ETSI GS ZSM 012 V1.1.1 (2022-12)

GROUP SPECIFICATION

Zero-touch network and Service Management (ZSM); Enablers for Artificial Intelligence-based Network and Service Automation

Disclaimer

The present document has been produced and approved by the Zero-touch network and Service Management (ZSM) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference

DGS/ZSM-012_AI_Enablers

Keywords

artificial intelligence, automation, network

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 Group Specification (GS) has been produced by ETSI Industry Specification Group (ISG) Zero-touch network and Service Management (ZSM).

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 goal of ZSM is to enable zero-touch automated network and service management in a multi-vendor environment. Current techniques (e.g. rule-based management) require significant involvement of the operator. In order to achieve zero-touch automation, the involvement of the operator in network management tasks must be reduced. One way to achieve this is through Artificial Intelligence (AI). Through the union of AI and network and service management, the effort of network management operations can be significantly reduced. AI can be applied to several high potential areas such as:

- Network planning, optimization, and Service provisioning.
- Service assurance by prediction, anomaly detection, and correlation of events.
- Transforming the operator experience in adapting control and supervision interactions through machine reasoning, human/AI interaction, and scalability.

- Certain security aspects e.g. AI-based threat detection and mitigation.
- Intent fulfilment, e.g. learn what actions are more efficient and impactful to realize intents in given contexts with self-evaluation and self-measurement capabilities.

To maximize the full potential of AI in network and service automation, enabling seamless AI integration and evolution from operation to mission autonomy is required. Moreover, a comprehensive set of AI enablers should be specified to increase the scope of interoperability and to ensure that AI is trusted and capable of delivering - continuously and reliably - required business targets. Such enablers include capabilities to:

- Ensure the infrastructure supports the AI application execution requirements and constraints.
- Provide access to the right data, at the right place, and at the right time.
- Support AI techniques to interpret, recommend and act, while shifting operators' role towards formulation of higher-level declarative behavioural requirements and goals for the AI solutions.
- Govern the operation of AI applications.
- Support coordination for AI solutions.

1 Scope

The present document specifies extensions and new capabilities (so-called "AI enablers") for the ZSM framework reference architecture providing support for the automation of management functionalities and operations based on Artificial Intelligence (AI), applicable to end-to-end and per management domain. The set of AI-enabling capabilities is specified as management services, complementing the existing management services defined in ETSI GS ZSM 002 [2]. The focus is on AI-related areas such as data (including data handling and analytics), action, interoperation, governance and execution environment. Furthermore, the use and integration in the ZSM framework of externally provided AI-based capabilities are taken into account. Security and privacy aspects of AI-enabled network and service automation are taken into account, where the details would be addressed in a Security related WI.

The present document considers AI-related scenarios defined in ETSI GS ZSM 001 [1], as well as new scenarios, in order to derive AI-specific requirements. The present document also documents deployment aspects of the above scenarios to validate the applicability of the AI enablers. Related work from standard development organizations, open-source projects and other sources are considered and re-used, where applicable, in the development of the specifications.

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 GS ZSM 001: "Zero-touch network and Service Management (ZSM); Requirements based on documented scenarios".

[2] ETSI GS ZSM 002: "Zero-touch network and Service Management (ZSM); Reference Architecture".

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] European Commission (21/04/2021): "Proposal for a Regulation laying down harmonised rules on artificial intelligence".

[i.2] Kamiran, Faisal, Asim Karim, and Xiangliang Zhang: Reject Option Classification: "Decision theory for discrimination-aware classification". In 2012 IEEE 12th International Conference on Data Mining, pp. 924-929. IEEE, 2012.

[i.3] ETSI GR ZSM 013: "Zero-touch network and Service Management (ZSM); Automation of CI/CD for ZSM services and managed services".

[i.4] ETSI GS ZSM 009-2: "Zero-touch network and Service Management (ZSM); Closed-Loop Automation; Part 2: Solutions for automation of E2E service and network management use cases".

[i.5] ETSI GR ZSM 009-3: "Zero-touch network and Service Management (ZSM); Closed-Loop Automation; Part 3: Advanced topics".

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

explainable Machine Learning: Machine Learning model that can explain its decisions to humans in a comprehensible manner

fair Machine Learning: Machine Learning model which ensures biases in the data and/or model inaccuracies do not result in unwanted preferences towards individuals or groups

Quality of Trustworthiness (QoT): metric that describes or measures the trustworthiness aspects in Machine Learning

NOTE 1: Trustworthiness aspects may include explainability, fairness, robustness, etc.

