ETSI TS 103 779 V1.1.1 (2022-05)

SmartM2M;
Requirements and Guidelines for cross-domain data usability of IoT devices

Reference
DTS/SmartM2M-103779

Keywords
artificial intelligence, data usability, IoT, oneM2M, use case

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

If you find a security vulnerability in the present document, please report it through our Coordinated Vulnerability Disclosure Program:
https://www.etsi.org/standards/coordinated-vulnerability-disclosure

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM®** and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Smart Machine-to-Machine communications (SmartM2M).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

The growth of the number of IoT devices and the data generated by those devices has led to the need to process a large amount of data. Processing such large amounts of data has become more challenging and has led to increased use of automated processing such as machine learning. Effective use of this data for decisions has been seen to depend on the quality of information used for modelling and how the systems work and interact together. The use of machine learning to process data has led to a debate on data gathering, data ownership, data transparency, and data bias that is going well beyond technical matters (privacy, regulation, remuneration schemes). The (negative) impact of poor-quality training data is obvious, especially in health applications, road travel, etc.

IoT devices and platforms also provide data that are used directly by human and very often non-technical users. This is the case for example for medical teams and their patients in the medical sector, mechanics in the automotive sector or first responders in the emergency sector. Trust in the IoT system can be ensured only if these data bring in a real added-value and are delivered in a non-ambiguous manner to these users.

In AI, in many cases, the source of poor or incorrect results is because of machine learning models that have biased outputs which can be traced back to lack of sufficient or poorly classified training data. Developing trained models is time and compute intensive and poor data used in training can results in the need to retrain which can take time (and therefore money). Models based on poor data can have unintended consequences from incorrectly classifying new data that can lead to expensive failures or negative social outcomes. As they become used in more critical use cases the results can be catastrophic, such as could be the case of failure in an autonomous vehicle. Similar impact would arise from poor data when IoT devices provide information to non-technical persons or to monitoring algorithms.

Creating more accurate machine learning models can be greatly enhanced by improving the quality and quantity of classified training sets. To emphasize the point, it is not a lack of data but the lack of classified data that impacts the machine learning algorithms. Recommendations in the present document include clearly describing the generated data at all stages of a machine learning pipeline, including:

- a description of the data from an IoT sensor with a common ontology
- a description of the environment the sensor data was collected with a common ontology
- storage in manner that makes the collected data shareable and discoverable
- classification of the data (either manually or by machine learning algorithms) with a common ontology
- traceability of all the sources of classification

The recommendations captured in the present document address the full machine learning pipeline. For maximum benefit the entire system should apply these recommendations, but each individual component or actor in the system can implement the relevant guidelines to provide a better outcome for the usability of the data generated from sensors and machine learning based solutions.

The intended audience of the present document are IoT sensor module developers, IoT platform and service providers, machine learning model developers, application developers and IoT consumers.

IoT sensor module developers are at the start of the pipeline and improvements in the characterization of data generated by the sensor can have a significant impact on its use throughout the pipeline. The data generated should be described with an appropriate common ontology that will make discovery and use of the data easier.

IoT platform and service providers can make data easily available and easy to annotate with the information needed for proper utilization through the lifecycle of the data generation, classification, and consumption.

Machine Learning algorithm developers will be able to find good data easier and subsequently they will generate better models. Additionally, machine learning models should generate labels with an appropriate common ontology that will make discovery and use of the data easier.

Application developers will be able to find and use machine learning models that are relevant to their application use case and know exactly what data is suitable for a discovered model and which models are suitable for their data sources.

IoT consumers in this context are those intending to make use of a solution that includes an IoT device or system. These recommendations can be used as a checklist for any solution considered for deployment.

1 Scope

The present document has the objective:

- to define minimum requirements for data and services usability on professional and general public IoT devices and platforms, whether they are critical or not;
- to develop a horizontal cross-domain specification encompassing these requirements.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] ETSI TS 103 264: "SmartM2M; Smart Applications; Reference Ontology and oneM2M Mapping".

[2] ETSI EN 303 645: "CYBER; Cyber Security for Consumer Internet of Things: Baseline Requirements".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI TR 103 778:"SmartM2M; Use cases for cross-domain data usability of IoT devices".

