SESLAB
Software Engineering & Security Lab

AIStar
Enabling AI to IoT

ETSI IoT Week 2022

세종대학교
SEJONG UNIVERSITY

AI-enablement to oneM2M

Prof. Song JaeSeung
Sejong University
jssong@sejong.ac.kr

Contents

- Motivation and background
- How to support common AI/ML features via IoT service layer platform?
- Introduction to AI enabled oneM2M system
- Other data centric work items in oneM2M

Motivation and Background

- Many Artificial Intelligence (AI) and Machine Learning (ML) applications use data collected in IoT platforms to train their model.
- Depending on the quality and quantity of collected dataset for model training, the performance of AI models are different.
- IoT platform (including oneM2M) is a place holder to collect and manage various data (image, text, sensory, etc.)
- In order to build a good model, it is very important to have good data management.
- As AI technologies are now being used in many network systems (such as telco core network, smart factory platform, including IoT platforms), it is good to consider providing necessary AI enablement features to IoT platforms.
- If AI applications use IoT platforms that support proper AI data management, the applications can provide various intelligent services more easily.

Steps of Machine Learning (ML)

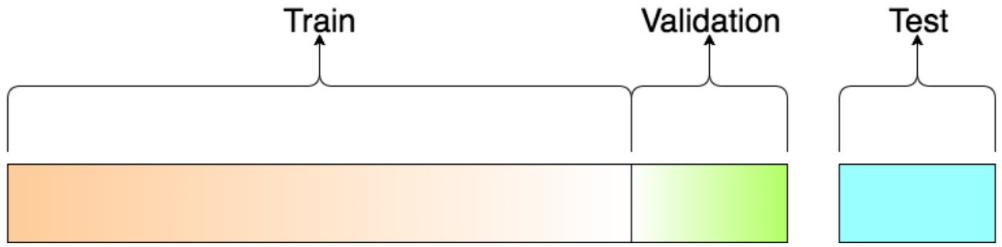
- Typically it is known that there are seven main steps involved in ML as follows:
  - Step 1: Collect data
    → The quality and quantity of data is very important to have a good predictive model
  - Step 2: Prepare data
    → Use collected data to prepare for use in a machine learning training (a dataset is classified into three parts, i.e., train, validate, and test. Data can also be augmented to have enough dataset for training)
  - Step 3: Choose a model
    → There exist many models. Some are suitable for image data, others for text. Depending on a situation, a proper model should be selected.
  - Step 4: Train the model
    → Use collected data set to improve the model.
  - Step 5: Evaluation
    → The trained model is evaluated to see the model is good or not. (a collected dataset should be split into three parts, i.e., train, validation and test)
  - Step 6: Parameter tuning
    → If possible and there is a room for improving a tested model, tune parameters used by the model
  - Step 7: Make predictions
    → Once we have a good model, the model is now used for prediction.

Required data set for ML

- In order to perform ML, three types of dataset are required as follows:

- Training dataset: *The sample of data used to fit the model.*
- Validation dataset: *The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.*
- Test dataset: *The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset*

- There exist a ratio splitting a dataset for training, validation and testing
- For example, 70% train, 15% validation, and 15% test



A visualisation of the splits

Nothing NEW

- ETSI:
  - SmartM2M STF 584
    - Objectives: The objective of STF 584 is to provide an initially validated architecture that describes how IoT systems can make use of Artificial Intelligence (AI) and Machine Learning (ML) for the management and interpretation of IoT devices data over a large variety of deployment models (e.g., edge or cloud-based) while remaining interoperable, secure and manageable.
  - ETSI TR 103 674 (v1.1.1) – Artificial Intelligence and the oneM2M architecture
  - ETSI TR 103 675 (v1.1.1) – AI for IoT: A Proof of Concept

- South Korea:
  - A new government project, called ‘AIStar’, has been granted to enable AI features to oneM2M platform
  - An open-source project based on oneM2M
  - Huge interest in AI/ML
  - AI + Edge + IoT

What oneM2M can do?

- oneM2M can consider to support data management for AI/ML
- For example,
  - A resource for AI/ML algorithms?
  - Classification and management of ML dataset stored in oneM2M?
  - A new CSF for various AI/ML models?
  - Management of parameters for AI/ML models?
  - Management of trained AI/ML models?

- A new WID for the enablement of AI/ML features to oneM2M?