NOTE 2: ML QoT may apply for ML data or ML model.

robust Machine Learning: Machine Learning model that is resilient to adversarial attacks (e.g. data poisoning, model leakage), that can handle unintentional errors (e.g. missing data, data drift), that have safeguard mechanisms (e.g. fallback to rule-based algorithms) put in place to deal with unexpected outcomes and that are reproducible

trustworthy Machine Learning: Machine Learning model that respects applicable laws, regulations, ethical principles, values, and is robust from a technical perspective while considering its social environment (see [i.1])

NOTE 1: The proposed EU regulation [i.1] for Machine Learning divides Machine Learning systems into three categories:

i) unacceptable-risk Machine Learning systems;

ii) high-risk Machine Learning systems; and

iii) limited- and minimal-risk Machine Learning systems.

NOTE 2: Based on those risk levels, the proposed EU regulation for Machine Learning has put forward a set of seven key requirements that Machine Learning systems should meet for them to be considered trustworthy:

i) human agency and oversight;

ii) technical robustness and safety;

iii) privacy and data governance;

iv) transparency;

v) diversity, non-discrimination, and fairness;

vi) accountability; and

vii) societal and environmental well-being (see [i.1]).

The details on each of those seven requirements are presented in annex C.

3.2 Symbols

Void.

3.3 Abbreviations

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

ACK	Acknowledge
AI	Artificial Intelligence
AIML	Artificial Intilligence and Machine Learning
AIOp	Aritificial Intelligence Operations
API	Application Interface
AppLCM	Application Life Cycle Management
CI/CD	Continuous Integration/Continuous Development
CPU	Central Process Unit
DataOp	Data Operations
DevOp	Development Operations
DML	Decentralized Machine Learning
E2E	End to End
E2ESMD	End to End Service Management Domain
EU	European Union
FFS	For Future Study
FL	Federated Learning
GDPR	General Data Protection Regulation
KPI	Key Performance Index
MD	Management Domain
ML	Machine Learning
MnS	Management Service
QoE	Quality of Experience
QoS	Quality of Service
QoT	Quality of Trustworthiness
RAN	Radio Access Network
RL	Reinforcement Learning
SL	Supervised Learning
SMD	Service Management Domain
WI	Work Item

4 Enabling areas

4.1 Overview

This clause specifies enabling areas to support the broad use of AI in a multi-vendor network and service management environment. These enablers relate to each other and together facilitate the use of AI in achieving zero touch network and service management automation. As depicted in Figure 4.1-1, the areas that are important to facilitating and enabling AI based management and automation are:

- Execution: The execution enabling area is critical for supporting deployment and operation of AI/ML applications. It addresses specific execution requirements e.g. computational requirements, time constraints.

- Data: Data is the lifeblood of AI/ML empowered automation. Providing data access across domains, ensuring the integrity and trustworthiness of the data and whether the data satisfies the required training and inference needs are of high importance for AI/ML applications to ensure correct management and orchestration decisions.

- Action: AI/ML applications play a crucial role in providing optimal control decisions and recommendations. These outputs may target machines, network entities, management domains, or other management functions and understanding AI/ML outputs is important to correctly apply these decisions.

- Governance: The governance enabling area is crucial for ensuring the trustworthiness of AI/ML applications by designing them to respect applicable laws, regulations, ethical principles, and values and be robust from a technical perspective while considering its social environment.
- Inter-AI: The Inter-AI enabling area focuses on supporting the functionalities and interactions between AI/ML applications and application components.

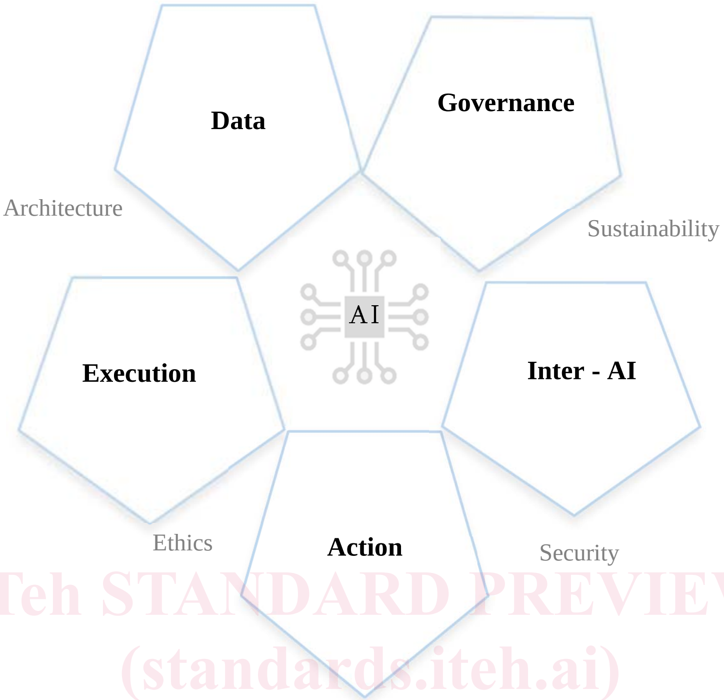

Figure 4.1-1: AI Enabling Areas

4.2 Enabling area: Execution

4.2.1 Description

The execution enabling area is critical for the deployment and operation of AI/ML applications in an operator's network empowered by AI/ML. Each AI/ML application has specific execution requirements that change depending on the operational environment where it is deployed (e.g. on-cloud, on-premises, etc.). These requirements range from computational requirements to time constraints. Matching the AI/ML application with the correct environment/infrastructure capable of meeting these requirements is very important to the successful operation of the application.

