration and interfaces that enable users to experience the value in their business operations or as individual customers.

In the AI value platform, core AI functionalities are integrated into parts that can be organically used



Fig. 1. Examples of domains we have worked on.

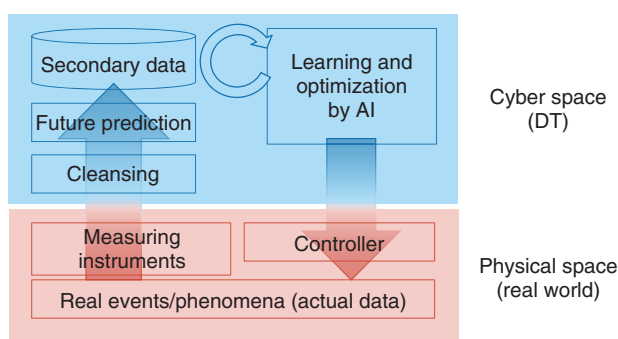


Fig. 2. Conceptual diagram of cyber physical system.

within services offered across multiple industry domains. By combining these parts in accordance with the characteristics of the obtained value, it becomes possible to accelerate valorization by facilitating the selection of necessary functionalities for new services, automating initial data processing, and minimizing trial and error.

We introduce two main innovations among various efforts to accelerate valorization within the functionalities of the AI value platform:

- (a) aggregating and providing refined problem designs as references, derived from past achievements, to ensure operability in society
- (b) extracting preprocessing tasks, such as data cleansing and future forecasting, which precede optimization, as common analysis func-

tions of the AI value platform, making them applicable to other datasets.

2. How the AI value platform accelerates data valorization

We explain how the aforementioned functionalities (a) and (b) contribute to the valorization process.

The valorization process at SDSC mostly follows the typical steps of data analysis shown in **Fig. 3**. In the problem-design phase (1), customer problems and surrounding knowledge are decomposed, leading to the formulation of optimization problems. In the data-analysis phase (2), data are measured in a real environment, and optimization methods are implemented in the DT. If these steps are deemed effective, the process moves to the validation phase (3), where actual equipment is operated for field validation.

Functionalities (a) and (b) in particular accelerate the problem-design and data-analysis phases. This not only speeds up valorization but also helps customers realize the significance of the initiatives by accelerating steps (1) and (2) to demonstrate results promptly, especially in SDSC's efforts, which require the cooperation of customers and using real data. While the AI value platform is also being developed to be used as a DT, we omit discussion on this aspect as it pertains to different areas than those covered by (a) and (b).

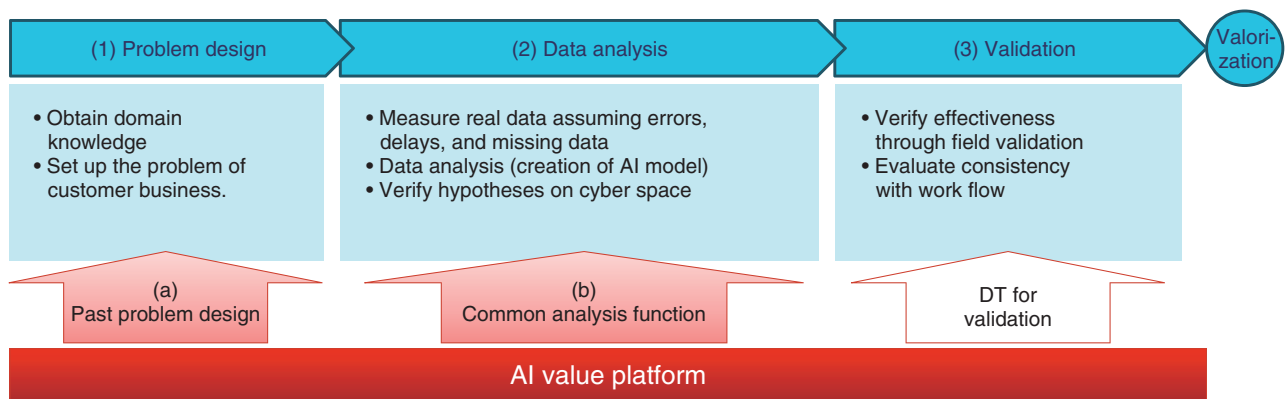


Fig. 3. Valorization flow and AI value platform's contribution.