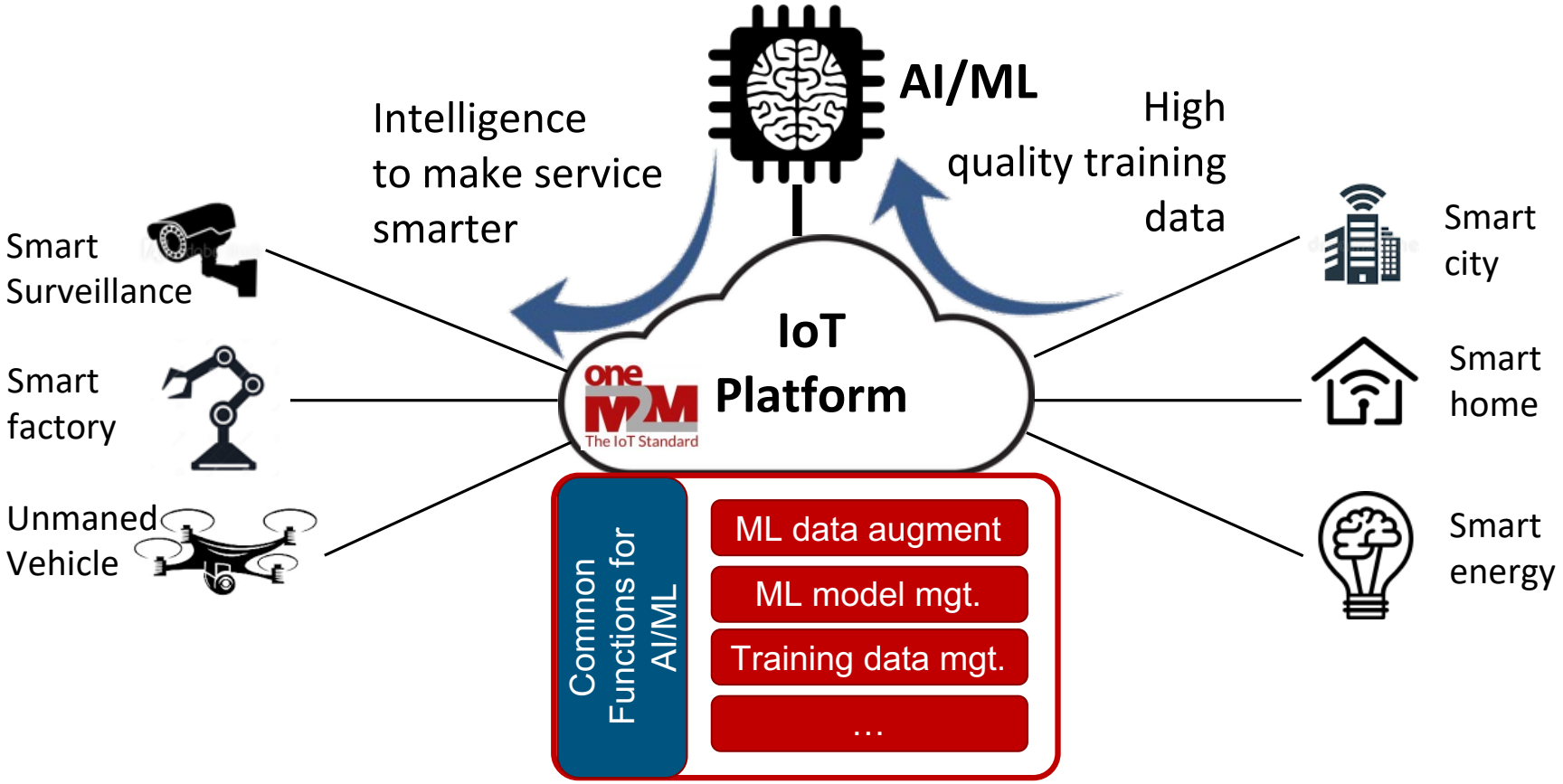
AI/ML as a common service

AE In oneM2M terminology, an AE represents an Application Entity

CSE In oneM2M terminology, an CSE represents a Common Services Entity

High-level concept

- oneM2M system should be enhanced with
  - A new common service function (CSF) to support AI capabilities
  - A set of new resources
  - Various AI/ML use cases

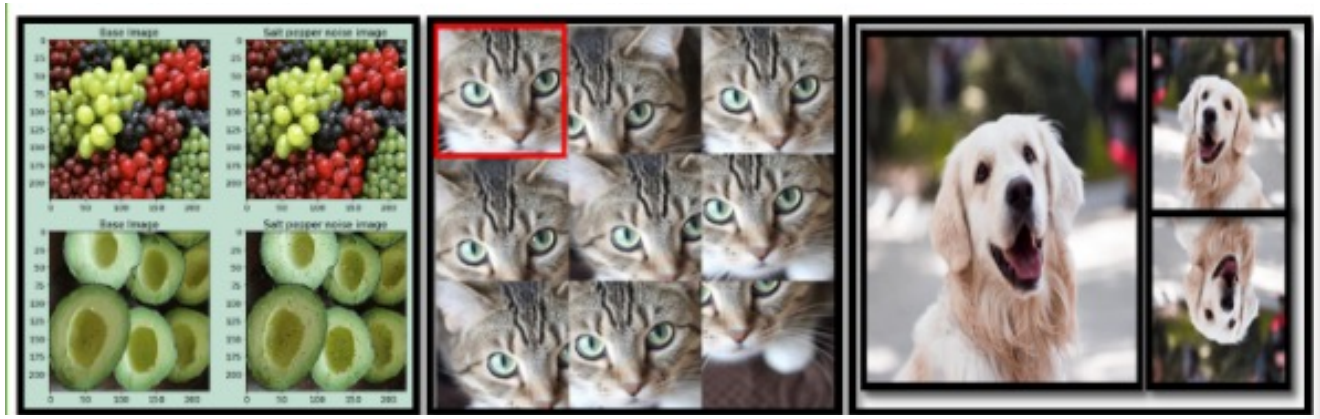


Use case #1. Data Augmentation

- The size of the dataset to be used for training is a big problem
- A technique to obtain a larger dataset from original dataset is required ➔ **Data Augmentation**

> *"Data augmentation techniques artificially generate different versions of a real dataset by adding slightly modified copies of already existing data or newly created artificial data from existing data to increase its size."*