[i.2] E Goldstein, U Gasser, and B Budish: "Data Commons Version 1.0: A Framework to Build Toward AI for Good", 2018.

NOTE: Available at https://medium.com/berkman-klein-center/data-commons-version-1-0-a-framework-to-build-toward-ai-for-good-73414d7e72be (Accessed 15 November 2021).

[i.3] 3GPP TS 22.891 (V14.2.0): "Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers", September 2016.

[i.4] 3GPP R1-162204: "Numerology requirements", April 2016.

[i.5] M Chen, Y Miao, Y Hao, and K Hwang: "Narrow band internet of things", IEEE Access, vol. 5, pp. 20557-20577, 2017.

[i.6] Z He: "Automatic cooking system", US Patent App. 16/155,895, February 2019.

[i.7] F Adelantado, X Vilajosana, P Tuset-Peiro, B Martinez, J Melia-Segui, and T Watteyne: "Understanding the limits of lorawan", IEEE Communications Magazine, vol. 55, pp. 34-40, September 2017.

[i.8] C Yi, J Cai, and Z Su: "A multi-user mobile computation offloading and transmission scheduling mechanism for delay-sensitive applications", IEEE Transactions on Mobile Computing, 2019.

[i.9] A Pal and K Kant: "Nfmi: Connectivity for short-range iot applications", Computer, vol. 52, pp. 63-67, February 2019.

[i.10] M Merry: "Environmental problems that batteries cause", Sciencing, March 2019.

[i.11] A Froytlog, T Foss, O Bakker, G Jevne, M A Haglund, F Y Li, J Oller, and G Y Li: "Ultra-low power wake-up radio for 5g iot", IEEE Communications Magazine, vol. 57, no. 3, pp. 111-117, 2019.

[i.12] Z Qin, F Y Li, G Y Li, J A McCann, and Q Ni: "Low-power wide-area networks for sustainable iot", IEEE Wireless Communications, 2019.

[i.13] B Safaei, A M H Monazzah, M B Bafroei, and A Ejlali: "Reliability side-effects in internet of things application layer protocols", in 2nd International Conference on System Reliability and Safety (ICSRS), pp. 207-212, IEEE, 2017.

[i.14] N A Mohammed, A M Mansoor, and R B Ahmad: "Mission-critical machine-type communications: An overview and perspectives towards 5G", IEEE Access, 2019.

[i.15] M B Mollah, S Zeadally, and M A K Azad: 'Emerging wireless technologies for internet of things applications: Opportunities and challenges', 2019.

[i.16] J Wu and P Fan: "A survey on high mobility wireless communications: Challenges, opportunities and solutions", IEEE Access, vol. 4, pp. 450-476, 2016.

[i.17] M Ryu, J Yun, T Miao, I-Y Ahn, S-C Choi, and J Kim: "Design and implementation of a connected farm for smart farming system', in IEEE SENSORS, pp. 1-4, IEEE, 2015.

[i.18] L F Ochoa, G P Harrison: "Minimizing energy losses: optimal accommodation and smart operation of renewable distributed generation", IEEE Trans Power Syst, 26 (1), pp. 198-205, 2011.

[i.19] T Hedberg Jr, S Krima, J A Camelio: "Embedding X.509 digital certificates in three-dimensional models for authentication, authorization, and traceability of product data", Journal of Computing and Information Science in Engineering 17(1):11008-11011, 2016.

NOTE: Available at https://doi.org/10.1115/1.4034131.

[i.20] T Hedberg Jr, S Krima, J A Camelio: "Method for enabling a root of trust in support of product data certification and traceability", Journal of Computing and Information Science in Engineering 19(4):041003, 2019.

NOTE: Available at https://doi.org/10.1115/1.4042839.

[i.21] D Yaga, P Mell, N Roby, K Scarfone: "Blockchain technology overview", National Institute of Standards and Technology, Gaithersburg, MD, 2018.

NOTE: Available at https://doi.org/10.6028/NIST.IR.8202.

[i.22] S Krima, T Hedberg Jr, A Barnard Feeney: "Securing the digital threat for smart manufacturing", National Institute of Standards and Technology, Gaithersburg, MD, AMS 300-6, 2019.

