ETSI TS 123 482 V19.5.0 (2026-04)

TECHNICAL SPECIFICATION

5G;
Functional architecture and information flows for AIML Enablement Service
(3GPP TS 23.482 version 19.5.0 Release 19)

Reference
RTS/TSGS-0623482vj50

Keywords
5G

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

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 IPR online database.

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™**, **LTE™** and **5G™** logo 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.

Legal Notice

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found at 3GPP to ETSI numbering cross-referencing.

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.

Contents

Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

shall indicates a mandatory requirement to do something

shall not indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

should indicates a recommendation to do something

should not indicates a recommendation not to do something

may indicates permission to do something

need not indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

can indicates that something is possible

cannot indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

will indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

will not indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

might indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

might not indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

is (or any other verb in the indicative mood) indicates a statement of fact

is not (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

Introduction

Data analytics is a useful tool for the operator to help optimizing the service offering by predicting events related to the network or slice or UE conditions. 3GPP introduced data analytics function (NWDAF) [2] to support network data analytics services in 5G Core network, management data analytics service (MDAS) [3] to provide data analytics at the OAM, and application data analytics service (ADAES) [4].

In this direction, the support for AI/ML services in 3GPP core network has been studied for providing AI/ML enabled analytics in NWDAF, as well as for assisting the ASP/3rd party AI/ML application service provider for the AI/ML model distribution, transfer, training for various applications (e.g., video/speech recognition, robot control, automotive).

Considering vertical-specific applications and edge applications as the major consumers of 3GPP-provided data analytics services, the AIML enablement (AIMLE) service plays role on the exposure of AI/ML services from different 3GPP domains to the vertical/ASP in a unified manner on top of 3GPP core network and OAM; and on defining, at a SEAL layer, value-add support services for assisting AI/ML services provided by the VAL layer, while being complementary to AI/ML support solutions provided in other 3GPP domains.

This technical specification provides architecture and procedures for enabling AIMLE service over 3GPP networks.

1 Scope

The present document specifies the procedures and information flows necessary for AIML Enablement SEAL Service.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 23.288: "Architecture enhancements for 5G System (5GS) to support network data analytics services".

[3] 3GPP TS 28.104: "Management and orchestration; Management Data Analytics".

[4] 3GPP TS 23.436: "Functional architecture and information flows for Application Data Analytics Enablement Service".

[5] 3GPP TS 23.434: "Service Enabler Architecture Layer for Verticals (SEAL); Functional architecture and information flows".

[6] 3GPP TS 23.501: "System Architecture for the 5G System; Stage 2".

[7] 3GPP TS 23.401: "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access".

[8] 3GPP TS 28.105: "Management and orchestration; Artificial Intelligence/ Machine Learning (AI/ML) management".

[9] 3GPP TS 23.502: "Procedures for the 5G System (5GS)".

[10] 3GPP TS 22.261: "Service requirements for the 5G system".

[11] 3GPP TS 26.531: "Data Collection and Reporting; General Description and Architecture".

[12] 3GPP TS 23.558: "Architecture for enabling Edge Applications".

[13] 3GPP TS 23.273:"5G System (5GS) Location Services (LCS); Stage 2".

[14] 3GPP TS 33.434: "Security aspects of Service Enabler Architecture Layer (SEAL) for verticals".

3 Definitions of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

AIMLE client set identifier: an identifier of the set of selected AIMLE clients.

AI/ML intermediate model: For federated learning, members need to train models for multiple rounds, intermediate models indicate the model which do not meet the required training rounds and/or meet the requirements of the federation training.

AI/ML operation: also known as AI/ML task, it refers to the specific execution step in AI/ML lifecycle, it can include data management, model training etc.

AIMLE service: An AIMLE service is an AIMLE capability which aims assisting in performing or enabling one or more AIML operations.

FL client: An FL member which locally trains the ML model as requested by the FL server. Such FL client functionality can be at the network (e.g. AIMLE server with FL client capability) or at the device side (e.g. AIMLE client with FL client capability).

FL member: An FL member or participant is an entity which has a role in the FL process. An FL member can be an FL client performing ML model training, or an FL server performing aggregation/collaboration for the FL process.

FL server: An FL member which generates global ML model by aggregating local model information from FL clients.

ML model: According to 3GPP TS 28.105 [8], mathematical algorithm that can be "trained" by data and human expert input as examples to replicate a decision an expert would make when provided that same information.

ML model inference: According to 3GPP TS 28.105 [8], ML model inference includes capabilities of an ML model inference function that employs an ML model and/or AI decision entity to conduct inference.

ML model lifecycle: The lifecycle of an ML model aka ML model operational workflow consists of a sequence of ML operations for a given ML task / job (such job can be an analytics task or a VAL automation task). This definition is aligned with the 3GPP definition on ML model lifecycle according to 3GPP TS 28.105 [8].

ML model training: According to 3GPP TS 28.105 [8], ML model training includes capabilities of an ML training function or service to take data, run it through an ML model, derive the associated loss and adjust the parameterization of that ML model based on the computed loss.

Split AI/ML operation pipeline: A Split AI/ML operation pipeline is a workflow for ML model inference in which AI/ML endpoints are organized and collaborate to process ML models in sequential stages, where processing at each stage involves ML model inference on the output of the previous stage.

3.2 Symbols

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

<symbol> <Explanation>

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

ADAES	Application Data Analytics Enablement Server
AIMLE	AI/ML enablement
ASP	Application Service Provider
FL	Federated Learning
NEF	Network Exposure Function
NWDAF	Network Data Analytics Function
OAM	Operation, Administration and Maintenance
SEAL	Service Enabler Architecture Layer
SEALDD	SEAL Data Delivery
VAL	Vertical Application Layer
VFL	Vertical FL

4 Architectural requirements

4.1 General requirements

[AR-4.1-a] The AIML enablement layer shall be able to support one or more VAL applications.

[AR-4.1-b] Supported AIML enablement capabilities shall be offered as APIs to the VAL applications.

[AR-4.1-c] The AIML enablement layer shall support interaction with 3GPP network system to consume network and AI/ML support services.

[AR-4.1-d] The AIMLE client shall be capable to communicate with one or more AIMLE servers of the same AIMLE service provider.

4.2 AIML capability related requirements

[AR-4.2-a] The AIMLE server shall be capable of provisioning and exposing ML client information.

[AR-4.2-b] The AIMLE server shall be capable of supporting the registration, discovery, and selection of AIMLE clients which participate as ML members in AIML service lifecycle.

[AR-4.2-c] The AIMLE layer shall be capable of supporting ML service lifecycle operations (e.g., ML model training).

[AR-4.2-d] The AIMLE server shall be capable of supporting discovery and provisioning of AIML models.

5 Application architecture for enabling AI/ML services

5.1 General

The functional architecture enhancements for the AIML enablement service are based on the generic functional model specified in clause 6.2 of 3GPP TS 23.434 [5]. The architecture enhancements are organized into functional entities to describe a functional architecture enhancement which addresses the support for AIML enablement aspects for vertical applications.

5.2 Application enablement architecture

5.2.1 On-Network AIML Enablement (AIMLE) Functional Architecture

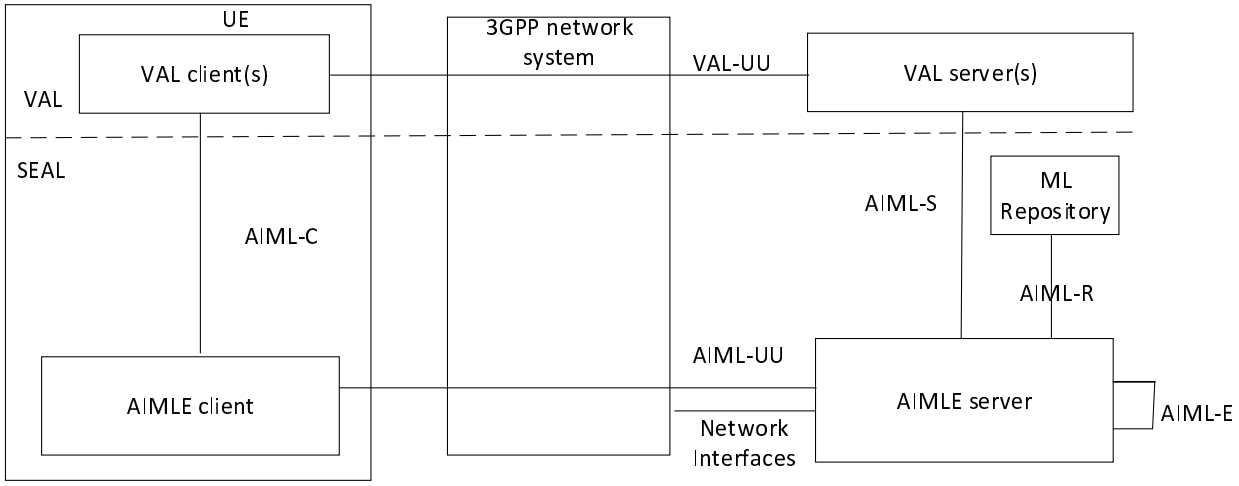

Figure 5.2.1-1: On-network AIMLE functional model

Figure 5.2.1-1 illustrates the on-network functional model of AIMLE. In the vertical application layer, the VAL client communicates with the VAL server over VAL-UU reference point. VAL-UU supports both unicast and multicast delivery modes. The AIMLE functional entities on the UE and the server are grouped into AIMLE client(s) and AIMLE server(s) respectively.

The AIMLE includes of a common set of services for comprehensive enablement of AIML functionality, including federated and distributed learning (e.g., FL client registration management, FL client discovery and selection), and reference points. The AIMLE services are offered to the vertical application layer (VAL).

The AIMLE client communicates with the AIMLE server(s) over the AIML-UU reference points. The AIMLE client provides functionality to the VAL client(s) over AIML-C reference point. The VAL server(s) communicate with the AIMLE server(s) over AIML-S reference points. The AIMLE servers communicate with the underlying 3GPP network systems using the respective 3GPP interfaces specified by the 3GPP network system. AIML-E reference point enables interactions between two AIMLE servers (e.g., central and edge AIMLE servers).

NOTE: AIMLE client can be implemented as a separated software and provide APIs to VAL client over AIML-C as the above. It can also be implemented as part of VAL client.

The AIMLE server interacts with the ML repository which serves as repository for ML model and ML participants over AIML-R.

5.2.1.1 Service-based AIMLE architecture representation

Figure 5.2.1.1-1 exhibits the service-based interfaces for providing and consuming AIMLE services. The AIMLE server could provide service to VAL server and AIMLE client through interface SAiml.

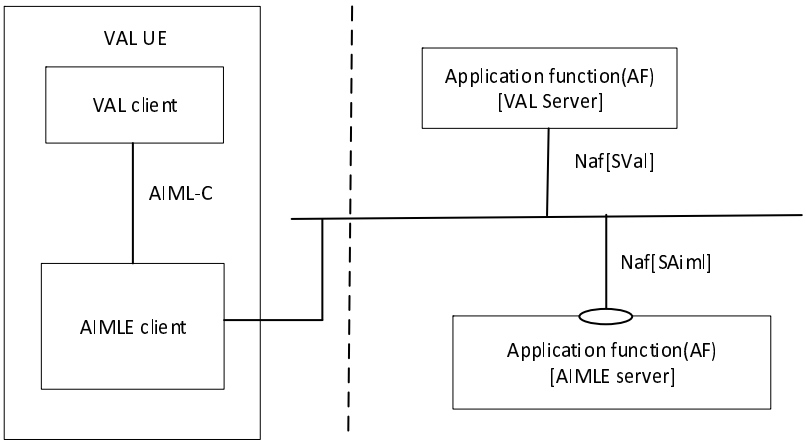

Figure 5.2.1.1-1: Architecture for AIML enablement – Service based representation.

Figure 5.2.1.1-2 illustrates the service-based representation for utilization of the 5GS network services based on the 5GS SBA specified in 3GPP TS 23.501 [6].

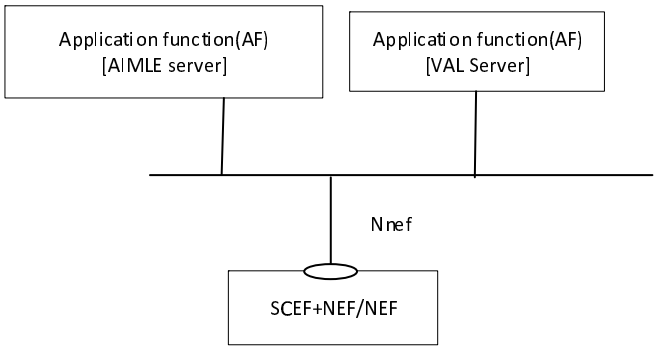

Figure 5.2.1.1-2: Architecture for AIMLE utilizing the 5GS network services based on the 5GS SBA – Service based representation,

The AIMLE server as well as ADAES is deployed as a SEAL server; hence enhancements to SEAL architecture (as specified in 3GPP TS 23.434 [5]) are needed to incorporate the AIMLE service. Figure 5.2.1.1-3 illustrates the service-based representation including AIMLE server as part of the SEAL framework.

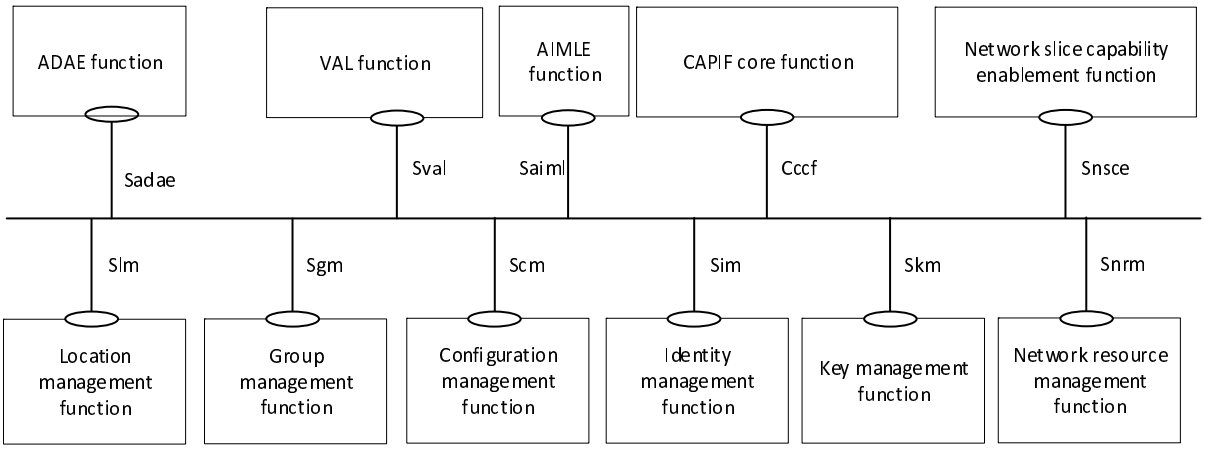

Figure 5.2.1.1-3: SEAL functional model representation using service-based interfaces and including AIMLE function.

5.2.2 Off-Network AIMLE Functional Architecture

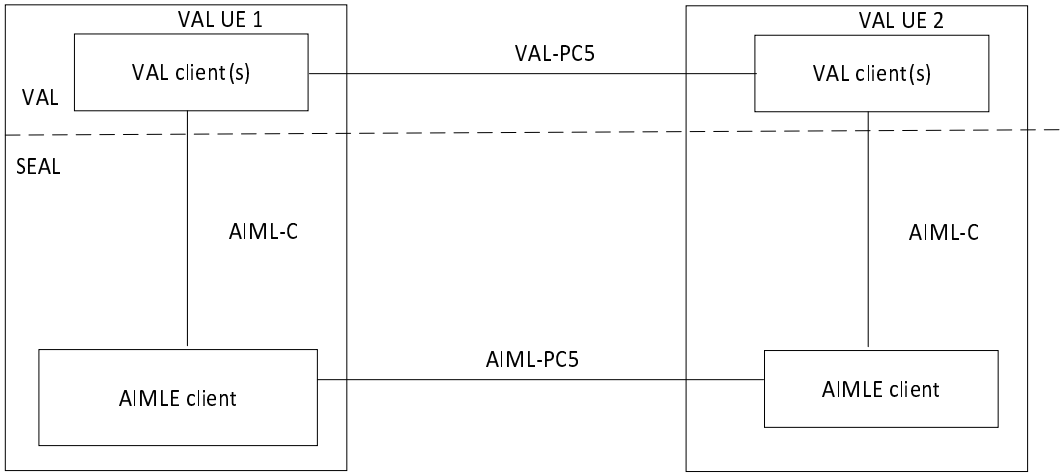

Figure 5.2.2-1: Off-network AIMLE functional model

Figure 5.2.2-1 illustrates the off-network (UE-to-UE) functional model of AIML enablement. In the vertical application layer, the VAL client communicates with a further VAL client over VAL-PC5 reference point. VAL-PC5 supports both unicast and multicast delivery modes. The UE1, if connected to the network via Uu reference point, can also act as a UE-to-network relay, to enable UE2 to access the VAL server(s) over the VAL-UU reference point.

The AIMLE client communicates with a further AIMLE client(s) over the AIML-PC5 reference points. The AIMLE client provides functionality to the VAL client(s) over AIML-C reference point. Such communication is performed for supporting local ML operations (training, distribution, inference) in a coordinated manner.

The off network functional architecture is similar to SEAL off-network architecture (as specified in 3GPP TS 23.434 [5]).

5.2.3 Functional Entities Description

5.2.3.1 General

Each subclause is a description of a functional entity and does not imply a physical entity.

5.2.3.2 AIMLE client

The AIMLE client functional entity acts as the application client supporting AIMLE services. It interacts with the AIMLE server.

5.2.3.3 AIMLE server

The AIMLE server functional entity provides AIMLE services supported within the vertical application layer. It interacts with the AIMLE client, the 3GPP network, other SEAL services, and VAL server.

5.2.3.4 ML repository

The ML repository is a logical entity that serves both as a registry for ML/FL members and as a repository for application layer ML model related information. It can be accessed by the AIMLE server.

5.2.4 Reference Points Description

5.2.4.1 General

The reference points for the functional model related to AIMLE are described in the following subclauses.

5.2.4.2 AIML-UU

The interactions related to AIML enablement functions between the AIMLE client and AIMLE server are supported by AIML-UU reference point. This reference point utilizes Uu reference point as described in 3GPP TS 23.401 [7] and 3GPP TS 23.501 [6].

5.2.4.3 AIML-S

The interactions related to AIML enablement functions between the VAL server(s) and the AIMLE server are supported by AIML-S reference point.

5.2.4.4 AIML-C

The interactions related to AIML enablement functions between the VAL client(s) and the AIMLE client within a VAL UE are supported by AIML-C reference point.

5.2.4.5 AIML-R

The interactions related to AIML enablement functions between the AIMLE server, and the ML repository are supported by AIML-R reference point.

5.2.4.6 AIML-E

The interactions related to AIML enablement functions between AIMLE servers (e.g., central and edge AIMLE servers).

5.2.4.7 AIML-PC5

The interactions related to AIML enablement functions between AIMLE clients in off-network deployments.

6 AIMLE Functional Description

6.1 Support for ML model retrieval

This functionality covers the retrieval of ML models by an AIMLE client or a VAL server via the AIMLE server. The AIMLE server provides the functionality required to retrieve ML models stored in a ML repository based on a filtering criteria. The functionality is provided to the ML model consumer via an API following a request/response model or a subscribe/notify model.

6.2 Support for ML model training

This functionality enables the AIMLE server to support ML model training based on requests from a VAL server. The AIMLE server provides the functionality to train a specific ML model or assist the ML model training at VAL server/client(s).

6.3 Support for FL member registration

This functionality covers the registration, registration update and de-registration of the candidate FL member to the ML repository which is keeping the FL member registrations. Such candidate member can be a VAL server functionality or an enabler layer functionality (e.g. AIMLE server) which is registering to the ML repository/registry to act as FL member for a given application event (analytics event or event triggered by a VAL layer application server).

6.4 Support for FL events subscription and notification

This functionality enables a consumer (who can be the AIMLE server or a VAL server e.g. acting as FL server) to subscribe for FL related events and getting notified on changes on the availability of the FL members which are to be used for the FL-related task (e.g., training). This capability at the ML repository acting as an AIML service registry supports the subscription for events related to FL members and the notification to the consumer in case of changes. This feature assumes that such FL members (AIMLE or VAL server or AIMLE clients) have previously registered to this registry their availability and capabilities.

6.5 Support AI/ML task transfer

This functionality covers the AIMLE support for ML task transfer, which is applicable to scenarios where an AI/ML member cannot finish the assigned AI/ML task during the performing process. In this scenario, the AIMLE server assists the source AI/ML member by support transferring the intermediate AI/ML information (e.g., the intermediate AI/ML operation status and results) to another AI/ML member (target AI/ML member) for further operations to complete the AI/ML task.

6.6 Support for AIMLE client registration

This functionality allows an AIMLE client (e.g., AI/ML capable UEs) to register with an AIMLE server. The AIMLE server stores the client information for future interactions. This functionality is crucial for enabling the AIMLE client to participate in AI/ML operations.

6.7 Support for AIMLE client discovery

This functionality enables the VAL server to discover available AIMLE clients for AI/ML operations, such as training or inference. The AIMLE server provides the functionality to select suitable AIMLE clients that fulfill the discovery criteria.

NOTE: AIMLE client discovery can be also used to discover VAL clients associated with the AIMLE clients. The VAL client may integrate the AIMLE clients as part of the VAL client software.6.8 Support for AIMLE client selection

This functionality enables the selection of AIMLE clients to participate in AI/ML operations. There are two modes for client selection: VAL server selection and AIMLE server selection. In VAL server selection, the functionality is provided by the AIMLE Server selecting candidate AIMLE clients from the client list provided by the VAL Server. In AIMLE server selection, the functionality is provided by the AIMLE Server selecting candidate AIMLE clients based on the client selection criteria provided by the VAL Server from the AIMLE clients in the ML repository.

6.9 Support for AIMLE client participation

This functionality enables the AIMLE server to verify and manage the participation of AIMLE clients in AI/ML operations. The AIMLE client responds with its willingness to perform AI/ML operations based on the information provided by the AIMLE server.

6.10 Support for ML model management

This functionality enables the AIMLE server to manage ML models through interaction with the ML repository. The AIMLE server provides the storage and discovery functionality for ML model information. The storage functionality

allows the AIMLE server to store ML model information in the repository. The discovery functionality enables the AIMLE server to retrieve information about available ML models via an API following a request/response model.

6.11 Support HFL training

This functionality provides the AIMLE server support for horizontal federated learning. This support is applicable to the case where multiple AIMLE clients are expected to locally train the ML model, and the AIMLE server is required to select, configure and coordinate the HFL clients.

6.12 Support AIMLE client selection subscription and notification

This functionality is related to the AIMLE client selection subscription request and notification to enable VAL Servers to subscribe for monitoring AIML members who meet criteria for performing an AIMLE service, selecting AIMLE clients (associated with the VAL clients) and receiving notification when there is an update on the selected and re-selected AIML client's status when re-selection is performed according to AIML member selection criteria.

NOTE: The AIMLE client selection subscription and notification to VAL server is also used for the case where VAL server needs to find some VAL clients (integrating AIMLE client), which can be used by VAL server to request certain AI task operation.

6.13 Support for Split AI/ML Operation

Split AI/ML operation is a type of AI/ML operation that allows distributed processing related to ML models into multiple stages on different processing nodes. The intention is to offload the computation-intensive and energy-intensive AI/ML stages to network endpoints, whereas leave the privacy-sensitive and delay-sensitive stage at the end device as described in 3GPP TS 22.261 [10].

The following functionality is provided to support split AI/ML operation:

- An application consuming services from the AI/ML application enablement layer can discover or manage (e.g., create, update, delete) a split operation profile with the AIMLE server for the purpose of consuming results from corresponding instance of a split AI/ML operation pipeline.
- A VAL server can register with the AIMLE server to indicate its capabilities for acting as a processing node of an instance of a split AI/ML operation pipeline.- An application consuming services from the AI/ML application enablement layer can subscribe with the AIMLE server to receive event notifications related to an instance of a split AI/ML operation pipeline.

NOTE: How to split a ML model is out of scope of this release and the ML model used in a stage needs to be available in the ML repository.

6.14 Support data management assistance

This functionality covers the AIMLE assistance in data management related operations in the ML model lifecycle. AIMLE data management assistance is the process of the AIMLE server assisting AIMLE service consumers with managing data operations (data preparation and processing) performed by VAL clients.

6.15 Support for Transfer Learning enablement

This functionality covers the support for discovering and selecting pre-trained models transfer learning operations in application enablement layer, where the support is based on the request for either an ML task from VAL layer or for an analytics task from ADAES. Transfer Learning enablement allows the consumer to discover the similar ML models to be used as base models for the TL, as well as to support the selection of the best model to be used as pre-trained model.

6.16 Support for FL member grouping

This functionality covers the AIMLE capability to enable the group management of the entities serving as FL clients at the application enablement layer. Such group management is about the creation, monitoring and update of the FL member groups based on the AI/ML operations, which are based on 1) the analytics event/service by ADAES or 2) the VAL requirement for FL support services.

6.17 Support vertical federated learning

This functionality is related to the AIMLE support for vertical FL (VFL) among AIMLE clients serving as VFL members. This capability involves the determination of employing VFL based on the ML model training request, the feature alignment and the decision on the AIMLE clients serving as VFL clients, based on the training capability evaluation functionality.

6.18 Support for ML model training capability evaluation

This functionality enables AIMLE server to request AIMLE client for ML model training capability evaluation to support FL training (e.g. HFL, VFL). AIMLE client evaluates its capability and availability to join the FL training process and responds to the AIMLE server with the evaluation result (e.g. join the FL training process with test result, or not join with fail reason). The ML model training capability evaluation result can be used by the AIMLE server to select FL members for FL training process (e.g. HFL, VFL).

6.19 Support AIML service operations control and management

This functionality enables the VAL server (as AIMLE service consumer) to control the operation mode of an AIMLE service for a given AI/ML operation.

6.20 Support for ML model update

This functionality enables the AIMLE server to update trained and deployed ML models by detecting model performance degradation, and triggering ML model re-training and update to fix the observed degradation.

6.21 Support for ML model performance monitoring

This functionality covers a capability of AIMLE server for monitoring detecting a degradation relation to an ML operation / analytics operation and translating to an ML model degradation (expected or predicted) and performing an action to alleviate this issue (new model training or re-training).

6.22 Support for AIMLE assisted ML model selection

This functionality enables an AIMLE service consumer to request assistance with ML model selection from an AIMLE server. The AIMLE server returns a list of ML models with corresponding model performance.

6.23 Support for AIMLE context transfer in edge data networks

This functionality provides support for AIMLE operations performed by AIMLE clients spread across multiple edge service areas in edge data networks. The functionality allows the context transfer between edge AIMLE servers when an AIMLE client moves to a different edge service area.

6.24 Support for Assisting Hierarchical Computing

This functionality provides support for assisting hierarchical computing by AIMLE servers. The functionality allows the consumer to request hierarchical computing assistance information to support its operations.

7 Identities and commonly used values

7.1 General

The common identities for SEAL and ADAES refer to 3GPP TS 23.434 [5] and 3GPP TS 23.436 [4] respectively. The following clauses list the additional identities and commonly used values for AIMLE Service.

7.2 AIMLE server ID

The AIMLE server ID uniquely identifies the AIML enablement server.

7.3 AIMLE client ID

The AIMLE client ID uniquely identifies the AIML enablement client.

7.4 ML repository ID

The ML repository ID uniquely identifies the ML-related repository function, which is used for storing the ML models and ML related information, as well as for serving as registry for the ML / FL members.

7.5 ML model ID

The ML model ID uniquely identifies the application-layer ML model.

7.6 FL member ID

The FL member ID uniquely identifies the participant entity in a Federated Learning process which is supported by AIMLE service, e.g., AIMLE server ID, VAL server ID or EAS ID.

7.7 AIMLE service area

The AIMLE service area is the area where the AIML Enablement server owner provides its AIML support services. The AIMLE service area can be expressed as a Topological Service Area (e.g., a list of TA), a Geographical Service Area (e.g., geographical coordinates) or both.

7.8 ML model profile ID

The ML model profile ID uniquely identifies a ML model profile that is created by the ML repository as a result of a successful ML model storage request by a AIMLE server.

7.9 AIMLE client set ID

The AIMLE client set ID is a unique identifier assigned by an AIMLE server that is associated with the set of AIMLE clients that have been selected and have agreed to participate in performing AI/ML related operations, e.g., ML model training.

8 Procedures and information flows

The security aspects for the AIMLE services are specified in 3GPP TS 33.434 [14].

8.1 General

8.2 ML model retrieval

8.2.1 General

ML model retrieval enable applications consuming services from the AI/ML application enablement layer to retrieve ML models available with the enablement layer.

ML model retrieval enables an AIMLE client or VAL server to obtain ML models of interest. The retrieval of ML models is based on matching ML model retrieval filters provided in the request.

ML model retrieval subscription enables an AIMLE client or VAL server to receive notifications about ML models of interest. The subscription is based on matching ML model retrieval filters provided in the subscription request.

8.2.2 Procedure

8.2.2.1 General

Following procedures are supported for ML model retrieval:

- ML model retrieval request-response procedure;
- ML model subscribe-notify procedures for retrieval of ML models, comprising:
  - Subscription procedure;
  - Subscription update procedure; and
  - Unsubscribe procedure;

8.2.2.2 ML model retrieval

Figure 8.2.2.2-1 illustrates the procedure for an AIMLE client or VAL server to discover and obtain ML models from the AIMLE server.

Pre-conditions:

1. The AIMLE client or VAL server has received information (e.g., URI, IP address) related to the AIMLE Server.
2. The AIMLE client or VAL server has received security credentials authorizing it to communicate with the AIMLE server.

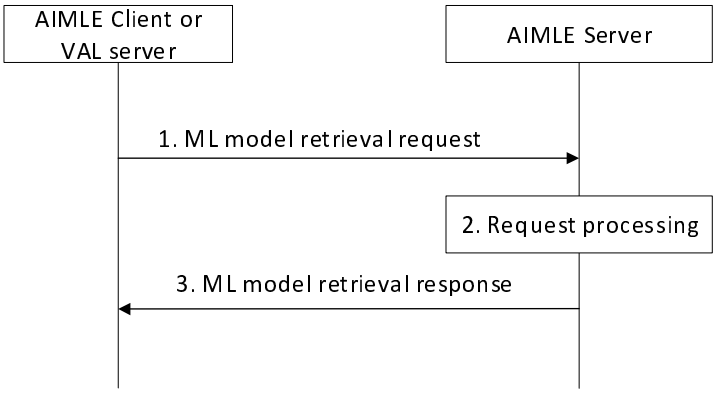

Figure 8.2.2.2-1: ML model retrieval

1. The requestor (e.g., AIMLE client or VAL server) sends a ML model retrieval request to the AIMLE server. The request includes the requestor identifier along with security credentials and may include ML model retrieval filters.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request.

   If the requestor is authorized, the AIMLE server may determine if the requested ML model is available with the AIMLE server based on ML model retrieval filters. If the ML model is not available with the AIMLE server, the AIMLE server performs the ML model information discovery procedure with the ML repository as described in clause 8.11.3.

   If the request is received from an AIMLE client that has registered with the AIMLE server, the AIMLE server determines the ML model based on the AIMLE client profile (Table 8.7.3.2-2), e.g. supported AI/ML model types, application layer AIMLE client capabilities.

3. The AIMLE server sends a ML model retrieval response to the requestor. If the AIMLE server has determined ML models (e.g., locally, or from the ML repository), the response includes an indication of success and may include the ML models. Otherwise, the response includes an indication of failure and may include a reason for failure.

Upon receiving a response including ML models, the requestor can store the ML models and may provide the ML model to the VAL client(s) if the requestor is the AIMLE client.

8.2.2.3 ML model retrieval subscription

8.2.2.3.1 General

Clause 8.2.2.3.2 and clause 8.2.2.3.3 together illustrate the ML model retrieval subscribe-notify ML model.

Clause 8.2.2.3.4 illustrates the ML model retrieval subscription update procedure.

Clause 8.2.2.3.5 illustrates the ML model retrieval unsubscribe procedure.

8.2.2.3.2 Subscribe

Figure 8.2.2.3.2-1 illustrates the procedure for an AIMLE client or VAL server to subscribe with the AIMLE server to be notified of retrieval of ML models.

Pre-conditions:

1. The AIMLE client or VAL server has received information (e.g. URI, IP address) related to the AIMLE Server.

2. The AIMLE client or VAL server has received security credentials authorizing it to communicate with the AIMLE server.

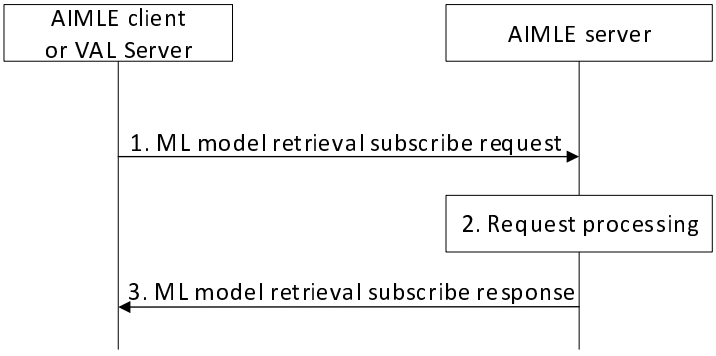

Figure 8.2.2.3.2-1: ML model retrieval subscription

1. The requestor (e.g., AIMLE client or VAL server) sends a ML model retrieval subscribe request to the AIMLE server. The request includes the requestor identifier, security credentials and may include ML model retrieval filters and expiration time.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server creates the subscription and stores the subscription information.