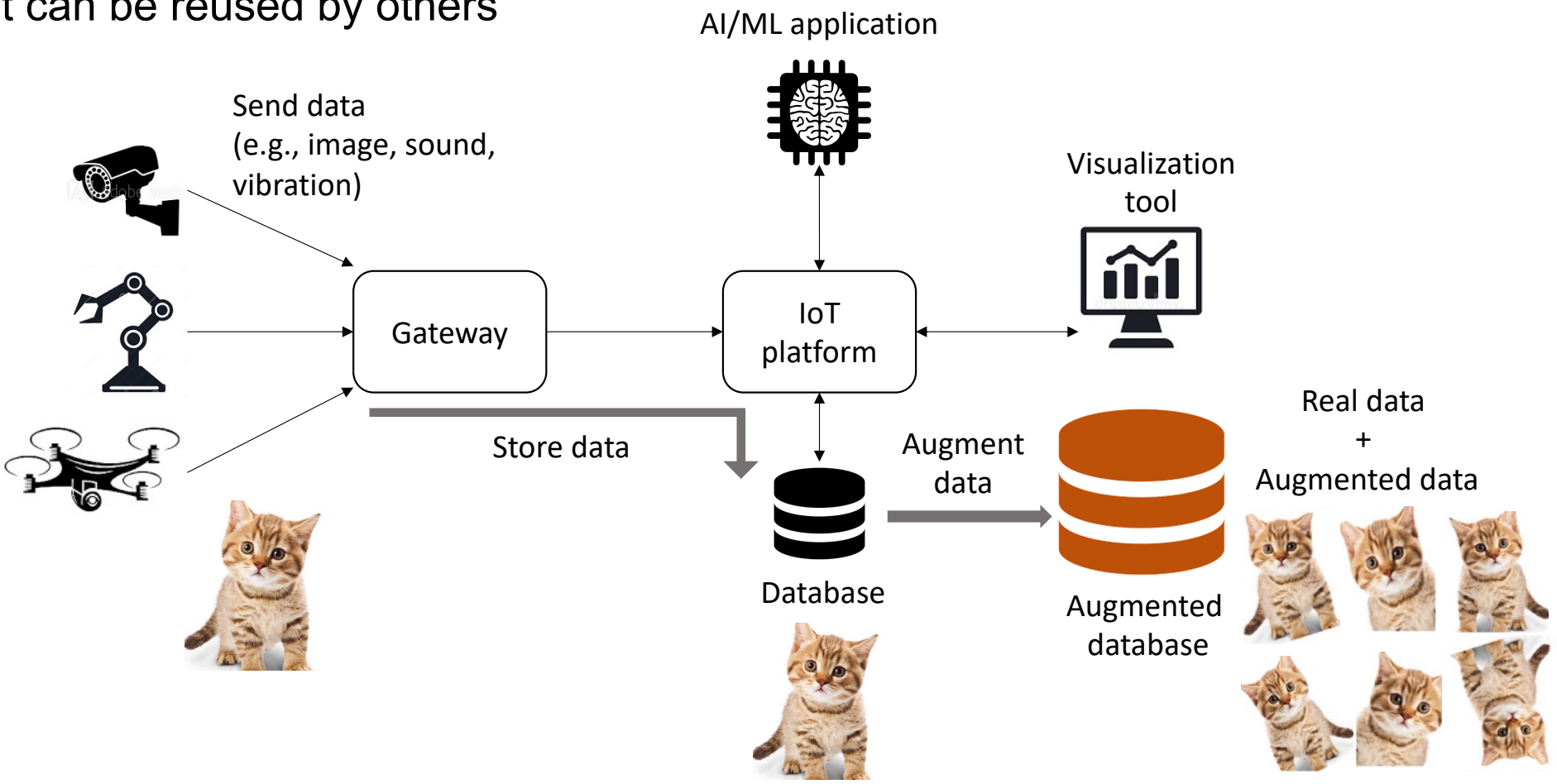
- Well know data augmentation techniques for image data:
  - Flipping → The image is flipped horizontally and vertically
  - Rotation → The image is rotated by a degree between 0 and 360 degree
  - Cropping → A section of the image is selected, cropped and then resized



What if an IoT platform provides a function to to generate different versions of a dataset?

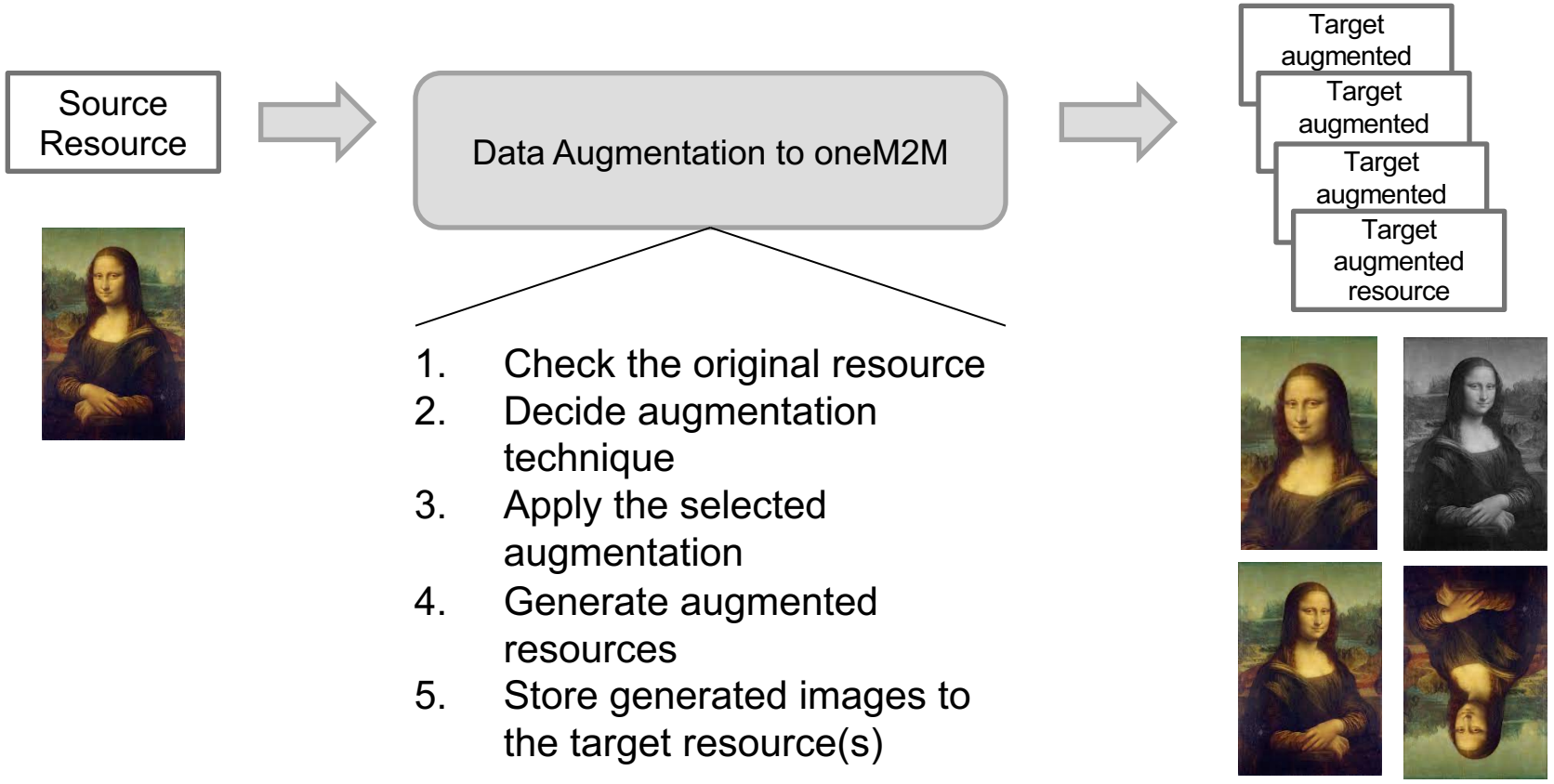
Use case #1. Data Augmentation

- Introduce a new resource to oneM2M for data augmentation
- AI/ML developers simply request data augmentation to oneM2M platform
- Larger training dataset is given to AI/ML developers for training and build a model
- Augmented dataset can be reused by others