(a) Provide examples of optimization problem design as references

In the real world, addressing challenges can often be translated into mathematical optimization problems that minimize (or maximize) specific metrics while meeting various real-world constraints. Example challenges that SDSC has addressed include:

- to reduce energy consumption in air conditioning, minimize the total energy consumption of air conditioning operations for the day while maintaining visitor comfort
- to deliver products to as many customers as possible on time, plan routes for multiple robots for product delivery to maximize the number of deliveries in the shortest time
- to respond promptly to emergency calls, minimize the time required for ambulances to arrive at the scene by changing the fire stations where ambulances wait.

These examples represent mathematical optimization problems with constraints. However, it is important to note that translating real-world problems into mathematical problems can be challenging and is often considered the most crucial and difficult process in valorization.

Ensuring the practicality of this problem design requires the ability to correctly understand and decompose the customer's problems and surrounding environment then effectively align and redesign methods for solving them within the allotted time. For this problem design to be feasible, all the following issues must be resolved:

- tolerating errors, delays, and missing data in measurements
- developing analytical logic to derive optimal

solutions from measurement data

- controlling within an acceptable time frame from the analysis results to execution
- incorporating the execution procedure into operational workflows.

It is rare to anticipate and resolve all these issues completely in a single attempt. In reality, it often involves iterating through multiple problem designs, operationalizing them, and gathering various feedback to revise and refine the necessary areas. This iterative process is essential for refining problem designs and ensuring their practical applicability.

Experienced data scientists, who have accumulated a wealth of experience in data analysis, have encountered various problems that can arise in each stage of measurement, analysis, control, and operation in various situations. They are familiar with numerous cases of both success and failure, enabling them to reach realistic solutions in fewer cycles. To become such experienced data scientists, it is essential not only to gain personal experience but also to seek advice from experienced professionals and review past deliberations and materials to relive the trial and error of predecessors. This is the aim of accumulating optimization problem design cases as references.

The aim with the AI value platform is to accumulate problem designs that data scientists have solved in the past so that they can be referenced as guides when facing new problems. This provides a means for data analysts facing new challenges to relive past experiences and examples. Even if the domains are different, being able to reference the process of problem decomposition and reduction to mathematical optimization problems in one's valorization process for the problems they encounter, it is expected that they

will be able to quickly carry out practical problem designs.

(b) Make optimization preprocessing a common analysis function that can be applied to other data

Temperature, brightness, air-conditioning settings, and the number of people in the area are quantified through sensors and become data. Using these data, we can understand or predict real-world phenomena and optimize various situations by controlling equipment or recommending actions. However, various factors, such as sensor malfunctions, movements, obstructions, or sudden increases in the number of people due to events, can introduce errors, missing data, or anomalies in the data. These errors, gaps, and anomalies hinder our understanding of phenomena and decrease prediction accuracy. Therefore, before optimization, data cleansing is necessary to process the data into a suitable form for optimization.

The data that can be measured are past events. Therefore, even if equipment control or behavioral recommendations are based on past data, it may not be optimal when equipment or people move. For example, in building air conditioning, changes in set temperature and airflow may take several dozens of minutes to several hours to be reflected in the space as changes in temperature and humidity, due to the large area managed by the air-conditioning system. During this time, it is common for crowded periods to end, resulting in excessive air conditioning. To prevent this, predicting when the optimization results are reflected can also be considered as a form of preprocessing.

At SDSC, we sometimes use AI, which requires a large amount of data, as an optimization method. However, in the real world, it is often not possible to conduct parallel or accelerated experiments, making it difficult to collect large amounts of data. On the basis of the examples from SDSC, it has been found that in such cases, specific data-processing techniques unique to machine learning are used, including:

- various data-augmentation techniques to supplement small amounts of data
- rebalancing to address extreme frequency differences between positive and negative examples
- typification of phenomena when there is a shortage of data due to a multitude of rare events (clustering).

These processing techniques can also be considered preprocessing, as they involve generating secondary

data on the basis of the data measured before optimization.