3. The AIMLE server sends a ML model retrieval subscribe response to the requestor. If the AIMLE server has created the subscription, the response includes an indication of success, the subscription identity and may include an expiration time; to maintain the subscription, the requestor shall send a subscription update request before the expiration time, otherwise the ML model retrieval subscription expires. If the AIMLE server has not created the subscription, the response includes an indication of failure and may include a reason for failure.

8.2.2.3.3 Notify

Figure 8.2.2.3.3-1 illustrates the ML model retrieval notify operation between the AIMLE server and an AIMLE client or VAL server.

Pre-conditions:

1. The AIMLE client or VAL server has subscribed for ML model retrieval with the AIMLE Server.

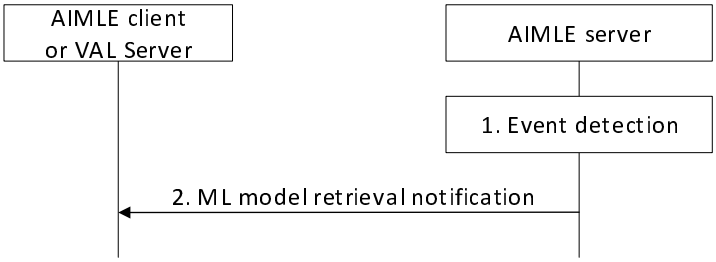

Figure 8.2.2.3.3-1: ML model retrieval notification

1. An event occurs at the AIMLE server that satisfies trigger conditions for notifying a subscriber (e.g., AIMLE client or VAL server). The event may be the detection of availability of new ML model(s) (e.g., new ML model or new revision of a ML model) that satisfies the ML model retrieval filters of the subscription, or changes in the AIMLE client profile. For example, the AIMLE server receives an updated AIMLE client profile with updated "supported AI/ML model type" which indicates the existing model is no longer supported and a new model needs to be retrieved. The detection of availability of new ML models by the AIMLE server is based on ML model storage in the ML repository as described in clause 8.11.2.

2. The AIMLE server sends a ML model retrieval notification to the requestor indicating newly available ML models. The notification includes a subscription identity and newly available ML model(s).

8.2.2.3.4 Subscription update

Figure 8.2.2.3.4-1 illustrates the procedure for an AIMLE client or VAL server to update a subscription with the AIMLE server.

Pre-conditions:

1. The AIMLE client or VAL server has subscribed for ML model retrieval with the AIMLE Server.

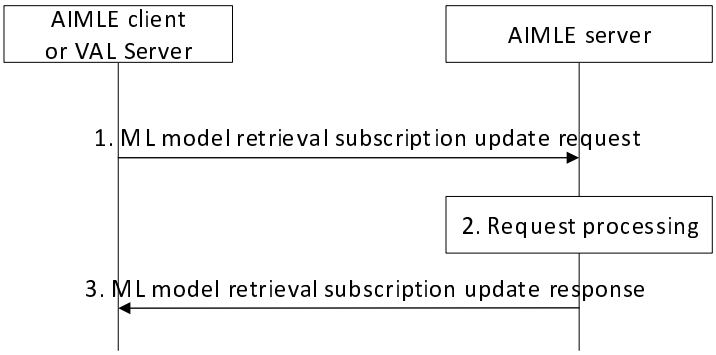

Figure 8.2.2.3.4-1: ML model retrieval subscription update

1. The requestor (e.g., AIMLE client or VAL server) sends a ML model retrieval update request to the AIMLE server. The request includes the requestor identifier, security credentials and the subscription identifier and may include ML model retrieval filters and expiration time.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server updates the subscription information.

3. The AIMLE server sends a ML model retrieval subscription update response to the requestor. If the AIMLE server has updated the subscription, the response includes an indication of success and may include an expiration time. To maintain the subscription, the requestor shall send a subscription update request before the expiration time, otherwise the ML model retrieval subscription expires. If the AIMLE server has not updated the subscription, the response includes an indication of failure and may include a reason for failure.

8.2.2.3.5 Unsubscribe

Figure 8.2.2.3.5-1 illustrates the procedure for an AIMLE client or VAL server to unsubscribe with the AIMLE server.

Pre-conditions:

1. The AIMLE client or VAL server has subscribed for ML model retrieval with the AIMLE Server.

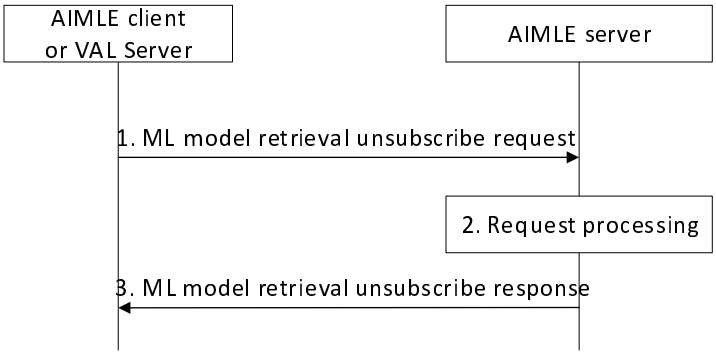

Figure 8.2.2.3.5-1: ML model retrieval unsubscribe

1. The requestor (e.g., AIMLE client or VAL server) sends a ML model retrieval unsubscribe request to the AIMLE server. The request includes the requestor identifier, security credentials and the subscription identifier.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server cancels the subscription.

3. The AIMLE server sends a ML model retrieval unsubscribe response to the requestor. If the AIMLE server has canceled the subscription, the response includes an indication of success. If the AIMLE server has not canceled the subscription, the response includes an indication of failure and may include a reason for failure.

8.2.3 Information flows

8.2.3.1 ML model retrieval request

Table 8.2.3.1-1 describes information elements for the ML model retrieval request from the AIMLE client or VAL server to the AIMLE server.

Table 8.2.3.1-1: ML model retrieval request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor (e.g., AIMLE client or VAL server).
Security credentials	M	The security credentials of the requestor.
Filtering criteria	O	Represents the filtering criteria, which can be any of the ML model information as in Table 8.11.4.1-2.

8.2.3.2 ML model retrieval response

Table 8.2.3.2-1 describes information elements for the ML model retrieval response from the AIMLE server.

Table 8.2.3.2-1: ML model retrieval response

Information element	Status	Description
Successful response	O (NOTE 2)	Indicates that the ML model retrieval request was successful.
>List of ML models	M	List of retrieved ML models; the nested information elements are provided for each ML model in the list.
>> ML model identifier	M	The ML model identifier.
>> ML model retrieval endpoint (NOTE 1)	O	The endpoint (e.g., URL, URI, IP address) where the ML model file can be retrieved.
>> ML model (NOTE 1)	O	The ML model containing a retrieved ML model.
Failure response	O (NOTE 2)	Indicates that the request has failed.
> Cause	O	The cause for the request failure.
NOTE 1: At least one of these information elements is needed. NOTE 2: One of the IEs shall be present.		

8.2.3.3 ML model retrieval subscribe request

Table 8.2.3.3-1 describes information elements for the ML model retrieval subscription request from the AIMLE client or VAL server to the AIMLE server.

Table 8.2.3.3-1: ML model retrieval subscription request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor (e.g., AIMLE client or VAL server).
Security credentials	M	The security credentials of the requestor.
Notification Target Address	M	Notification target address (e.g. URL) where the notifications should be sent.
Proposed expiration time	O	Proposed expiration time for the subscription.
Filtering criteria	O	Represents the filtering criteria, which can be any of the ML model information as in Table 8.11.4.1-2.

8.2.3.4 ML model retrieval subscription response

Table 8.2.3.4-1 describes information elements for the ML model retrieval subscription response from the AIMLE server.

Table 8.2.3.4-1: ML model retrieval subscription response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the ML model retrieval request was successful.
> Subscription ID	M	Subscription identifier corresponding to the subscription.
> Expiration time	O	Indicates the expiration time of the subscription. To maintain an active subscription, a subscription update is required before the expiration time.
Failure response	O (NOTE)	Indicates that the request has failed.
> Cause	O	The cause for the request failure.
NOTE: One of the IEs shall be present.		

8.2.3.5 ML model retrieval notification

Table 8.2.3.5-1 describes information elements for the ML model retrieval subscription response from the AIMLE server.

Table 8.2.3.5-1: ML model retrieval notification

Information element	Status	Description
Subscription ID	M	Subscription identifier corresponding to the subscription.
List of ML models	O	List of retrieved ML models; the nested information elements are provided for each ML model in the list.
> ML model identifier	O	The ML model identifier.
> ML model retrieval endpoint (NOTE)	O	The endpoint (e.g., URL, URI, IP address) where the ML model file can be retrieved.
> ML model (NOTE)	O	The ML model containing a retrieved ML model.
NOTE: At least one of these elements is needed.		

8.2.3.6 ML model retrieval subscription update request

Table 8.2.3.6-1 describes information elements for the ML model retrieval subscription update request from the AIMLE client or VAL server to the AIMLE server.

Table 8.2.3.6-1: ML model retrieval subscription update request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor (e.g., AIMLE client or VAL server).
Security credentials	M	The security credentials of the requestor.
Subscription ID	M	Subscription identifier corresponding to the subscription to be updated.
Notification Target Address	O	Notification target address (e.g. URL, URI, IP address) where the notifications should be sent.
Proposed expiration time	O	Proposed expiration time for the subscription.
Filtering criteria	O	Represents the filtering criteria, which can be any of the ML model information as in Table 8.11.4.1-2.

8.2.3.7 ML model retrieval subscription update response

Table 8.2.3.7-1 describes information elements for the ML model retrieval subscription update response from the AIMLE server.

Table 8.2.3.7-1: ML model retrieval subscription update response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the ML model retrieval subscription update request was successful.
> Expiration time	O	Indicates the expiration time of the subscription. To maintain an active subscription, a subscription update is required before the expiration time.
Failure response	O (NOTE)	Indicates that the request has failed.
> Cause	O	The cause for the request failure.
NOTE: One of the IEs shall be present.		

8.2.3.8 ML model retrieval unsubscribe request

Table 8.2.3.8-1 describes information elements for the ML model retrieval unsubscribe request from the AIMLE client or VAL server to the AIMLE server.

Table 8.2.3.8-1: ML model retrieval unsubscribe request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor (e.g., AIMLE client or VAL server).
Security credentials	M	The security credentials of the requestor.
Subscription ID	M	Subscription identifier corresponding to the subscription to be unsubscribed.

8.2.3.9 ML model retrieval unsubscribe response

Table 8.2.3.9-1 describes information elements for the ML model retrieval unsubscribe response from the AIMLE server.

Table 8.2.3.9-1: ML model retrieval unsubscribe response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the ML model retrieval unsubscribe request was successful.
Failure response	O (NOTE)	Indicates that the request has failed.
> Cause	O	The cause for the request failure.
NOTE: One of the IEs shall be present.		

8.3 ML model training

8.3.1 General

This clause describes the procedure for supporting the ML model training by the AIMLE server. The VAL server requests the AIMLE server for the ML model training. VAL server requests the AIMLE server to train a ML model by specifying the type of learning, details about the ML model to be trained or requirements to select a ML model for training and criteria for selecting the members who can participate in the ML model training.

8.3.2 Procedure for ML model training

Figure 8.3.2-1 illustrates the procedure for AIMLE server to support ML model training based on the request from the VAL server.

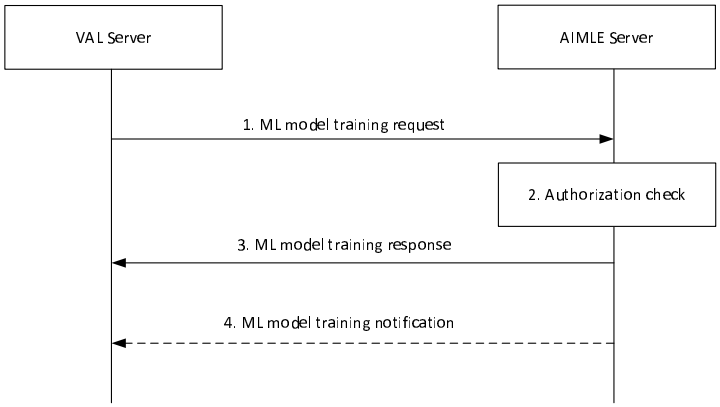

Figure 8.3.2-1: ML model training

1. The VAL server sends an ML model training request to AIMLE server, requesting to assist in its ML model training. This request consists of ML model information or ML model requirement information, etc.
2. The AIMLE server checks whether the VAL server is authorized to perform the ML model training request. If no model information is provided but only the model requirement information is provided in step 1, the AIMLE server identifies and selects the appropriate ML model for training based on the ML model requirement information using the procedure as specified in clause 8.11.3.
3. If the VAL server is authorized, AIMLE server returns the success response, otherwise a failure response indication the reason for failure.
4. The AIMLE server notifies the VAL server to update the list FL/ML clients selected or de-selected for the ML model training or to share the training output or any errors during the training process.

8.3.3 Information flows

8.3.3.1 ML model training request

Table 8.3.3.1-1 describes the information flow from the VAL server to the AIMLE server for requesting the ML model training.

Table 8.3.3.1-1: ML model training request

Information element	Status	Description
Requestor identity	M	Identity of the VAL server performing the request.
Training type	M	Identifies whether the VFL or HFL training to be performed.
List of member clients	O (NOTE 1)	List of member clients or AIMLE client set identifier to be utilized for training the ML model.
Member selection criteria	O (NOTE 1)	Identifies the criteria that needs to be continuously monitored for selecting the member clients (e.g., AIMLE clients in a particular location, member availability duration, new member clients registering with the required capabilities).
Number of required AIML clients	O	Indicates the requested number of AIML clients to be selected based on member selection criteria.
ML model information	O (NOTE 2)	Identifies the ML model that has to be distributed to the selected member clients for training. This information consists of the model identifier, address (e.g., a URL or an FQDN) of the ML model file or address of the model repository where the ML model resides
ML model requirement information	O (NOTE 2)	Identifies the requirement for selecting a model to be trained and this information contains the filtering criteria for selecting the model as specified in Table 8.11.4.1-2.
Dataset identifiers	M	Identifier of dataset used to HFL or VFL training. For VFL training, multiple dataset identifiers is specified for different data domains.
Number of data samples	O	The number of data samples required for a round of HFL or VFL training.
Operational schedule	O	A schedule for when training is to occur.
VFL specific parameters	O (NOTE 3)	Parameters specific to VFL training
> Dataset common features	O (NOTE 3)	A list of one or more common features is specify for VFL training. Common features can be UE identifiers, AIMLE client identifiers, group identifier, VAL service identifier, area of interest, and VAL service area.
> Data domain feature ID lists	O (NOTE 3)	List of features for each data domain(s) of the datasets at the UE for VFL training.
> Feature alignment information	O	Information provided to align features from dataset of different domains for VFL training. Alignment information is also provided for data domain features with ground truth data.
> Data labels	O	Ground truth data provided for VFL training.
Training objective	O	Identifies the termination condition for the ML model training. Table 8.3.3.1-2 describes the information related to training objective.
Members update notification	O	Indicates whether requestor needs to be notified whenever there is update related to new member clients selected or de-selected.
Notification Target Address	O	Notification target address (e.g. URL) where the notifications should be sent.
NOTE 1: At least one of these IE shall be present. NOTE 2: At least one of these IE shall be present. NOTE 3: These IE shall be present for VFL training.		

Table 8.3.3.1-2: ML model training objective

Information element	Status	Description
Objective Type	M	Identifies the metric to be optimized by the ML model (e.g. accuracy.)
Target Value	O	Identifies the threshold to be reached for the objective type.
Early stopping Criteria	O	Identifies the metric to be used for early stopping.
Maximum number of Epochs	O	Identifies the maximum number of training epochs
Acceptable training errors	O	Maximum error acceptable with training.
Inference Latency	O	Inference latency requirements for the trained model.

8.3.3.2 ML model training response

Table 8.3.3.2-1 describes the information flow of ML model training response from the AIMLE server to the VAL server.

Table 8.3.3.2-1: ML model training response

Information element	Status	Description
Success response	O (NOTE)	Indicates that the ML model training request was successful.
> ML model training identifier	M	An identifier for the ML model training request.
> ML model identifier	O	Identifies the ML model selected by AIMLE server for training.
Failure response	O (NOTE)	Indicates that the ML model training request was failure.
> Cause	M	Reason for the failure.
NOTE: Only one of these information elements shall be present		

8.3.3.3 ML model training notification

Table 8.3.3.3-1 describes the information flow of ML model training notification from the AIMLE server to the VAL server.

Table 8.3.3.3-1: ML model training notification

Information element	Status	Description
ML model training identifier	M	An identifier for the ML model training request the notification is for.
List of AIMLE clients	O (NOTE)	Indicates the list of AIMLE clients selected or de-selected for the ML model training.
Training output	O (NOTE)	Output of training, e.g., ML model parameters for the training.
Percentage completion	O	Indicates a completion percentage for the training.
Errors list	O (NOTE)	A list of errors, if any, encountered during training process.
NOTE: At least one of these IEs shall be present.		

8.4 Support for FL member registration

8.4.1 General

This clause provides the procedures for the registration, registration update, and deregistration of candidate FL members in the ML repository which serves as service registry for the FL members undertaking a task related to the ML model lifecycle, such as ML model local training. Candidate FL members can be application layer entities at the server side (e.g., VAL server, AIMLE server), which can potentially be selected as FL clients or FL server for a particular ASP

/vertical requirement, or at the client side (e.g. AIMLE client), which can potentially be selected as FL clients. If VAL server or the AIMLE client is the candidate FL member, it registers indirectly via AIMLE server to the ML repository.

8.4.2 Procedure on FL member registration

Figure 8.4.2-1 illustrates the procedure where the registration of a candidate FL member happens via the ML repository, serving as AIML service registry.

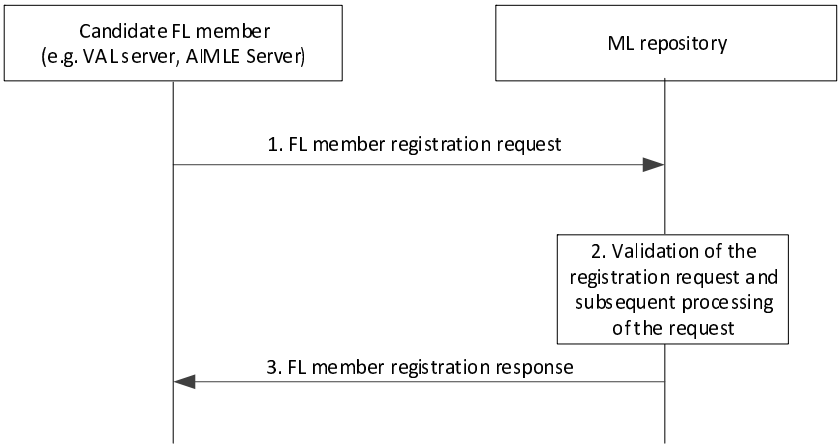

Figure 8.4.2-1: Procedure for registration on FL member registry

1. The candidate FL member (e.g., VAL server via AIMLE server or AIMLE server) sends an FL member registration request to the ML repository for registering to the ML repository which acts as the AIML service registry.
2. The ML repository validates the received request and generates the identity and other security related information for all the FL members listed in the registration request.
3. The ML repository sends the generated information in the FL member registration response message to the candidate FL member.

   The procedure for AIMLE client to register to ML repository as FL member is introduced in clause 8.7.2.2. The procedure for VAL server to register to the AIMLE server is introduced in clause 8.4.3.b.

8.4.3 Procedure on FL member registration update

Figure 8.4.3-1 illustrates the procedure where the registration update of a candidate FL member happens via the ML repository.

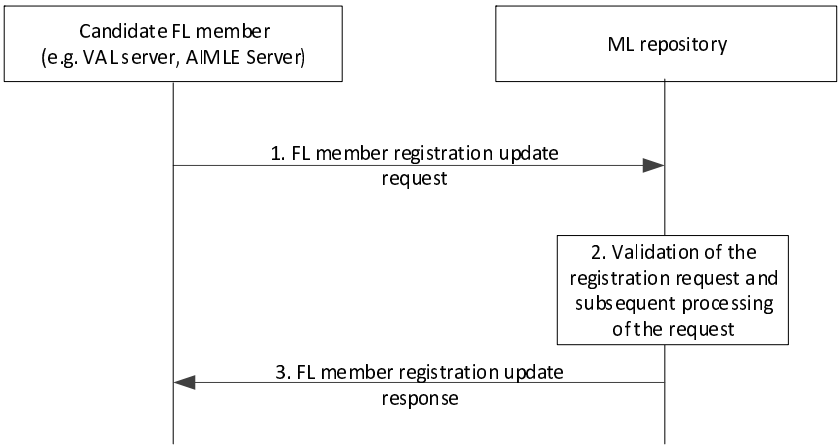

Figure 8.4.3-1: Procedure for registration update

1. The candidate FL member (VAL server via AIMLE server or AIMLE Server) sends an FL member registration update request to the ML repository acting as AIML service registry for updating the registration of the FL member.
2. The ML repository validates the received update request and generates the identity and other security related information for the FL member.
3. The ML repository sends the generated information in the FL member registration update response message to the candidate FL member.

The procedure for AIMLE client to update its registration to ML repository as FL member is introduced in clause 8.7.2.3.

8.4.3a Procedure on FL member deregistration

Figure 8.4.3a-1 illustrates the procedure where the deregistration of a candidate registered FL member happens via the ML repository, serving as AIML service registry.

The deregistration may be triggered by the VAL server or AIMLE server in case of an expected or predicted unavailability of the candidate FL member e.g. due to load / energy, access limitation in different VAL service area based on the vertical/ASP requirement.

Such deregistration may be provided by the AIMLE server on behalf of a VAL server e.g. after offboarding of the VAL server from the SEAL platform, or upon termination of the AIMLE/VAL service (FL task).

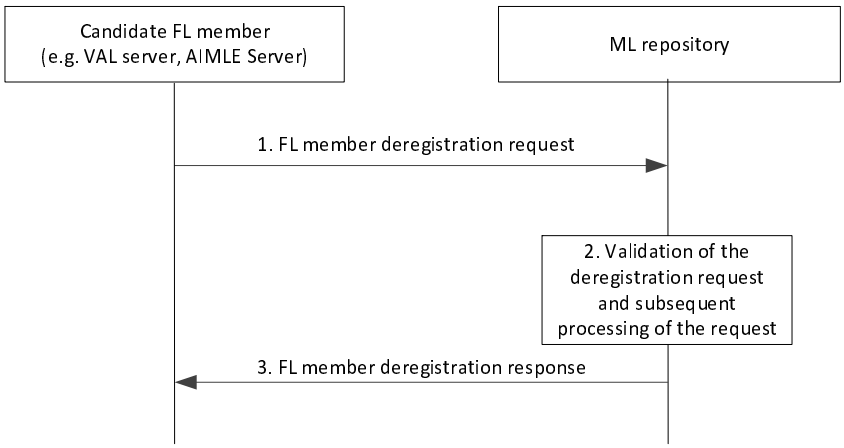

Figure 8.4.3a-1: Procedure for deregistration on FL member registry

1. The candidate registered FL member (e.g., VAL server via AIMLE server or AIMLE server) sends an FL member deregistration request to the ML repository for registering to the ML repository which acts as the AIML service registry.
2. The ML repository validates the received request for all the registered FL members listed in the deregistration request.
3. The ML repository sends the generated information in the FL member deregistration response message to the candidate registered FL member.

The procedure for AIMLE client to deregister to ML repository as FL member is introduced in clause 8.7.2.4.

8.4.3b Procedure on VAL server registration to AIMLE server

If the candidate FL member is a VAL server, the FL member registration requires that the VAL server registers itself to the AIMLE server, and subsequently the procedure in clause 8.4.2 to follow to enable the registration information storage to the ML repository.

Figure 8.4.3b-1 illustrates the procedure where the registration of a candidate FL member who is a VAL server happens to the AIMLE server.

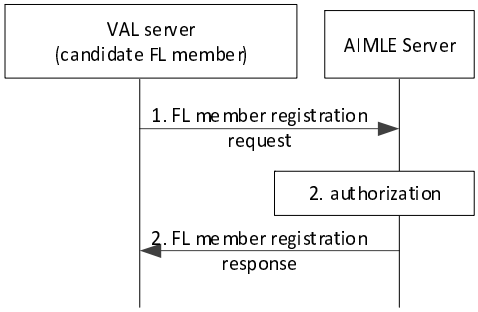

Figure 8.4.3b-1: Procedure for registration of VAL server as FL member

1. VAL server sends an FL member registration request to the AIMLE server to register as FL member to the ML repository.
2. The AIMLE server authorizes the request and stores the registration information to the ML repository based on procedure in clause 8.4.2.
3. The AIMLE server sends the FL member registration response message to the VAL server.8.4.4 Information flows

8.4.4.1 General

The following information flows are specified for supporting the FL member registration and registration update based on clause 8.4.2 and clause 8.4.3.

8.4.4.2 FL member registration request

Table 8.4.4.2-1 describes information elements for the FL member registration request from the candidate FL member (VAL server, AIMLE server) to the ML repository/registry.

Table 8.4.4.2-1: FL member registration request

Information element	Status	Description
Security information	M	Information for ML repository to validate the registration request.
FL member ID	M	The identifier of the candidate FL member. This can be the VAL server ID, the EAS ID, the AIMLE server ID based on which entity is the candidate FL member.
FL member supported AI/ML role	M	The supported AI/ML role of FL member which can be used as FL server or FL client.
FL member capabilities	O (NOTE)	FL member capability information (e.g. ML application type, allowed resource usage level).
FL member velocity	O	Indicates the FL member velocity (e.g., mobile or static).
FL member Location information	O	Indicates the location information of the UE (e.g., Cell Identity, Tracking Area Identity, GPS Coordinates or civic addresses).
FL member Availability schedule	O (NOTE)	Availability schedule of the FL member for a certain FL member.
Supported ML model ID list	O	The list of ML model IDs for which the FL member can be used.
Area of Interest	O	The area of interest (i.e. location co-ordinates) for which the registration applies.
Time validity	O	The time validity for the registration of the FL member.
NOTE: At least one of these shall be present.		

8.4.4.3 FL member registration response

Table 8.4.4.3-1 describes information elements for the FL member registration response to the candidate FL member (VAL server, AIMLE server) from the ML repository/registry.

Table 8.4.4.3-1: FL member registration response

Information element	Status	Description
Result	M	The result of the registration request (positive or negative acknowledgement).
Registration ID	M	The generated ID for the registration.

8.4.4.4 FL member registration update request

Table 8.4.4.4-1 describes information elements for the FL member registration update request from the candidate FL member (VAL server, AIMLE server) to the ML repository/registry.

Table 8.4.4.4-1: FL member registration update request

Information element	Status	Description
Registration ID	M	Identifier of the existing registration for which the update request applies.
Security information	M	Information for ML repository to validate the registration update request
FL member ID	M	The identifier of the candidate FL member for which the update applies. This can be the VAL server ID, the EAS ID, the AIMLE server ID based on which entity is the candidate FL member.
Updated registration parameters	M	The updates of the registration parameters.
> FL member supported AI/ML role	O (see NOTE)	The updated supported AI/ML role of FL member which can be FL server, FL client.
> FL member Availability schedule	O (see NOTE)	Updated Availability schedule of the FL member.
> supported ML model ID list	O (see NOTE)	The updated list of ML model IDs for which the FL member can be used
> area of interest	O (see NOTE)	The updated area of interest for the registration update.
> time validity	O (see NOTE)	The update time validity for the FL member registration update.
NOTE: At least one of these information elements shall be present.		

8.4.4.5 FL member registration update response

Table 8.4.4.5-1 describes information elements for the FL member registration update response to the candidate FL member (VAL server, AIMLE server) from the ML repository/registry.

Table 8.4.4.5-1: FL member registration update response

Information element	Status	Description
Result	M	The result of the registration update request (positive or negative acknowledgement).
Registration ID	O	The generated ID for the new registration based on the update (if needed).

8.4.4.6 FL member deregistration request

Table 8.4.4.6-1 describes information elements for the FL member deregistration request from the candidate FL member (VAL server, AIMLE server) to the ML repository/registry.

Table 8.4.4.6-1: FL member deregistration request

Information element	Status	Description
Registration ID	M	The identity of the registered FL member

8.4.4.7 FL member deregistration response

Table 8.4.4.7-1 describes information elements for the FL member deregistration response to the candidate registered FL member (VAL server, AIMLE server) from the ML repository/registry.

Table 8.4.4.7-1: FL member deregistration response

Information element	Status	Description
Result	M	The result of the deregistration request (positive or negative acknowledgement).

8.4.4.8 FL member registration request

Table 8.4.4.8-1 describes information elements for the FL member request from the VAL server to AIMLE server.

Table 8.4.4.8-1: FL member registration request

Information element	Status	Description
VAL server ID	M	The identifier of the VAL server as candidate FL member.
FL member supported AI/ML role	M (NOTE)	The supported AI/ML role of FL member which can be used as FL server or FL client.
FL member capabilities	O (NOTE)	FL member capability information (e.g., ML application type, allowed resource usage level).
FL member Availability schedule	O (NOTE)	Availability schedule of the FL member for a certain FL member.
Supported ML model ID list	O	The list of ML model IDs for which the FL member can be used.
Area of Interest	O	The area of interest for which the registration applies.
Time validity	O	The time validity for the registration of the FL member.
NOTE: At least one of these shall be present.		

8.4.4.9 FL member registration information storage response

Table 8.4.4.9-1 describes information elements for the FL member registration response to VAL server from the AIMLE server.

Table 8.4.4.9-1: FL member registration response

Information element	Status	Description
Result	M	The result of the registration request (positive or negative acknowledgement).

8.5 Support for FL events subscription and notification

8.5.1 General

This clause consists of the following procedures 1) the subscription for the FL related events in clause 8.5.2 and 2) the event notification procedure in clause 8.5.3.

This capability targets the enhancements to a global ML repository acting as AIML service registry which monitors the status and changes to the availability and capabilities of the FL members. The subscriber can be the AIMLE server or VAL server which requires to get notified on the FL member availability, since this information is vital for enabling the subscriber to perform the FL-related AIML operations (e.g. AIMLE server selecting FL members for a certain FL task). If VAL server is the subscriber, it registers indirectly via AIMLE server for FL-related events.

8.5.2 Procedure on subscription for FL related events

This procedure, as illustrated in Figure 8.5.2-1, describes the subscription for events related to FL member availability.

Pre-conditions:

1. The AIMLE server has the authorization to subscribe for the FL-related events (events described in clause 8.5.4).

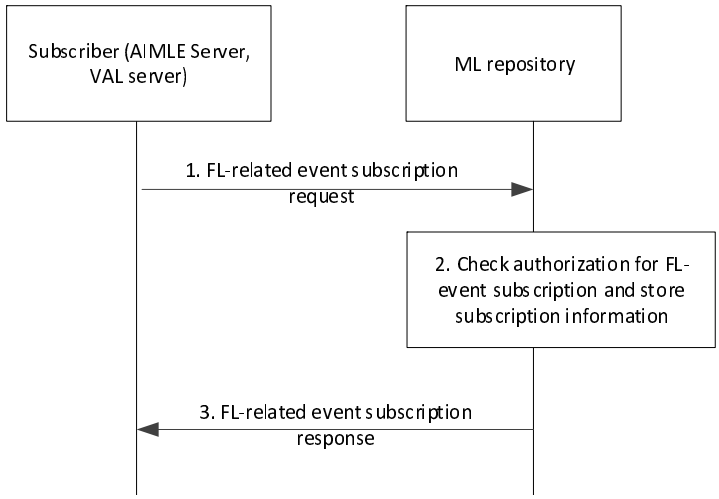

Figure 8.5.2-1: Procedure for FL-related event subscription

1. The subscriber (AIMLE server or VAL server (via AIMLE server)) sends an FL-related event subscription request to the ML repository to receive notification of FL related events and in particular the availability of FL member for a target area and time.

2. Upon receiving the event subscription request from the AIMLE server, the ML repository checks for the relevant authorization for the event subscription. If the authorization is successful, the ML repository stores the subscription information.

3. The ML repository sends an FL-related event subscription response to the subscriber indicating successful operation.

8.5.3 Procedure on FL related event notification

This procedure describes the event notification procedure for the FL member availability. This can be triggered based on a change at the availability or capabilities of the candidate or selected FL members.

Pre-conditions:

1. The subscription procedure in clause 8.5.2 was performed by the AIMLE server or VAL server (via AIMLE server).

2. The FL member has registered its capabilities or availability to the ML repository based on the support for FL registration capability.

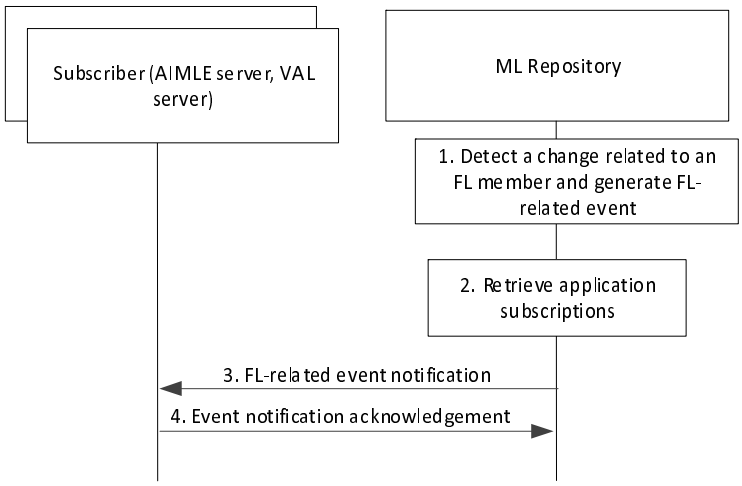

Figure 8.5.3-1: Procedure for FL-related event notifications

1. The ML repository detects a change related to the availability of an FL member and generates events to be consumed by the AIMLE server, based on the notification, or based on a received updated registration from FL member, who has already registered at the ML repository.
2. For the generated event, the ML repository retrieves the list of corresponding subscriptions.
3. The ML repository sends FL-related event notification to all entities (e.g., AIMLE servers) that have subscribed for the FL-related event matching the criteria. If a notification reception information is available as part of the FL-related event subscription, then the notification reception information is used by the ML repository to send event notifications to the corresponding AIMLE servers.
4. The notified subscriber (s) sends an event notification acknowledgement to the ML repository.