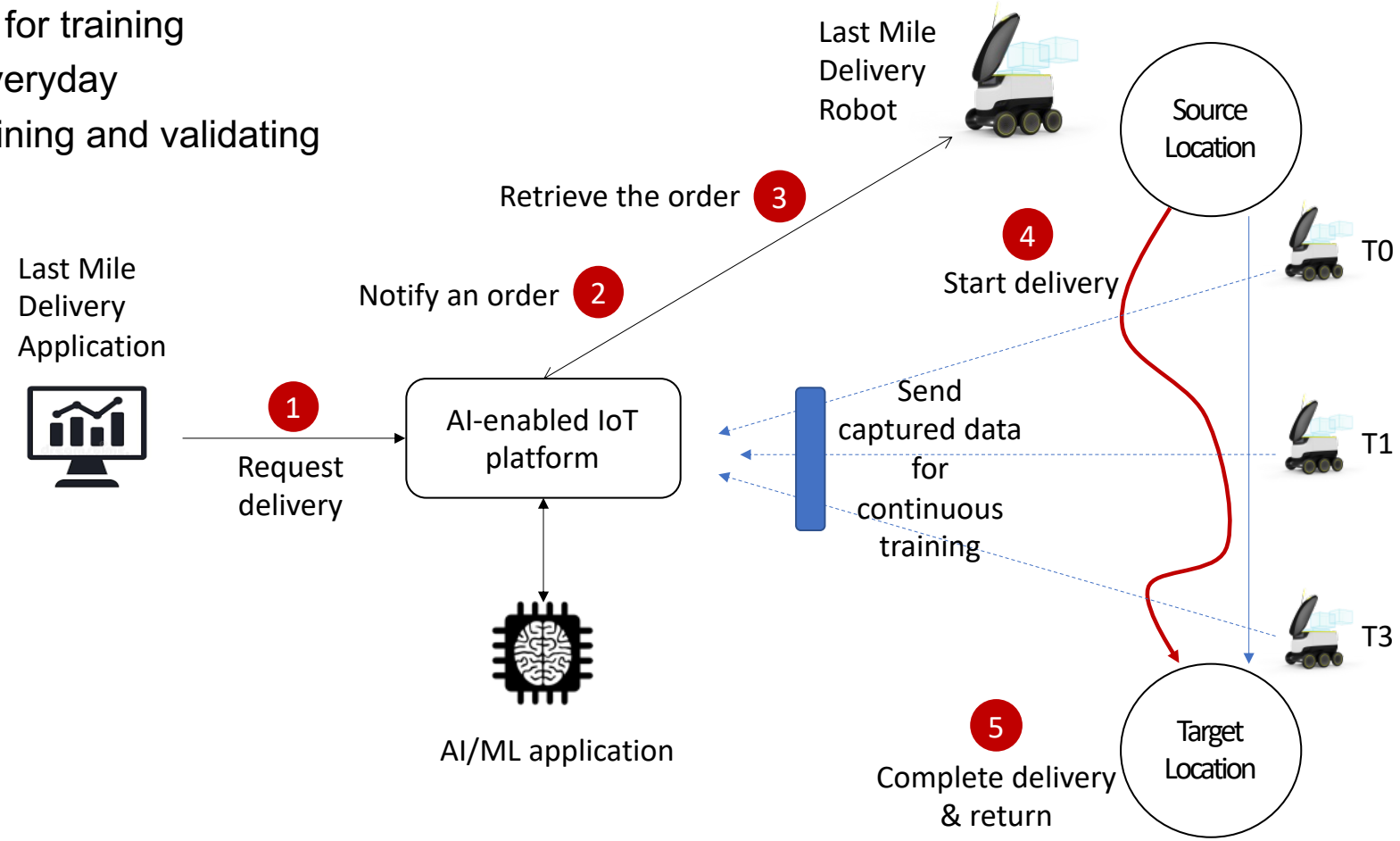
Use case #1. Data Augmentation for Autonomous Driving

- Step 1: AI application (oneM2M AE) sends request to the <dataAugmentation> resource
- Step 2: CSE stores received input to the <dataAugementation> resource
- Step 3: Based on the given parameters, CSE tries to get the source image, apply the given data augmentation technique (e.g., resize), and generate target resources containing generated augment images