The preprocessing tasks depend not only on the problem design introduced in (a) but also on the characteristics of the measuring instrument, measurement environment, and the measured data (phenomenon). This means that even if the problem designs are different, the preprocessing can still be effective. Therefore, in the AI value platform, the decision was made to extract the preprocessing part as a common analysis function. By modularizing the common analysis function in the same format, it is possible to try various methods simultaneously when analyzing new data and compare the results, enabling the quick selection of suitable preprocessing methods. Therefore, (a) and (b) support and accelerate the valorization process at SDSC.

3. Difference from conventional technology

In today's digital age where various real-world phenomena can be digitized, data analysis has become a challenge tackled by many individuals and companies. For this reason, numerous analysis technologies, libraries, and platforms that aggregate them are being provided. We describe the differences between common data-analysis platforms, various libraries, and the AI value platform.

We first explain the characteristics of representative analysis environments, such as R [1] and Python [2].

The R language is a language and software environment known for statistical analysis, developed with reference to the S language environment originally created at AT&T. It provides an interactive interpreter environment and is capable of processing large amounts of data rapidly. It also includes implementations of numerous statistical methods, enabling users to begin data analysis immediately upon installation. Additionally, advanced processes not implemented by default can be achieved by importing external packages.

Python is a general-purpose programming language with procedural aspects. While the programming language aspect of Python is not our main focus, like R, it includes many statistical processing, vector, and matrix calculation functions in its standard library. Despite being an interpreter, Python is known for its fast numerical computation library. Being a general-purpose programming language, it is well-suited for implementing complex processes. It also has robust support for graphics processing units

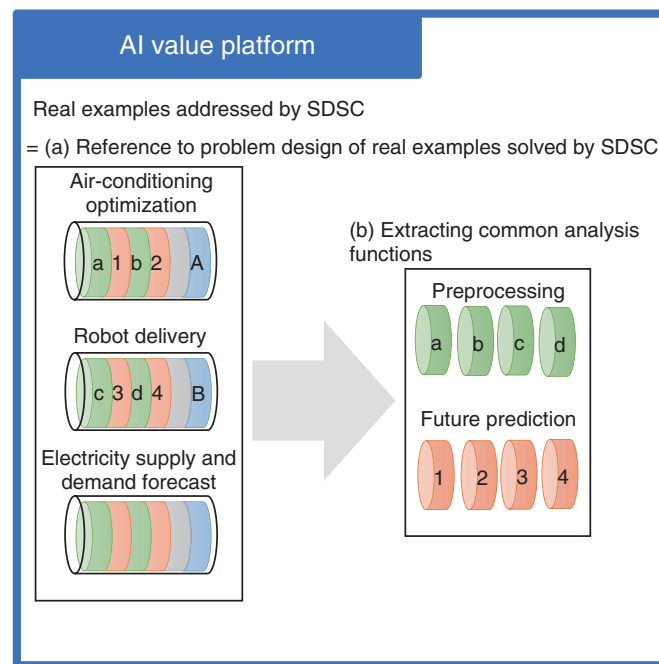


Fig. 4. Image of AI value platform implementation.

(GPUs), making it highly usable for advanced machine-learning tasks such as deep learning using neural networks. Python is widely chosen as an option for data-analysis implementation due to its capability to use cutting-edge algorithms by importing external libraries.

In addition to the two environments mentioned as examples, there are numerous programming languages and analysis platforms that aggregate algorithms implemented in these languages. In a broader sense, business intelligence tools for visualizing data from various aspects and relational databases with data accumulation and aggregation functions can also be considered as part of the analysis environment.

The common feature worth highlighting among these environments is their provision of versatile analytical functions that do not limit data formats or domains. This is, of course, expected in the context of providing analytical solutions. However, when seeking to solve real-world challenges using such versatile analytical environments, data analysts with expertise in problem design and preprocessing must spend time and effort selecting the appropriate method from a multitude of analytical techniques. Bridging the gap between cyber and physical spaces and optimizing them in a practical way still requires problem design and preprocessing, necessary in the

valorization process, and which remain significant areas involving considerable time and effort through trial and error.

To address this, in the AI value platform, preprocessing techniques proven to solve real-world challenges are modularized to accelerate this trial-and-error process. Using this approach aims to provide support for these processes, which are difficult to support in general analytical environments, thus significantly differing from conventional technologies.