8.5.4 Definition of FL-related Events

The definition of the FL related events is provided in Table 8.5.4-1.

Table 8.5.4-1: Definition of FL related events

Events	Events Description
Availability changes of FL member	The event type relates to the availability change of a FL member. Such availability change can be an indication of availability or unavailability for a given service area, or for a given ML model ID/profile.
FL model information	Information such as accuracy, time schedule and latency for the FL training. Latency is considering the target latency, i.e., when the FL model training shall be completed.
FL member load information	The event type relates to the monitoring of the computational load for the requested FL member. Such load can be the estimated (based on measurements) or expected/predicted based on the tasks that the FL member undertakes. This event may be triggered on-demand or can be monitored periodically.

8.5.5 Information flows

8.5.5.1 General

The following information flows are specified for supporting the FL events subscription and events notification based on clause 8.5.2 and clause 8.5.3.

8.5.5.2 FL-related event subscription request

Table 8.5.5.2-1 describes information elements for FL-related event subscription request from the subscriber (AIMLE server, or VAL server (via AIMLE server)) to the ML repository.

Table 8.5.5.2-1: FL-related event subscription request

Information element	Status	Description
Requestor Identifier	M	The identifier to determine the identity of the requestor entity. This can be either the AIMLE server ID, or AIMLE server ID and VAL server ID.
Security Credentials	M	The security credentials of the requestor.
Time of validity	M	The time validity for the subscription.
ML model information	M	Information of the ML model for, specified in Table 8.11.4.1-2, which the subscription is needed.
FL member information	M	Information on the FL member (candidate or selected).
>FL member type	O (NOTE)	The type of FL member which can be an FL server or FL client.
>FL member ID	O (NOTE)	The ID of the entity serving as FL member. This can be used for events where the FL member is known and possibly selected, and the availability needs to be checked.
Event Information	M	Information on the FL-related event.
> Event ID	O	The identity of the FL-related event (if known by the requestor).
> Event type	M	The event type requested, based on the Definition of FL related events (shown in Table 8.5.4-1).
Notification endpoint	M	The information of the endpoint for receiving the notifications for the event.
NOTE: At least one of these shall be present.		

8.5.5.3 FL-related event subscription response

Table 8.5.5.3-1 describes information elements for the FL-related event subscription response to the subscriber (e.g. AIMLE server or VAL server (via AIMLE server)) from the ML repository.

Table 8.5.5.3-1: FL-related event subscription response

Information element	Status	Description
Result	M	Indicates the success or failure of the event subscription operation.
Subscription Identifier	M	The unique identifier for the event subscription.

8.5.5.4 FL related event notification

Table 8.5.5.4-1 describes information elements for the FL-related event notification from the ML repository to the requestor (e.g., AIMLE server).

Table 8.5.5.4-1: FL-related event notification

Information element	Status	Description
Subscription identifier	M	The unique identifier of the event subscription.
Event identifier	M	The unique identifier for the event. For the definition of events, refer to the Table 8.5.4-1 with the list of FL related events.
Event related information	M	The event related information (e.g., time at which the event originated or is planned to be executed, location of event).
Event content	M	The content of the event information.
> List of FL member ID /address	O	The list of the FL members and their addresses.
>> FL member enter / leave	O	The indication of an FL member entering or leaving the available list.
>> FL member updated capability	O	The update of capabilities for the FL members.
>> Time availability of FL member	O	The time period for which the FL member is available.
>> Area of interest	O	The area of interest or exact or predicted location for which the event applies.

8.5.5.5 FL-related event notification acknowledgement

Table 8.5.5.5-1 describes information elements for the FL-related event notification acknowledgement from the AIMLE server to the ML repository.

Table 8.5.5.5-1: FL-related event notification acknowledgement

Information element	Status	Description
Result	M	The positive or negative acknowledgement of the notification reception.

8.6 Support AI/ML Task Transfer

8.6.1 General

The following clauses specify procedures to support AI/ML task transfer.

8.6.2 Procedures for AI/ML task transfer

Due to various reasons (e.g., changes of available resource, changes of available time), an AI/ML member (e.g., AIMLE Client, VAL Client) finds itself cannot finish the assigned AI/ML task during the performing process. Here AI/ML member means the participant of an AI/ML task. To save resource and time, the AI/ML member (source AI/ML member) decides to transfer the intermediate AI/ML information (e.g., the intermediate AI/ML operation status and results) to another AI/ML member (target AI/ML member) for further operations to complete the AI/ML task.

8.6.2.1 Request AIMLE Server to assist AI/ML task transfer

Pre-conditions:

1. The information of target AI/ML member (e.g., another AIMLE Client or VAL Client different from the source AI/ML member) is unknown at the source AI/ML member. The source AI/ML member decides that assistance from the AIMLE Server is needed.

2. The source determines which AI/ML tasks to transfer. For example, if the AI/ML task is model training, AI/ML model are obtained by source AI/ML member performing AI/ML model training. Then the source AI/ML member determines the remaining AI/ML model training requirement based on the intermediate AI/ML operation status and results and AI/ML model training requirement.

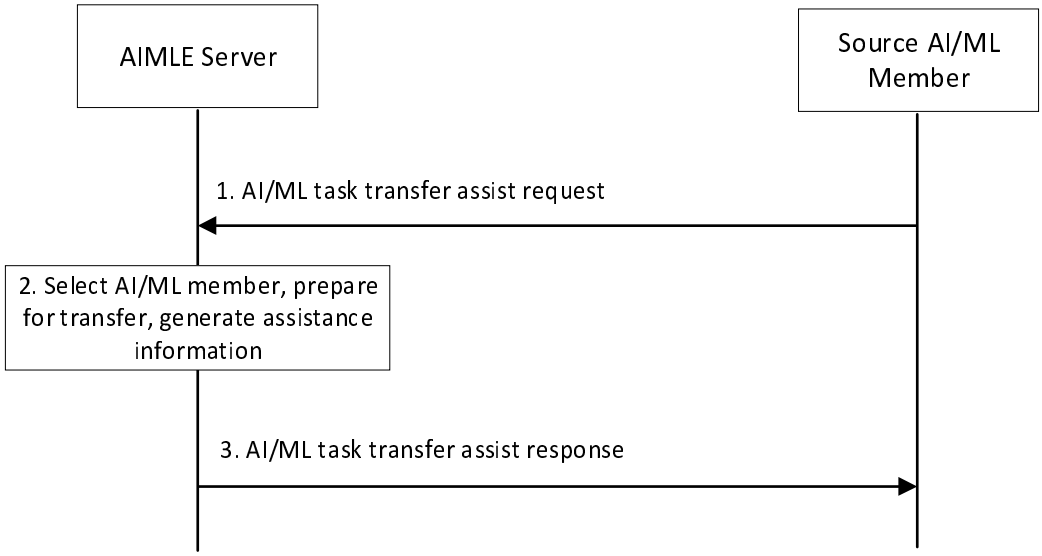

Figure 8.6.2.1-1: Procedure for requesting AIMLE Server assist AI/ML task transfer

Figure 8.6.2.1-1 illustrates the procedure for requesting AIMLE Server to assist the AI/ML task transfer. The detailed corresponding procedure is detailed as follows:

1. The source AI/ML member (e.g., an AIMLE Client) sends an AI/ML task transfer assist request to AIMLE Server for assisting the AI/ML task transfer. The request message contains the information as specified in Table 8.6.3.2-1.

2. The AIMLE Server discovers other AI/ML members (e.g., other AIMLE Clients) directly or via ML repository, and selects one or more target AI/ML member(s) based on the request from the source AI/ML member. The AIMLE Server sends request to the selected target AI/ML member(s) for the AI/ML task transfer, as defined in clause 8.6.2.2. Then, the AIMLE Server determines the transfer mode, i.e., transfer with server-controlled or transfer directly from the source AI/ML member to the target AI/ML member. For example, if the task is model training, the AI/MLE server determines the target AI/ML member(s) based on AI/ML member(s)'s capability in the AIMLE client profile and task's AI/ML model training requirement. The task's AI/ML model training requirement indicates the AI/ML Information.

   If target AI/ML member(s) is discovered, the AIMLE Server generates assistance information for the AI/ML task transfer from the source AI/ML member to the target AI/ML member(s) (e.g., time window for the transfer).

3. The AIMLE Server sends an AI/ML task transfer assist response to the source AI/ML member with the information of the target AI/ML member(s), and assistance information generated in step 2. The response message contains the information as specified in Table 8.6.3.3-1.

8.6.2.2 Request target AI/ML member for AI/ML task transfer

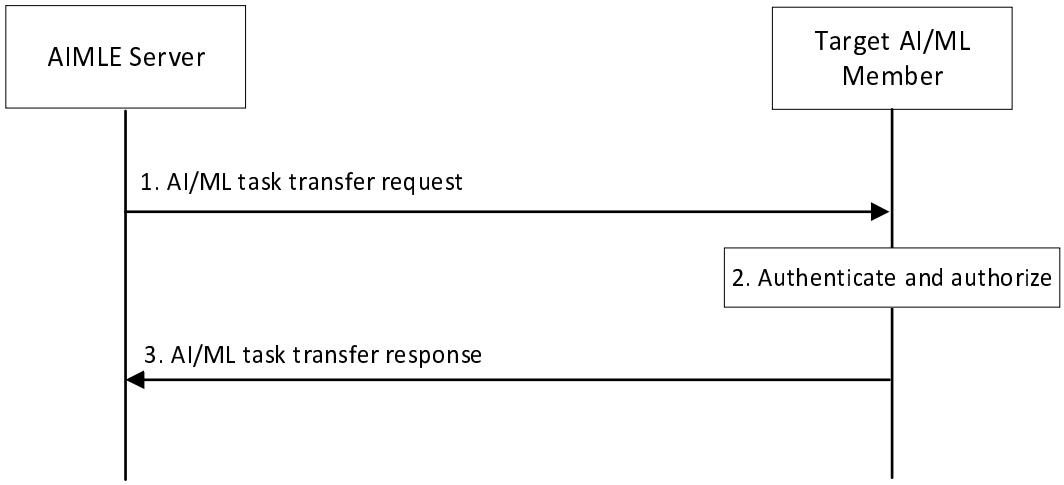

Figure 8.6.2.2-1: Procedure for requesting target AI/ML member for AI/ML task transfer

Figure 8.6.2.2-1 illustrates the procedure for requesting target AI/ML member for AI/ML task transfer. The corresponding procedure in detail is as follows:

1. The AIMLE Server sends an AI/ML task transfer request to the selected target AI/ML member(s) for AI/ML task transfer. The request message contains the information as specified in Table 8.6.3.4-1.

2. The target AI/ML member(s) authenticates and authorizes the request from the AIMLE Server. If the request is authorized, the target AI/ML member(s) performs step 3.

3. The target AI/ML member(s) sends an AI/ML task transfer response to the AIMLE Server with information as specified in Table 8.6.3.5-1.

8.6.2.3 Direct AI/ML task transfer

Pre-conditions:

1. The information of target AI/ML member (e.g., another AIMLE Client or VAL Client different from the source AI/ML member) is assumed to be known at the source AI/ML member, or the source AI/ML member may get the target AI/ML member information via ML repository. The source AI/ML member decides to transfer the AI/ML task to the target AI/ML member.

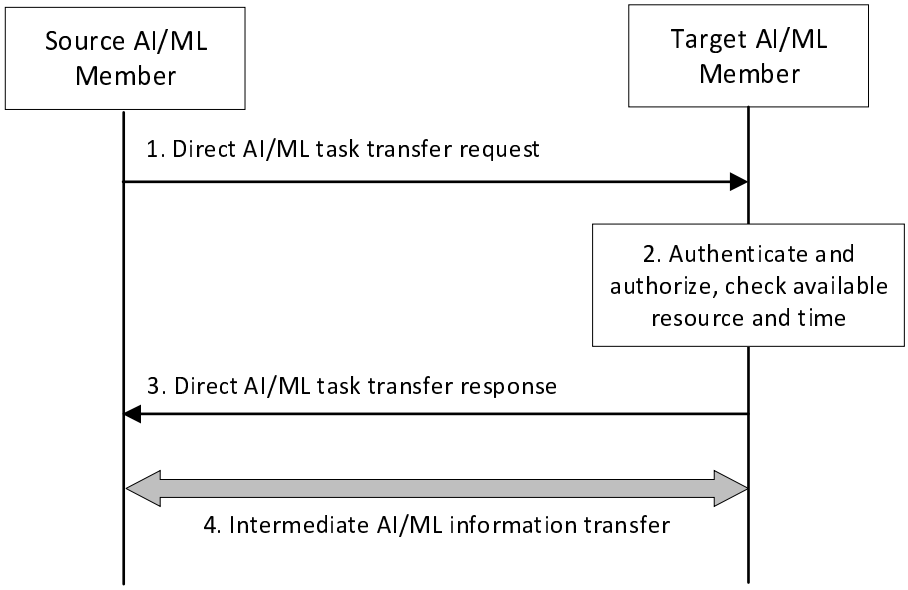

Figure 8.6.2.3-1: Procedure for direct AI/ML task transfer

Figure 8.6.2.3-1 illustrates the procedure for direct AI/ML task transfer. The corresponding procedure in detail is as follows:

1. The source AI/ML member (e.g. an AIMLE Client) sends a direct AI/ML task transfer request to the target AI/ML member for direct AI/ML task transfer. The request message contains the information as specified in Table 8.6.3.6-1.

2. The target AI/ML member authenticates and authorizes the request from the source AI/ML member. If the request is authorized, the target AI/ML member checks its availability for continue the AI/ML operations (e.g., available resource and time).

3. The target AI/ML member sends a direct AI/ML task transfer response to the source AI/ML member with information as specified in Table 8.6.3.7-1.

4. The source AI/ML member interacts with the target AI/ML member to perform AI/ML task transfer to the target AI/ML member.

8.6.2.4 AIMLE server-controlled AI/ML task transfer

Pre-conditions:

1. The information of target AI/ML member (e.g. another AIMLE Client or VAL Client different from the source AI/ML member) is unknown at the source AI/ML member. The source AI/ML member decides that AIMLE server-controlled AI/ML task transfer is needed.

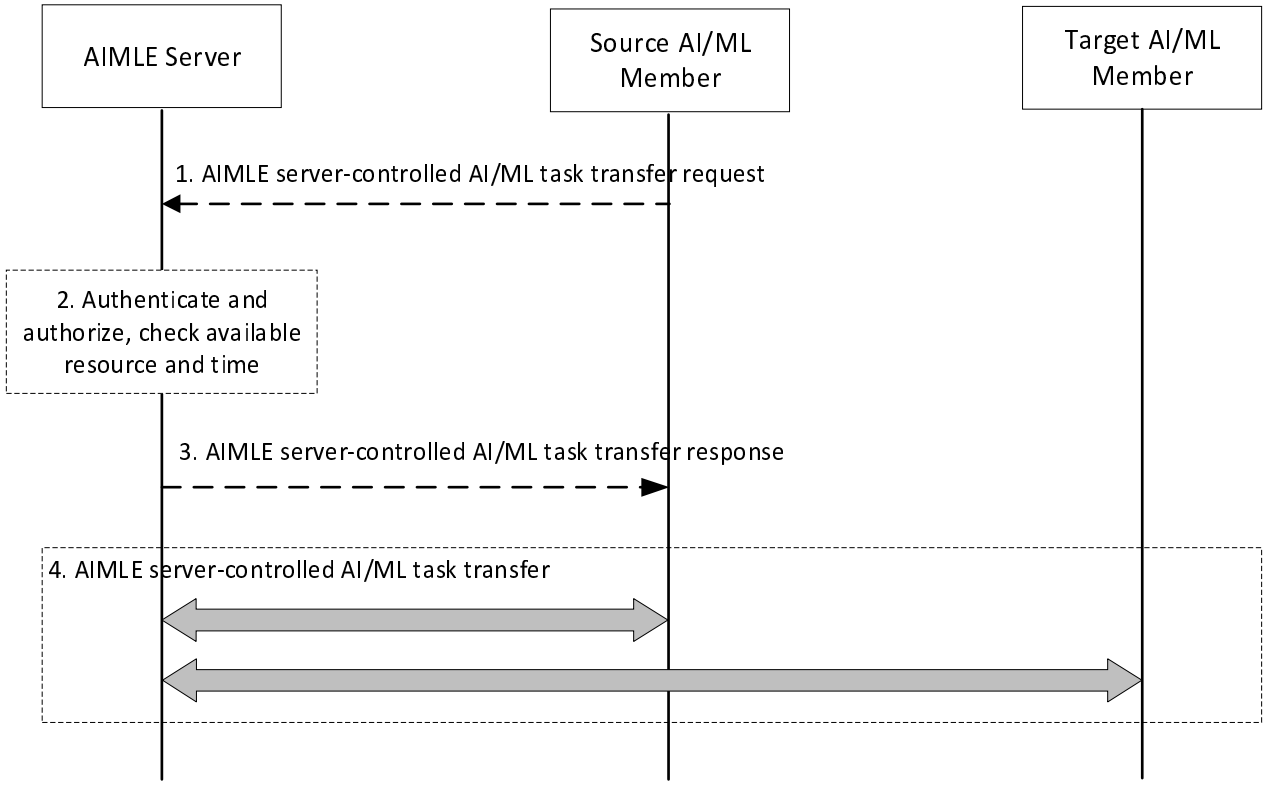

Figure 8.6.2.4-1: Procedure for AIMLE server-controlled AI/ML task transfer

Figure 8.6.2.4-1 illustrates the procedure for AIMLE server-controlled AI/ML task transfer. The corresponding procedure in detail is as follows:

1. The source AI/ML member (e.g., an AIMLE Client) may send an AIMLE server-controlled AI/ML task transfer request to AIMLE Server. The request message contains the information as specified in Table 8.6.3.8-1. Or, AIMLE server-controlled AI/ML task transfer may be decided by the AIMLE server as described in step 2 of clause 8.6.2.1.

2. If received the request for AIMLE server-controlled AI/ML task transfer from the source AI/ML member, the AIMLE sever authenticates and authorizes the request from the source AI/ML member. If the request is authorized, the AIMLE server checks the availability of the target AI/ML member(s) for AI/ML task transfer as defined in clause 8.6.2.2. The AIMLE server generates assistance information for the transfer of AI/ML task the source AI/ML member to the target AI/ML member(s) via AIMLE server (e.g. time window for the transfer).

3. The AIMLE server sends response to the source AI/ML member.

4. The source AI/ML member performs AI/ML task transfer to the target AI/ML member(s) via the AIMLE server based on the information received from the AIMLE server in step 3 if AIMLE server-controlled AI/ML task transfer is requested by the source AI/ML member, or based on the assistance information generated in step 2 of clause 8.6.2.1 if AIMLE server-controlled AI/ML task transfer is decided by the AIMLE server.

8.6.3 Information flows

8.6.3.1 General

The following information flows are specified for supporting AI/ML task transfer. In the following clauses 8.6.3.2 to 8.6.3.9, the source AI/ML member and the target AI/ML member could be e.g., VAL UE, AIMLE Client.

8.6.3.2 AI/ML task transfer assist request

Table 8.6.3.2-1 describes information elements for the AI/ML task transfer assist request from the source AI/ML member to the AIMLE Server.

Table 8.6.3.2-1: AI/ML task transfer assist request

Information element	Status	Description
Requestor identity	M	The identifier of the source AI/ML member (e.g. identifier of the source AIMLE client or the VAL UE).
VAL service ID	O	The identifier of the VAL service for which the assistance information is requested.
AI/ML Task Type	M	The type of the AI/ML operation (e.g. ML model training).
AI/ML Information Type	M	The type of the AI/ML information in the AI/ML task need be transferred (e.g. intermediate AI/ML operation status, intermediate AI/ML operation results).
> AI/ML model remaining training requirement (NOTE)	O	Indicate the requirement for AI/ML model training including, required remaining training resource, required remaining training number of iterations,
> AI/ML intermediate information (NOTE)	O	Indicate the AI/ML intermediate information for intermediate AI/ML operation and result (e.g. ML model training), including AI/ML intermediate model, AI/ML intermediate model used training time, used training resource, used training number of iterations.
Time validity	O	The time validity of the request.
NOTE: The IE only present when AI/ML Task Type is ML model training.		

8.6.3.3 AI/ML task transfer assist response

Table 8.6.3.3-1 describes information elements for the AI/ML task transfer assist response from the AIMLE Server to the source AI/ML member.

Table 8.6.3.3-1: AI/ML task transfer assist response

Information element	Status	Description
Transfer Mode	O	Indication the transfer mode (e.g. direct transfer).
Target AI/ML member(s)	M	The identifier of the target AI/ML member (e.g. identifier of the target AIMLE client or the VAL UE).
Assistance Information	M	Assistance information for the AI/ML task transfer (e.g. time window for the transfer).

8.6.3.4 AI/ML task transfer request

Table 8.6.3.4-1 describes information elements for the AI/ML task transfer request from the AIMLE Server to the target AI/ML member.

Table 8.6.3.4-1: AI/ML task transfer request

Information element	Status	Description
Requestor identity	M	The identifier of the AIMLE server.
Source AI/ML member	M	The identifier of the source AI/ML member (e.g. identifier of the source AIMLE client or the VAL UE).
AI/ML Task Type	M	The type of the AI/ML operation (e.g. ML model training).
AI/ML Information Type	M	The type of the AI/ML information in the AI/ML task need be transferred (e.g. intermediate AI/ML operation status, intermediate AI/ML operation results).
AI/ML Task Transfer Time	O	Information on time or time window for the AI/ML task transfer.
Time validity	O	The time validity of the request.

8.6.3.5 AI/ML task transfer response

Table 8.6.3.5-1 describes information elements for the AI/ML task transfer response from the target AI/ML member to the AIMLE Server.

Table 8.6.3.5-1: AI/ML task transfer response

Information element	Status	Description
Status	M	The status for the request: success or fail.
AI/ML Task Transfer Time	O	Information on time or time window for the AI/ML task transfer.

8.6.3.6 Direct AI/ML task transfer request

Table 8.6.3.6-1 describes information elements for the direct AI/ML task transfer request from the source AI/ML member to the target AI/ML member.

Table 8.6.3.6-1: Direct AI/ML task transfer request

Information element	Status	Description
Requestor identity	M	The identifier of the source AI/ML member (e.g. identifier of the source AIMLE client or the VAL UE).
AI/ML Task Type	M	The type of the AI/ML operation (e.g. ML model training).
AI/ML Information Type	M	The type of the AI/ML information in the AI/ML task need be transferred (e.g. intermediate AI/ML operation status, intermediate AI/ML operation results).
AI/ML Task Transfer Time	O	Information on time or time window for the AI/ML task transfer.
Time validity	O	The time validity of the request.

8.6.3.7 Direct AI/ML task transfer Response

Table 8.6.3.7-1 describes information elements for the direct AI/ML task transfer response from the target AI/ML member to the source AI/ML member.

Table 8.6.3.7-1: Direct AI/ML task transfer response

Information element	Status	Description
Status	M	The status for the request: success or fail.

8.6.3.8 AIMLE server-controlled AI/ML task transfer request

Table 8.6.3.8-1 describes information elements for the AIMLE server-controlled AI/ML task transfer request from the source AI/ML member to the AIMLE server.

Table 8.6.3.8-1: AIMLE server-controlled AI/ML task transfer request

Information element	Status	Description
Requestor identity	M	The identifier of the source AI/ML member (e.g. identifier of the source AIMLE client or the VAL UE).
AI/ML Task Type	M	The type of the AI/ML operation (e.g. ML model training).
AI/ML Information Type	M	The type of the AI/ML information in the AI/ML task need be transferred (e.g. intermediate AI/ML operation status, intermediate AI/ML operation results).
AI/ML Task Transfer Time	O	Information on time or time window for the AI/ML task transfer.
Time validity	O	The time validity of the request.

8.6.3.9 AIMLE server-controlled AI/ML task transfer Response

Table 8.6.3.9-1 describes information elements for the AIMLE server-controlled AI/ML task transfer response from the AIMLE server to the source AI/ML member.

Table 8.6.3.9-1: AIMLE server-controlled AI/ML task transfer response

Information element	Status	Description
Status	M	The status for the request: success or fail.
Assistance Information	M	Assistance information for the AI/ML task transfer (e.g. time window for the transfer).

8.7 AIMLE client registration

8.7.1 General

Prior to participation in AI/ML operations, AI/ML capable UEs register to an AIMLE server and provide an AIMLE client profile and optionally a list of supported services. The AIMLE server uses information provided by the AIMLE client to discover and select suitable AIMLE clients for requested AI/ML operations.

The following clauses specify procedures, information flows, and APIs for AIMLE client registration.

8.7.2 Procedures

8.7.2.1 General

The following are supported for AIMLE client registration:

- AIMLE client registration procedure;
- AIMLE client registration update procedure; and
- AIMLE client de-registration procedure

8.7.2.2 AIMLE client registration

Pre-conditions:

1. The AIMLE client has been pre-configured or has discovered the address (e.g., URI) of the AIMLE server.
2. The AIMLE client has been pre-configured with an AIMLE client profile.

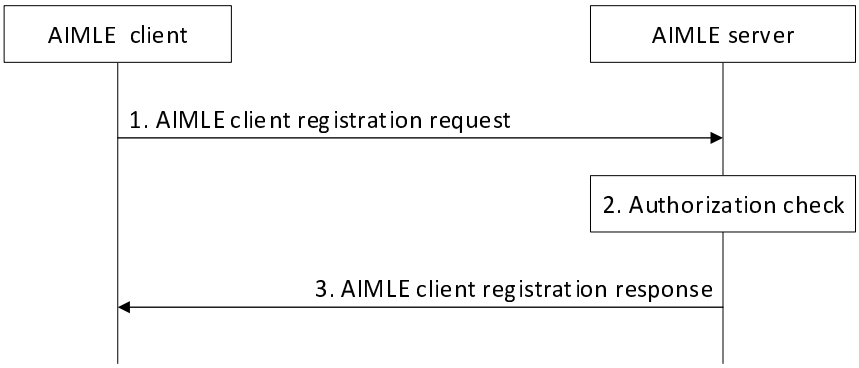

Figure 8.7.2.2-1: AIMLE client registration

1. The AIMLE client sends an AIMLE client registration request to the AIMLE server, the registration request includes information as described in Table 8.7.3.2-1. The AIMLE client indicates in the registration request its AI/ML capabilities such as supported ML model types and supported AI/ML operations, supported AIMLE client task capabilities with compute and task performance capabilities to assist with performing AIMLE client discovery and AIMLE client selection.

2. The AIMLE server validates the registration request and performs an authentication and authorization check to determine if the AIMLE client is permitted to register to the AIMLE server and participate in AI/ML operations. Upon successful authorization, the AIMLE server saves the context of the AIMLE client registration in the ML repository.

3. The AIMLE server returns an AIMLE client registration response to the AIMLE client with the status of the request.

8.7.2.3 AIMLE client registration update

Pre-conditions:

1. The AIMLE client has already registered with the AIMLE Server.

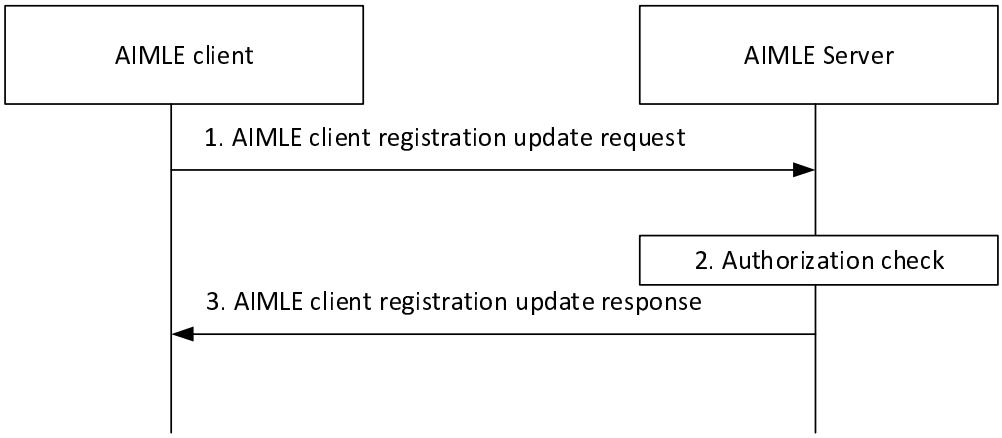

Figure 8.7.2.3-1: AIMLE client registration update

1. The AIMLE client sends an AIMLE client registration update request to the AIMLE server, the registration update shall include registration identifier, updated list of supported services and updated AIMLE client profile as described in Table 8.7.3.4-1. The AIML client removes services from the list of services identified by Service ID(s) when it ceases to provide a service and adds to the service when it starts to provide a service.

2. The AIMLE server validates the registration update and performs an authentication and authorization check. Upon successful authorization, the AIMLE server updates the context of the AIMLE client registration in the ML repository.

3. The AIMLE server returns an AIMLE client registration update response to the AIMLE client with the status of the update.

8.7.2.4 AIMLE client de-registration

Pre-conditions:

1. The AIMLE client has already registered with the AIMLE Server.

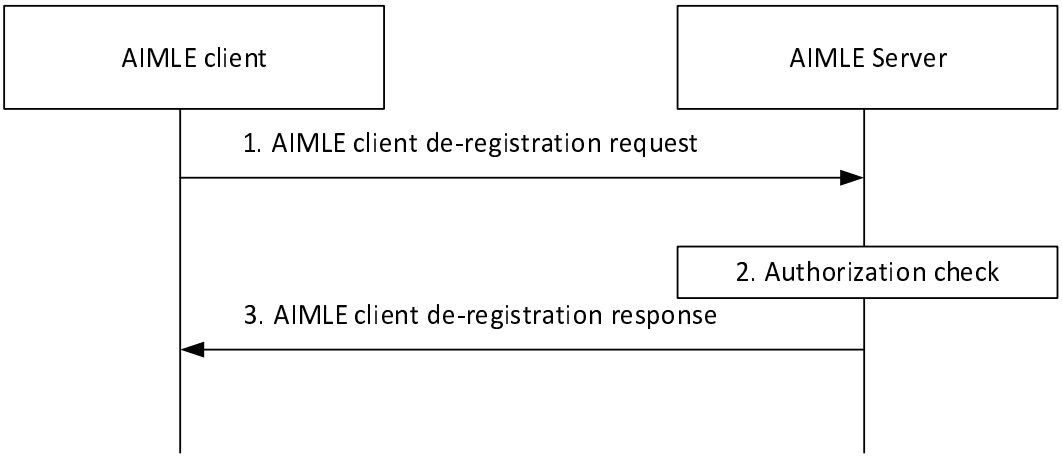

Figure 8.7.2.4-1: AIMLE client de-registration

1. The AIMLE client sends an AIMLE client de-registration request to the AIMLE server. The request shall include registration identifier as described in Table 8.7.3.6-1.
2. The AIMLE server validates the deregistration request and performs an authentication and authorization check to determine whether the AIMLE client can deregister from AI/ML operations. Upon successful authorization, the AIMLE server removes the context of the AIMLE client registration from the ML repository.
3. The AIMLE server returns an AIMLE client de-registration response to the AIMLE client with the status of the request.

8.7.3 Information flows

8.7.3.1 General

The following information flows are specified for AIMLE client registration:

- AIMLE client registration request and response;
- AIMLE client registration update request and response; and
- AIMLE client de-registration request and response

8.7.3.2 AIMLE client registration request

Table 8.7.3.2-1 shows the request sent by an AIMLE client to an AIMLE server for the AIMLE client registration request.

Table 8.7.3.2-1: AIMLE client registration request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
List of supported profiles	M	Supported AIML client profile(s). For each client profile provided in the list, the supported service information is provided.
> AIMLE client profile	M	Information about the capability of the AIMLE client to support AI/ML operations for the VAL service ID as detailed in Table 8.7.3.2-2.
> List of supported services	M	List of VAL service IDs with corresponding permissions.
> VAL service ID	M	The identifier of the VAL service.
> Service permission level	O	Service permission level (e.g., premium resource usage, standard resource usage, limited resource usage).

Table 8.7.3.2-2: AIMLE client profile

Information element	Status	Description
Supported AI/ML model types	O	AI/ML model types supported by the AIMLE client (e.g., decision trees, linear regression, neural networks).
Supported AI/ML operations	M	AI/ML operations supported by the AIMLE client such as ML model training, model transfer, model inference, model offload, model split, continue perform intermediate AI/ML operation/task).
AIMLE client time schedule configurations	O	Indicates the availability schedule of the AIMLE client for the AIML service, e.g., the AIMLE client is (not) available to participate in the AIML operations (e.g. ML model training) in the given time slot(s) and/or day(s) of the week.
AIMLE client location configurations	O	Indicates the location-based configurations of the AIMLE client for the AIML service, e.g., the AI/ML member is (not) available to participate in the AI/ML operations in the given locations represented by coordinates, civic addresses, network areas, or VAL service area ID.
AIMLE client capabilities	M	AIMLE client capability information (e.g. ML application type, allowed resource usage level).
Dataset availability	O	Dataset availability such as dataset size, age, list of dataset features, and dataset identifiers.
Data capabilities	O	A list of data capabilities such as the type of data that can be collected (e.g. raw data), supported data processing capabilities (e.g. processed data), and supported exploratory data analysis functions.
AIMLE client task capability	O See NOTE	Indicates the AIML task performing capabilities. It includes compute capabilities (e.g., high, low), task performance preference capabilities (e.g., Green task, energy-efficient, low costs)
NOTE: The Green and Energy-efficient task performance may not be applicable to a UE.		