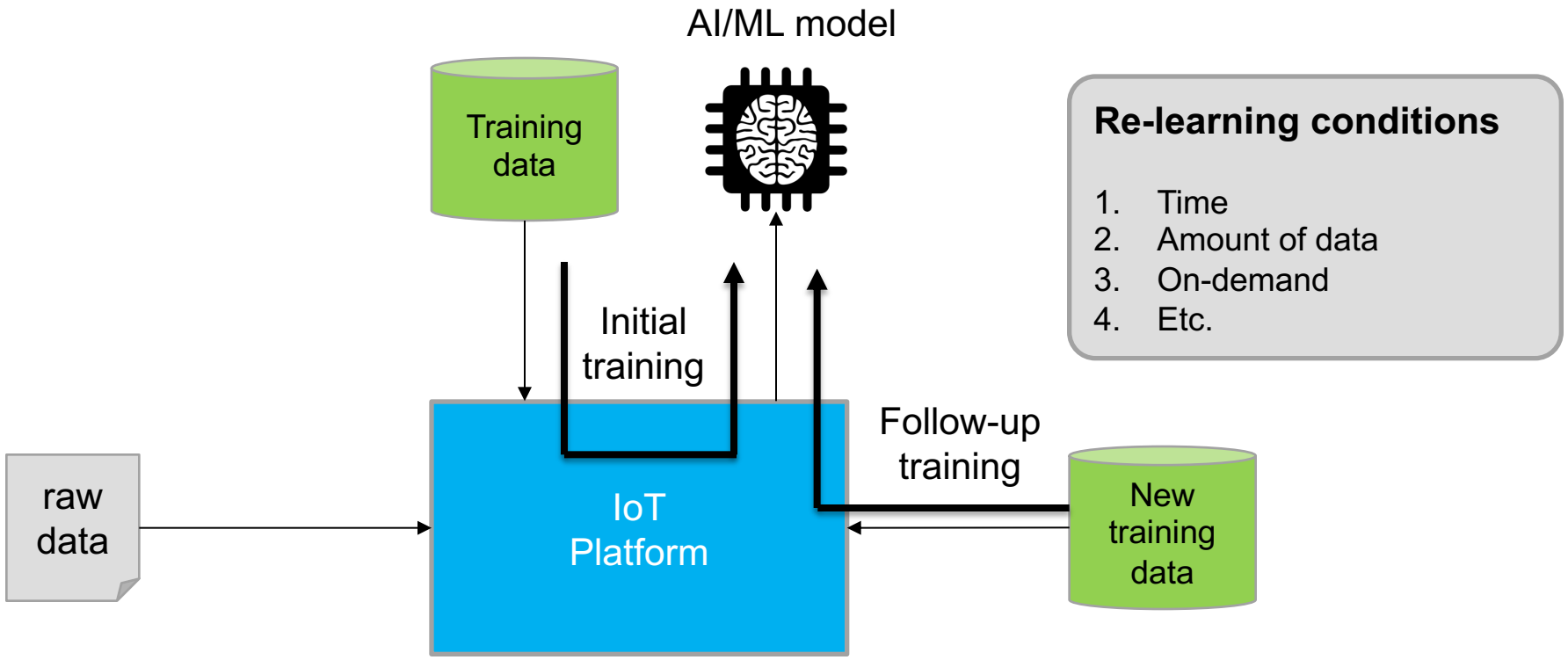
Use case #2 – Last Mile Delivery

- Last Mile is a term used in supply chain management and transportation planning to describe the last leg of a journey comprising the movement of people and goods from a transportation hub to a final destination.
- IoT platform can provide the following functions:
  - » Manage structured and unstructured data for training
  - » Update trained model using new inputs everyday
  - » Classify AI/ML data into two parts, i.e., training and validating



Use case #2 – Last Mile Delivery

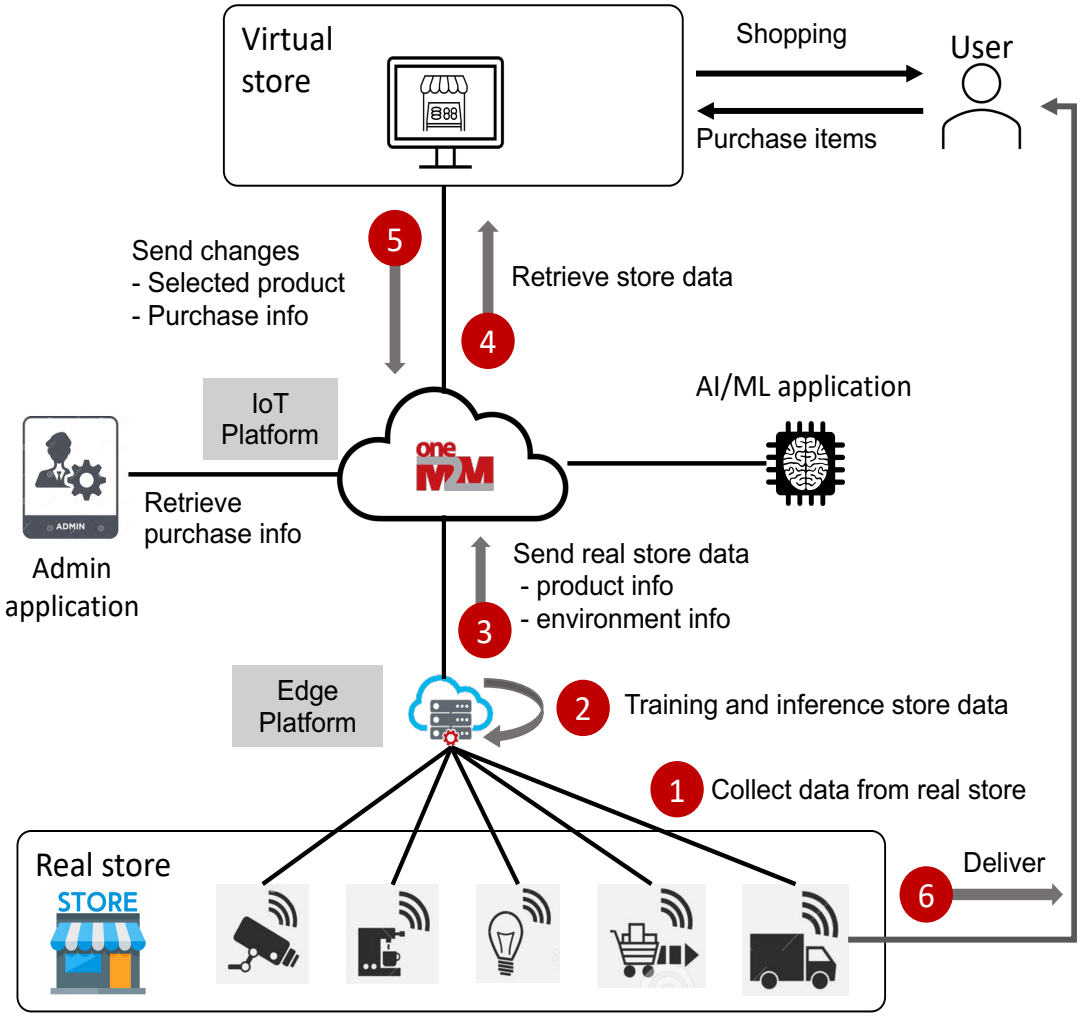
- After initial training to build a model, IoT platform collects data.
- Collected real time data is used for prediction.
- IoT platform collects several data for future training (re-learning)
- When a criteria is satisfied, IoT platform performs re-learning based on the new training data.
  - » Time based
  - » Amount of data
  - » On-demand
  - » Etc.