4. Design policy for the AI value platform

The following is a design policy (**Fig. 4**) for the AI value platform based on the ideas introduced thus far.

(a) Providing references to past problem design cases

As we have mentioned, to reduce a real-world problem to a mathematical optimization problem, it is necessary to have experience in solving many constraints and designing problems. The AI value platform supports analysts by allowing them to relive past problem designs, thus assisting in problem design through experience. In other words, the knowledge we want to provide with this functionality is how to resolve real-world constraints in a way that

does not distort the mathematical optimization problem. Specifically, within the AI value platform, past examples are analyzed as mathematical optimization problems, where

- goals → key performance indicators (KPIs) set for those goals → formulation as optimization problems
- measurement data → preprocessing executed
- optimization problems → optimization algorithms used.

We aim to list these real-world constraints and adopted methods as sets, enabling reference and search functionalities.

In this process, we consider successful cases to be more important than failure cases. While learning from failures is valuable, when it comes to accelerating the value realization with the AI value platform, the best approach is often to emulate successful cases. Although SDSC's initiatives appear to solve problems across various domains, fundamentally, the real world is driven by human social activities. This commonality forms the characteristics of measurement data and the operational conditions that solution methods must meet. Therefore, we believe that leveraging past successful cases is possible across different domains, as they share this common foundation.

(b) Standardization of preprocessing as common analysis function

The aforementioned commonalities also apply to preprocessing, which aims to eliminate constraints in the physical space.

SDSC's approach to data valorization involves handling real-world phenomena as time-series numerical data, capturing them in human social cycles such as time, day, day of the week, and week. Using periodicity and similarity, techniques like smoothing, interpolation, and data augmentation are implemented. The methods used to measure physical phenomena, such as temperature, brightness, power consumption, and population, are similar regardless of the domain.

Preprocessing methods leveraging these commonalities can be extracted by decomposing the cases SDSC has solved. The AI value platform was developed to standardize effective preprocessing in the same format, enabling uniform application to new data and accelerating the trial-and-error process of preprocessing.

new value by carrying out data preprocessing, future prediction, optimization, visualization, and optimal control tailored to the specific challenges faced by customers. The AI value platform provides an environment like the one shown in **Fig. 5**, where users can analyze sample data or their own data to confirm their usefulness. For each service, a dedicated DT is provided, offering independent execution environments for each customer (tenant). If the results of data valorization or services meet the customer's satisfaction, they can choose to use the environment on the AI value platform as it is, integrate it with their own systems and facilities, or customize it in accordance with the connected systems and facilities.

By using the aforementioned common analysis functions, it becomes possible to try and compare various data-analysis algorithms, enabling the selection of the optimal method.

Specific services are introduced below.

For instance, in efforts aimed at reducing food loss in stores, using customer data, we conduct future predictions such as forecasting the number of visitors and sales of each menu item. By visualizing these data, we can reduce excessive purchasing and preparation, thus reducing food loss (see **Fig. 6**).

When implementing this initiative at the customer's facility, quick results can be obtained using cameras, sensors to measure visitor numbers, and point-of-sales (POS) data (sales per menu item), as depicted in **Fig. 7**, which summarizes the necessary data and analysis functions for the service.

SDSC has been working on data-value-creation initiatives such as energy-efficient and comfortable air conditioning control, personal air conditioning in offices, achieving zero food loss in restaurants, optimizing store operations using customer-behavior analysis, efficient robot delivery to achieve effective distribution, optimal placement of emergency teams, high-precision power-generation forecasting and demand matching, detection and prediction of traffic congestion, nationwide optimization of agricultural product distribution, hospitality services in smart homes and urban areas based on predictive behavior, among others, as shown in **Fig. 1**. With the support of the AI value platform facilitating value creation, we will contribute to a future society where people's comfortable living standards are significantly improved.