8.7.3.3 AIMLE client registration response

Table 8.7.3.3-1 shows the AIMLE client registration response sent by the AIMLE server to the AIMLE client.

Table 8.7.3.3-1: AIMLE client registration response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the registration was successful.
> Registration ID	M	The identifier of the registration.
> Expiration time	O	Indicates the expiration time of the updated registration. To maintain an active registration status, a registration update is required before the expiration time. If the Expiration time IE is not included, it indicates that the updated registration never expires.
Failure response	O (NOTE)	Indicates that the registration request failed.
> Cause	O	Indicates the cause of registration request failure
NOTE: One of the IEs shall be present.		

8.7.3.4 AIMLE client registration update request

Table 8.7.3.4-1 shows the request sent by an AIMLE client to an AIMLE server for the AIMLE client registration update request.

Table 8.7.3.4-1: AIMLE client registration update request

Information element	Status	Description
Registration identifier	M	Identifier of the existing registration for which the update request applies.
AIMLE client profile	O	Update information about the capability of the AIMLE client to support AI/ML operations as detailed in Table 8.7.3.2-2.
List of supported services	O	Update to supported service information. Services identified by their VAL Service ID are either removed from the list when the AIMLE client ceases to provide a service and added when the AIMLE client starts to provide a service.
> VAL service ID	M	The identifier of the VAL service.
> Service permission level	O	Service permission level (e.g., premium resource usage, standard resource usage, limited resource usage).

8.7.3.5 AIMLE client registration update response

Table 8.7.3.5-1 shows the AIMLE client de-registration update response sent by the AIMLE server to the AIMLE client.

Table 8.7.3.5-1: AIMLE client registration update response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the registration update request was successful.
> Expiration time	O	Indicates the expiration time of the updated registration. To maintain an active registration status, a registration update is required before the expiration time. If the Expiration time IE is not included, it indicates that the updated registration never expires.
Failure response	O (NOTE)	Indicates that the registration update request failed.
> Cause	O	Indicates the cause of registration update request failure
NOTE: One of the IEs shall be present.		

8.7.3.6 AIMLE client de-registration request

Table 8.7.3.6-1 shows the request sent by an AIMLE client to an AIMLE server for the AIMLE client de-registration request.

Table 8.7.3.6-1: AIMLE client de-registration request

Information element	Status	Description
Registration ID	M	Identifier of the existing registration for which the update request applies.

8.7.3.7 AIMLE client de-registration response

Table 8.7.3.7-1 shows the AIMLE client de-registration response sent by the AIMLE server to the AIMLE client.

Table 8.7.3.7-1: AIMLE client de-registration response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the de-registration request was successful.
Failure response	O (NOTE)	Indicates that the de-registration request failed.
> Cause	O	Indicates the cause of de-registration request failure
NOTE: One of the IEs shall be present.		

8.8 AIMLE client discovery

8.8.1 General

Discovery of AIMLE clients is an important step in the AI/ML process for distributed, federated, split AI/ML, and transfer learning. Due to the nature of such learning, VAL servers need to discover suitable AIMLE clients to fulfil the requirements for the AI/ML application. The VAL server can then use the discovered AIMLE clients to select a set of AIMLE clients to perform AI/ML operations.

The following clauses specify procedures, information flows, and APIs for AIMLE client discovery.

8.8.2 Procedure

8.8.2.1 AIMLE client discovery

Pre-conditions:

1. AIMLE clients that support AI/ML operations have registered with the AIMLE server and included their AIMLE client profiles and optionally a list of supported services.

2. The AIMLE server can access a ML repository to obtain AIMLE client profiles and supported services associated with AIMLE clients.

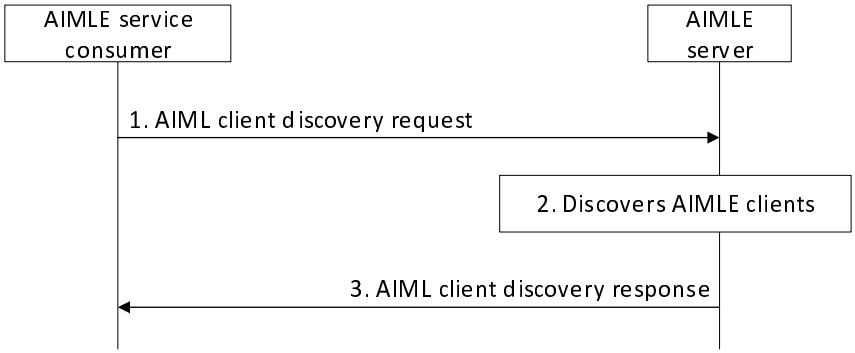

Figure 8.8.2.1-1: AIMLE client discovery

1. A VAL server sends an AIMLE client discovery request to an AIMLE server to discover a list of AIMLE clients that are available to participate in AI/ML operations (e.g., is available and have required data to train an ML model). The request may also include AIML client task capability requirements to discover clients who can perform the AIML tasks like AIML model training/offload/split with compute requirements and task performance preference like green task performance. The AIMLE client discovery request includes information as described in Table 8.8.3.1-1.

2. The AIMLE server performs authentication and authorization checks to determine if the requestor is able to discover AIMLE clients. If the requestor is authorized, the AIMLE server performs the following to discover AIMLE clients fulfilling the provided AIMLE client discovery criteria.

   The AIMLE server also obtains candidate UEs from the ML repository.

   From the list of candidate UEs, the AIMLE server discovers a list of UEs (i.e., AIMLE clients) that fulfils the discovery criteria based on supported services and AIMLE client profiles.

   The AIMLE server determines whether the discovered UEs fulfil the location information in the discovery criteria by using their identifiers to determine their location and includes in the discovery response only those UEs that meet the requirements. The AIMLE server may use SEAL-LMS (as in 3GPP TS 23.434 [5] clause 9.3.9, or 3GPP TS 23.434 [5] clause 9.3.10) or 3GPP 5G Core Network Services (such as GMLC as in 3GPP TS 23.273 [13] and NEF as in 3GPP TS 23.273 [13] or 3GPP TS 23.502 [9]) to determine the UEs which fulfil the location requirement.

   If the available AIMLE clients (as determined by the AIMLE server) that fulfil discovery criteria are less than the required minimum number of AIMLE clients for AI/ML operation (e.g., split AI/ML), then the AIMLE server may discover the remaining required AIMLE clients from other AIMLE servers over AIML-E reference point and include them in the response message.

3. The AIMLE server sends an AIMLE client discovery response that includes information in Table 8.8.3.2-1.

   If the required number of AIMLE clients are not included in the response message from the AIMLE server (i.e., the AIMLE clients in the response message are less than the required minimum number of AIMLE clients), then the AIML service consumer may discover the remaining required AIMLE clients from other AIML enablement server indicating the required number of AIMLE clients. The response may also include information related to supported tasks with guaranteed task KPIs, available AIML tasks service duration, AIML task expected KPI like latency.

8.8.3 Information flows

8.8.3.1 AIMLE client discovery request

Table 8.8.3.1-1 shows the request sent by a VAL server to an AIMLE server for the AIMLE client discovery procedure.

Table 8.8.3.1-1: AIMLE client discovery request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor (e.g., VAL server).
AIMLE client discovery criteria	M	Discovery criteria for finding suitable AIMLE clients for AI/ML operations as detailed in Table 8.8.3.1-2.
Number of required AIMLE clients	O	Indicates the requested number of AIMLE clients to be discovered based on the discovery criteria.

Table 8.8.3.1-2: AIMLE client discovery criteria

Information element	Status	Description
Service requirement	M	Information about the required service
> VAL Service ID	M	VAL Service ID that the client is required to support. This identifies the service associated with the requester.
> Service permission level	O	Required corresponding service permission level (e.g., premium resource usage standard resource usage, limited resource usage).
Requested ML model types	O	Requested ML model types (decision trees, linear regression, neural networks).
Requested AIML operations	M	Requested role for AI/ML operations such as model training, model transfer, model inference, model offload, model split.
Application layer AIMLE client capabilities	M	Application layer AIMLE client capability information (e.g., ML application type like FL/TL/SL, client availability to support AIML operations at the UE, AIMLE drop off rate).
Dataset requirements	O	Information about dataset.
> Dataset availability	O	Dataset availability, including dataset identifiers, dataset size, age, list of dataset features.
> Data capabilities	O	A list of data capabilities such as the type of data that can be collected (e.g., raw data), supported data processing capabilities (e.g., processed data) and supported exploratory data analysis functions.
AIML client task capability requirements	O	Indicates the AIML task requirements to discover the AIML clients for performing AIML tasks. It includes compute capabilities (e.g., high, low), task performance preference capabilities (e.g. Green task, Energy-efficient, low costs)
AIMLE client velocity	O	Indicates the AIMLE client velocity. It includes mobile (e.g., high, low), static.
Location information	O	Indicates the location information (e.g., Cell Identity, Tracking Area Identity, GPS Coordinates or civic addresses, VAL service area ID) to discover the AIMLE clients.
AIMLE client QoS requirements	O	Indicates the AIMLE client QoS information (like PLR, bandwidth, latency jitter) with the corresponding threshold(s) and threshold matching direction(s) (e.g., above or below) to discover the AIMLE clients.

8.8.3.2 AIMLE client discovery response

Table 8.8.3.2-1 shows the response sent by the AIMLE server to the VAL server for the AIMLE client discovery procedure.

Table 8.8.3.2-1: AIMLE client discovery response

Information element	Status	Description
Status	M	The status for the request: success or fail.
List of AIMLE client IDs	M	A list of AIMLE client IDs that matches the AIML discovery criteria.
AIML supported tasks	O	Indicates the discovered AIML task providers fulfilling the AIML task and AIML task requirements. It includes expected AIML task KPIs like latency, available AIML tasks service duration, list of AIML task providers like IDs or URI, and supported tasks.

8.9 AIMLE client selection

8.9.1 General

The AIMLE client selection procedure has two modes of operation: VAL server selection and AIMLE server selection. In VAL server selection, a List of AIMLE client IDs is provided to form an AIMLE client set, which can be used for AI/ML operations (e.g., training). In AIMLE server selection, a VAL server provides AIMLE client selection criteria and number of the required AIMLE clients for the AIMLE server to select the AIMLE client set.

The following clauses specify procedures, information flows, and APIs for AIMLE client selection.

8.9.2 Procedure

8.9.2.1 AIMLE client selection

Pre-conditions:

1. AIMLE clients that support AI/ML operations have registered with the AIMLE server and included their AIMLE client profiles and optionally a list of supported services.

2. The AIMLE server can access a ML repository to obtain AIMLE client profiles and supported services associated with the AIMLE clients.

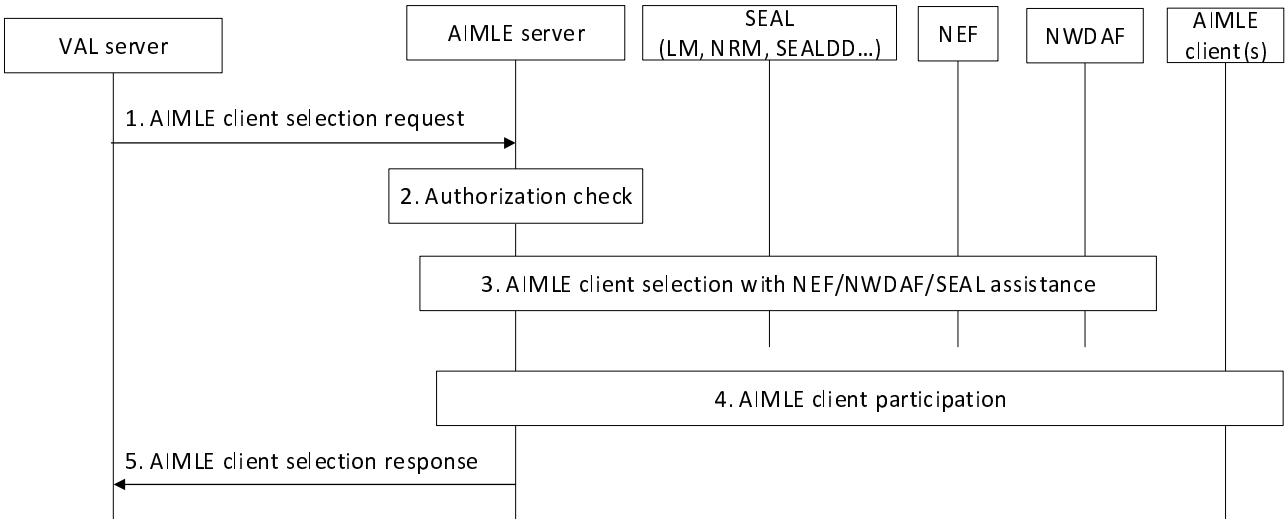

Figure 8.9.2.1-1: AIMLE client selection

1. A VAL server sends an AIMLE client selection request to an AIMLE server to select AIMLE clients available for participation in AI/ML operations (e.g., is available and have required data to train an ML model). The AIMLE client selection request includes information as described in Table 8.9.3.1-1.

2. The AIMLE server performs authentication and authorization checks to determine if the requestor is able to select AIMLE clients.

3. If the requestor is authorized, the AIMLE server performs AIMLE client selection based on the provided inputs as described below.

   For VAL server selection, the AIMLE server receives a list of AIMLE client IDs in the request and selects the AIMLE clients in the list as candidate AIMLE clients.

   For AIMLE server selection, the AIMLE server retrieves a list of clients with AI/ML capabilities from the ML repository. From the list of clients, the AIMLE server selects a list of candidate AIMLE clients, whose client profiles fulfil the AIMLE client selection criteria. The AIMLE server may use SEAL (LM service), NEF (e.g. MonitoringEvent API), and NWDAF (e.g. UE mobility analytics as defined in clause 6.7.2 of 3GPP TS 23.288 [2]) capabilities to assist the AIMLE client selection.

   The AIMLE server determines UEs by using their identifiers to determine their location and then selects only those that fulfil the location requirements specified in the AIMLE client selection criteria. The AIMLE server may use SEAL-LMS (as in 3GPP TS 23.434 [5] clause 9.3.9, or 3GPP TS 23.434 [5] clause 9.3.10) or 3GPP 5G Core Network Services (such as GMLC as in 3GPP TS 23.273 [13] and NEF as in 3GPP TS 23.273 [13] or 3GPP TS 23.502 [9]) to determine the UEs which fulfil the location requirement.

   The AIMLE server may then determine the QoS parameters for the AIML traffic session between the requestor and the VAL client associated with the candidate AIMLE client(s) and configure the AI/ML traffic session(s) via SEALDD (Sdd_RegularTransmission API) or NEF services (AfSessionWithQoS API). The AIMLE server determines the application QoS parameters based on the VAL Service ID.

NOTE 1: To determine the application QoS parameters based on the VAL service ID, the AIMLE server can use the VAL service ID associated AIMLE client QoS requirements in the AIMLE client selection criteria provided in step 1.

NOTE 2: If the available AIMLE clients (as determined by the AIMLE server) that fulfil selection criteria are less than the required number of AIMLE clients for AI/ML operation (e.g., split AI/ML), then the AIMLE server can discover and select the remaining required AIMLE clients from other AIMLE servers over AIML-E reference point and include them in the AIMLE client selection response message.

NOTE 3: The AIMLE server can reuse SEAL group management for any necessary group management.

4. The AIMLE server performs AIMLE client participation procedure with each candidate AIMLE client as described in clause 8.10.

   For all candidate AIMLE clients that agreed to participate in AI/ML operations, the AIMLE server selects the AIMLE clients and assigns an AIMLE client set identifier for the selected clients. The AIMLE client set may then be used for training a ML model.

NOTE 4: If a required minimum number of AIMLE clients for AI/ML operation is provided in the AIMLE client selection request and the number of AIMLE clients that agreed to participate in AI/ML operations is less than this number, an AIMLE client set identifier will not be assigned and the status will be set to fail in the AIMLE client selection response in step 5.

5. The AIMLE server sends an AIMLE client selection response that includes information in Table 8.9.3.2-1.

8.9.3 Information flows

8.9.3.1 AIMLE client selection request

Table 8.9.3.1-1 shows the request sent by a VAL server to an AIMLE server for the AIMLE client selection procedure.

Table 8.9.3.1-1: AIMLE client selection request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor (e.g., VAL server).
VAL service identifier	O	An identifier for the VAL service associated with the requestor.
List of AIMLE client IDs	O (NOTE 1) (NOTE 3)	A list of AIMLE client IDs that was previously discovered for inclusion into an AIMLE client set.
AIMLE client selection criteria	O (NOTE 2)	Selection criteria for finding suitable AIMLE clients for AI/ML operations as detailed in Table 8.8.3.1-2.
Number of required AIMLE clients	O (NOTE 2) (NOTE 3)	Indicates the requested number of AIMLE clients to be selected based on selection criteria.
NOTE 1: This information element is only present for VAL server selection. NOTE 2: These information elements are only present for AIMLE server selection. NOTE 3: At least one of the information elements are present.		

8.9.3.2 AIMLE client selection response

Table 8.9.3.2-1 shows the response sent by the AIMLE server to the VAL server for the AIMLE client selection procedure.

Table 8.9.3.2-1: AIMLE client selection response

Information element	Status	Description
Status	M	The status for the request: success or fail.
AIMLE client set identifier	O (NOTE 1)	An identifier to associate with the set of AIMLE clients selected and agreed to participate in AI/ML operations. The AIMLE client set can be updated by using this identifier.
List of AIMLE client IDs	O (NOTE 1) (NOTE2)	A list of AIMLE client IDs that matches the AIML selection criteria.
NOTE 1: This is present if the response status is success. NOTE 2: This information element is present for dynamic selection.		

8.10 AIMLE client participation

8.10.1 General

The following clauses specify procedures, information flows, and APIs for AIMLE client participation.

8.10.2 Procedure

8.10.2.1 AIMLE client participation

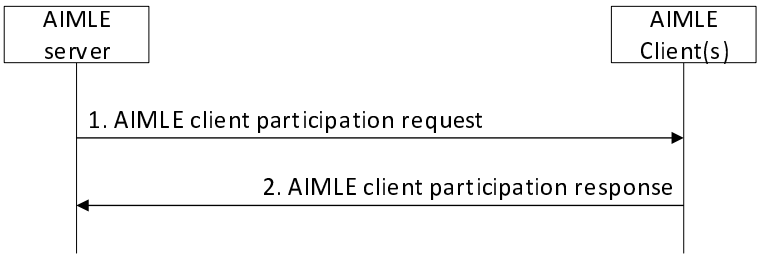

Figure 8.10.2.1-1: AIMLE client participation

1. An AIMLE server sends an AIMLE client participation request to an AIMLE client to verify participation in AI/ML operation(s). The request includes information as described in Table 8.10.3.1-1.

2. The AIMLE client acknowledge its willingness to be part of (or removed from) the AIMLE client set in an AIMLE client participation response sent to the AIMLE server. The response includes information as described in Table 8.10.3.1-2.

8.10.3 Information flows

8.10.3.1 AIMLE client participation request

Table 8.10.3.1-1 shows the request sent by the AIMLE server to each AIMLE client selected for AIMLE client participation procedure.

Table 8.10.3.1-1: AIMLE client participation request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor.
AIMLE client set identifier	M	An identifier for the AIMLE client set.
Add/remove indicator	M	Indicator for adding/removing the AIMLE client to/from the AIMLE client set.
ML model ID	M	The ML model ID to use for the AI/ML operation.
AI/ML operations	M	A list of AI/ML operations (e.g., training) required to be performed.
Operational schedule	O	Schedule for the AI/ML operations.
Dataset requirement	M	Dataset requirements for the AI/ML operations. Requirements includes dataset identifier, dataset size and age, and/or dataset features.
Service requirement	M	Information about the required Service including its VAL service ID and its service permission level (e.g., premium resource usage standard resource usage, limited resource usage)

8.10.3.2 AIMLE client participation response

Table 8.10.3.2-1 shows the response sent by the AIMLE clients to the AIMLE server for the AIMLE client participation procedure.

Table 8.10.3.2-1: AIMLE client participation response

Information element	Status	Description
Status	M	The status for the request: success or failure.
Participation agreement	O (NOTE)	Indicated whether the AIMLE client is willing to be part or be removed from the AIMLE client set.
Cause	O	The cause for the request failure.
NOTE: This is present if the response status is success.		

8.11 ML model management

8.11.1 General

In this functionality, two prodedures are described in more detail in clause 8.11.2 and clause 8.11.3 accordingly:

- Clause 8.11.2 for ML model information storage;

- Clause 8.11.3 for ML model information discovery.

8.11.2 ML model information storage

8.11.2.1 AIMLE server-initiated ML model information storage

Figure 8.11.2.1-1 illustrates the procedure of the AIMLE server-initiated ML model information storage.

Pre-condition:

1. The AIMLE server has either received application specific model details from AIML consumer or produced analytics model.

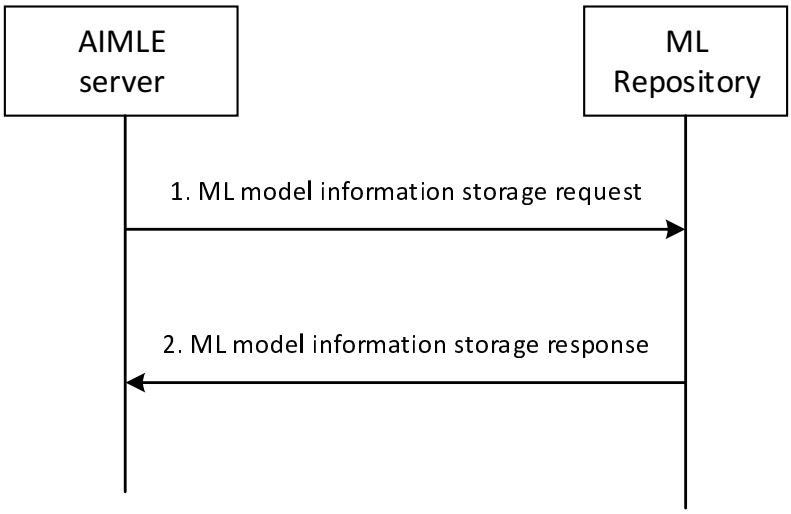

Figure 8.11.2.1-1: AIMLE server-initiated ML model information storage

1. The AIMLE server sends a ML model information storage request to the ML repository to store an ML model. The ML model included in the request is one trained by the ML repository consumer or its information is provisioned to the ML repository consumer. The request contains information elements as described in Table 8.11.4.1-1. The request message may contain Indication of continuous training and continuous training model parameter.

NOTE: The security requirements within the ML model storage requirements are out of scope.

2. Upon receiving the ML model information storage request, the ML repository verifies if the AIMLE server is authorized to store the ML model identified by ADAE Analytics ID and/or list of the allowed vendors provided within the ML model information attribute. If the AIMLE server is authorized, the ML repository processes the request and records information of the ML model (e.g., by creating a ML model profile). The ML repository sends an ML model information storage response to the AIMLE server with an identifier of the created ML model profile.

8.11.2.2 AIMLE consumer-initiated ML model information storage

Figure 8.11.2.2-1 illustrates the procedure of AIMLE consumer-initiated ML model information storage.

Pre-condition:

1. AIMLE consumer such as AIMLE client or VAL server has ML model or address to ML model.

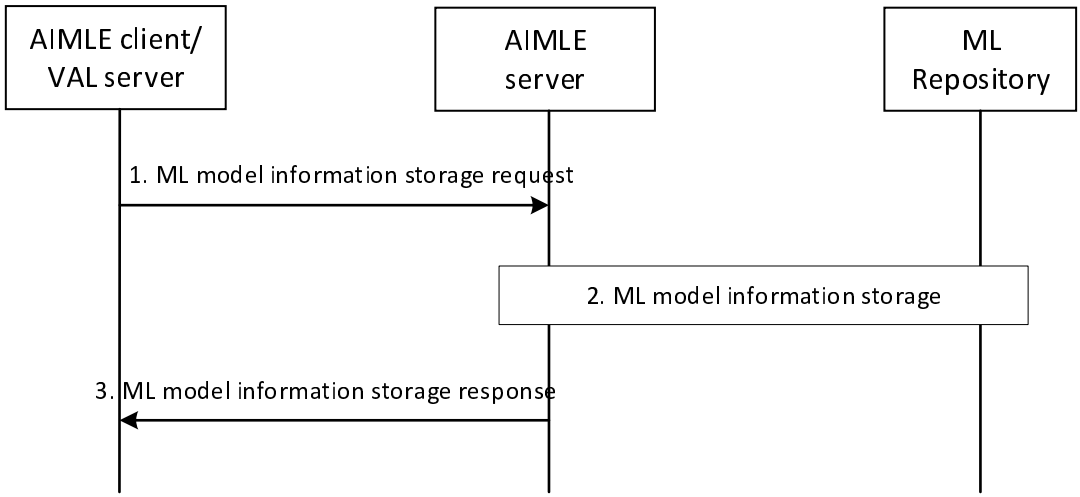

Figure 8.11.2.2-1: AIMLE consumer-initiated ML model information storage

1. An AIMLE client or a VAL server sends an ML model information storage request to an AIMLE server to store an ML model in the ML repository. The request includes the information elements as described in Table 8.11.4.1-1.
2. The AIMLE server sends an ML model information storage request as described in clause 8.11.2.1 to store the ML model in the ML repository.
3. The AIMLE server sends an ML model information storage response to the AIMLE client or VAL server with the information elements as described in Table 8.11.4.2-1.

8.11.3 ML model information discovery

Figure 8.11.3-1 illustrates the procedure of ML model information discovery.

Pre-condition:

1. Either AIMLE consumer has requested to discover ML model or AIMLE server decides to discover model for training.

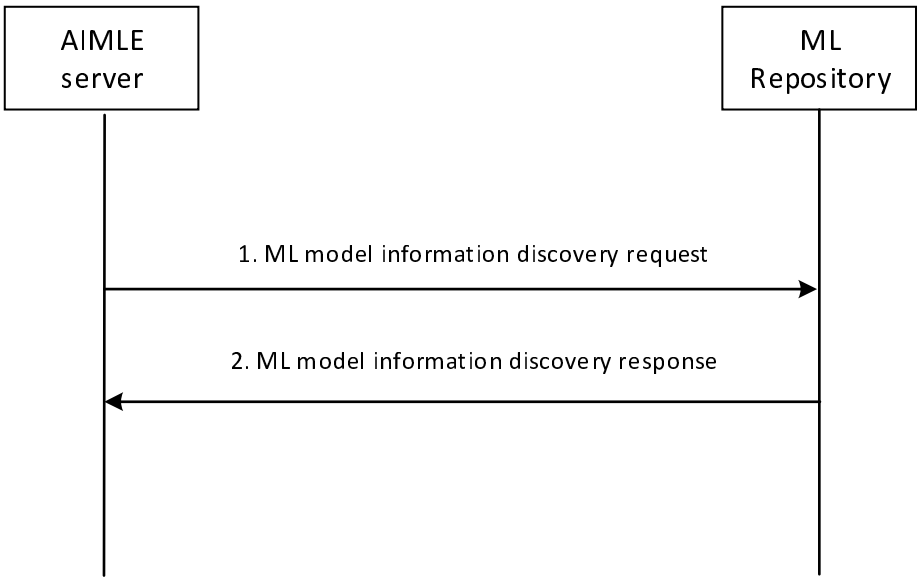

Figure 8.11.3-1: ML model information discovery

1. The AIMLE server sends an ML model information discovery request to the ML repository. The request contains information elements as described in Table 8.11.4.3-1.

2. Upon receiving the ML model information discovery request, the ML repository verifies if the requestor is authorized to discover the ML model(s) identified by the ML model ID. Further, it verifies whether the requestor is present in the list of allowed vendors or not. If the ML repository consumer is authorized, the ML repository processes the request and discovers the model based on ML model ID, ADAE Analytics ID, Base Model ID, and/or ML model interoperability information.

   If the discovery request is for transfer of learning, the ML repository identifies all candidate models matching the domain and required accuracy level.

   The ML repository sends the response message to the ML repository consumer. The response includes information of the discovered ML model.

NOTE: When the ML model is within training or re-training phase, the ML model is not available for discovery. The details are indicated within the Result IE in Table 8.11.4.4-1. The training and re-training phase is used during federated learning so that AIML enablement consumer can rely on model repository to get the model details.

After the AIMLE server receives the response from the ML repository, the AIMLE server determines whether the received ML model satisfy the ML model requirement of the AIMLE server. If not, the AIMLE server needs to continuously train the model to satisfy the ML model requirement based on the AIMLE server capability of supporting continuous training and ML model profile (Indication of continuous model training, continuous training model parameter).

8.11.4 Information flows

8.11.4.1 ML model information storage request

Table 8.11.4.1-1 describes the information flow from the AIMLE server to the ML repository or from the AIMLE client/VAL server to the AIMLE server as a request for the ML model information storage.

Table 8.11.4.1-1: ML model information storage request

Information element	Status	Description
Requestor Identity	M	The identity of the requestor performing the request.
Security Credentials	O (NOTE 1)	Security credentials of the requestor performing the request.
ML model	O (NOTE 2)	The ML model to be stored in the ML repository.
ML model address	O (NOTE 2)	The address (e.g., a URL or an FQDN) of the ML model file.
ML model information	M	Provides information of the ML model, as described in Table 8.11.4.1-2.
NOTE 1: This information is needed if the requestor is at the VAL service provider domain. NOTE 2: At least one of these information elements shall be provided.		

Table 8.11.4.1-2: ML model information

Information element	Status	Description
ML model identifier	M	An identifier for the ML model
ADAE Analytics ID	O	Represents ADAE analytics ID for which the model can be used.
ML model size	O	Indicates the size of the ML model.
ML model source identifier	O	The identifier of ML model source (e.g., VAL server ID, VAL client ID) that stored the model in the ML repository.
VAL service ID(s)	O	Identify the VAL service ID(s).
Domain	O	Specifies domain for which the model can be used (e.g., for speech recognition, image recognition, video processing, location prediction, etc.).
List of allowed vendors	O (NOTE 1)	Indicates which vendors that are allowed to use the ML model and thereby also are interoperable to the model.
ML model interoperability information	O (NOTE 1)	Represents the vendor-specific information that conveys, e.g., requested model file format, model execution environment, input/output parameters of the ML model, etc. The encoding, format, and value of ML Model Interoperable Information is not specified since it is vendor specific information, and is agreed between vendors, if necessary for sharing purposes.
ML Model phase	O (NOTE 1)	Represents the ML model phase, e.g., in training, trained, re-training, deployed.
> Observed performance	O (NOTE 2)	Provides information on the performance of the model e.g. accuracy, or application-specific performance metrics (if ML model is in trained or deployment phase).
> Training information	O (NOTE 2)	If the ML model is in trained or deployed phase: Information on the data that has been used to train the model (e.g. data sources, volume, freshness), and the base model ID in case of Transfer Learning.
> Indication of continuous model training	O	Indicates whether the model can be continuously trained or not.
> Continuous model training parameter	O	Parameters required for continuous model training.
ML model storage and discovery requirements	O (NOTE 1)	Represents the requirements for the ML repository for the ML model storage and discovery.
> Storage duration	O	Represents the ML model storage duration time. When the storage duration time is expired, the stored ML model and the related information shall be deleted.
> Security and access requirements	O	Represents the information on security requirements for storing the ML model information and the ML model access requirements (e.g., publicly available, private use only, or available for the list of VAL server IDs or VAL client IDs, time period and location access). If the access requirement is private use only, then the model is not discoverable by other consumers.
ML model usage requirements	O	Represents the requirements for using the ML model (e.g. for inference or for training). The requirements are used by the AIMLE server to determine whether an AIMLE client is capable of using the model based on comparing the requirements with information in the AIMLE client profile in Table 8.7.3.1-2.
NOTE 1: At least one of these information elements shall be provided. NOTE 2: This IE is included only if trained ML model is available.		

8.11.4.2 ML model information storage response

Table 8.11.4.2-1 describes the information flow from the ML repository to the AIMLE server or from the AIMLE server to the AIMLE client/VAL server as a response for the ML model information storage request.

Table 8.11.4.2-1: ML model information storage response

Information element	Status	Description
Result	M	Indicates success or failure of the request.
ML model profile identifier	O	The identifier of the ML model profile created as a result of a successful ML model storage request.

Table 8.11.4.2-2: ML model profile

Information element	Status	Description
ML model profile identifier	M	The identifier of the ML model profile.
AIMLE server identifier	M	The identifier of the AIMLE server that stored the ML model.
ML repository identifier	M	The identifier of the ML repository where the ML model is stored.
ML model information	M	The information about the ML model, as described in Table 8.11.4.1-2.
ML model retrieval endpoint	O	Represents the ML model retrieval endpoint (e.g., URL, URI, IP address and Port) that can be used to download the ML model.

8.11.4.3 ML model information discovery request

Table 8.11.4.3-1 describes the information flow from the AIMLE server to the ML repository as a request for the ML model information discovery.

Table 8.11.4.3-1: ML model information discovery request

Information element	Status	Description
Requester Identity	M	The identity of the ML repository consumer performing the request.
Filtering criteria	M	Represents the filtering criteria, which can be any of the ML model information as in Table 8.11.4.1-2.