Use case #3 – Smart Virtual Store using Metaverse

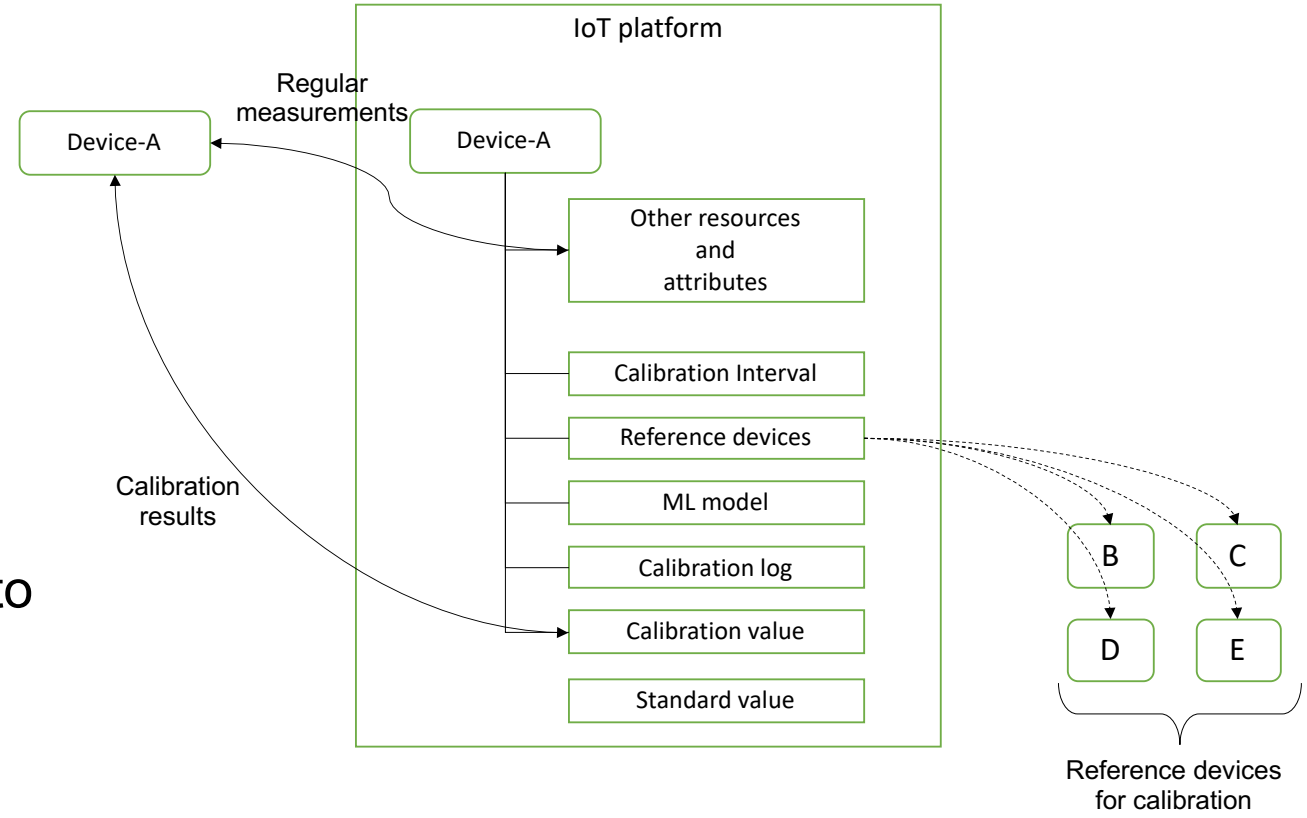
The word "Metaverse" is made up of the prefix "meta" (meaning beyond) and the stem "verse" (a back-formation from "universe");

- Step 1: The Edge node collects data from sensors in the real world store.
- Step 2: The Edge node infers product information from the collected data.
- Step 3: The Edge node send inferred product data to the IoT platform.
- Step 4: The virtual store application in the metaverse retrieves information from the IoT platform about the products in the real world.
- Step 5: A user picks up products from the virtual store. The purchased information is sent to the IoT platform.
- Step 6: An admin application gets a notification for the purchase, and delivers real products to the user.



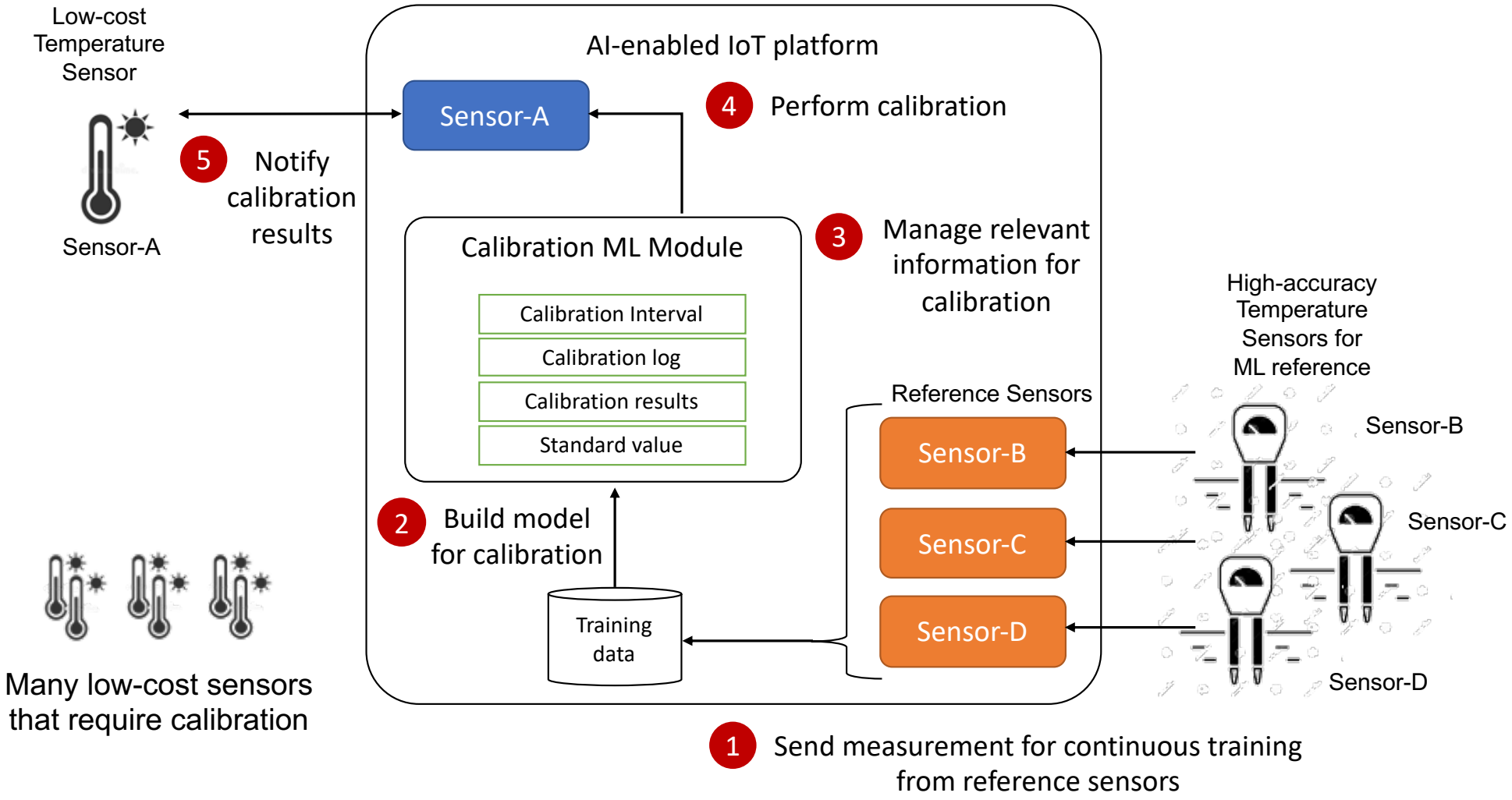
Use case #4 – IoT Device Calibration using Machine Learning