NOTE: Available at https://doi.org/10.6028/NIST.AMS.300-6.

[i.23] D Wu, M J Greer, D W Rosen, D Schaefer: "Cloud manufacturing: Strategic vision and state-of-the-art", Journal of Manufacturing Systems 32(4):564-579, 2013.

NOTE: Available at https://doi.org/10.1016/j.jmsy.2013.04.008.

[i.24] X Vincent Wang, X W Xu: "An interoperable solution for cloud manufacturing", Robotics and Computer-Integrated Manufacturing 29(4):232-247, 2013.

NOTE: Available at https://doi.org/10.1016/j.rcim.2013.01.005.

[i.25] L Zhang, Y Luo, F Tao, B H Li, L Ren, X Zhang, H Guo, Y Cheng, A Hu, Y Liu: "Cloud manufacturing: a new manufacturing paradigm", Enterprise Information Systems 8(2):167-187, 2014.

NOTE: Available at https://doi.org/10.1080/17517575.2012.683812.

[i.26] L Ren, L Zhang, L Wang, F Tao, X Chai: "Cloud manufacturing: key characteristics and applications", International Journal of Computer Integrated Manufacturing 30(6):501-515, 2017.

NOTE: Available at https://doi.org/10.1080/0951192X.2014.902105.

[i.27] High Priority IoT Standardisation Gaps and Relevant SDOs, Release 2.0, Alliance for Internet of Things Innovation (AIOTI), January 2020.

NOTE: Available at https://aioti.eu/wp-content/uploads/2020/01/AIOTI-WG3-High-Priority-Gaps-v2.0-200128-Final.pdf.

[i.28] ETSI TR 103 582: "EMTEL; Study of use cases and communications involving IoT devices in provision of emergency situations".

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in ETSI TR 103 778 [i.1] and the following apply:

data consumer: AI, monitoring algorithm or human that uses the data provided by an IoT platform or device

NOTE: After the data consumer has used the data, they remain available for further usage.

ML algorithms: specific algorithms used to analyse data as well as any pre-processing or post-processing performed on the data before use in the ML algorithm

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

4G/5G	4th/5th Generation (mobile networks)
AI	Artificial Intelligence
AI/ML	Artificial Intelligence/Machine Learning
AIOTI	Alliance for Internet of Things Innovation
API	Application Programming Interface
CPU	Central Processing Unit
CSV	Comma Separated Value
DCAT	Data CATalogue vocabulary
HTTP	Hyper Text Transfer Protocol
IoT	Internet of Things
IP	Intellectual Property
JSONLD	JavaScript Object Notation for Linked Data
MIMO	Multiple-Input and Multiple-Output

ML	Machine Learning
mMIMO	massive MIMO
NOMA	Non-Orthogonal Multiple Access
RDF	Resource Description Framework
ROI	Return Of Investment
SAREF	Smart Applications REFerence ontology
SDMX	Statistical Data and Metadata eXchange
SKOS	Simple Knowledge Organization System
UAV	Unmanned Aerial Vehicle
URI	Universal Resource Identifier

4 Recommendations for data usability

ETSI TR 103 778 [i.1] identifies and describes use cases where the IoT data and services require data usability for humans and for machines consuming data for AI (for example machine learning). The data that IoT devices and platforms provide should be easily accessed to all authorized users, understood and acted upon by a large non-technical public in the case of humans (e.g. medical teams and their patients in the medical sector, mechanics in the automotive sector, first responders in the emergency sector, etc.) and by machines and processes when the data are fed to the AI components of a system (e.g. machine learning). Its main objective is to analyse these use cases to derive requirements and guidelines towards a horizontal cross-domain standard, with the specification of minimum requirements for data usability of professional and general public IoT services, whether they are critical or not. In that aim, ETSI TR 103 778 [i.1] analyses the impact of these use cases from the data usability point of view for both machines (algorithms and AI/ML) and humans. The present document fulfils one of the standardization gaps identified in the AIOTI report published in 2020 [i.27]. It also includes part of the recommendations that were produced in the use case analysis of ETSI TR 103 582 [i.28].