8.11.4.4 ML model information discovery response

Table 8.11.4.4-1 describes the information flow from the model repository to the AIMLE server as a response for the ML model information discovery request.

Table 8.11.4.4-1: ML model information discovery response

Information element	Status	Description
Successful response	O (NOTE 1)	Indicates that the request was successful.
> ML model profile list	O (NOTE 2)	Represents the ML model profile(s) of the discovered list of ML models, as described in Table 8.11.4.2-2.
>> ML model profile ID	M	Represents the ML model profile ID of the ML model profile.
> ML model list	O (NOTE 2)	Represents the ML model(s) of the discovered list of ML models.
>> ML model ID	M	Represents the ML model ID of the ML model.
> Indication of continuous model training	O	Indicate whether the model needs to be continuously trained or not.
Failure response	O (NOTE 1)	Indicates that the request failed.
> Cause	O	Indicates the failure cause.
NOTE 1: Only one of these information elements shall be provided. NOTE 2: Only one of these information elements shall be provided.		

8.12 HFL training

8.12.1 General

AI/ML training is a highly iterative process and is performed over many training rounds. In the case of federated and distributed learning, the training is performed with many AIMLE clients. Due to the repetitive nature of AI/ML training, the AIMLE server can be configured to manage the training process over multiple training rounds for VAL servers. Note that the HFL training procedure supports horizontal federated learning, distributed learning, transfer learning, and split AI/ML.

The following clauses specify procedures, information flows, and APIs to support HFL training. The procedure for the HFL training is supported for AIMLE server:

- subscription and notification;
- subscription update; and
- unsubscription.

8.12.2 Procedure

8.12.2.1 HFL training subscription and notification

Pre-conditions:

1. AIMLE client discovery and selection have been performed.
2. Datasets are available and prepared for AI/ML training at the AIMLE clients. The datasets are assigned identifiers for use in HFL training.

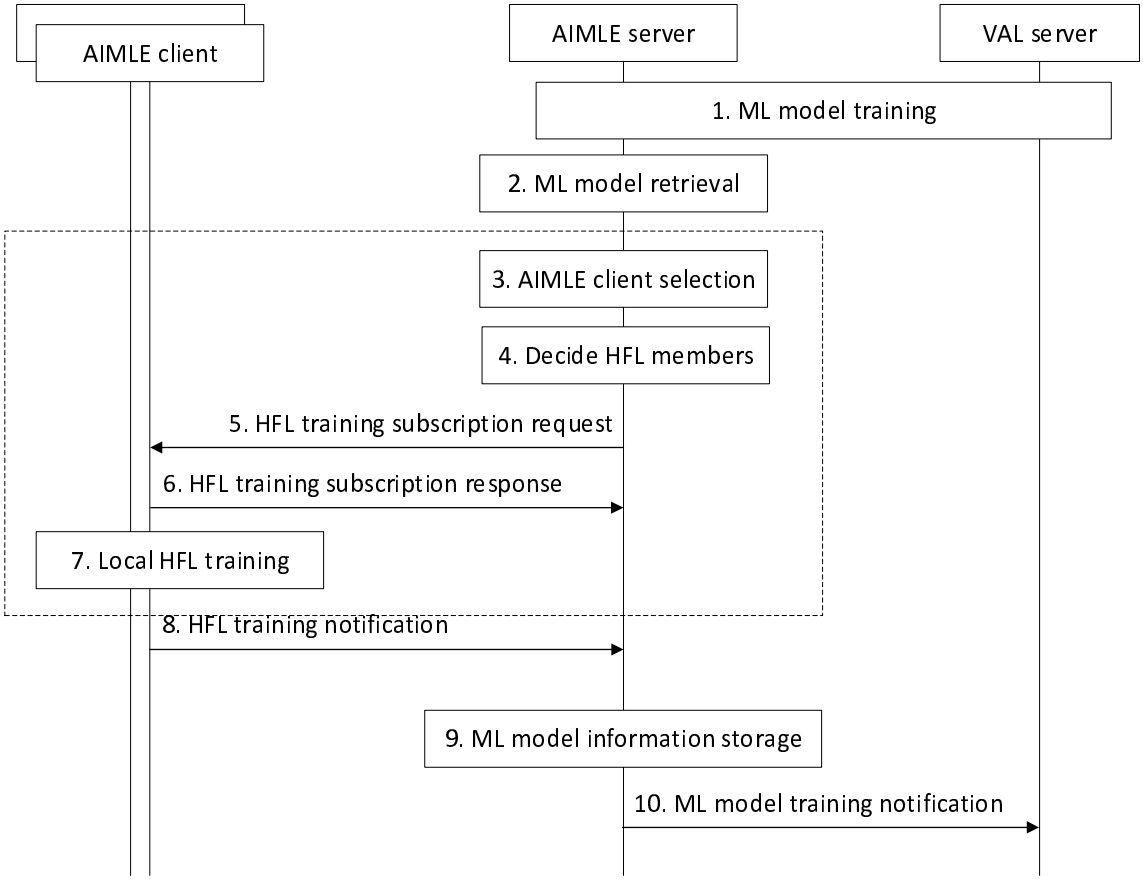

Figure 8.12.2-1: HFL training subscription and notification

1. An AIMLE server receives an ML model training request from a VAL server as described in clause 8.3.2. If the AIMLE server determines to use HFL training, the procedure continues to step 2.

2. The AIMLE server retrieves the indicated ML model using the ML model retrieval procedure as described in clause 8.1.

3. For the HFL training, the AIMLE server performs AIMLE client selection using the AIMLE client selection criteria or a list of AIMLE clients provided in step 1. If AIMLE client selection criteria are provided in step 1, then the AIMLE server continuously monitors and selects AIMLE clients for the HFL training. If a list of AIMLE clients is provided in step 1, the AIMLE server selects the provided AIMLE clients for the HFL training.

4. Based on the AIMLE client set determined in step 3, the AIMLE server checks with the selected AIMLE clients for their capability and participation in the HFL training as described in the clause 8.10.

5. The AIMLE server configures training schedule for each of the AIMLE client and sends a HFL training subscription request with information as described in Table 8.12.3.1-1.

6. Each AIMLE client sends a HFL training subscription response with information as described in Table 8.12.3.2-2. If the AIMLE client is not able to grant the subscription (e.g., is not able to perform the training), the AIMLE client sends a response with a failure status and the procedure skips to step 8.

7. Each AIMLE client performs local training using the configured AI/ML model, model parameters, and the prepared local data associated with the dataset identifier for the specified number of samples according to the operational schedule.

8. Upon completion or due to errors in the training, each AIMLE client sends a HFL training notification to the AIMLE server. The notification includes information as described in Table 8.12.3.1-3. The AIMLE client provides an VAL service ID for the HFL training operation.

9. If errors were encountered, the procedure skips to step 10.

If training was successful, the AIMLE server aggregates (e.g. averages) the model parameters received from the AIMLE clients.

If the training schedule is not complete (e.g., there are remaining training rounds), the AIMLE server configures the next set of training schedules and steps 3 to 8 are repeated for the next training round.

When the training schedule has been exhausted (e.g. there are no remaining training round) and if configured, the AIMLE server makes a ML model information storage request as described in clause 8.11.2 to store the AI/ML model in the ML repository.

10. The AIMLE server sends a ML model training notification to the VAL server as described in clause 8.3.2.

8.12.2.2 HFL training subscription update

Pre-conditions:

- The AIMLE server has subscribed for the HFL training with the AIMLE client.

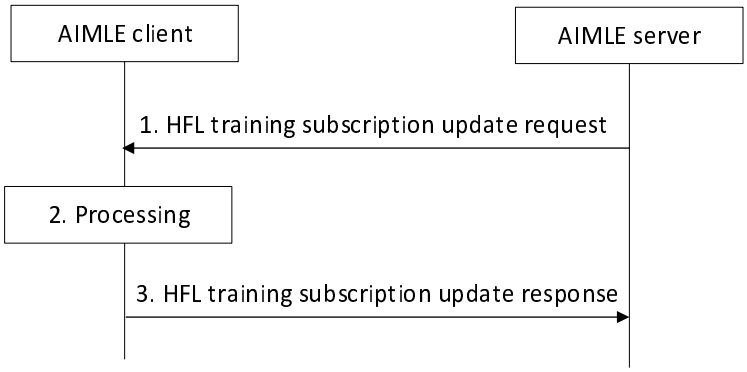

Figure 8.12.2-2: HFL training subscription update

1. The AIMLE server sends the HFL training subscription update request containing the subscription identifier and new information as described in Table 8.12.3.1-4 to the AIMLE Client.
2. Upon receiving the request from the AIMLE server, the AIMLE client validates if the AIMLE server is authorized for the request. If the AIMLE server is authorized, the AIMLE client updates the HFL training subscription.
3. The AIMLE client sends the HFL training subscription update response to the AIMLE server. If the AIMLE client has updated the HFL training subscription, the response includes an indication of success. Otherwise, the response includes an indication of failure and may include a reason for failure.

8.12.2.3 HFL training unsubscription

Pre-conditions:

- The AIMLE server has subscribed for the HFL training with the AIMLE client.

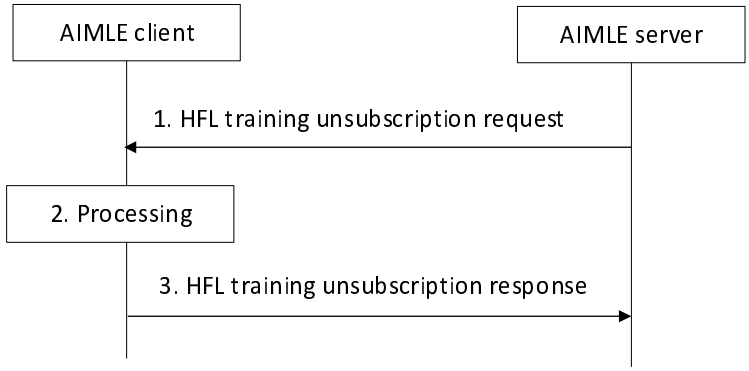

Figure 8.12.2-3: HFL training unsubscription

1. The AIMLE server sends the HFL training unsubscription request containing the subscription identifier to the AIMLE Client.

2. Upon receiving the request from the AIMLE server, the AIMLE client validates if the AIMLE server is authorized for the request. If the AIMLE server is authorized, the AIMLE client unsubscribe the HFL training subscription.

3. The AIMLE client sends the HFL training unsubscription response to the AIMLE server. If the AIMLE client has unsubscribed the HFL training subscription, the response includes an indication of success. Otherwise, the response includes an indication of failure and may include a reason for failure.

8.12.3 Information flows

8.12.3.1 HFL training subscription request

Table 8.12.3.1-1 shows the request sent by the AIMLE server to AIMLE clients for the HFL training subscription and HLF training subscription update procedure.

Table 8.12.3.1-1: HFL training subscription request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
AI/ML model and model parameters	M (NOTE)	Information about the AI/ML model and model parameters for use in HFL training.
Dataset identifier	M (NOTE)	The dataset identifier associated with the dataset used for HFL training.
Number of data samples	M (NOTE)	The number of data samples required for a round of HFL training.
Operational schedule	O	A schedule for when training is to occur.
Notification settings	O	Settings for how often to send notifications providing status of HFL training. For example, periodic, event-triggered (e.g. based on percentage completion), upon completion of each training round.
NOTE: These IEs are optional in HFL training subscription update procedure.		

8.12.3.2 HFL training subscription response

Table 8.12.3.2-1 shows the response sent by AIMLE clients to the AIMLE server for the HFL training subscription procedure.

Table 8.12.3.2-1: HFL training subscription response

Information element	Status	Description
Status	M	The status for the subscription request: success, fail.
Subscription ID	O	An identifier for the subscription only if the status is success.
VAL service ID	M	The VAL service identifier for the AIMLE HFL training operation.

8.12.3.3 HFL training subscription notification

Table 8.12.3.3-1 shows the notification sent by AIMLE clients to the AIMLE server for the HFL training subscription procedure.

Table 8.12.3.3-1: HFL training subscription notification

Information element	Status	Description
Status	M	Status for the request: success, fail.
VAL service ID	M	The VAL service identifier for the AIMLE HFL training operation.
HFL training output	M	ML model parameters from HFL training.
Errors list	O	A list of errors encountered during a HFL training round.
Timestamp	O	A timestamp for the notification.

8.13 AIMLE client selection subscription and notification

8.13.1 General

AIMLE client selection subscription request and notification enable VAL Servers to subscribe for monitoring and selection of AIMLE clients and receive notification when there is an update on the selected and re-selected AIMLE client’s status when re-selection is performed according to AIML member selection criteria.

The AIMLE server interacts with the NEF and/or SEAL services to monitor AIML members who meet the selection criteria and obtain their identifiers and configuring the AIML traffic session(s) between the VAL Server and the VAL client associated with the selected AIMLE Client(s) who meet the criteria. When the AIMLE clients no more meet the criteria, the QoS adjustment is reversed.

The subscription may also be updated when the VAL Server's requirements change, ensuring that it remains relevant and accurate. The VAL server may also unsubscribe when the subscription is no longer needed so that the AIMLE Server terminates monitoring of AIMLE clients and reverses QoS adjustments.

8.13.2 Procedures

8.13.2.1 General

The following are supported for AIMLE client selection subscription and notification :

- AIMLE client selection subscription and notification;
- AIMLE client selection subscription update; and
- AIMLE client selection unsubscribe

8.13.2.2 AIMLE client selection subscription and notification

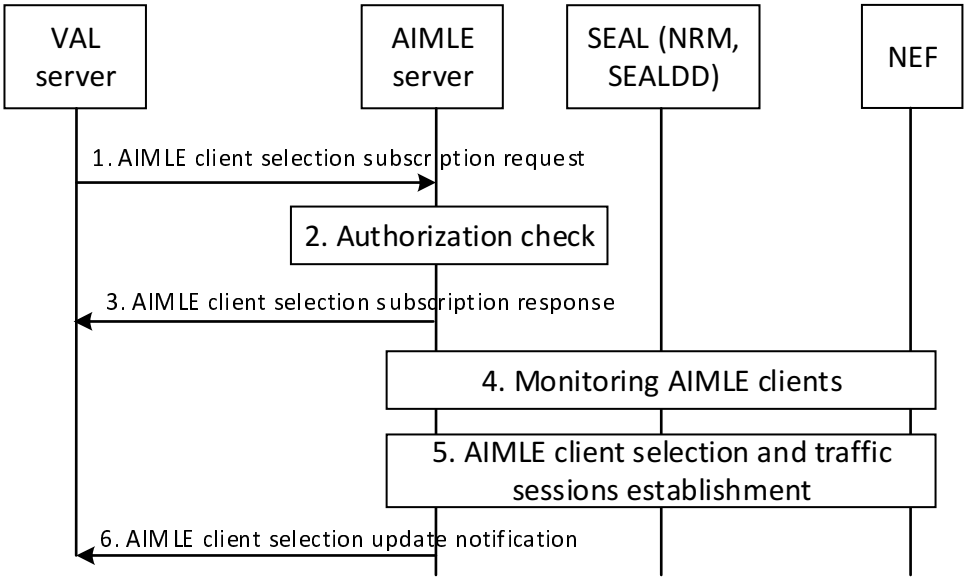

Figure 8.13.2.2-1: AIMLE client selection subscription and notification

1. A VAL server sends an AIMLE client selection subscription request to the AIMLE Server. The AIMLE client selection subscription request includes information as described in Table 8.13.3.2-1 which includes selection criteria.

2. The AIMLE server validates the AIMLE client selection subscription request. The AIMLE server further performs authentication and authorization checks to determine if the requestor is able to subscribe to the selected AIMLE client selection subscription request.

3. The AIMLE server sends the AIMLE client selection subscription response to the VAL server.

4. The AIMLE server monitors AIMLE clients whether they fulfil the selection criteria as provided in step 1. The AIMLE server interacts with the NEF and/or SEAL services (including SEALDD) to establish monitoring. The AIMLE server utilizes SEAL-LMS (as in 3GPP TS 23.434 [5] clause 9.3.11, or 3GPP TS 23.434 [5] clause 9.3.12) or 3GPP 5G Core Network Services (such as GMLC as in 3GPP TS 23.273 [13] and NEF as in 3GPP TS 23.273 [13] or 3GPP TS 23.502 [9]) to establish monitoring of UEs entering or present in the target location provided in the location information in the selection criteria.

5. The AIMLE Server obtains the identifiers of the AIMLE clients from the monitoring and selects the clients that fulfil the selection criteria and remove the AIMLE clients which do not fulfil the selection criteria. The AIMLE server uses the location monitoring for selecting UEs that fulfil the location criteria and removing UEs which cease to fulfil the location criteria as provided in the location information in the AIMLE client selection criteria. The AIMLE Server may determine the application QoS parameters (e.g. bandwidth, latency, jitter) for the AIML traffic session between the VAL server and the VAL client associated with the selected AIMLE client and configure the AIML traffic session(s) via SEALDD (Sdd_RegularTransmission API) or NEF services (AfSessionWithQoS API). When the AIMLE clients no more meet the criteria, the QoS adjustment is reversed.

   The AIMLE server may determine the application QoS parameters based on the VAL Service ID.

   If a desired service in the selection criteria is ceased to be provided by the client or its profile change so that it no longer meets the selection criteria, the AIMLE server removes the AIMLE clients which ceases to fulfil the criteria and selects other clients that fulfil selection criteria.

NOTE 1: To determine the application QoS parameters based on the VAL service ID, the AIMLE server can use the VAL service ID associated AIMLE client QoS requirements in the AIMLE client selection criteria provided in step 1.

NOTE 2: In case of AIMLE server determining the application QoS parameters based on the VAL service ID, how it does this is up to implementation.

6. The AIMLE Server notifies the VAL server about the selected and re-selected AIMLE clients e.g., the AIMLE Client A is de-selected and replaced by AIMLE Client B.

8.13.2.3 AIMLE client selection subscription update

Figure 8.13.2.3-1 illustrates the procedure for an VAL server to update a subscription with the AIMLE server.

Pre-conditions:

1. The AIMLE client or VAL server has subscribed for AIMLE client selection with the AIMLE Server.

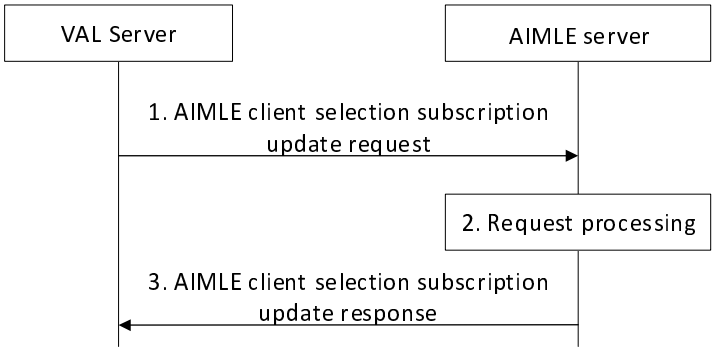

Figure 8.13.2.3-1: AIMLE client selection subscription update

1. A VAL server sends a AIMLE client selection subscription update request to the AIMLE server. The AIMLE client selection subscription update request includes the subscription identifier and may include information as described in Table 8.13.3.w-1 for updated subscription.

2. Upon receiving the request from the VAL server, the AIMLE server validates if the VAL server is authorized for the request. If the VAL server is authorized, the AIMLE server updates the subscription.

3. The AIMLE server sends a AIMLE client selection subscription update response to the VAL server. If the AIMLE server has updated the subscription, the response includes an indication of success. If the AIMLE server has not updated the subscription, the response includes an indication of failure and may include a reason for failure.

   If the subscription update request in step 1 include updated information as in Table 8.13.3.w-1, the AIMLE server also adjusts step 4 and 5 as in clause 8.13.2.2 based on the updated information in the subscription update.

8.13.2.4 AIMLE client selection unsubscribe

Figure 8.13.2.4-1 illustrates the procedure for a VAL server to unsubscribe with the AIMLE server.

Pre-conditions:

1. The VAL server has subscribed for AIMLE client selection with the AIMLE Server.

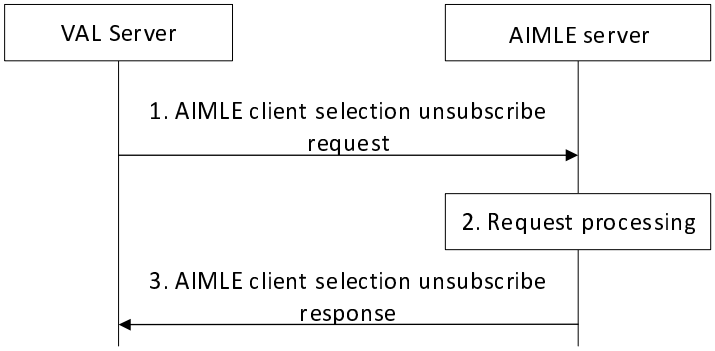

Figure 8.13.2.4-1: AIMLE client selection unsubscribe

1. A VAL server sends a AIMLE client selection unsubscribe request to the AIMLE server. The request includes the subscription identifier.

2. Upon receiving the request from the requestor, the AIMLE server validates if the VAL server is authorized for the request. If the VAL is authorized, the AIMLE server unsubscribes the subscription.

3. The AIMLE server sends a AIMLE client selection unsubscribe response to the VAL server. If the AIMLE server has unsubscribed the subscription, the response includes an indication of success. If the AIMLE server has not unsubscribed the subscription, the response includes an indication of failure and may include a reason for failure.

   If the AIMLE server has successfully unsubscribed the subscription, it also cancels its corresponding subscriptions with the NEF and/or SEAL services (including SEALDD) associated with the subscription. Additionally, the AIMLE server reverses the QoS adjustments performed as described in step 5 of clause 8.13.2.2.

8.13.3 Information flows

8.13.3.1 General

The following information flows are specified for AIMLE client registration:

- AIMLE client selection subscription and notification request and response;
- AIMLE client selection subscription update request and response; and
- AIMLE client selection unsubscribe request and response

8.13.3.2 AIMLE client selection subscription request

Table 8.13.3.2-1 shows the request sent by a VAL server to an AIMLE server for the Selected AIMLE Client selection subscription.

Table 8.13.3.2-1: AIMLE client selection subscription request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor.
AIMLE client selection criteria	M	Selection criteria for finding suitable AIMLE clients for AIML operations as detailed in per clause 8.8.3.1-2.
Number of the required AIML clients	O	Indicates the requested number of AIML clients to be selected based on member selection policies.
Notification endpoint for the selected AIMLE Client's	M	Represents the endpoint at the VAL server for receiving the notifications on the selected AIMLE Client's status update.

8.13.3.3 AIMLE client selection subscription response

Table 8.13.3.3-1 shows the response sent by the AIMLE server to the VAL server for the selected AIMLE client selection subscription response.

Table 8.13.3.3-1: AIMLE client selection subscription response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the subscription was successful.
> Subscription ID	M	The identifier of the subscription.
> Expiration time	O	Indicates the expiration time of the updated subscription. To maintain an active subscription status, a subscription update is required before the expiration time. If the Expiration time IE is not included, it indicates that the updated subscription never expires.
Failure response	O (NOTE)	Indicates that the subscription request failed.
> Cause	O	Indicates the cause of subscription request failure
NOTE: One of the IEs shall be present.		

8.13.3.4 AIMLE client selection update notification

Table 8.13.3.4-1 shows the notification sent by the AIMLE Server to the VAL server for the AIMLE client selection update notification.

Table 8.13.3.4-1: AIMLE client selection update notification

Information element	Status	Description
Requestor ID	M	The identifier of the requestor.
List of the selected AIMLE Client status update events	M	Represents the list of selected AIMLE Client status update events, e.g., the AIMLE Client A is de-selected and replaced by AIMLE Client B.

8.13.3.5 AIMLE client subscription update request

Table 8.13.3.5-1 shows the request sent by a VAL server to an AIMLE server for the AIMLE client selection subscription update request.

Table 8.13.3.5-1: AIMLE client selection subscription update request

Information element	Status	Description
Subscription ID	M	Identifier of the existing subscription for which the update request applies.
AIMLE client selection criteria	O	Selection criteria for finding suitable AIMLE clients for AIML operations as detailed in per clause 8.8.3.1-2.
Number of the required AIML clients	O	Indicates the requested number of AIML clients to be selected based on member selection policies.
Notification endpoint for the selected AIMLE Client's	O	Represents the endpoint at the VAL server for receiving the notifications on the selected AIMLE Client's status update.

8.13.3.6 AIMLE client selection subscription update response

Table 8.13.3.6-1 shows the AIMLE client selection subscription update response sent by the AIMLE server to the VAL server.

Table 8.13.3.6-1: AIMLE client selection subscription update response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the subscription update request was successful.
> Expiration time	O	Indicates the expiration time of the updated subscription. To maintain an active subscription status, a subscription update is required before the expiration time. If the Expiration time IE is not included, it indicates that the updated subscription never expires.
Failure response	O (NOTE)	Indicates that the subscription update request failed.
> Cause	O	Indicates the cause of subscription update request failure.
NOTE: One of the IEs shall be present.		

8.13.3.7 AIMLE client selection unsubscribe request

Table 8.13.3.7-1 shows the request sent by a VAL server to an AIMLE server for the AIMLE client selection unsubscribe request.

Table 8.13.3.7-1: AIMLE client selection unsubscribe request

Information element	Status	Description
Subscription ID	M	Identifier of the existing subscription for which the unsubscribe request applies.

8.13.3.8 AIMLE client selection unsubscribe response

Table 8.13.3.8-1 shows the AIMLE client selection unsubscribe response sent by the AIMLE server to the VAL server.

Table 8.13.3.8-1: AIMLE client selection unsubscribe response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the unsubscribe request was successful.
Failure response	O (NOTE)	Indicates that the unsubscribe request failed.
> Cause	O	Indicates the cause of unsubscribe request failure.
NOTE: One of the IEs shall be present.		

8.14 Support for Split AI/ML Operation

8.14.1 General

Split AI/ML operation (e.g., split operation) procedures enable applications consuming services from the AI/ML application enablement layer to discover and create instance(s) of split operation pipeline and addresses aspects related to consumption of split operation results.

Split operation pipeline discovery enables an AIMLE client to obtain information about available instances of split operation pipeline or processing nodes of interest. The discovery of split operation or processing node instances is based on matching split operation discovery filters provided in the request.

Split operation pipeline creation enables an AIMLE client to create an instance of a split operation pipeline. The creation of an instance of a split operation pipeline is based on discovery of available ML models and processing nodes.

Split operation pipeline update and delete enable an AIMLE client to manage an instance of a split operation pipeline.

Split operation node registration enables a VAL server to register with the AIMLE server to indicate split operation capabilities. Registered VAL servers can be notified by the AIMLE server for participating in an instance of a split operation pipeline.

Split operation event subscription enables an AIMLE client or VAL server to receive notifications about split operation events. The subscription is based on matching split operation events filters provided in the request.

8.14.2 Procedure

8.14.2.1 General

Following procedures are supported for split AI/ML operation:

- Split operation pipeline discovery request-response procedure;
- Split operation pipeline creation, update and delete request-response procedures;
- Split operation node registration, registration update and de-registration request-response procedures;
- Split operation event subscribe-notify procedures for information, comprising:
  - Subscription procedure;
  - Subscription update procedure; and
  - Unsubscribe procedure;

8.14.2.2 Split operation pipeline discovery

Figure 8.14.2.2-1 illustrates the procedure for an AIMLE client to discover instance(s) of split AI/ML operation pipeline or processing nodes from the AIMLE server.

Pre-conditions:

1. The AIMLE client has received information (e.g. URI, IP address) related to the AIMLE Server;
2. The AIMLE client has received security credentials authorizing it to communicate with the AIMLE server;

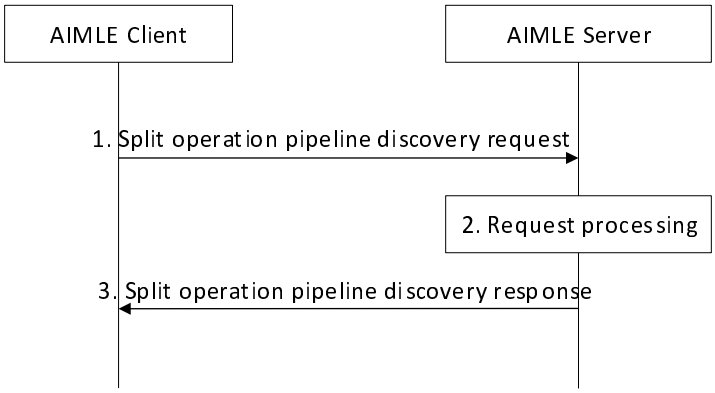

Figure 8.14.2.2-1: Split operation pipeline discovery

1. The AIMLE client sends split operation pipeline discovery request to the AIMLE server; the request may be based on a VAL client requirement(s) and the split operation request includes information defined in Table 8.14.3.2-1.
2. Upon receiving the request from the AIMLE client, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server may determine if existing instance(s) of a split operation pipeline satisfy the request parameters. If the AIMLE server determines existing instance(s) of split operation pipeline, the AIMLE server may proceed to step 3.

If no instance of a split operation pipeline satisfies the request parameters, the AIMLE server determines if an instance of a split operation pipeline can be created. The determination for split operation pipeline creation is based on availability of ML models and determination of available processing nodes (e.g., VAL server(s), NWDAF(s), etc.) to execute operations for such ML models (e.g., inference). VAL server nodes register with the AIMLE server to indicate their availability and capabilities as described in clause 8.14.2.4. If the AIMLE server determines that an instance of a split operation pipeline can be created, the AIMLE server creates a split operation profile accordingly as defined in Table 8.14.3.3-2.

3. The AIMLE server sends a split operation pipeline discovery response to the requestor. If the AIMLE server has determined instance(s) of split operation pipeline (e.g., existing, or new), the response includes an indication of success and includes the corresponding split operation profile(s). Otherwise, the response includes an indication of failure and may include a reason for failure.

Upon receiving a response including split operation profile(s), the AIMLE client stores the discovered split operation profile(s). The AIMLE client provides split operation profile information (e.g. endpoints, usage, etc.) to VAL client(s). The VAL client may use the provided information to access the split operation pipeline.

If the VAL client is a data provider, the VAL client may use the endpoint information to send data (e.g., intermediate inference data) towards the split operation pipeline instance's head endpoint.

If the VAL client is a data consumer, the VAL client may use the endpoint information to request results or subscribe to results from the split operation pipeline instance's tail endpoint.

8.14.2.3 Split operation pipeline creation

Figure 8.14.2.3-1 illustrates the procedure for an AIMLE client to create an instance of a split operation pipeline at the AIMLE server. If the required number of nodes are not discovered in using the discovery procedure (i.e., the number of nodes in the discovery response message are less than the required minimum number of nodes), then the AIML enablement client can discover the remaining required processing nodes from other AIMLE server(s).

Pre-conditions:

1. The AIMLE client has received information (e.g. URI, IP address) related to the AIMLE Server;
2. The AIMLE client has received security credentials authorizing it to communicate with the AIMLE server;
3. The AIMLE client has discovered information about processing nodes available with the AIMLE server;
4. The VAL client (along with the AIMLE client) selects the nodes, which can take part in the split operation and decides the order in which nodes should perform the split operation inference.

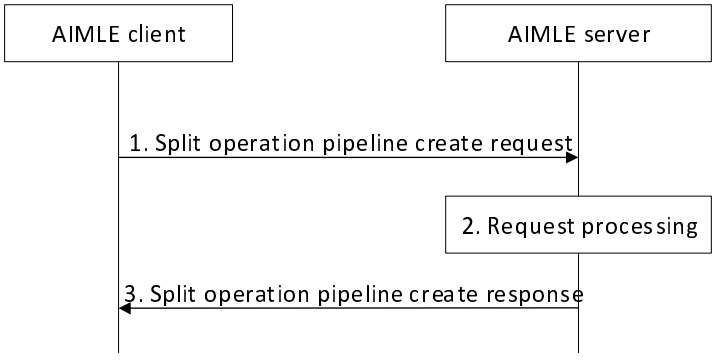

Figure 8.14.2.3-1: Split operation pipeline creation

1. The AIMLE client sends a split operation pipeline create request to the AIMLE server. The request includes information defined in Table 8.14.3.4-1.
2. Upon receiving the request from the AIMLE client, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server validates if the requested split operation pipeline can be created based on the information included in the request.

If the requestor is authorized and the AIMLE server has determined that an instance of a split operation pipeline can be created, the AIMLE server creates a split operation profile along with its identifier and notifies the appropriate processing nodes indicated in the request about their inclusion in the split operation pipeline as described in clause 8.14.2.5.

3. The AIMLE server sends a split operation pipeline create response message to the AIMLE client. If the AIMLE server has created an instance of a split operation pipeline, the response includes an indication of success, and the corresponding split operation profile. Otherwise, the response includes an indication of failure and may include a reason for failure.

Upon receiving a response including a split operation profile, the AIMLE client can store the split operation profile. The AIMLE client provides split operation profile information to VAL client(s). Split operation profile information can be used to access the split operation as described in clause 8.14.2.2.

8.14.2.4 Split operation node registration

8.14.2.4.1 General

Clause 8.14.2.4.2 illustrates the split operation node registration procedure.

Clause 8.14.2.4.3 illustrates the split operation node registration update procedure.

Clause 8.14.2.4.4 illustrates the split operation node de-registration procedure.

8.14.2.4.2 Split operation node registration

Figure 8.14.2.4.2-1 illustrates the procedure for VAL server to register with the AIMLE server to indicate split operation capabilities.