Moreover, AI/ML applications can be deployed in multiple locations, layers, and domains of the network. For example, in an operator's network, an AI/ML application may be deployed in the core or RAN. Depending on the use-case and specific solution, one deployment option might be favourable than others. Supporting all possible deployment options, as well as providing a level of coordination across domains (E2E) between different AI/ML applications, is paramount to the possible integration of a wide range of AI/ML applications with different operational, executional, and deployment specific requirements.

Finally, AI/ML applications may have different learning types. For example, a supervised learning application may need historical data sets for training purposes. Once this training is complete, deployment of the application is possible expecting accurate performance. Another example is unsupervised learning and Reinforced learning applications. These type of AI/ML applications learn by experience (e.g. performing actions and observing the effect on the environment). Depending on the use-case and solution specific implementation, an AI/ML application may have one learning type or another. It is of value to the operator's network to be able to support a wide range of possible learning types as well as provide a controlled environment where reinforcement learning based solutions can optimize their performance before being deployed in an operational environment. These controlled environments are usually referred to as Sandboxes and are typically used for testing purposes. Sandboxes are further discussed in ETSI GR ZSM 013 [i.3] Automation of CI/CD for ZSM services and managed services.

4.2.2 Requirements

The following requirements need to be fulfilled by the ZSM framework to enable the above-described aspects of AI/ML empowered network and service automation:

Req-1: ZSM framework shall support the capability to deploy AI/ML instances in a controlled testing environment (sandbox). The sandbox can be a dedicated part of the network, test network, simulation environment or a digital twin of the network.

Req-2: ZSM framework shall support the capability for the seamless integration of AI/ML applications within the ZSM fabric. AI/ML applications should be able to operate smoothly in a single domain or cross domains (E2E) through a properly defined and flexible AppLCM.

Req-3: ZSM framework shall support the capability to instantiate, integrate, chain, decommission and aggregate Data/action pipelines.

Req-4: ZSM framework shall support the capability to dynamically orchestrate and manage data/action pipelines.

4.2.3 Provided management services

4.2.3.1 ML model validation service

The ML model validation service is used to assess the performance or the trustworthiness of the trained ML model under specified conditions (e.g. operational requirements) before deployment. The trained ML models may be validated based on different aspects using a predefined sandboxing environment. Moreover, the validation of ML models may continue after the deployment, if necessary.

NOTE: The aspects of validation may be assessment of ML model performance, ML model trustworthiness or trade-off between ML model performance (e.g. accuracy, utilization of network resources) and ML model trustworthiness (e.g. explainability), etc.

Table 4.2.3.1-1: ML model validation service

Service name		ML model validation service
External visibility		Optional
Service capabilities		
	Request ML model validation (O)	Trigger model validation in predefined sandboxing environment based on defined performance and trustworthiness requirements
	Provide result of the ML model validation (O)	Provide information on ML model validation results

4.2.3.2 Sandbox Configuration Service

The sandbox configuration service enables the consumer to configure sandbox environment and provide reports on the tasks executed with and inside the sandbox, sandbox usage and status. Sandboxing environment supports different types of tasks.

Examples for tasks supported by sandbox:

- Online learning (exploration) in case of reinforcement learning.
- Validation and testing during training as well as inference.

Table 4.2.3.2-1: Sandbox configuration service

Service name		Sandbox configuration Service
External visibility		Optional
Service capabilities		
	Manage sandbox (M)	Manage (create, read, update, delete, list) sandbox environment Update allows to modify configuration parameters of the sandbox environment including CPU, memory, etc.
	Request sandbox report (M)	Request sandbox report on the tasks executed with the sandbox, sandbox usage and status. The request may specify aspects to report on e.g. memory usage, CPU status, task status, etc.
	Provide sandbox report (M)	Provide sandbox report on the tasks executed with the sandbox, sandbox usage and status according to the specification in the request

4.3 Enabling area: Data

4.3.1 Description

In an automated network and service management environment empowered by AI/ML, Data plays a crucial role. Providing data access across domains while satisfying the required data for training and inference ensures correct management and orchestration decisions. Moreover, the integrity and trustworthiness of the data are of high importance before distribution.