Potential solutions build up a list of what can mitigate the identified issues with the intent of decreasing the likelihood of these issues. Each use case has been analysed again to determine which potential solutions could be applied and then identify the residual impact assessment, with a goal to have the minimal residual impacts for each use case.

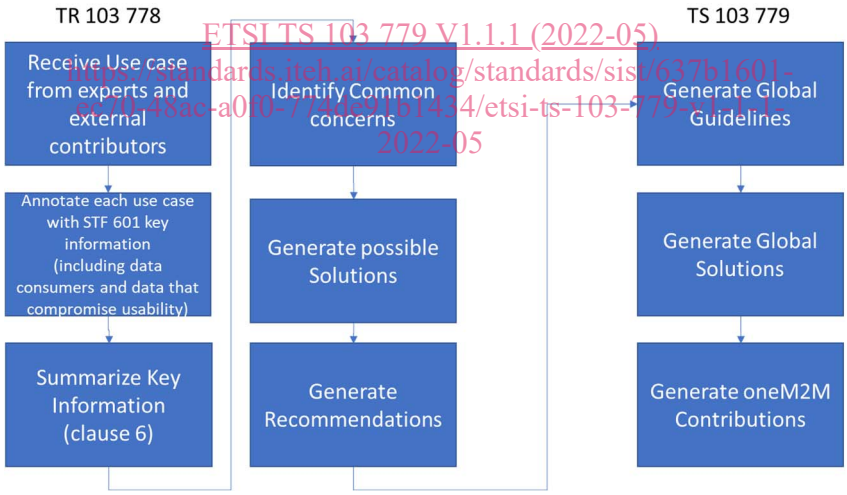

Figure 1: Link between the use cases and the specifications

This clause contains a summary describing the major points of attention to consider when an AI system is deployed. It provides a table describing a list of recommendations grouped by type and, for each of them, the recommendation that may be addressed to handle some of the impact to issues raised under the use cases that have been described in ETSI TR 103 778 [i.1]. The aim of this clause is to connect the outcomes of the work performed in ETSI TR 103 778 [i.1] with the set of requirements provided in clause 5.

Table 1: Summary of recommendations in ETSI TR 103 778 [i.1]

Category	Recommendation	Description
Setup	IoT infrastructure/devices bootstrap.	Easy way for sensor data to be directed to a data consumer (human or ML algorithm). Each deployed IoT infrastructure/device has to be properly setup in order to grant an efficient and effective flow of involved data. During the bootstrap operation it is necessary to check if all data gathered by sensors are easily provided to the target data consumers. Target data consumers may be either humans or ML algorithms.
	Data format description and intelligibility.	Data formats used within a deployed IoT infrastructure/device have to be properly described in order to avoid ambiguity for the target data consumers using such data. Target data consumers may be either humans or ML algorithms.
Configuration	Mitigation of data heterogeneity.	A complex IoT infrastructure/device may include data produced by means of different data formats (e.g. different sensor manufacturers, external API services). It may be necessary to foresee operations to mitigate the data heterogeneity. Such an operation is necessary to standardize the input data format exploited by ML algorithms and/or humans. Use of ontologies thought for specific domains (e.g. SAREF [1]) can be foreseen.
	Data quality.	Each IoT infrastructure/device has to be accompanied with appropriate metadata (e.g. accuracy) for each data source used, of the granularity and frequency with which each data source provides data. Such information is exploited for determining the reliability and suitability of data sources in different scenarios as well as for understanding how to configure ML algorithms to better exploit such data.
Machine Learning or monitoring output	Explainability.	Transparency is one of the most important challenges to address in ML field. Associated with the output produced by a ML algorithm (e.g. classification of an object based on the features provided as input), it is important to reconstruct the classification process through the meaning provided to the data of interest generated such a classification.
	Terminology.	Misunderstanding concerning the usage of terms is common. The definition of a precise vocabulary associated with the output produced by ML algorithms and with the meaning of each data feature is recommended. The usage of an ontology may be a proper way for providing such a terminology.
	Output management.	Output provided by ML algorithms has to be stored and described within an effective and efficient repository. Such a repository works as an enabler for making data easy to find for target data consumers and for supporting the retrieval and understanding of important information linked with them.
	Data duplication.	Data duplication is an issue that may affect the effectiveness of ML algorithms. This may happen when multiple instances of the same raw data are stored within the same repository. This fact may lead to the generation of biases during the building/update of classification models due to the usage of same data instance more than once.
	Traceability.	It is necessary to reconstruct the classification process through the identification of the ML modules providing specific outputs. This need is the basis for preserving the traceability of the data flow within the entire infrastructure.
IoT system operation	Data coordinates.	IoT infrastructure/device has to label data provided with both timing and location information when they are used in scenarios exploiting such information.
	Data access.	The deployment of an IoT infrastructure/device has to ensure a precise policy for managing the authorization to access data by all authorized data consumers and not authorized data consumers.
	IoT data interoperability.	IoT infrastructure/device may include IoT devices provided by different manufacturers adopting, in turn, different data format and exporting methods. It is recommended the integration of data interoperability modules for supporting the effective and efficient sharing of data provided between different IoT systems.