Pre-conditions:

1. The VAL server has received information (e.g. URI, IP address) related to the AIMLE Server;
2. The VAL server has received security credentials authorizing it to communicate with the AIMLE server;

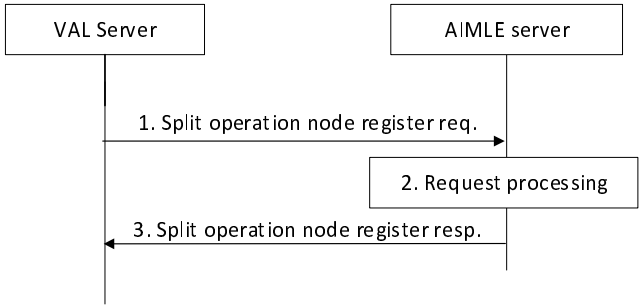

Figure 8.14.2.4.2-1: Split operation node registration

1. The VAL server sends a split operation node register request to the AIMLE server to indicate the split operation capabilities of the VAL server. The request includes information defined in Table 8.14.3.6-1.
2. Upon receiving the request from the VAL server, the AIMLE server validates if the requestor is authorized for the request, the AIMLE server stores the registration information.
3. The AIMLE server sends a split operation node register response to the requestor. If the AIMLE server has registered the requestor, the response includes an indication of success and may include an updated expiration time. The requestor shall send a registration update request before the expiration time to maintain the registration; otherwise, the split operation node registration expires. If the requestor was not registered, the response includes an indication of failure and may include a reason for failure.

8.14.2.4.3 Split operation node registration update

Figure 8.14.2.5-1 illustrates the procedure for VAL server to update a split operation node registration with the AIMLE server.

Pre-conditions:

1. The VAL server has registered with the AIMLE Server.

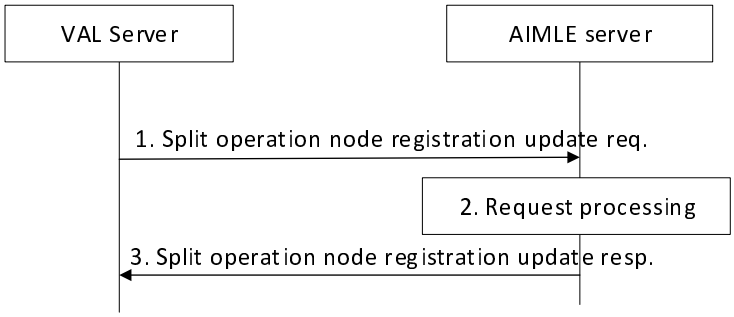

Figure 8.14.2.4.3-1: Split operation node registration update

1. The VAL server sends a split operation node registration update request to the AIMLE server to update an existing split operation registration. The request includes information defined in Table 8.14.3.8-1.
2. Upon receiving the request from the VAL server, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server updates the registration information.
3. The AIMLE server sends a split operation node registration update response to the requestor. If the AIMLE server has successfully updated the registration, the response includes an indication of success and may include an updated expiration time. The requestor shall send a registration update request before the expiration time to maintain the registration; otherwise, the split operation registration expires. If the registration was not updated, the response includes an indication of failure and may include a reason for failure.

8.14.2.4.4 Split operation node de-registration

Figure 8.14.2.4.4-1 illustrates the procedure for VAL server to de-register for split operation with the AIMLE server.

Pre-conditions:

1. The VAL server has registered with the AIMLE Server.

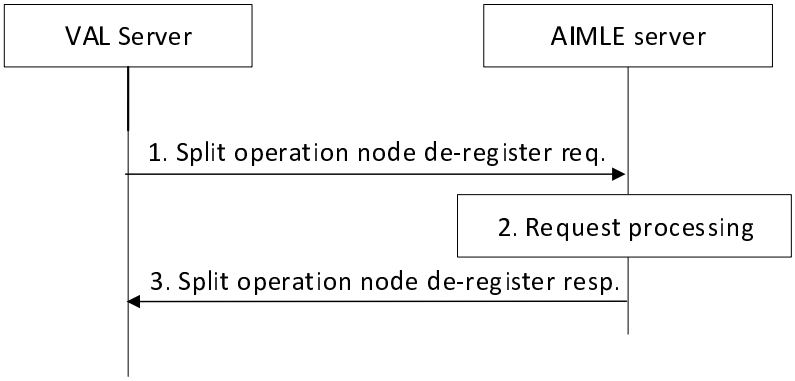

Figure 8.14.2.4.4-1: Split operation node de-registration

1. The VAL server sends a split operation node de-register request to the AIMLE server. The request includes information defined in Table 8.14.3.10-1.
2. Upon receiving the request from the VAL server, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server terminates the registration for split operation.

3. The AIMLE server sends a split operation node de-register response to the requestor. If the AIMLE server has successfully terminated the registration for split operation node, the response includes an indication of success. If the registration for split operation was not terminated, the response includes an indication of failure and may include a reason for failure.

8.14.2.5 Split operation event subscription

8.14.2.5.1 General

Clause 8.14.2.5.2 and clause 8.14.2.5.3 together illustrate the split operation subscribe/notify model.

Clause 8.14.2.5.4 illustrates the split operation subscription update procedure.

Clause 8.14.2.5.5 illustrates the split operation unsubscribe procedure.

8.14.2.5.2 Subscribe

Figure 8.14.2.5.2-1 illustrates the procedure for an AIMLE client or VAL server to subscribe with the AIMLE server to be notified of events related to split AI/ML operation.

Pre-conditions:

1. The AIMLE client or VAL server has received information (e.g. URI, IP address) related to the AIMLE Server;
2. The AIMLE client or VAL server has received security credentials authorizing it to communicate with the AIMLE server;

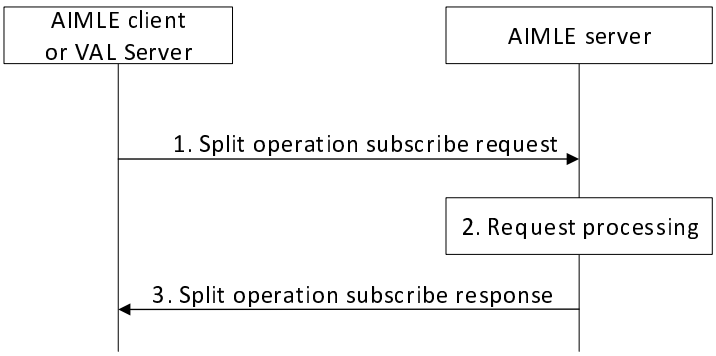

Figure 8.14.2.5.2-1: Split operation subscription

1. The requestor (e.g., AIMLE client or VAL server) sends a split operation subscribe request to the AIMLE server. The request includes information defined in Table 8.14.3.12-1.

   a. The requestor (e.g., AIMLE client or VAL server) may include the "split operation pipeline information" event to indicate the AIMLE server to notify the requestor when a split operation profile is created or an existing pipeline is updated (e.g., node is added or removed, stage order or ML models are modified, etc.) or deleted and satisfy the discovery filters of the subscription request.

   b. The requestor (e.g., AIMLE client or VAL server) may include the " split operation assistance information" event to indicate the AIMLE server to notify the requestor when the AIMLE server determines assistance information related to the split operation usage.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server creates the subscription and stores the subscription information.

3. The AIMLE server sends a split operation subscribe response to the requestor. If the AIMLE server has created the subscription, the response includes an indication of success, the subscription identity and may include an expiration time; to maintain the subscription, the requestor shall send a subscription update request before the expiration time, otherwise the split operation subscription expires. If the AIMLE server has not created the subscription, the response includes an indication of failure and may include a reason for failure.

If the subscription is for "assistance information", the AIMLE Server considers the assistance information included in the subscription request and can perform the following:

The AIMLE Server aggregates the collected assistance information from NEF, NWDAF, and/or ADAES to generate assistance information, e.g. Time (time point(s) or time window(s)) to deliver the task or data for the split operation. Or the AIMLE server performs inference using ML model to generate assistance information, e.g. achievable QoS with current configuration for task or data delivery, or suggestion of QoS for task or data delivery.

- For assistance information from NEF, the AIMLE Server, acting as AF, sends a request to NEF for assistance information (e.g. PDTQ on recommended time windows for AI/ML operations with QoS, QoS monitoring) as described in clause 4.16.15 of 3GPP TS 23.502 [3]. The request may include the information of the AI/ML task/ML model/data (e.g. size of the task or ML model, or data volume), information of the receiving nodes, requirement on the delivery/distribution (e.g. time budget, time critical or not, QoS), maximum time for complete the distribution/delivery.
- For assistance information from NWDAF, the AIMLE Server can send a request or subscription request to NWDAF for analytics (e.g. E2E data volume transfer time, DN performance, Network performance, UE mobility). The details parameters in the requests to NWDAF for analytics are given in 3GPP TS 23.288 [2].
- For assistance information from ADAES, the AIMLE Server can send a request or subscription request to ADAES for analytics (e.g. slice-specific application performance analytics, UE-to-UE application performance analytics). The details parameters in the requests to ADAES for analytics are given in 3GPP TS 23.436 [4].

8.14.2.5.3 Notify

Figure 8.14.2.5.3-1 illustrates the split operation notify operation between the AIMLE server and an AIMLE client or VAL server.

Pre-conditions:

2. The AIMLE client or VAL server has subscribed for split operation with the AIMLE Server;

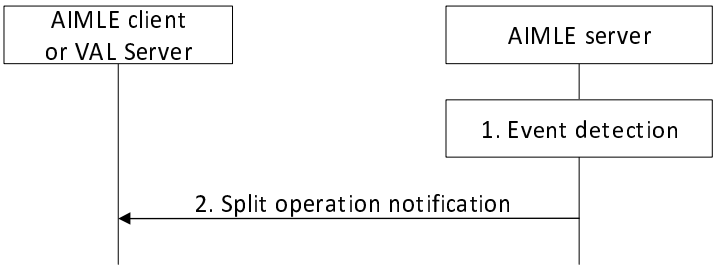

Figure 8.14.2.5.3-1: Split operation notification

1. The AIMLE server detects event(s) that satisfies trigger conditions for notifying a subscriber (e.g., AIMLE client or VAL server) according to subscribed events.

   a. If the subscribed event is for "split operation pipeline information" and the AIMLE server detects that a new split operation profile is created or an existing pipeline is updated (e.g., node is added or removed, stage order or ML models are modified, etc.) or deleted and satisfy the discovery filters of the subscription, the AIMLE server notifies the subscriber (e.g., AIMLE client or VAL server) accordingly;

   b. If the subscribed event is for "split operation assistance information" and the AIMLE server has determined assistance information related to split operation usage, the AIMLE server notifies the subscribers (e.g., AIMLE client or VAL server) accordingly.

2. The AIMLE server sends a split operation notification to the requestor indicating the event. The notification includes a subscription identity and may include a split operation profile(s), split operation node information, or split operation assistance information dependent on the associated event.

8.14.2.5.4 Subscription update

Figure 8.14.2.5.4-1 illustrates the procedure for an AIMLE client or VAL server to update a subscription with the AIMLE server.

Pre-conditions:

2. The AIMLE client or VAL server has subscribed for split operation with the AIMLE Server;

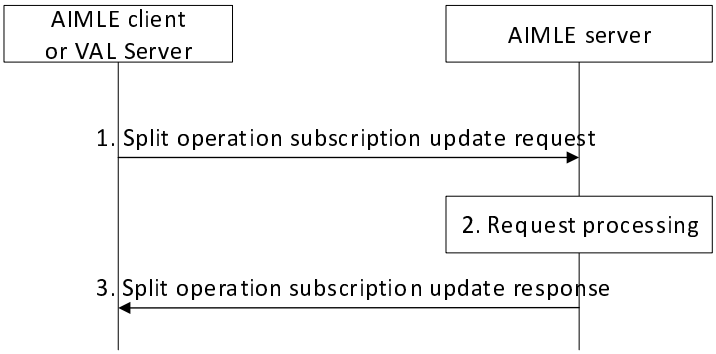

Figure 8.14.2.5.4-1: Split operation subscription update

1. The requestor (e.g., AIMLE client or VAL server) sends a split operation subscription update request to the AIMLE server. The request includes information defined in Table 8.14.3.15-1.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server updates the subscription information.

3. The AIMLE server sends a split operation subscription update response to the requestor. If the AIMLE server has updated the subscription, the response includes an indication of success and may include an expiration time. To maintain the subscription, the requestor shall send a subscription update request before the expiration time, otherwise the split operation subscription expires. If the AIMLE server has not updated the subscription, the response includes an indication of failure and may include a reason for failure.

8.14.2.5.5 Unsubscribe

Figure 8.14.2.5.5-1 illustrates the procedure for an AIMLE client or VAL server to unsubscribe with the AIMLE server.

Pre-conditions:

1. The AIMLE client or VAL server has subscribed for split operation with the AIMLE Server;

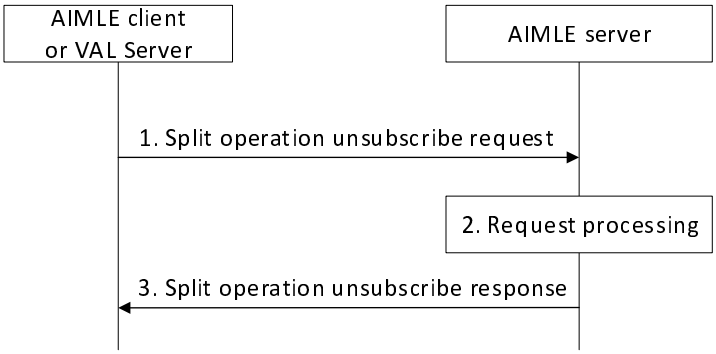

Figure 8.14.2.5.4-1: Split operation unsubscribe

1. The requestor (e.g., AIMLE client or VAL server) sends a split operation unsubscribe request to the AIMLE server. The request includes information defined in Table 8.14.3.17-1.

2. Upon receiving the request from the requestor, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server cancels the subscription associated with the subscription identifier.

3. The AIMLE server sends a split operation unsubscribe response to the requestor. If the AIMLE server has canceled the subscription, the response includes an indication of success. If the AIMLE server has not canceled the subscription, the response includes an indication of failure and may include a reason for failure.

8.14.2.6 Split operation pipeline update

Figure 8.14.2.6-1 illustrates the procedure for an AIMLE client to update an instance of a split operation pipeline at the AIMLE server. The AIMLE client or VAL client can determine that some of the nodes in the split operation pipeline may not be able to perform the required operation and decides to modify the pipeline either to add or remove the nodes.

Pre-conditions:

1. The AIMLE client has received information related to split operation pipeline profile as specified in clause 8.14.2.3.

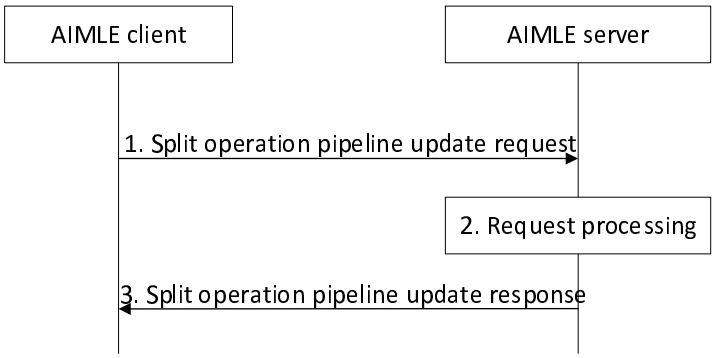

Figure 8.14.2.6-1: Split operation pipeline update

1. The AIMLE client sends a split operation pipeline update request to the AIMLE server. The request includes information defined in Table 8.14.3.19-1.

2. Upon receiving the request from the AIMLE client, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server validates if the requested split operation pipeline can be updated based on the split operation pipeline identifier included in the request.

   If the requestor is authorized and a split operation pipeline profile is determined, the AIMLE server updates the split operation pipeline profile and notifies the appropriate processing nodes indicated in the request about their inclusion or exclusion in the split operation pipeline as described in clause 8.14.2.5. The AIMLE server also notifies the already existing nodes about modification of the existing pipeline as described in clause 8.14.2.5.

3. The AIMLE server sends a split operation pipeline update response message to the AIMLE client. If the AIMLE server has updated an instance of a split operation pipeline profile, the response includes an indication of success, and the corresponding split operation profile. Otherwise, the response includes an indication of failure and may include a reason for failure.

8.14.2.7 Split operation pipeline delete

Figure 8.14.2.7-1 illustrates the procedure for an AIMLE client to delete an instance of a split operation pipeline at the AIMLE server. The AIMLE client or VAL client can determine to delete a split operation pipeline when it is no longer needed.

Pre-conditions:

1. The AIMLE client has received information related to split operation pipeline profile as specified in clause 8.14.2.3.

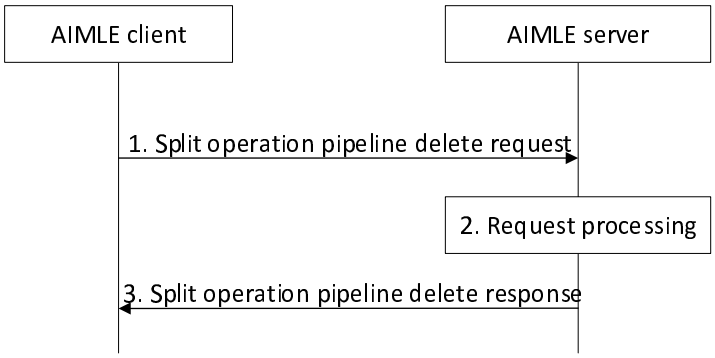

Figure 8.14.2.7-1: Split operation pipeline delete

1. The AIMLE client sends a split operation pipeline delete request to the AIMLE server. The request includes information defined in Table 8.14.3.21-1.

2. Upon receiving the request from the AIMLE client, the AIMLE server validates if the requestor is authorized for the request. If the requestor is authorized, the AIMLE server validates if the requested split operation pipeline can be deleted based on the split operation pipeline identifier included in the request.

   If the requestor is authorized and a split operation pipeline is determined, the AIMLE server deletes the split operation pipeline profile and notifies the appropriate processing nodes about deletion of the split operation pipeline as described in clause 8.14.2.5.

3. The AIMLE server sends a split operation pipeline delete response message to the AIMLE client. If the AIMLE server has deleted a split operation profile, the response includes an indication of success. Otherwise, the response includes an indication of failure and may include a reason for failure.

8.14.3 Information flows

8.14.3.1 General

The following information flows are specified for operation splitting.

8.14.3.2 Split operation pipeline discovery request

Table 8.14.3.2-1 shows the request sent by a AIMLE client to the AIMLE server for split operation pipeline discovery request.

Table 8.14.3.2-1: Split operation pipeline discovery request

Information element	Status	Description
Requestor identifier	M	The identity of the requestor (e.g., VAL client ID, AIMLE client ID, UE identifier)
Security credentials	M	The security credentials of the requestor.
Split operation discovery filters	M	Split operation discovery filters.
> stage information	M	Information about the split operation stages (e.g., number, order, etc.)
> model information	M	Information about the ML models to be used in each stage (e.g., identifiers, versions, etc.)
> usage information	O	Information about the planned usage of the split operation (e.g., inputs frequency/size, output frequency/size, etc.)
> number of nodes	O	Minimum number of nodes required to support AIML operation splitting

8.14.3.3 Split operation pipeline discovery response

Table 8.14.3.3-1 shows the response sent by sent by the AIMLE server after processing of split operation pipeline discovery request.

Table 8.14.3.3-1: Split operation pipeline discovery response

Information element	Status	Description
Successful response	O (NOTE 2)	Indicates that the request was successful.
> List of nodes (NOTE 1)	O	The list of discovered nodes.
> List of split operation profiles (NOTE 1)	O	The list of split operation profiles as specified in Table 8.14.3.3-2.
Failure response	O (NOTE 2)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE 1: Only one of the IE shall be present for successful response. NOTE 2: One of the IEs shall be present.		

Table 8.14.3.3-2: Split operation profile

Information element	Status	Description
Split operation pipeline identifier	M	The identifier of the split operation pipeline.
Head endpoint	M	The endpoint information of the head node for providing intermediate data.
Tail endpoint	M	The endpoint information of the tail node for obtaining results.
Usage information	O	The usage information of the AI/ML split operation (e.g., inputs frequency/size, output frequency/size, etc.).
Stage information	M	The list of stage(s) of the AI/ML split operation, the nested information is provided for each stage.
> Stage identifier	M	The identifier of the stage.
> number of nodes	O	Number of nodes included in the stage
> head node	M	Endpoint of the head node - for providing initial the inference data.
> tail node	M	Endpoint of the head node - for obtaining inference results.
> order of the nodes	M	The order of the nodes in the AI/ML split operation (including head node, tail node).
>> list of nodes	M	List of all discovered node in the order in which they process the data.
>> notification target	O	Endpoint information where the inference result of the split operation is sent by the nodes
> Model information	M	The ML model information used in the stage (e.g., identifiers, versions, etc.).

8.14.3.4 Split operation pipeline create request

Table 8.14.3.4-1 shows the request sent by a AIMLE client to the AIMLE server for split operation pipeline create request.

Table 8.14.3.4-1: Split operation pipeline create request

Information element	Status	Description
Requestor identifier	M	The identity of the requestor (e.g., VAL client ID, AIMLE client ID, UE identifier).
Security credentials	M	The security credentials of the requestor.
Split operation requirements	M	Split operation requirements.
> model information	M	Information about the ML models to be used in each stage (e.g., identifiers, versions, etc.).
> usage information	O	Information about the planned usage of the split operation (e.g., inputs frequency/size, output frequency/size, etc.).
> notification target	O	Endpoint information where the result of the split operation is sent by the tail node.
> stage information	M	Information about the split operation stages (e.g., number, order, etc.).
>> head node	M	Endpoint of the head node - for providing initial the inference data.
>> tail node	M	Endpoint of the head node - for obtaining inference results.
>> node information	M	List of Information about the nodes in order of the stage (e.g., identifier, endpoint, etc.).

8.14.3.5 Split operation pipeline create response

Table 8.14.3.5-1 shows the response sent by a AIMLE server after processing of split operation pipeline create request.

Table 8.14.3.5-1: Split operation pipeline create response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
> Split operation profile	O	The split operation profile as specified in Table 8.14.3.3-2.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.6 Split operation node register request

Table 8.14.3.6-1 shows the request sent by a VAL server to the AIMLE server for split operation node register request.

Table 8.14.3.6-1: Split operation node register request

Information element	Status	Description
Requestor identifier	M	The identity of the requestor (e.g., VAL server ID).
Security credentials	M	The security credentials of the requestor.
Node information	M	Information about the VAL server node (e.g., identifier, endpoints, etc.).
Split operation capabilities	M	Split operation capabilities of the VAL server.
> Model information	M	Information about ML model capabilities of the VAL server for split operation (e.g., identifiers, versions, etc.).
> Usage information	O	Information about usage capabilities of the VAL server for split operation (e.g., inputs frequency/size, output frequency/size, etc.).
Expiration time	O	The proposed expiration time of the registration.

8.14.3.7 Split operation node register response

Table 8.14.3.7-1 shows the response sent by an AIMLE server after processing of split operation node register request.

Table 8.14.3.7-1: Split operation node register response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
> Registration identifier	M	The identifier of the registration.
> Expiration time	O	The expiration time of the registration. To maintain an active registration, a registration update is required before the expiration time. If the Expiration time IE is not included, it indicates that the registration never expires.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.8 Split operation node registration update request

Table 8.14.3.8-1 shows the request sent by a VAL server to the AIMLE server for split operation node registration update request.

Table 8.14.3.8-1: Split operation node registration update request

Information element	Status	Description
Registration identifier	M	The identifier of the registration.
Security credentials	M	The security credentials of the requestor.
Node information	O	Information about the VAL server node (e.g., identifier, endpoints, etc.).
Split operation capabilities	O	Split operation capabilities of the VAL server.
> Model information	O	Information about ML model capabilities of the VAL server for split operation (e.g., identifiers, versions, etc.).
> Usage information	O	Information about usage capabilities of the VAL server for split operation (e.g., inputs frequency/size, output frequency/size, etc.).
Expiration time	O	The proposed expiration time of the registration.

8.14.3.9 Split operation node registration update response

Table 8.14.3.9-1 shows the response sent by an AIMLE server after processing of split operation node registration update request.

Table 8.14.3.9-1: Split operation node registration update response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
> Expiration time	O	The expiration time of the registration. To maintain an active registration, a registration update is required before the expiration time. If the Expiration time IE is not included, it indicates that the registration never expires.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.10 Split operation node de-register request

Table 8.14.3.10-1 shows the request sent by a VAL server to the AIMLE server for split operation de-register request.

Table 8.14.3.10-1: Split operation node de-register request

Information element	Status	Description
Registration identifier	M	The identifier of the registration.
Security credentials	M	The security credentials of the requestor.

8.14.3.11 Split operation node de-register response

Table 8.14.3.11-1 shows the response sent by an AIMLE server after processing of split operation de-register request.

Table 8.14.3.11-1: Split operation de-register response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.12 Split operation subscribe request

Table 8.14.3.12-1 shows the request sent by an AIMLE client or a VAL server to the AIMLE server for split operation subscribe request.

Table 8.14.3.12-1: Split operation subscribe request

Information element	Status	Description
Requestor identifier	M	The identity of the requestor (e.g., VAL server ID).
Security credentials	M	The security credentials of the requestor.
Split operation pipeline identifier	M	Identifier of the split operation pipeline for which the requestor is subscribing.
Notification endpoint	M	The notification endpoint (e.g. URL/URI/IP address) where the notifications should be sent to.
Event identifier	M	The event identifier for the subscription: - Split operation pipeline information - Split operation assistance information
Discovery filters	O	The set of characteristics to determine matching split operation profiles or nodes (as detailed in Table 8.14.3.2-1). Applicable for "Split operation pipeline information".
Assistance information	O	The assistance information for the subscription: - Aggregate the collected assistance information from NEF, NWDAF, and/or ADAES to generate assistance information, e.g. Time (time point(s) or time window(s)) to deliver the task or data for the split operation. - Perform inference using ML model to generate assistance information, e.g. achievable QoS with current configuration for task or data delivery, or suggestion of QoS for task or data delivery. Applicable for "Split operation assistance information".
Expiration time	O	The proposed expiration time of the subscription.

8.14.3.13 Split operation subscription response

Table 8.14.3.13-1 shows the response sent by an AIMLE server after processing of split operation subscription request.

Table 8.14.3.13-1: Split operation subscription response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
> Subscription identifier	M	The identifier of the subscription.
> Expiration time	O	The expiration time of the subscription. To maintain an active subscription, a subscription update is required before the expiration time. If the Expiration time IE is not included, it indicates that the subscription never expires.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.14 Split operation notification

Table 8.14.3.14-1 shows the request sent by an AIMLE server to notify a subscriber of detected split operation events.

Table 8.14.3.14-1: Split operation notification

Information element	Status	Description
Subscription identifier	M	The identifier of the subscription.
Event identifier	M	The identifier for the detected event, either: - Split operation pipeline information - Split operation assistance information
Availability information	O	The split operation availability information. Applicable for "Split operation availability" event.
> List of split operation profiles	O	The list of newly available split operation profiles as specified in Table 8.14.3.3-2.
> List of nodes	O	The list of newly available nodes.
Split operation pipeline information	O	The split operation pipeline information. Applicable for "Split operation pipeline information" event.
> Sub-event	O	Indicates the possible sub-event. - Created - Updated - Deleted
> Split operation profile	O	The split operation profile that the VAL server participates to as specified in Table 8.14.3.3-2.
Assistance information	O	The split operation assistance information. Applicable for "Split operation assistance information" event.
> Delivery time	O	Time (time point(s) or time window(s)) to deliver the task or data for the split operation.
> Achievable QoS	O	The achievable QoS with current configuration for task or data delivery.
> Suggestion of QoS	O	The suggestion of QoS for task or data delivery.

8.14.3.15 Split operation subscribe update request

Table 8.14.3.15-1 shows the request sent by an AIMLE client or a VAL server to the AIMLE server for split operation subscription update request.

Table 8.14.3.15-1: Split operation subscription update request

Information element	Status	Description
Subscription identifier	M	The identifier of the subscription.
Security credentials	M	The security credentials of the requestor.
Notification endpoint	O	The notification endpoint (e.g. URL/URI/IP address) where the notifications should be sent to.
Event identifier	O	The event identifier for the subscription: - Split operation pipeline information - Split operation assistance information
Split operation discovery filters	O	The set of characteristics to determine matching split operation profiles or nodes (as detailed in Table 8.14.3.2-1). Applicable for "Split operation availability" event.
Expiration time	O	The proposed expiration time of the subscription.

8.14.3.16 Split operation subscribe update response

Table 8.14.3.16-1 shows the request sent by an AIMLE server after processing of split operation subscription update request.

Table 8.14.3.16-1: Split operation subscription update response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
> Expiration time	O	The expiration time of the subscription. To maintain an active subscription, a subscription update is required before the expiration time. If the Expiration time IE is not included, it indicates that the subscription never expires.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.17 Split operation unsubscribe request

Table 8.14.3.17-1 shows the request sent by an AIMLE client or a VAL server to the AIMLE server for split operation unsubscribe request.

Table 8.14.3.17-1: Split operation unsubscribe request

Information element	Status	Description
Subscription identifier	M	The identifier of the subscription.
Security credentials	M	The security credentials of the requestor.

8.14.3.18 Split operation unsubscribe response

Table 8.14.3.18-1 shows the response sent by an AIMLE server after processing of split operation unsubscribe request.

Table 8.14.3.18-1: Split operation unsubscribe response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
Failure response	O (NOTE)	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of the IEs shall be present.		

8.14.3.19 Split operation pipeline update request

Table 8.14.3.19-1 shows the request sent by an AIMLE client to the AIMLE server for split operation pipeline update request.

Table 8.14.3.19-1: Split operation pipeline update request

Information element	Status	Description
Requestor identifier	M	The identity of the requestor (e.g., VAL client ID, AIMLE client ID, UE identifier).
Security credentials	M	The security credentials of the requestor.
Split operation pipeline identifier	M	The identifier of the AI/ML split operation pipeline.
Split operation pipeline information	M	Split operation pipeline information to be updated; this may include information elements defined in the split operation profile in Table 8.14.3.3-2.

8.14.3.20 Split operation pipeline update response

Table 8.14.3.20-1 shows the response sent by an AIMLE server after processing of split operation pipeline update request.

Table 8.14.3.20-1: Split operation pipeline update response

Information element	Status	Description
Successful response (NOTE)	O	Indicates that the request was successful.
> Split operation profile	O	The split operation profile as specified in Table 8.14.3.3-2.
Failure response (NOTE)	O	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of these IE is included.		

8.14.3.21 Split operation pipeline delete request

Table 8.14.3.21-1 shows the request sent by an AIMLE client to the AIMLE server for split operation pipeline delete request.

Table 8.14.3.21-1: Split operation pipeline delete request

Information element	Status	Description
Requestor identifier	M	The identity of the requestor (e.g., VAL client ID, AIMLE client ID, UE identifier).
Security credentials	M	The security credentials of the requestor.
Split operation pipeline identifier	M	The identifier of the AI/ML split operation pipeline.

8.14.3.22 Split operation pipeline delete response

Table 8.14.3.22-1 shows the response sent by an AIMLE server after processing of split operation pipeline delete request.

Table 8.14.3.22-1: Split operation pipeline delete response

Information element	Status	Description
Successful response (NOTE)	O	Indicates that the request was successful.
Failure response (NOTE)	O	Indicates that the request failed.
> Failure cause	O	Indicates the failure cause.
NOTE: One of these IE is included.		

8.15 AIMLE data management assistance

8.15.1 General

AIMLE data management assistance is the process of the AIMLE server assisting AIMLE service consumers with managing data operations performed by VAL clients. The data operations include data preparation and data analysis. The AIMLE server offloads the AIMLE service consumer from interacting with AIMLE clients to manage the data operations. The VAL clients perform the actual data preparation and data analysis operations and send the outputs to the AIMLE server for aggregation.

The following clauses specify procedures, information flows, and APIs for AIMLE data management assistance.

8.15.2 Procedure

Pre-conditions:

1. AIMLE clients have registered with AIMLE server.
2. UE application data for AI/ML operations have been collected and dataset identifier have been assigned to each dataset. EVEX mechanism can be reused for data collection as described in 3GPP TS 26.531 [10].

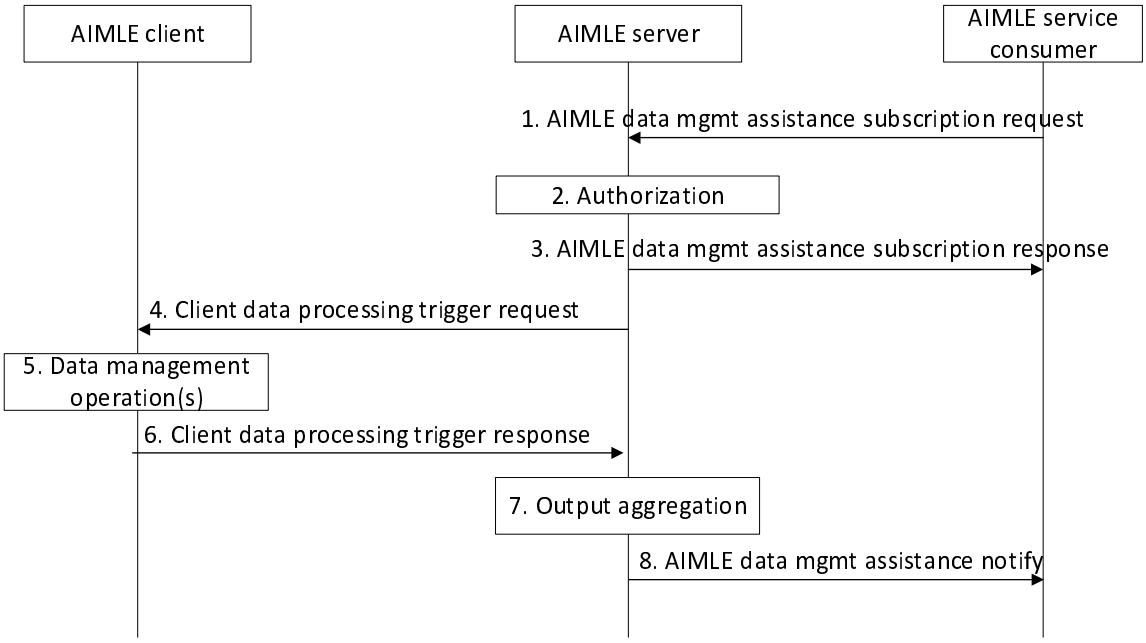

Figure 8.15.2-1: AIMLE data management assistance

1. An AIMLE service consumer (e.g., VAL server) makes a request for AIMLE data management assistance. The request includes information as described in Table 8.15.3.1-1.