5. Commercialization initiative

As mentioned earlier, SDSC has been providing

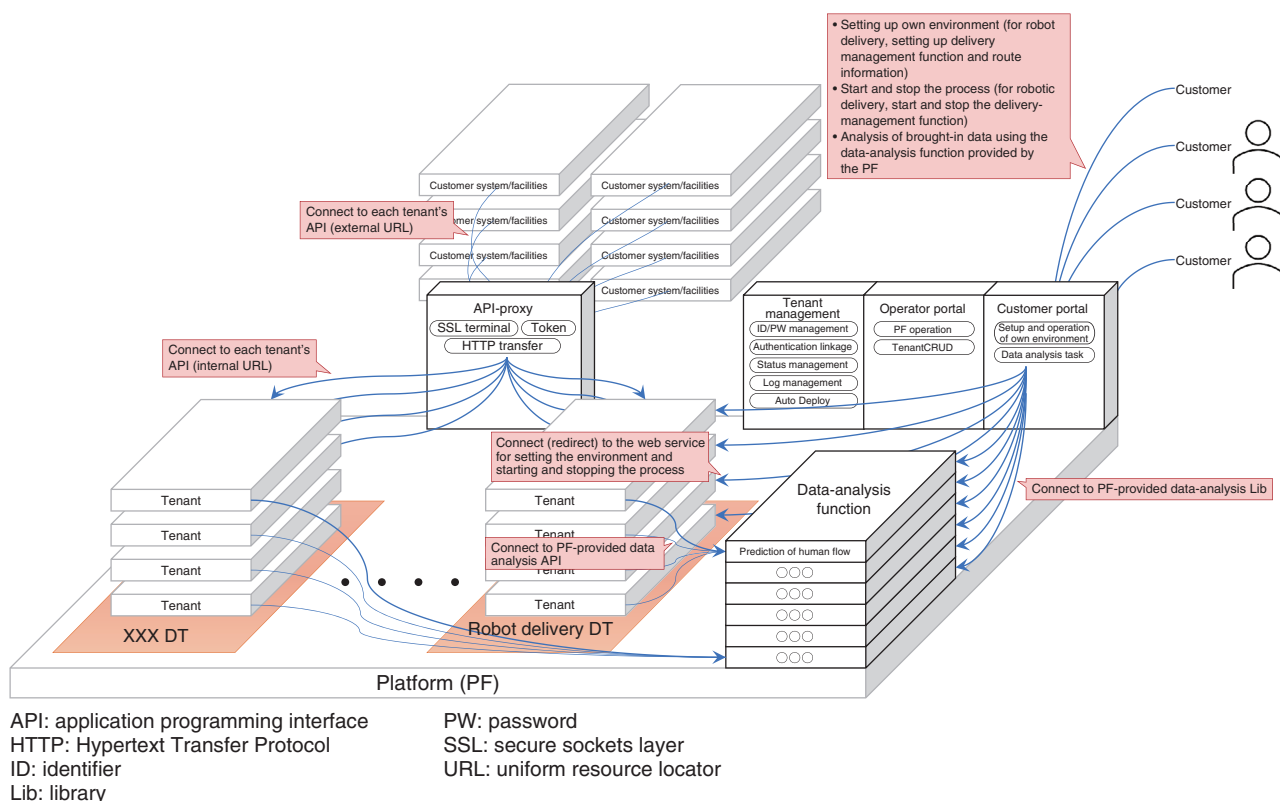


Fig. 5. Grand design of AI value platform services.