To reduce the load on an Operator's network as well as the management framework, data collection techniques can be optimized. This can be done through aggregation of data sources and supporting data pools for enabling the re-use of collected data by multiple AI/ML instances. Moreover, this elevates the need to provide consumers with direct access to data sources. In addition, the correct description of data in the form of metadata facilitates discovery of required data by AI/ML instances or other authorized consumers. More expressive descriptions, containing aspects such as type of data, version, and sampling frequency as well as AI/ML aspects such as labelled vs unlabelled and data statistics, ensures an easier search of required data.

Data privacy and security aspects are paramount for an automated network management environment. Providing data access only to authorized entities/consumers is critical for data governance. Additionally, anonymization and encryption of domain data provides a needed extra layer of privacy for domain specific data. Finally, observing region specific data privacy laws and regulations is important.

4.3.2 Requirements

The following requirements need to be fulfilled by the ZSM framework to enable the above-described aspects of AI/ML empowered network and service automation:

Req-1: ZSM framework shall support the capability to aggregate and reuse multiple data sources to ensure an efficient data distribution mechanism.

Req-2: ZSM framework shall support the capability to collect data, based on AI/ML model data requirements, through qualitative criteria or prediction capabilities.

Req-3: ZSM framework shall support the capability to automatically process collected data in a way that increases data quality and trustworthiness.

Req-4: ZSM framework shall support the capability to provide access to and distribute data under requirements in the same domain and across multiple domains.

NOTE 1: The requirements can be consistency requirements, time requirements such as low latency, real-time, etc.

Req-5: ZSM framework shall support the capability to describe data sets using metadata representation.

NOTE 2: The metadata representation can include data type, version, sampling frequency, labelled/unlabelled data, etc.

Req-6: ZSM framework shall support the capability to pre-process data sets according to the AI/ML model specific requirements.

NOTE 3: The AI/ML model specific requirements can be feature extraction, data labelling, etc.

Req-7: ZSM framework may support region specific laws regarding data privacy during data distribution across domains.

4.4 Enabling area: Inter AI

4.4.1 Description

The Inter-AI enabling area focuses on supporting the functionalities and interactions between AI/ML applications and application components. First, supporting a variety of AI/ML deployment schemes enables the use of AI/ML applications with different requirements and constraints. For example, one AI/ML solution may be completely centrally located while another solution may be distributed across multiple locations/domains. Each of these solutions will have use-case specific constraints deciding the exact deployment schemes (e.g. latency constraints). Another example is Federated learning, in which multiple AI/ML applications are being trained in multiple domains and then aggregated in a central location. Supporting AI/ML specific information exchange such as model parameter and training information enables operators to deploy and make use of AI/ML solutions based on Federated learning.

A very effective method to reduce the load on the operator and the management environment is to reuse existing knowledge and exploit task and domain similarity for different or similar use-cases. For example, an AI/ML application in one domain providing a solution for a specific use-case might have insight and knowledge that can be exploited by another AI/ML application in another domain providing a solution for a similar use-case.

Additionally, Multiple AI/ML applications may cooperate to solve a common problem or provide a common ML enabled solution. For example, the output of one AI/ML application can be used as input to another AI/ML application (i.e. forming a chain or sequence of interlinked modular ML applications). Alternatively, multiple ML models might provide the same type of output in parallel, and their outputs may be merged (e.g. using weights).

Enabling such and other examples of AI/ML application cooperation requires a level of coordination on the domain or E2E level to ensure consistency and concurrency. Finally, it is critical to provide means of proper authentication and trust for access control to AI/ML applications operations.

Some network scenarios require the adaptation of AI/ML applications based on domain information and data to obtain a domain specific AI/ML application. Using transfer learning methods, pre-trained AI/ML applications can be used as starting point for obtaining a domain specific AI/ML application. This method reduces training time and computational resources needed which facilitates rapid deployment of AI applications.

Additionally, due to continuous change of the network and environment, model performance of AI applications may deteriorate over time. Therefore, AI application performance monitoring and evaluation is very important. Furthermore, it is crucial to monitor network and environment statistics to detect potential data drifts (i.e. data distribution changes over the time) and model reality changes. When the performance decreases or when data drift is detected, the AI/ML application should be retrained based on the newly collected data samples.

4.4.2 Requirements

The following requirements need to be fulfilled by the ZSM framework to enable the above-described aspects of AI/ML empowered network and service automation:

Req-1: ZSM framework shall support the capability to manage and orchestrate cross domain AI/ML application training schemes ranging from fully centralized to fully distributed while satisfying different training requirements.

NOTE 1: Example for distributed learning is federated learning. The training requirements can be training data, model parameter transfer, etc.