To prepare data for AIML operations, the AIMLE service consumer configures Data management operations to data preparation. Data preparation is performed to process data into a format that is required by ML models for the models to function properly, e.g.; before training or inferencing. The AIMLE service consumer includes data preparation requirements as described in Table 8.15.3.1-2.

To analyse data for AIML operations (i.e., to perform exploratory data analysis or EDA), the AIMLE service consumer configures Data management operations to data analysis. The AIMLE service consumer includes data analysis requirements as described in Table 8.15.3.1-3.

In addition to providing the data management requirements, the AIMLE service consumer provides either a list of AIMLE clients (if known) or AIMLE client selection criteria. When providing the AIMLE client selection criteria, the AIMLE service consumer indicates that the AIMLE server performs AIMLE client selection to select AIMLE clients for the data management operations.

2. The AIMLE server authenticates and authorizes the request. If authorized, the AMILE server assigns an identifier for the subscription.

3. The AIMLE server sends an AIMLE data management assistance subscription response to the AIMLE service consumer. The response includes information as described in Table 8.15.3.2-1.

4. The AIMLE server sends Client data processing trigger requests to AIMLE clients. The request can be for data preparation or data analysis. The request includes information as described in Table 8.15.3.4-1. If AIMLE client selection criteria were provided in step 1, the AIMLE server performs AIMLE client selection to select AIMLE clients that have Dataset ID and Dataset feature ID indicated in the data management requirements.

5. Each AIMLE client sends the requirements to trigger data management for the VAL client to perform the requested data operation. The VAL client performs the operation locally.

NOTE: The data preparation and data analysis functions operate in a similar manner as ML models. The AIMLE layer is able to transport the functions to the VAL clients and the VAL clients performs the associated function on the application data identified by the Dataset ID and Dataset feature ID as part of data preparation or data analysis in a similar manner as ML training.

6. After the data operation completes, each AIMLE client sends a response to the AIMLE server with information as described in Table 8.15.3.5-1.

7. The AIMLE server aggregates the output of each AIMLE client. Aggregation includes combining or performing a statistical operation on the received data. Data received from the AIMLE clients can be numerical or categorical, which allows the AIMLE server to aggregate the output data from the AIMLE clients. If necessary, steps 4–7 is repeated to complete all the required data operations.

8. The AIMLE server sends an AIMLE data management assistance notification to the AIMLE service consumer with information as described in Table 8.15.3.3-1.

8.15.3 Information flows

8.15.3.1 AIMLE data management assistance subscription request

Table 8.15.3.1-1 shows the request sent by an AIMLE service consumer to an AIMLE server for the AIMLE data management assistance subscription procedure.

Table 8.15.3.1-1: AIMLE data management assistance subscription request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
Data management operations	M	An indicator showing what data management type is being requested: data preparation, data analysis.
Data management requirements	M	Requirements for the data management request:
>Data preparation requirements	O (NOTE 1)	Data preparation requirements as detailed in Table 8.15.3.1-2.
>Data analysis requirements	O (NOTE 1)	Data analysis requirements as detailed in Table 8.15.3.1-3.
AIMLE clients list	O (NOTE 2)	A list of AIMLE clients for which data management should be performed. The list may be specified by AIMLE client set identifier.
AIMLE client selection criteria	O (NOTE 2)	Selection criteria for finding suitable AIMLE clients for AI/ML operations as detailed in Table 8.8.3.1-2.
NOTE 1: At least one of the information elements shall be provided. NOTE 2: At least one of the information elements shall be provided.		

Table 8.15.3.1-2: Data preparation requirements

Information element	Status	Description
Dataset identifier	M	An identifier for the dataset.
Data preparation requirements	M	Requirements for data preparation.
> Dataset ID	M	The identifier for the dataset.
> Dataset feature ID	M	Identifier or name of dataset feature to process.
> Data preparation function	M	This indicates the function which prepares the data, and it could be an identifier of a function if the function is available locally at the UE or an executable included in the request.

Table 8.15.3.1-3: Data analysis requirements

Information element	Status	Description
Dataset identifier	M	An identifier for the dataset.
Dataset analysis requirements	M	Requirements for data analysis.
> Dataset ID	M	The identifier for the dataset.
> Dataset feature ID	M	Identifier or name of dataset feature.
> Data analysis function	M	This indicates the function which performs the data analysis, and it could be an identifier of a function if the function is available locally at the UE or an executable included in the request.

8.15.3.2 AIMLE data management assistance subscription response

Table 8.15.3.2-1 shows the response sent by the AIMLE server to the AIMLE service consumer for the AIMLE data management assistance subscription procedure.

Table 8.15.3.2-1: AIMLE data management assistance subscription response

Information element	Status	Description
Status	M	The status for the data management operation
Subscription identifier	M	An identifier for the subscription.
Expiration time	O	Expiration time for the subscription

8.15.3.3 AIMLE data management assistance notify

Table 8.15.3.3-1 shows the notification sent by the AIMLE server to the AIMLE service consumer for the AIMLE data management assistance subscription procedure.

Table 8.15.3.3-1: AIMLE data management assistance notify

Information element	Status	Description
Status	M	The status for the data management operation
Aggregated data preparation outputs	O (NOTE 1) (NOTE 2)	Provides outputs for data preparation: dataset identifier, dataset features, and prepared data output.
Aggregated data analysis outputs	O (NOTE 1) (NOTE 2)	Provides outputs for data analysis: dataset identifier, statistical outputs for each feature, list of outlier and anomaly values, and feature correlation information.
Timestamp	O	Timestamp of the data management notification
NOTE 1: At least one of the information elements shall be provided in the output. NOTE 2: The output format can be numerical or categorical. If categorical, the format can be nominal or ordinal.		

8.15.3.4 Client data processing trigger request

Table 8.15.3.4-1 shows the request sent by the AIMLE server to AIMLE clients for the AIMLE client data processing trigger procedure.

Table 8.15.3.4-1: Client data processing trigger request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
Data management type	M	An indicator showing what data management type is being requested: data preparation, data analysis.
Data management requirements	M	Requirements for the data management request:
>Data preparation requirements	O (NOTE)	Data collection requirements as detailed in Table 8.15.3.1-2.
>Data analysis requirements	O (NOTE)	Data collection requirements as detailed in Table 8.15.3.1-3.
Operational schedule	O	A schedule to perform the requested data management operation.
NOTE: At least one of the information elements shall be provided.		

8.15.3.5 Client data processing trigger response

Table 8.15.3.5-1 shows the response sent by AIMLE clients to the AIMLE server for the Client data processing trigger procedure.

Table 8.15.3.5-1: Client data processing trigger response

Information element	Status	Description
Status	M	The status for the data management operation
Data preparation outputs	O (NOTE 1)	The output data after performing data preparation. One output data is generated for each requirement. (NOTE 2)
Data analysis outputs	O (NOTE 1)	The output data generated by data analysis. One output data is generated for each requirement. (NOTE 2)
Timestamp	O	Timestamp of the data management operation
NOTE 1: At least one of the information elements shall be provided in the output. NOTE 2: The output format can be numerical or categorical. If categorical, the format can be nominal or ordinal. The AIMLE server is able to process either type of data.		

8.16 Support for Transfer Learning enablement

8.16.1 General

This clause provides the procedures for the transfer learning (TL) enablement service, including the server-triggered procedure (in clause 8.16.2) and the client-triggered procedure (in clause 8.16.3).

8.16.2 Procedure for server-triggered transfer learning enablement

Figure 8.16.2-1 illustrates the procedure where the TL enablement is performed based on the request for either an ML task from VAL layer or for an analytics task from ADAES. Such TL enablement allows the consumer to discover the similar ML models to be used as base models for the TL, as well as to support the selection of the best model to be used as pre-trained model.

Pre-conditions:

1. VAL server is connected to AIMLE Server.

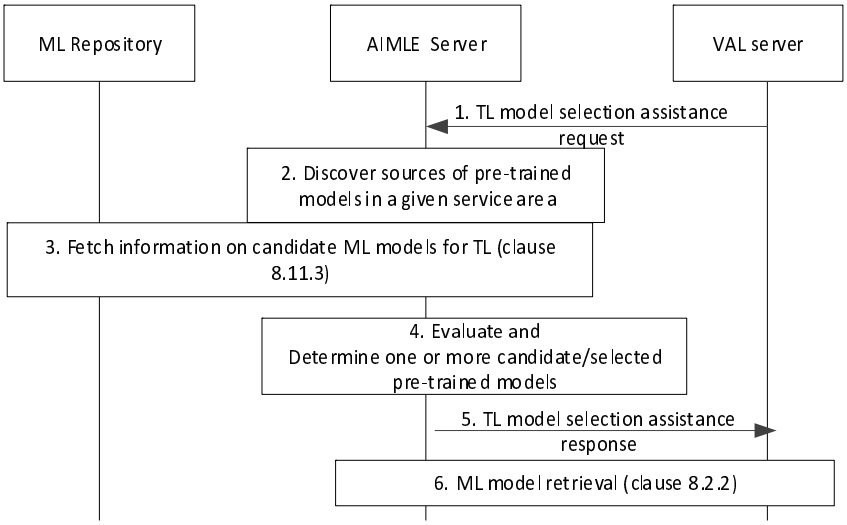

Figure 8.16.2-1: Procedure for Transfer Learning enablement

1. The VAL server sends a Transfer Learning model selection assistance request message to the AIMLE server to provide support for discovering and selecting the appropriate pre-trained model for a given ML task (for ADAE analytics ID or for a certain ML model ID).

2. The AIMLE Server discovers the possible entities which can provide a pre-trained model for this request. Such entities can be VAL servers or other ADAES or other AIMLE servers.

3. The AIMLE Server requests one or more pre-trained ML models which can be used for transfer learning for the target ML task. The ML repository identifies the base ML model as a pre-trained ML model that can be mapped to the target ML task and sends information on the models to AIMLE server which are candidate as pre-trained models available for the target task. This step reuses the ML model information discovery procedure as in clause 8.11.3.

4. The AIMLE Server evaluates with the support of the ML model repository, whether the pre-trained models are applicable to the ML task (ADAE analytics ID or model ID). This can be assisted using historical data or ML model rating based on previous utilization of these models for the certain ML task. Based on the evaluation (which can be based on the rating), the AIMLE Server determines one or more pre-trained models to be used for the ML task.

NOTE: In this step, the AIMLE Server can rate or set a weight to the pre-trained model or the source of the model.

5. The AIMLE Server sends to the VAL server a transfer learning selection assistance response to the VAL server, including the information for the pre-trained models (e.g., model ID, profile) which are identified for the ML

task. Also, this may include the rating/weight for the pre-trained model if the VAL server needs to select among a list of them.

6. Based on the selected pre-trained model information, the VAL server retrieves the selected ML model using the procedure as in clause 8.2.2.

8.16.3 Procedure for client-triggered transfer learning enablement

Figure 8.16.3-1 illustrates the procedure where the TL enablement is performed for a VAL UE task which is a UE analytics task (e.g. ADAEC-provided analytics). Such TL enablement allows the VAL UE to perform ML model training using a pre-trained model from the server side and is beneficial for minimizing the computational load at the VAL UE side.

Pre-conditions:

1. AIMLE client is connected to AIMLE Server.

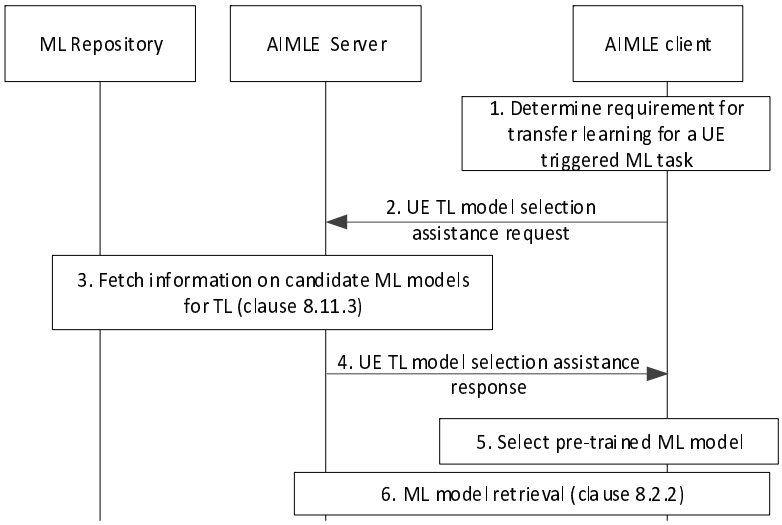

Figure 8.16.3-1 Procedure for client-triggered Transfer Learning enablement

1. The AIMLE client determines a requirement for using a pre-trained ML models for transfer learning for a VAL UE triggered ML task (e.g., UE analytics event).

2. The AIMLE client sends a UE transfer learning model selection assistance request to the AIMLE server to receive one or more pre-trained ML models which can be used for transfer learning for the target UE-triggered ML task.

3. The AIMLE Server requests one or more pre-trained ML models which can be used for transfer learning for the target ML task. The ML repository identifies the base ML model as a pre-trained ML model that can be mapped to the target ML task and sends information on the models to AIMLE server which are candidate as pre-trained models available for the target task.

4. The AIMLE Server sends a UE transfer learning model selection assistance response to the AIMLE client which includes information on the ML models which are candidate as pre-trained models available for the target UE ML task. Such models may be pre-trained for an ADAES analytics task (e.g., VAL server performance analytics) and are applicable to be used for the VAL UE analytics task (e.g. VAL session performance analytics).

5. The AIMLE client evaluates whether the pre-trained models are applicable to the target ML task. Based on the evaluation, the AIMLE client selects a base model to be used as pre-trained model for the ML task.

6. Based on the selected pre-trained model information, the AIMLE client retrieves the selected ML model using the procedure as in clause 8.2.2.

8.16.4 Information flows

8.16.4.1 Transfer learning model selection assistance request

Table 8.16.4.1-1 shows the request sent by the VAL server to AIMLE server for the server-triggered transfer learning support procedure.

Table 8.16.4.1-1: Transfer learning model selection assistance request

Information element	Status	Description
Requestor identity	M	Identity of the VAL server performing the request.
VAL service identity	M	The identity of the VAL service for which the request applies.
ML task identity	O (NOTE)	The ML task for which the transfer learning is to be used.
ADAE analytics ID	O (NOTE)	The ADAES analytics ID (as specified in TS 23.436) for which the transfer learning is to be used, in case when transfer learning is used per analytics task.
ML model profile	O (NOTE)	The ML model profile for which the transfer learning is to be used.
Transfer learning criteria	M	The criteria for identifying and selecting one or more pre-trained ML models. Such criteria include: - the required feature(s) of a pre-trained model. - training data requirements. - type of transfer learning. - the environment associated with the target ML task. - permissions / restrictions for the pre-trained model.
ML model requirement information	O	Identifies the requirement for selecting a base model to be trained as pre-trained.
List of VAL UE IDs	O	List of VAL UEs associated with the ML model task
Model rating requirement	O	Identifies the requirement for providing rating of the ML model(s) to serve as pre-trained model.
NOTE: At least one of these IE shall be present.		

8.16.4.2 Transfer learning model selection assistance response

Table 8.16.4.2-1 shows the response sent by the AIMLE server to the VAL server for the server-triggered transfer learning support procedure.

Table 8.16.4.2-1: Transfer learning model selection assistance response

Information element	Status	Description
Success response	O (NOTE)	Indicates that the transfer learning model selection assistance request was successful.
> List of ML models	O	List of ML models selected by AIMLE server for training as candidate pre-trained model.
>> ML repository identifier and address	O	Provides the ID and address of the ML repository which stores the pre-trained ML model selected by AIMLE server for training as pre-trained model.
>> ML model information	O	Information on the selected model, specified in Table 8.11.4.1-2.
>> ML model rating	O	If requested, a rating parameter for the ML model to serve as pre-trained. Such rating can be based on the ML task similarity score e.g. based on the feature.
Failure response	O (NOTE)	Indicates that the transfer learning model selection assistance request was failure.
> Cause	M	Reason for the failure.
NOTE: Only one of these information elements shall be present		

8.16.4.3 UE transfer learning model selection assistance request

Table 8.16.4.3-1 shows the request sent by the AIMLE client to AIMLE server for the client-triggered transfer learning support procedure.

Table 8.16.4.3-1: UE transfer learning model selection assistance request

Information element	Status	Description
VAL UE identity	M	Identity of the VAL UE (VAL client ID or AIMLE client ID) performing the request.
VAL service identity	M	The identity of the VAL service for which the request applies.
ML task identity	O (NOTE)	The ML task for which the transfer learning is to be used.
ADAE analytics ID	O (NOTE)	The ADAE analytics ID (as specified in 3GPP TS 23.436) for which the transfer learning is to be used, in case when transfer learning is used per UE analytics task.
ML model profile	O (NOTE)	The ML model profile for which the transfer learning is to be used.
Transfer learning criteria	M	The criteria for identifying and selecting one or more pre-trained ML models. Such criteria include: - the required feature(s) of a pre-trained model. - training data requirements. - type of transfer learning. - the environment associated with the target ML task. - permissions / restrictions for the pre-trained model.
ML model requirement information	O	Identifies the requirement for selecting a base model to be trained as pre-trained.
Model rating requirement	O	Identifies the requirement for providing rating of the ML model(s) to serve as pre-trained model.
NOTE: At least one of these IE shall be present.		

8.16.4.4 UE transfer learning model selection assistance response

Table 8.16.4.4-1 shows the response sent by the AIMLE server to the AIMLE client for the client-triggered transfer learning support procedure.

Table 8.16.4.4-1: UE transfer learning model selection assistance response

Information element	Status	Description
Success response	O (NOTE)	Indicates that the UE transfer learning model selection assistance request was successful.
> List of ML models	O	List of ML models selected by AIMLE server for training as pre-trained model.
>> ML model information	O	Information on the selected model, specified in Table 8.11.4.1-2.
>> ML model rating	O	If requested, a rating parameter for the ML model to serve as pre-trained.
Failure response	O (NOTE)	Indicates that the UE transfer learning model selection assistance request was failure.
> Cause	M	Reason for the failure.
NOTE: Only one of these information elements shall be present		

8.17 Support for FL member grouping

8.17.1 General

This clause provides the procedure for the grouping of the FL members using AIMLE. Such grouping for the given FL process, applicable to a specific VAL request or ML model ID or ADAE analytics ID. This grouping can be applicable also for a given service area in which one or more FL processes are expected to run. The grouping of FL members is performed for optimizing the process of selection and updating FL members which are entering or leaving the group, since due to dynamicity of FL member (e.g. AIMLE clients) changes this would be impose additional signalling / complexity.

8.17.2 Procedure

In this procedure, the AIMLE support capability is described for grouping the FL members, where the grouping is tailored to a specific ML task (VAL triggered task or analytics event/ID).

The grouping procedure covers the:

- creation of the FL member group;
- query for an individual FL Member whether it is part of the created group;
- change of the FL member group, including
  a) modification of the group member based on a change on the availability or capability of the FL member (e.g. due to high load or energy consumption the FL member may have limited capability to act as FL client for a given area and time);
  b) update of the group due to a new member entering or an existing member leaving the FL group; and
- deletion of the FL member group, and notification of the FL members regarding the deletion of the FL member group.

Figure 8.17.2-1 illustrates the procedure for supporting the FL member grouping.

Pre-conditions:

1. VAL Server is connected to AIMLE Server.
2. The candidate/selected FL member has registered to the FL member registry based on the capability in clause 8.4.

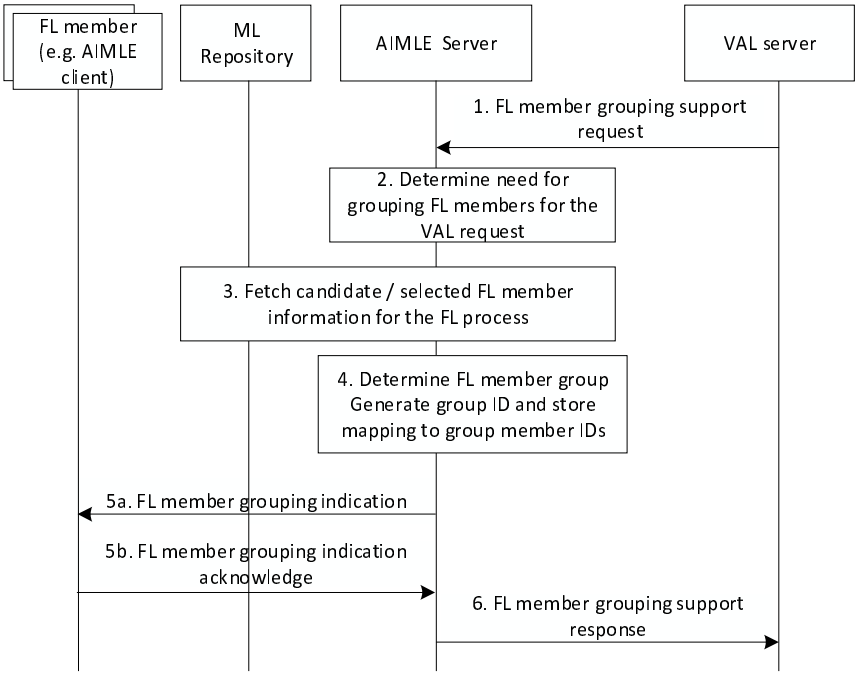

Figure 8.17.2-1: Procedure for FL member grouping

1. The VAL server sends an FL member grouping support request to the AIMLE Server for supporting an FL process. The initial request is to create the FL member grouping support as described in Table 8.17.3.1-1, which may be followed by other requests for querying (as described in Table 8.17.3.1-2), change (as described in Table 8.17.3.1-3), or deletion (as described in Table 8.17.3.1-4) of the FL member grouping support.

NOTE 1: The FL member grouping support request can be triggered by the AIMLE Server itself.

2. The AIMLE Server based on the request, determines the need for creating and processing a group consisting of the needed FL members for a given ML task (i.e., an ML model training/inference job ID). The need for creating an FL member group may be based on an ML task for a given AIMLE service area or for a given AIMLE service area where one or more ML tasks are expected to run, based on step 1 request.

3. To create, query, or change the FL member grouping support, the AIMLE Server fetches the available FL members for the given ML task (i.e., an ML model training/inference job ID) from the ML repository. Based on this information, AIMLE Server may select one or more FL members for the group for the ML task.

NOTE 2: If the FL member is an AIMLE client, this step re-uses procedure 8.9.2 on AIMLE client selection.

4. The AIMLE Server creates, configures, and processes the FL member group based on the available or selected FL members by the aggregator which may be the VAL Server or the AIMLE Server, generate group ID and store mapping to group member IDs. The criteria for determining the group members can be the capabilities of the FL participants, or whether the candidate participants are fixed or mobile nodes and their availabilities, the proximity of the participants among them.

   For the creation of the FL member group, AIMLE may utilize SEAL GMS capability for the FL member group ID generation.

5. The AIMLE Server interacts with the each candidate FL member from the configured FL member group as shown in step 5a.

NOTE 3: If the request is about the deletion of the FL member group, the FL member grouping indication includes notification for the deletion of the FL member group.

5a. The AIMLE Server sends indication to the candidate FL member (the AIMLE client which is deployed on UE) about the group ID and the group member identities for the ML model ID / ADAE analytics ID (based on the request in step 1.

5b. The candidate FL member sends to the AIMLE Server a FL grouping indication acknowledge.

6. The AIMLE Server sends an FL member grouping support response to the VAL request indicating the group creation or providing the query response, change or deletion (based on step 1 request) and the group information.

8.17.3 Information flows

8.17.3.1 FL member grouping support request

Table 8.17.3.1-1 shows the request sent by the VAL Server to AIMLE Server for the FL member grouping procedure to create the FL member grouping support.

Table 8.17.3.1-1: FL member grouping support create request

Information element	Status	Description
Requestor identity	M	Identity of the VAL Server performing the request.
VAL service identity	O (NOTE)	The identity of the VAL service for which the request applies.
ML model ID	O (NOTE)	The model ID for which the request applies
ADAE analytics ID	O (NOTE)	The ADAES analytics ID (as specified in TS 23.436) for which the FL grouping is to be used, in case when FL process is used for a given analytics task.
ML task information	O	Information related to the ML task / job for which the FL grouping is used. Such task can be an FL training task or FL inference task. This information may include an ML task ID which may be an FL process ID or correlation ID.
ML model profile	O	The ML model profile for which the FL grouping is to be used.
List of candidate/selected FL member IDs	O	The list of candidate or selected FL member identities (if known by the VAL Server) which are to be used in the grouping. If the FL member is a VAL UE, this is equivalent to the VAL UE ID.
> FL member type	O	The type of FL members (FL client, FL Server)
> FL member status	O	The status (selected, candidate) of the FL member.
NOTE: At least one of these information elements shall be present		

Table 8.17.3.1-2 shows the request sent by the VAL Server to AIMLE Server for the FL member grouping procedure to query the FL member within the group.

Table 8.17.3.1-2: FL member grouping support query request

Information element	Status	Description
FL member group identity	M	Identity of the FL member group which queried
FL member identity	O	Information on the queried FL member to be queried.

Table 8.17.3.1-3 shows the request sent by the VAL Server to AIMLE Server for the FL member grouping procedure to change the FL member grouping support.

Table 8.17.3.1-3: FL member grouping support change request

Information element	Status	Description
FL member group identity	M	Identity of the FL member grouping support
FL member group change	M	Information on the change type for the FL member group
>FL member update	O (NOTE)	Identifies the FL Members that are to be updated
>>Cause	O	The cause for the FL member group update (e.g. FL member enter or leave the group)
>FL member group modify	O (NOTE)	Identifies the FL Members that are to be modified
>>Cause	O	The cause for the FL member group modifies (e.g. change FL member availability, capability or FL member information).
NOTE: At least one of these shall be present		

Table 8.17.3.1-4 shows the request sent by the VAL Server to AIMLE Server for the FL member grouping procedure to delete the FL member grouping support.

Table 8.17.3.1-4: FL member grouping support delete request

Information element	Status	Description
FL member group identity	M	Identity of the FL member grouping support
Cause	O	Cause for the deletion of the group

8.17.3.2 FL member grouping support response

Table 8.17.3.2-1 shows the response sent by the AIMLE Server to the VAL Server for the FL member grouping procedure to create the FL member grouping support.

Table 8.17.3.2-1: FL member grouping support create response

Information element	Status	Description
Success response	O (NOTE)	Indicates that the FL process support request was successful.
> FL group identifier(s)	O	Identifies the AIMLE-created FL group for the FL process and ML task (FL training or inference).
> List of FL member IDs / addresses	O	Provides the ID and address of the FL members which are part of the FL group.
>> FL member information	O	Information on the FL members such as availability, constraints, role/type.
Failure response	O (NOTE)	Indicates that the FL process support request was failure.
> Cause	M	Reason for the failure.
NOTE: Only one of these information elements shall be present		

Table 8.17.3.2-2 shows the response sent to the VAL Server by the AIMLE Server for the FL member grouping procedure to query the FL member within the group.

Table 8.17.3.2-2: FL member grouping support query response

Information element	Status	Description
Success response	O (NOTE)	Indicates that the FL process support query request was successful.
> List of FL member IDs / addresses	O	Provides the ID and address of the FL members which are part of the FL group.
>> FL member information	O	Information on the FL members such as availability, constraints, role/type.
Failure response	O (NOTE)	Indicates that the FL process query support request was failure.
> Cause	M	Reason for the failure.
NOTE: Only one of these information elements shall be present		

Table 8.17.3.2-3 shows the response sent to the VAL Server by the AIMLE Server for the FL member grouping procedure to change the FL member grouping support.

Table 8.17.3.2-3: FL member grouping support change response

Information element	Status	Description
Success response	O (NOTE)	Indicates that the FL process support request was successful.
> FL group identifier(s)	O	Identifies the AIMLE updated or modified FL group for the FL process and ML task (FL training or inference).
> List of FL member IDs / addresses	O	Provides the ID and address of the update or modified FL members which are part of the FL group.
>> FL member information	O	Information on the updated or modified FL members such as availability, constraints, role/type.
Failure response	O (NOTE)	Indicates that the FL process support request was failure.
> Cause	M	Reason for the failure.
NOTE: Only one of these information elements shall be present		

Table 8.17.3.3-4 shows the response sent to the VAL Server by the AIMLE Server for the FL member grouping procedure to delete the FL member grouping support.

Table 8.17.3.2-4: FL member grouping support delete response

Information element	Status	Description
Result	M	Positive or negative acknowledgement for the deletion of the FL member group.

8.17.3.3 FL grouping indication

Table 8.17.3.3-1 shows the notification sent by the AIMLE Server to the FL members (AIMLE clients, VAL Servers) for the FL member grouping procedure.

Table 8.17.3.3-1: FL grouping indication

Information element	Status	Description
Requestor identity	M	Identity of the AIMLE Server performing the request.
VAL service identity	O (NOTE 1)	The identity of the VAL service for which the grouping indication applies.
ML model ID	O (NOTE 1)	The model ID for which the indication applies
ADAE analytics ID	O (NOTE 1)	The ADAES analytics ID (as specified in 3GPP TS 23.436) for which the FL grouping is to be used, in case when FL process is used for a given analytics task.
FL group identifier(s)	M	Identifies the AIMLE-created, changed FL group for the FL process.
> List of FL member IDs / addresses	O	Provides the ID and address of the FL members which are part of the FL group.
>> FL member information	O	Information on the FL members such as availability, constraints, role/type.
FL group deletion information	O (NOTE 2)	Indication that the FL group is going to be deleted based on VAL server request
> Cause	O	Cause for the expected deletion of the FL members group (e.g., due to AI/ML service termination or group UE mobility to different service area).
> Expiration time	O	Indicates the expiration time of the FL group deletion (in case the deletion of the FL group is expected in future time instance). If the Expiration time IE is not included, it indicates that the deletion of the group is instant.
NOTE 1: At least one of these information elements shall be present. NOTE 2: This IE is mandatory if the indication is related to an FL group deletion.		

8.17.3.4 FL grouping indication acknowledge

Table 8.17.3.4-1 describes the information flow FL grouping indication acknowledge from the FL members (AIMLE clients which are deployed on UEs) to the AIMLE server.

Table 8.17.3.4-1: FL grouping indication acknowledge

Information element	Status	Description
Success response	O (NOTE)	Acknowledgement of FL grouping indication.
Failure response	O (NOTE)	Indicates that the FL grouping indication was failure.
> Cause	O	Reason for the failure.
NOTE: Only one of these information elements shall be present		

8.18 Support Vertical FL

This clause describes procedure for supporting VFL among application layer multiple UEs.

8.18.1 General

The following clauses specify procedures, information flows and APIs to support VFL among Application Layer multiple UEs.

8.18.2 Procedure for supporting VFL

Pre-conditions:

1. VFL members registered their AIMLE client profile to the AIMLE Server. The VFL members may update their status or information in their AIMLE client profile to the AIMLE Server.

2. The datasets of each of the UEs (where the AIMLE Clients (as VFL members) are deployed on) belong to more than one different data domains.

3. The VAL server has successfully subscribed/registered with the AIMLE server for model training notifications.

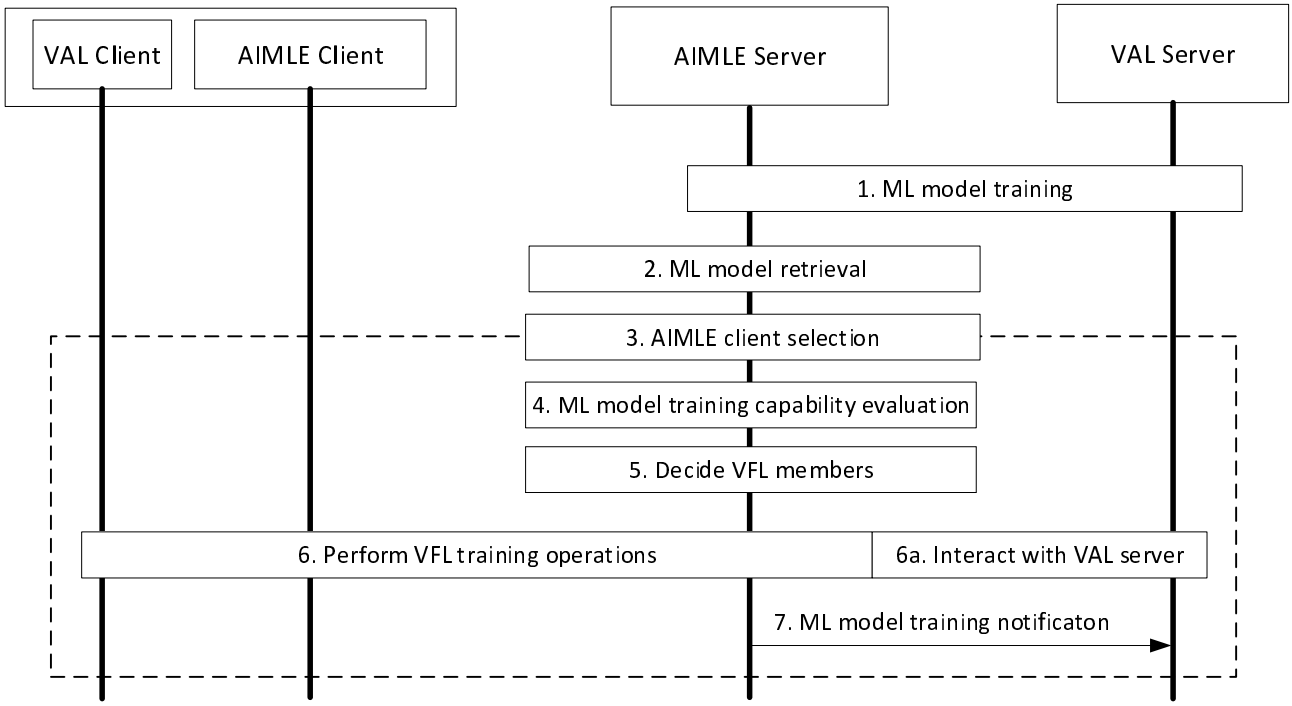

Figure 8.18.2-1: Procedure for supporting VFL

1. An AIMLE server receives an ML model training request from a VAL server as described in clause 8.3.2. If ML model training request is received and the AIMLE server determines to use VFL training (VFL training between different data domains), the procedure continues to step 2.