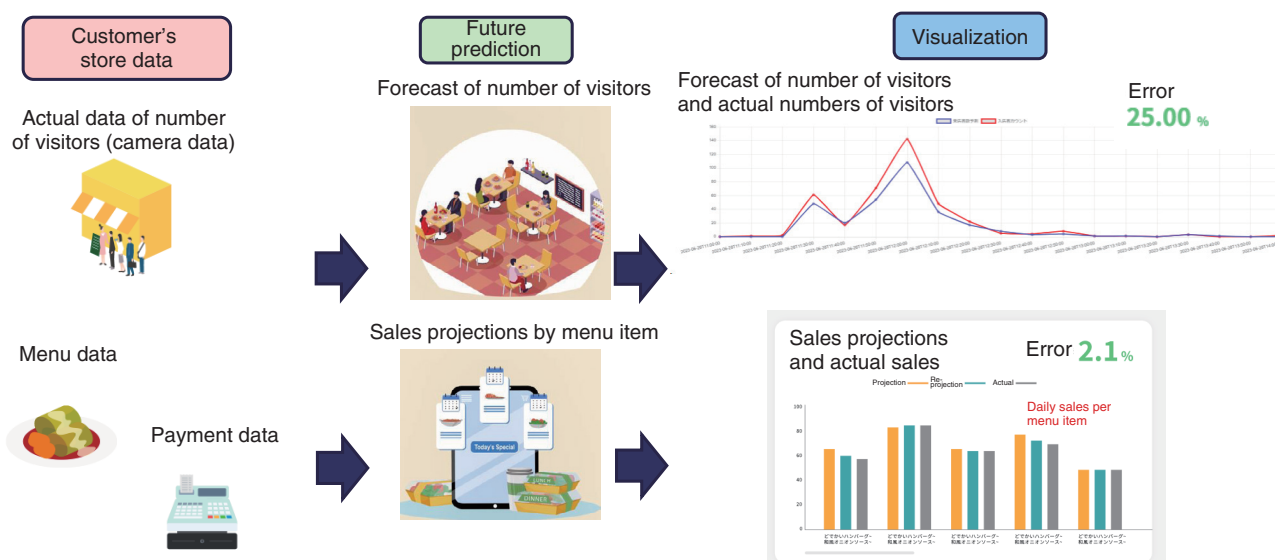
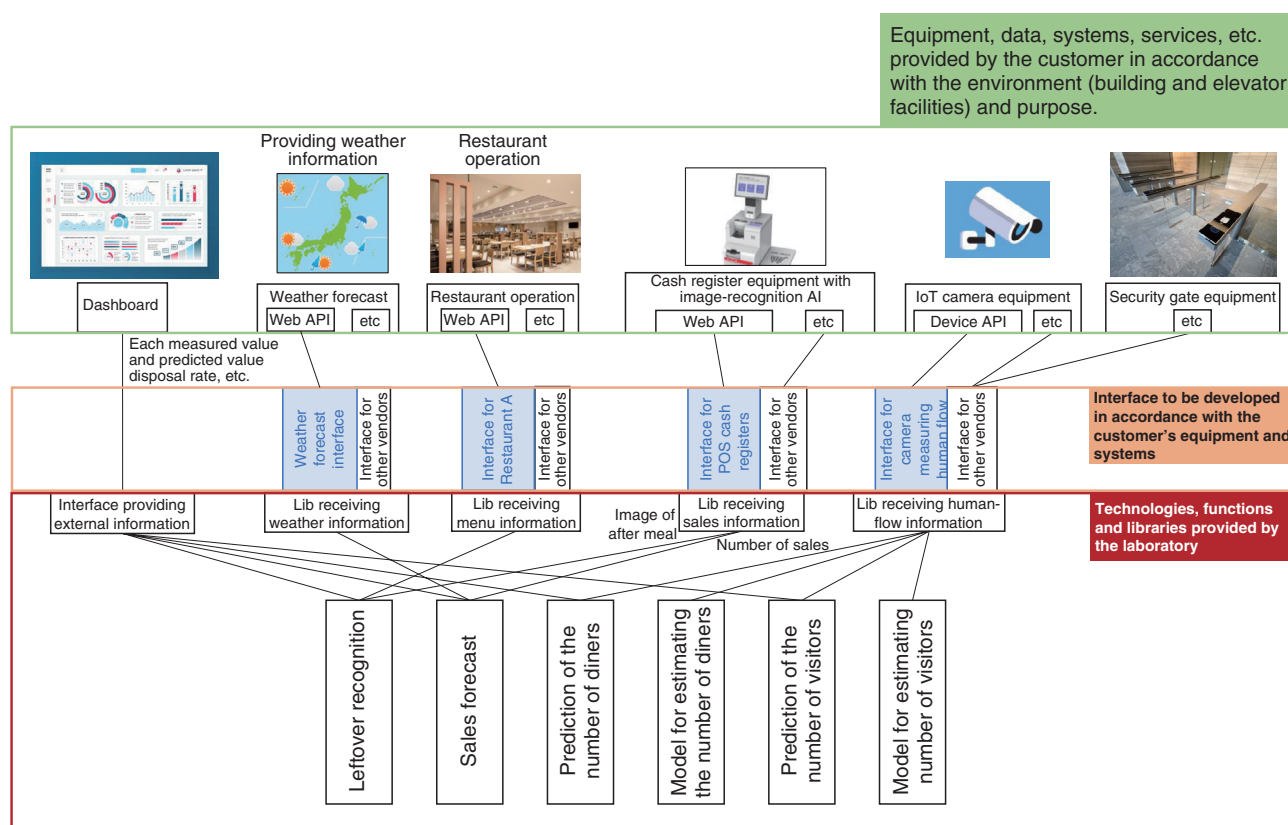