Category	Recommendation	Description
	Maintenance of IoT infrastructure/devices.	Complex IoT infrastructure/device has to define a maintenance policy ensuring the proper monitoring and maintenance of all components.
Security	Preservation of integrity, privacy and security.	All components of the deployed IoT infrastructure/device have to be compliant with standards and regulations related to privacy and security of data. Specific procedures have to be put in place for avoiding/managing data integrity breaches.

5 Requirements and guidelines for preserving data usability

5.1 General considerations

This clause contains the essential requirements to follow for preserving the data usability. Here, an abstract conceptual model is provided for giving unambiguous definitions of each guideline and, at the same time, to pave the way for future developments.

With respect to the content of Annex A, this clause provides the main contribution towards the objectives of the present document, i.e. the set of requirements to preserve the data usability aspect within IoT-based systems. While, in Annex A, the intention is to provide a more in-depth analysis of specific research-wise aspects associated with data usability that have been analysed in the literature and that are relevant for the topic treated in the present document.

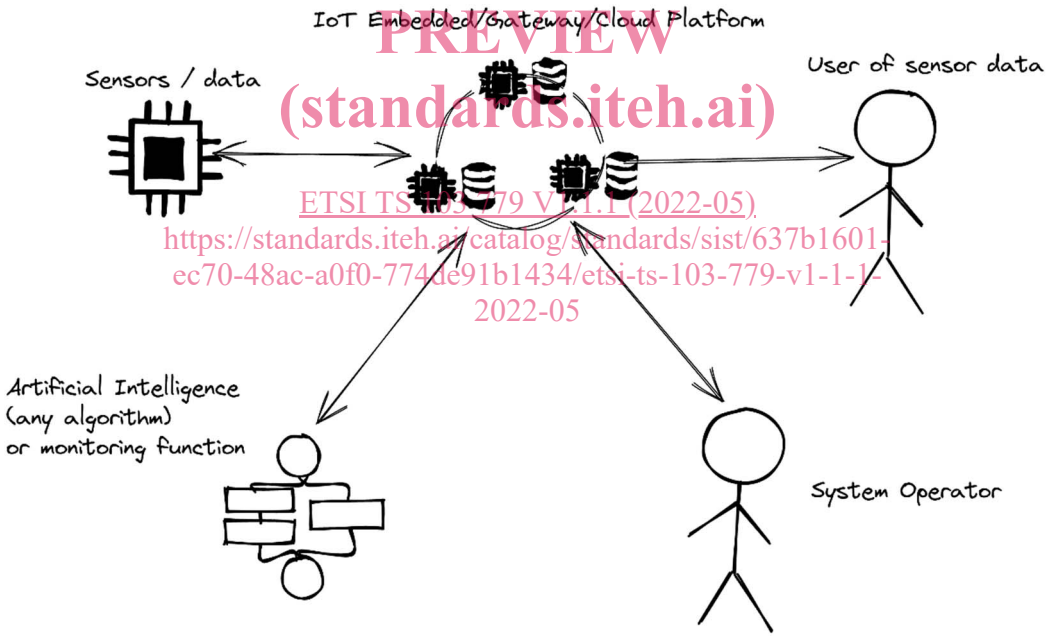

Figure 2: Architecture of an ML deployment