- Basic concept of this idea is to use Machine Learning for continuous IoT device calibration
  - The IoT platform performs a machine learning to generate a calibration value for an IoT device using data collected for a certain period of time from reference devices.
  - An output of the machine learning for calibration is used to calibrate the IoT device.
  - As the IoT device requires calibration regularly, IoT platform can continuously perform machine learning for calibration

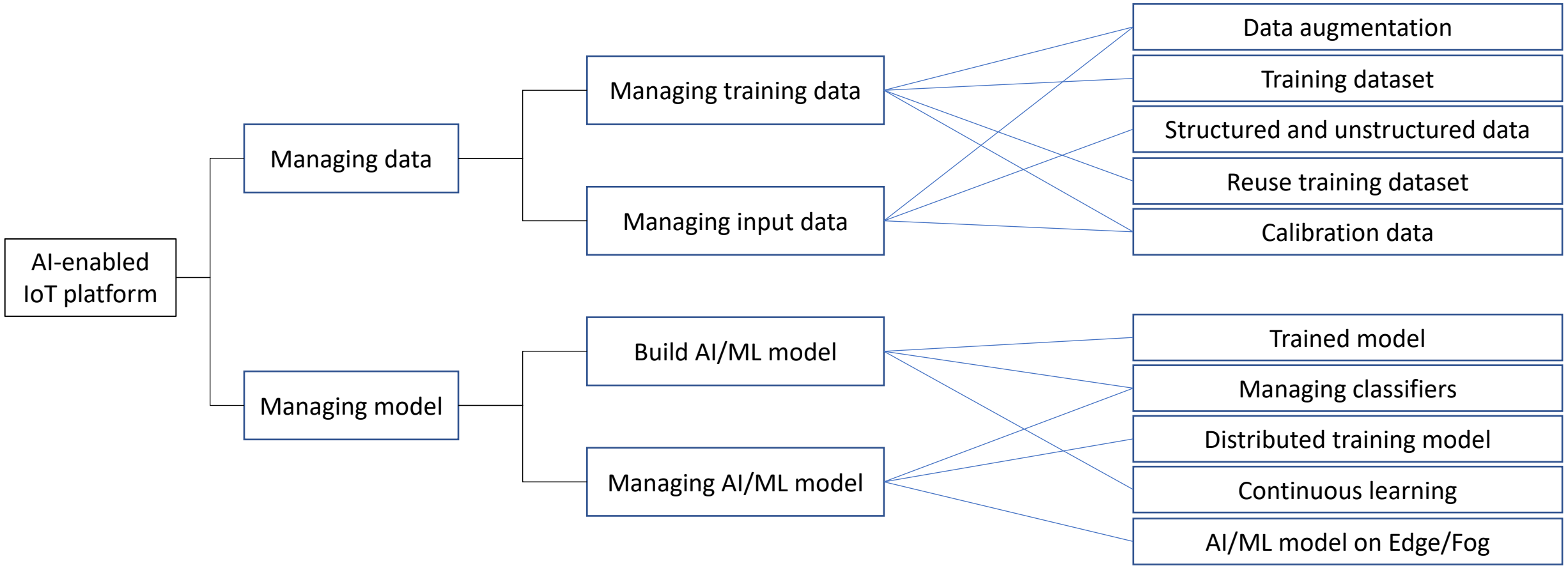


- In order to support the concept of IoT devices calibration using ML, additional information to a normal oneM2M resource and new behaviours to IoT platform are suggested in this patent.