Fig. 6. Zero food loss initiatives: Forecasting the number of store visitors and number of sales per menu item.



IoT: Internet of Things

Fig. 7. Image of the system for implementing zero food loss policies.

6. Challenges and prospects for the AI value platform

We introduced our attempt to support new data valorization by consolidating the knowledge of the actual cases we have solved into the AI value platform. However, there are several challenges in aggregating insights into this platform.

One challenge is that the functionalities used in actual cases are implemented in the most efficient way within those cases. To extract them as common functionalities, each functionality needs to be redesigned to be applicable in other domains, assuming the same execution capability.

Another challenge is that as the common analysis

functions increase, there is a need to select the appropriate analysis functionalities for application. For this purpose, metadata resembling semantics capable of understanding the characteristics and meanings of the data is necessary. These metadata may need to be reviewed whenever a new domain is added.

In fiscal year 2023, we completed the minimum implementation of the AI value platform. We will improve upon it on the basis of the feedback from actual users to address these challenges.

References

- [1] Website of R, <https://www.r-project.org>
- [2] Website of Python, <https://www.python.org>

**Takahiro Hata**

Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.Inf. from Kyoto University in 2006 and joined NTT the same year. He studied ubiquitous services at NTT Network Innovation Laboratories. From 2010 to 2013, he was involved in the development of public wireless local area network services at NTT WEST. Since 2013, he has been active in the development of smart-city applications and analysis of pedestrian flow and driving data of vehicles at NTT Smart Data Science Center.

**Midori Kodama**

Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.S. in biosciences and informatics from Keio University, Kanagawa, in 2015 and joined NTT the same year. She studied ubiquitous services and analysis of educational data at NTT Network Innovation Laboratories. Since 2017, she has been active in the development of smart-city applications and data analysis related to air-conditioning systems at NTT Smart Data Science Center.

**Miho Fujishima**

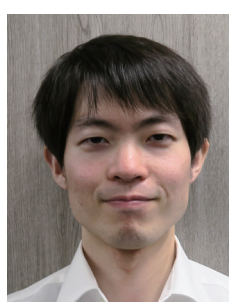
Researcher, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.S. in earth and planetary science from the University of Tokyo in 2017 and joined NTT WEST the same year. From 2018 to 2021, she was involved in the development of home routers and technology in the Internet-of-Things and AI fields. In 2022, she moved to NTT laboratories and studied multi-agent mobility simulation as part of DTC. Since October of 2023, she has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.

**Yu Adachi**

Senior Research Engineer, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. from the University of Electro-Communications, Tokyo, in 2009 and joined NTT the same year, where he studied distributed data management. From 2012 to 2014, he was involved in service development at NTT WEST. From 2015 to 2020, he was involved in software engineering at NTT Software Innovation Center. Since 2021, he has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.

**Kenichi Fukuda**

Researcher, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

He received an M.E. in control engineering from Osaka University in 2018 and joined NTT the same year. He studied optimization of ambulance allocation at NTT Network Innovation Laboratories. He has been active in the development of smart-city applications and data analysis at NTT Smart Data Science Center.

**Yuki Yokohata**

Senior Research Engineer, Supervisor, NTT Smart Data Science Center, NTT Computer and Data Science Laboratories.

She received an M.S. in earth and planetary science from Nagoya University, Aichi, in 2004 and joined NTT the same year. She studied ubiquitous services at NTT Network Service Systems Laboratories. From 2008 to 2015, she was involved in the development of public wireless local area network services at NTT Communications. Since 2016, she has been active in the development of smart-city applications and analysis of data at NTT Smart Data Science Center.