Use case #5 – IoT Device Calibration using Machine Learning

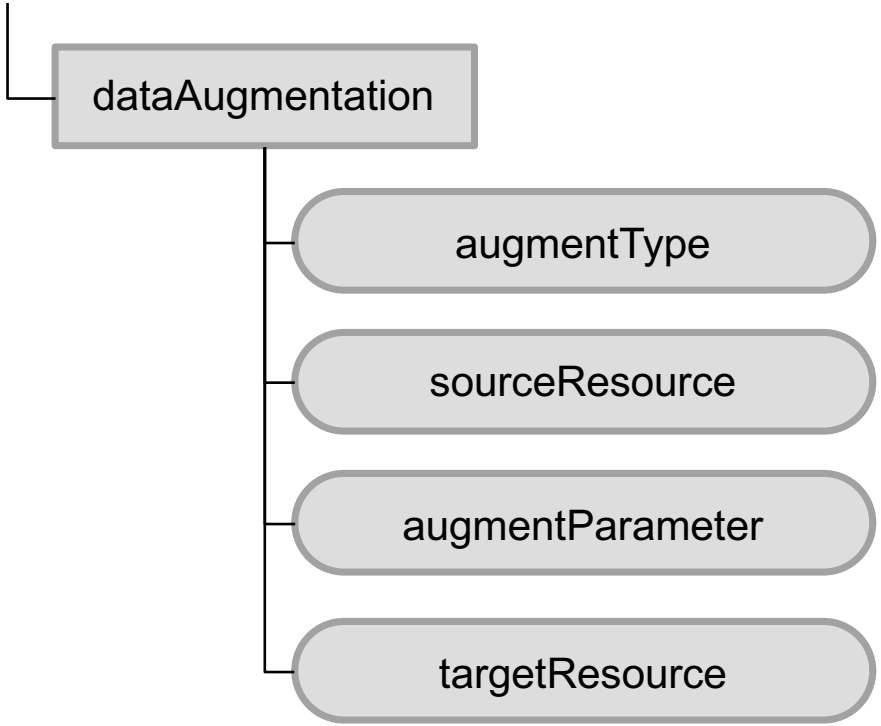


Potential Requirements



Potential solutions (Data Augmentation)

An example structure of [*dataAugmentation*] resource

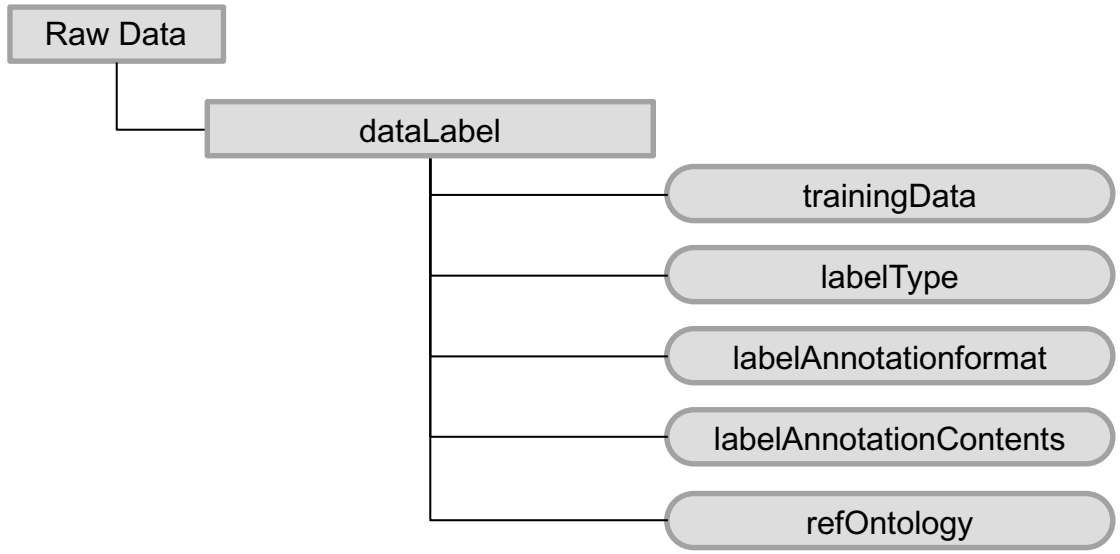


Description of attributes for [*dataAugmentation*] resource

Attribute Name	Description
autmentType	type of data augmentation (e.g., resize, crop, rotate)
sourceResource	a resource (or list) that contains the raw image
augmentParameter	required parameters for the selected augmentation type
targetResource	a resource or a set of resources to store generated images

Potential solutions (Data Labelling)

An example structure of [*dataLabel*] resource



Description of attributes for [*dataLabel*] resource

Attribute Name	Description
trainingData	confirms that this data is for training
labelType	describes labelling type, e.g. square, polygon, line
labelAnnotationFormat	there exist many labelling format such as COCO and YOLO
labelAnnotationContents	contains actual annotation contents following the given format (e.g., XML or JSON)
refOntology	reference ontology used in data label annotation

AI-Enabled IoT based on the oneM2M Standards
(Sejong University & KETI)

ETSI IoT Week 2022

Autonomous Counting System for Smart Store

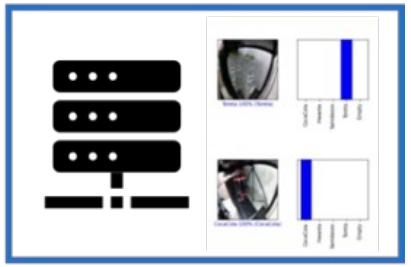
Sensor and Model Management UI

Service UI

Multi-Sensor

Mobius

AI Service Broker



AI Hub

AI-enabled oneM2M System

Make oneM2M platform to support data management for AI and provide AI/ML capabilities

WORK ITEM	
▪ Work Item Title:	System enhancements to support AI capabilities
▪ Document Number	WI-00XX
▪ Supporting Members or Partner type 2	KETI, Hyundai Motors, Exacta GSS, Deutsch Telecom, SBS, Nokia, Hansung University, Orange, Convida Wireless
▪ Date:	2021-05-27
▪ Abstract:	This work item aims to enable oneM2M to utilize Artificial Intelligence models and data management for AI services.
▪ Template Version:23 February 2015 (Do not modify)	

Data License Management

Allow oneM2M to manage data license so support

- **Linked Open Smart City Data**
- **License-based discovery**

WORK ITEM	
Work Item Title:	System enhancements to support Data License Management
Document Number	WI-00XX
Supporting Members or Partner type 2	Hyundai Motors, KETI, Deutsche Telecom, Telecom Italia, Convida Wireless, BT, Orange
Date:	2020-05-20
Abstract:	Proposes a work item to study oneM2M system enhancement to support data license management.

Data Protection Regulations

Make oneM2M platform to be compliant with Data Protection Regulations such as GDPR and PIPA

WORK ITEM	
Work Item Title:	System enhancements to support Data Protection Regulations
Document Number	WI-00XX
Supporting Members or Partner type 2	Hyundai Motor, KETI, BT, SyncTechno Inc., Hansung University, EGM
Date:	2019-09-27
Abstract:	Proposes a work item to study oneM2M system enhancement to support data protection regulations such as General Data Protection Regulation from EU.

Thank you