2. [Optional] The AIMLE server retrieves the indicated ML model in step 1 using the ML model retrieval procedure as described in clause 8.1. If the retrieved model is already trained and meets the machine learning requirements requested by the consumer, the procedure continues to step 7.

3. If AIMLE client selection criteria are provided in step 1, then the AIMLE server continuously monitors and selects AIMLE clients for the VFL training as described step 4 of the clause 8.13.2. If a number of the required AIML clients is provided, the AIMLE server ensures the requirement is met when selecting AIMLE clients for VFL training. If a list of AIMLE clients is provided in step 1, the AIMLE server selects the provided AIMLE clients for the VFL training.

4. The AIMLE Server gets VFL members information (the VFL members information may be included in the request in step 1 (Table 8.3.3.1-1) or be obtained through step 3. The AIMLE Server interacts with the VFL members (AIMLE Clients which are deployed on UEs) for each data domain for ML model training capability evaluation as described in clause 8.19.2.

5. The AIMLE Server determines the VFL members for this VFL training process based on the information received in step 4 and parameters received in step 1 (Table 8.3.3.1-1). The AIMLE Server determines VFL members for each data domain. The criteria used by the AIMLE Server include:

   - Available data and minimum number of data samples for the same sample among the VFL members.
   - Feature alignment of the sample/datasets with data labels among the VFL members.
   - Available time of the VFL members for support the VFL training operations.
   - Capability and minimum number of the VFL members for the VFL training operations.
   - AIML model information for the VFL members and for the AIMLE Server.

6. The AIMLE Server coordinates the selected VFL members for VFL training. During VFL training process, the VFL members send intermediate results to the AIMLE Server, and the AIMLE Server responds to the VFL members with the updated information (e.g. gradients). The information from the AIMLE server can be used to update the model parameters maintained at each VFL member for the different data domains.

6a. The AIMLE server may report to the VAL server with the training status, that includes intermediate training results, The VAL server may adjust its request on the ML model training. If the VAL Server is providing data labels to complete the training, the VAL Server sends updated training parameters for the AIMLE Server to distribute to the VFL members. The updated training parameters apply for models of VFL members associated with each data domain.

NOTE: How AIMLE Server coordinates the selected VFL members to perform VFL training is out of scope for this release.

7. The AIMLE server sends a ML model training notification to the VAL server as described in clause 8.3.2. If the training schedule is not complete (e.g., there are remaining training rounds), the AIMLE server configures the next set of training schedules and steps 3 to 6 are repeated for the next training round.

8.19 ML Model Training Capability Evaluation

This clause describes procedure for supporting ML model training capability evaluation for FL (e.g., HFL, VFL).

8.19.1 General

The following clauses specify procedures, information flows and APIs to support ML model training capability evaluation for FL (e.g., HFL, VFL). The ML model training capability result can be used by the AIMLE server to select FL members for FL training process (e.g. HFL, VFL).

8.19.2 Procedure for ML model training capability evaluation

Pre-conditions:

1. AIMLE server determines to use FL (e.g., HFL, VFL) training.
2. AIMLE Server knows the information of the FL (e.g., HFL, VFL) members (e.g., AIMLE Client which is deployed to a UE).

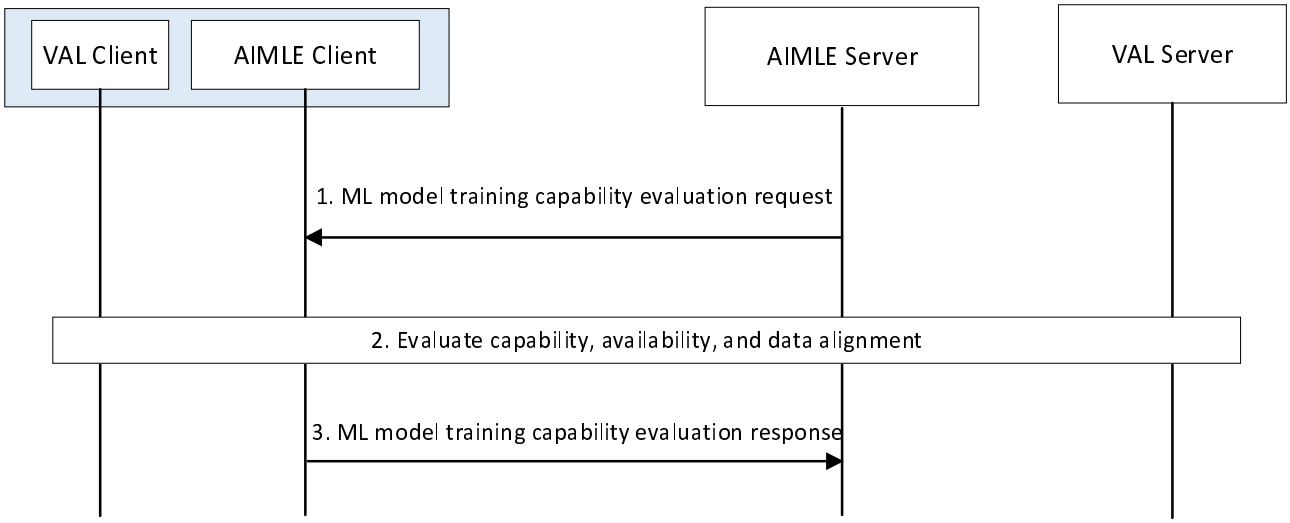

Figure 8.19.2-1: Procedure for ML model training capability evaluation

1. The AIMLE Server sends ML model training capability evaluation request to the FL members (AIMLE Clients which are deployed on UEs). The request message includes information as described in Table 8.19.3.1-1.

2. The FL members evaluate their capability and availability to join the FL training process. The FL members (AIMLE Clients which are deployed on UEs) run the test task contained in the request in step 1 and determine if it can join the FL process. For VFL, as part of the test task, data alignment between the datasets of the different domains are determined. The VAL server may also provide data labels for the data alignment.

NOTE: The procedures for data collection from UE need to take user consent into account.

3. The FL members (AIMLE Clients which are deployed on UEs) send ML model training capability evaluation response to the AIMLE Server. The response message contains the information as described in Table 8.19.3.2-1.

8.19.3 Information flows

8.19.3.1 ML model training capability evaluation request

Table 8.19.3.1-1 shows the request sent by AIMLE server to FL (e.g., HFL, VFL) members (AIMLE Clients which are deployed on UEs) for ML model training capability evaluation.

Table 8.19.3.1-1: ML model training capability evaluation request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
Available time	O	Requirement on available time for supporting FL operations.
Test task	O	The task for test ML model training capability.
AI/ML model and model parameter(s)	O	Information about the AI/ML model and model parameters for use in FL training. In VFL, AI/ML models for different data domains are provided.
Requirement on dataset	O	Requirements on dataset for FL training.
>Common feature ID(s)	O (NOTE 1)	Identifier(s) of the required features common to the dataset of the different data domains.
>Data domain feature ID lists	O (NOTE 1)	List of features for each data domain(s) of the datasets at the UE.
>Data source	O (NOTE 2)	Data source for the FL training.
NOTE 1: If Requirement on dataset is provided, at least one of these IEs shall be present when for VFL. NOTE 2: If Requirement on dataset is provided, the IE shall be present when for HFL.		

NOTE: The detail content of the Test task and AI/ML model is up to implementation.

8.19.3.2 ML model training capability evaluation response

Table 8.19.3.2-1 shows the response sent by FL members (AIMLE Clients which are deployed on UEs) to AIMLE server for the ML model training capability evaluation.

Table 8.19.3.2-1: ML model training capability evaluation response

Information element	Status	Description
Status	M	The status for the evaluation: success, fail. - Success means join the FL training process. - Fail means not join the FL training process.
Test result	O	The test result of the ML model training capability evaluation. The "test result" shall be provided when the "status" is "success".
Fail reason	O	The reason of the ML model training capability evaluation fail. The "fail reason" shall be provided when the "status" is "fail".

8.20 AIML service operations control and management procedure

8.20.1 General

The control and management of the AIML services is an essential requirement for the applications to manage the AIML services like model training, inference, discovery etc.

8.20.2 AIML service operations control and management procedure

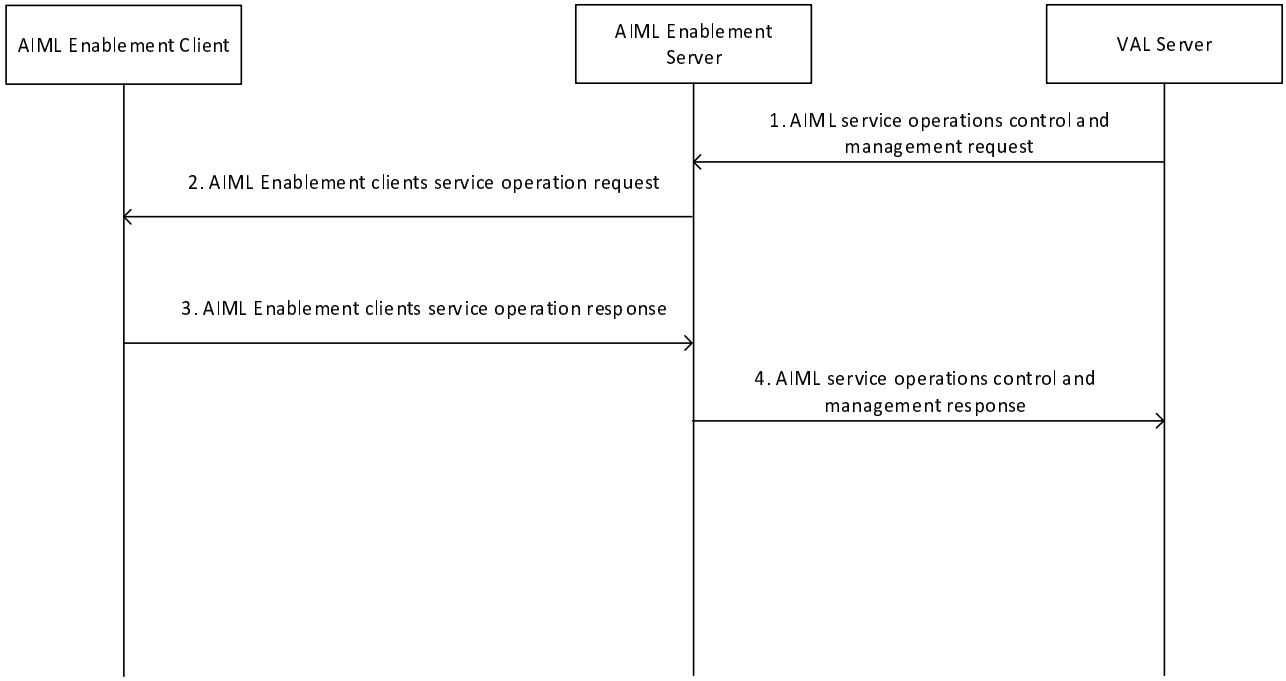

Figure 8.20.1-1: AIML service operations control and management procedure

1. The VAL server sends AIML service operations control and management request to the AIML Enablement server as per the Table 8.20.3.1-1.

2-3. The AIMLE server determines the required service operation mode to manage the AIML service operation lifecycle based on the AIML service operation mode, AIML service operation information. For example, the AIMLE server can perform client discovery/selection using the procedure defined in 3GPP TS 23.482 clause 9.3 and model training using the procedure defined in 3GPP TS 23.482 clause 8.3. Based on the AIML client identifier the AIMLE server sends the AIML Enablement client service operation request as per Table 8.20.3.3-1.

The AIML service operation mode includes start and stop operation. Start indicates the initiation of the AIML service and stop indicates termination of the AIML service.

The AIML Enablement client receiving the AIML service operation mode performs the service operation mode for the AIML service operation. The AIMLE client can configure and monitor the AIML service operation as per the AIML service operation mode configuration. Based on the AIML service operation mode status reporting configuration (periodic or event-triggered), the AIML client reports the service operation mode status to the AIML Enablement server. The AIML Enablement client sends a response indicating the success or failure of the AIML Enablement client service operation as per Table 8.20.3.4-1.

4. The AIML Enablement server provides the AIML service operations control and management response to the VAL server. The message includes AIML service operation ID and the reporting status of the AIML service operation.

8.20.3 Information flows

8.20.3.1 AIML service operations control and management request

Table 8.20.3.1-1 shows the request sent by an AIML service consumer (e.g., VAL server) to an AIMLE server for the AIML service operations control and management request.

Table 8.20.3.1-1: AIML service operations control and management request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor (e.g., VAL server).
VAL service identifier	O	An identifier for the VAL service associated with the requestor.
AIMLE client identifier(s)	O (NOTE)	Indicates the identifier(s) of AIML enablement client(s).
AIMLE client set identifier	O (NOTE)	An identifier for the AIMLE client set.
AIML service operation identifier	O	Indicates the AIML service operation identifier to identify the AIML service. (e.g., model training id, ml task id)
AIML service operation information	O	Indicates AIML service operation information. It includes AIML service model container, URI of the model to fetch the model from a repository, AIML service aggregator URI to send model updates, AIML service operation optimization assistance like maximum convergence time
AIML service operation mode	M	Indicates the required AIMLE service operation modes like start, stop. The start mode defines the initiation of the AIML service. The stop mode is defined to stop the AIML operation.
AIML service operation mode configuration	O	Indicates the configuration of the AIML service operation modes. It includes network utilization (like stop the AIML service when latency is worse than x milliseconds, time limit threshold (like stop the AIML service after 24 hours), model performance (like stop the AIML service when model accuracy is 99% achieved)
AIML service operation mode status reporting	O	Indicates the reporting configuration of the AIML service operation status like periodic (e.g. time interval) or event based (e.g., transition of AIML service operation from stop to start)
NOTE: One of the information elements is present.		

8.20.3.2 AIML service operations control and management response

Table 8.20.3.2-1 shows the request sent by an AIML Enablement server to an AIML service consumer (e.g., VAL server) for the AIML service operations control and management response.

Table 8.20.3.2-1: AIML service lifecycle management response

Information element	Status	Description
VAL service identifier	O	An identifier for the VAL service associated with the requestor.
AIML service operation ID	M	An identifier to identify the AIML service operation ID
AIML service operation mode report status	M	Indicates the current state of AIMLE service operation. E.g., start, stop

8.20.3.3 AIML Enablement client service operation request

Table 8.20.3.3-1 shows the request sent by an AIML Enablement server to an AIML Enablement client for the AIML Enablement client service operation request.

Table 8.20.3.3-1: AIML Enablement client service operation request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor (e.g. AIML service consumer).
VAL service identifier	O	An identifier for the VAL service associated with the requestor.
AIML service operation ID	M	An identifier to identify the AIML service operation ID
AIML service operation mode	M	Indicates the required AIMLE service operation modes like start, stop.
AIML service operation information	O	Indicates AIML service operation information. It includes AIML service model container, URI of the model to fetch the model from a repository, AIML service aggregator URI to send model updates, AIML service operation optimization assistance like maximum convergence time
AIML service operation mode configuration	O	Indicates the configuration of the AIML service operation modes. It includes network utilization (like stop the AIML service when latency is worse than x milliseconds, time limit threshold (like stop the AIML service after 24 hours), model performance (like stop the AIML service when model accuracy is 99% achieved)
AIML service operation mode status reporting	O	Indicates the reporting configuration of the AIML service operation status like periodic (e.g. time interval) or event based (e.g. transition of AIML service operation from stop to start)

8.20.3.4 AIML Enablement client service operation response

Table 8.20.3.4-1 shows the request sent by an AIML Enablement client to an AIML Enablement server for the AIML Enablement client service operation response or update response.

Table 8.20.3.4-1: AIML Enablement client service operation response

Information element	Status	Description
VAL service identifier	O	An identifier for the VAL service associated with the requestor.
AIML service operation ID	M	An identifier to identify the AIML service operation ID
AIML service operation mode status	M	Indicates the current state of AIMLE service operation. Possible values start, stop

8.21 ML model update

8.21.1 General

This clause provides the procedures to support ML model re-training and update when model performance degradation is observed by the AIML enablement layer. The model update procedure also supports using an existing model to re-train the model using Transfer Learning. Additionally, if the degraded model is related to other models due to e.g., Transfer Learning, the AIMLE server may trigger the update of those related models as well.

8.21.2 Procedure

Figure 8.21.2-1 depicts the procedure where the AIML enablement capability can trigger model update upon detecting model performance degradation.

Pre-conditions:

1. The AIMLE Server has provided a ML model to the AIMLE Consumer. The AIMLE consumer can be a VAL server, AIMLE client, or ADAE server.

2. The AIMLE Consumer detects a performance degradation of the ML model. If the consumer is an ADAE server, performance degradation may be detected as described in clause 8.17 in 3GPP TS 23.436 [4].

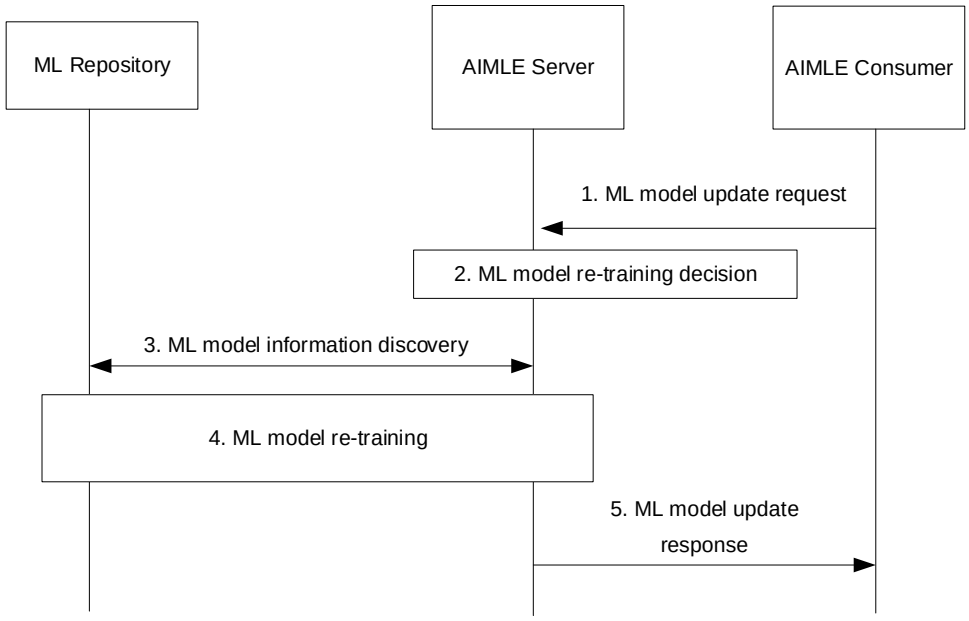

Figure 8.21.2-1: Support for ML model update

1. The AIMLE Consumer sends an ML model update request to the AIMLE Server that includes the ID of the model and performance degradation information.

2. Based on the performance degradation information, the AIMLE Server determines whether to update the ML model. If the AIMLE Server does not update the model, steps 3 and 4 are skipped.

3. The AIMLE Server retrieves the ML model information from the ML Repository as described in clause 8.11.3. The AIMLE Server may also perform ML model discovery to determine whether an existing ML model stored by the ML Repository can be used to replace the degraded model or train the new model (e.g., using Transfer Learning). If an existing model can be used to replace the degraded model, step 4 is skipped, and the identified model is provided in step 5.

   The AIMLE Server can also discover models that are related due to Transfer Learning or the use of the same training data, to identify additional models that may require an update.

4. The AIMLE Server performs ML model re-training, which corresponds to the ML model training procedure as described in clause 8.3. The updated model is stored in the ML repository once re-training is complete.

   If the degraded model is linked to other models (e.g., due to Transfer Learning, or the same training data has been used), the AIMLE Server may trigger the re-training and update of those related models.

5. The AIMLE Server provides the updated ML model to the AIMLE Consumer either by sending it directly, or by providing endpoint information to retrieve it from the ML Repository.

8.21.3 Information flows

8.21.3.1 ML model update request

Table 8.21.3.1-1 details the ML model update request IEs.

Table 8.21.3.1-1: ML model update request

Information element	Status	Description
Requestor Identity	M	The identity of the AIMLE Consumer sending the request.
ML model ID	M	Provides the ID of ML model for which the performance degradation has been detected.
Performance degradation information	O	Provides details about the detected performance degradation, such as the time, instances, or information on the degraded metrics (e.g. accuracy, recall, F1score).
ML model retrieval endpoint	O	The endpoint (e.g., URL, URI, IP address) where the ML model file can be retrieved.

8.21.3.1 ML model update response

Table 8.21.3.2-1 details the ML model update response IEs.

Table 8.21.3.2-1: ML model update response

Information element	Status	Description
Successful response	O (NOTE 1)	Indicates that the model has been updated.
> ML model	O (NOTE 2)	Provides the updated ML model.
> ML model retrieval endpoint	O (NOTE 2)	The endpoint (e.g., URL, URI, IP address) where the ML model file can be retrieved.
> ML model information	O	Provides information of the ML model, specified in Table 8.11.4.1-2.
Failure response	O (NOTE 1)	Indicates that the request has failed.
> Cause	O	Indicates the failure cause.
NOTE 1: Only one of these information elements shall be provided. NOTE 2: At least one of these information elements shall be provided.		

8.22 ML model performance monitoring

8.22.1 General

The following clauses specify procedures, information flows, and APIs for ML model performance monitoring and potential degradation detection.

8.22.2 Procedure

Pre-conditions:

1. One or more AIMLE services (at the AIMLE server or clients) using the given ML model are ongoing.

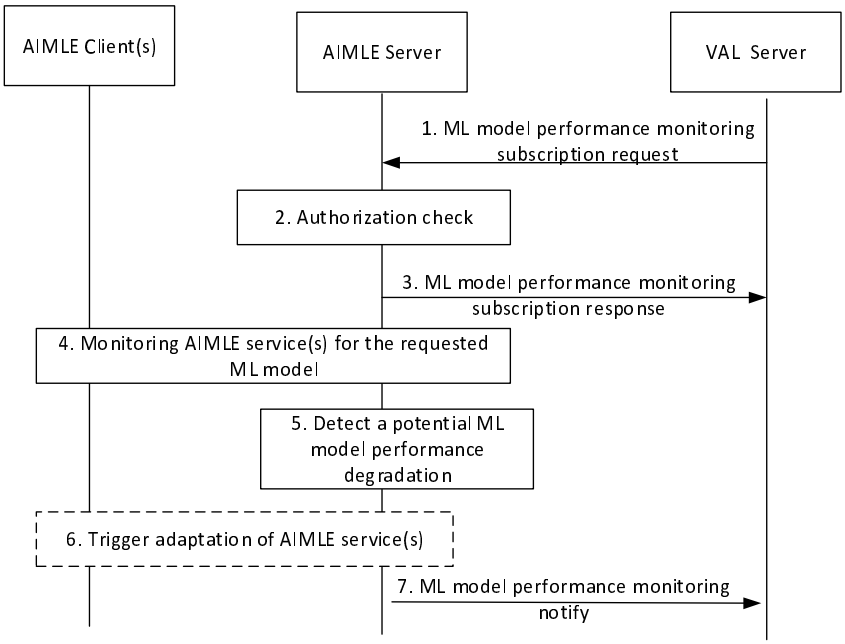

Figure 8.22.2-1: ML model performance monitoring

1. The VAL server sends an ML model performance monitoring subscription request to AIMLE server, requesting to assist in monitoring the ML model performance. This request consists of ML model information and optionally the AIML service information, etc.

2. The AIMLE server checks whether the VAL server is authorized to perform the ML model performance monitoring request.

3. If the VAL server is authorized, AIMLE server returns the success response, otherwise a failure response indication the reason for failure.

4. The AIMLE server identifies the AIMLE services which are utilizing the requested ML model. Such AIMLE services can be an ML model training service or an HFL service at the AIMLE server or clients.

The identification of the AIMLE services may be performed via fetching ML model information from the ML repository using ML model management procedure as in clause 8.11.3.

Then, the AIMLE server then starts monitoring the AIMLE service performance (e.g. accuracy, a KPI or QoS metric related to the AIML operation). This step includes receiving information from one or more AIMLE clients performing an operation based on the target ML model (e.g., based on step 7 of procedure in clause 8.12.2 or step 6 of procedure in clause 8.15.2), with an expected or experienced deviation of the required performance of the AIMLE service.

5. The AIMLE server detects an expected ML model degradation (e.g., model drift, data drift) based on the deviation of the performance of the AIMLE service as indicated in step 4.

6. The AIMLE server based on the expected ML model degradation, it may also indicate and execute a trigger action to ensure meeting the AIMLE service requirement.

Such trigger action may be either an adaptation of the AIMLE service, such as training of a new ML model for the AIMLE by the same or a different AIMLE client, or re-training of the ML model by the same or different AIMLE client; or termination of the AIMLE service and initiating a new AIMLE service with a new ML model.

NOTE: If this action involves re-selecting an AIMLE client for the AIMLE service, this is based on the procedure defined in clause 8.9.

7. The AIMLE server notifies the VAL server on the expected ML model degradation and if requested the triggered adaptation of the AIMLE service.

8.22.3 Information flows

8.22.3.1 ML model performance monitoring subscription request

Table 8.22.3.1-1 shows the request sent by a VAL server to an AIMLE server for the ML model performance monitoring.

Table 8.22.3.1-1: ML model performance monitoring subscription request

Information element	Status	Description
Requestor identity	M	The identifier of the requestor (e.g., VAL server).
ML model identifier	M	The identifier of the ML model for which the monitoring applies.
Notification endpoint	M	The notification endpoint (e.g. URL/URI/IP address) where the notifications should be sent to.
AIML operation information	O	The AIMLE operation (ML model training, HFL, VFL, TL) for which the ML model is used.
> VAL service ID	O	The VAL service identifier of the AIMLE service using the ML model (if known by the requestor).
> AIMLE client ID(s)	O	The identifier(s) of the AIMLE client(s) training the ML model (if known by the requestor).
> AIMLE service KPI	O	One or more KPIs for the AIMLE service performance (latency, accuracy, etc).
Monitoring report configuration	M	The reporting configuration for the monitoring service (thresholds for triggering a monitoring event, e.g. minimum accuracy, delay, whether the reporting is one time or periodical or event-triggered).
Area of interest	O	The geographical or service area for which the monitoring applies.
Time validity	O	The time validity for the monitoring subscription.
Trigger Action requirement	O	This requirement identifies policies for triggering an action based on a monitoring event (e.g. if degradation is detected, to train a new model or re-selecting AIMLE clients).

8.22.3.2 ML model performance monitoring subscription response

Table 8.22.3.2-1 shows the response sent by the AIMLE server to the VAL server for the ML model performance monitoring subscription.

Table 8.22.3.2-1: ML model performance monitoring subscription response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the ML model performance monitoring request was successful.
> Subscription ID	M	Subscription identifier corresponding to the subscription.
> Expiration time	O	Indicates the expiration time of the subscription. To maintain an active subscription, a subscription update is required before the expiration time.
Failure response	O (NOTE)	Indicates that the request has failed.
> Cause	O	The cause for the request failure.
NOTE: Only one of these information elements shall be present.		

8.22.3.3 ML model performance monitoring notify

Table 8.22.3.2-1 shows the notification sent by the AIMLE server to the VAL server for the ML model performance degradation.

Table 8.22.3.2-1: ML model performance monitoring notify

Information element	Status	Description
Subscription ID	M	Subscription identifier corresponding to the subscription.
ML model ID	M	Identity of the ML model.
ML model degradation indication	M	Identifies the degradation of the ML model.
>ML model degradation parameters	O	The performance metrics which are expected to be degraded (F1-score, recall, precision, accuracy).
> Cause	O	The cause for the degradation of the ML model.
Trigger Action	O	The trigger action, which is notified, and may be one of the following: - the adaptation of the AIMLE service, such as training of a new ML model for the AIMLE by the same or a different AIMLE client, - the re-training of the ML model by the same or different AIMLE client, - the termination of the AIMLE service and initiating a new AIMLE service with a new ML model.

8.23 AIMLE assisted ML model selection

8.23.1 General

ML model selection is an important consideration for successful ML training and deployment. Many ML models exist for ML applications and some models can generate better results than other models for a particular dataset. The procedure allows AIMLE service consumers to request assistance from an AIMLE server with the selection of appropriate ML models for a given dataset and for the provided requirements. The AIMLE server coordinates the selection of candidate ML models and training the ML models with the given dataset. A list of ML models with corresponding performance is return to the AIMLE service consumer.

The following clauses specify procedures, information flows, and APIs for ML model selection.

8.23.2 Procedure

Pre-conditions:

1. The AIMLE service consumer has identified datasets and ML requirements.

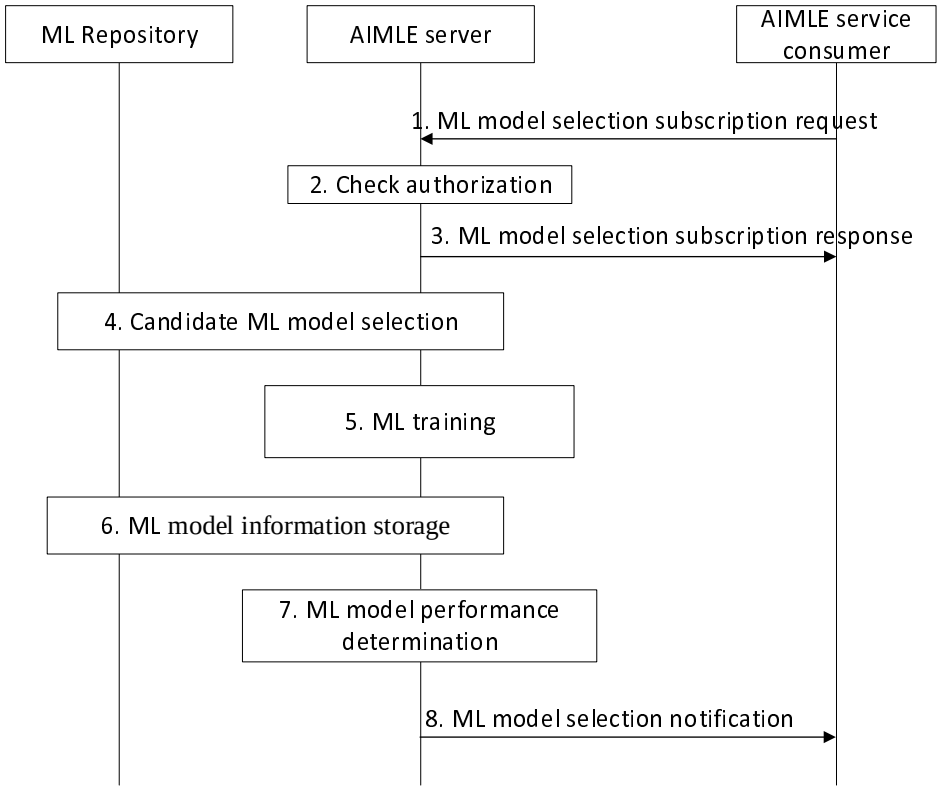

Figure 8.23.2-1: AIMLE assisted ML model selection

1. An AIMLE service consumer (e.g. VAL server) sends a subscription request to an AIMLE server. The request includes information as described in Table 8.23.3.1-1. The AIMLE service consumer provides a list of candidate ML models, dataset identifiers, and requirements for training the candidate ML models. The AIMLE service consumer can either provide AIMLE client set identifiers or AIMLE client selection criteria for selecting the AIMLE clients to train the candidate ML models.

2. The AIMLE server authenticates the requestor and checks authorization for the request. If authorized, the AMILE server assigns an identifier for the subscription.

3. The AIMLE server sends a ML model selection subscription response that includes information in Table 8.23.3.2-1.

4. The AIMLE server determines if additional ML models can be candidates for the type of ML application based on the ML model requirements provided in step 1 and selects additional candidate ML models to train with the given dataset. The AIMLE server can discover models from ML repository to determine the list of candidate ML models as described in clause 8.11.3.

5. The AIMLE server performs ML model training for each candidate model. ML model training can be for split AI/ML operation as described in clause 8.14, Transfer Learning as described in clause 8.16, or Federated Learning as described in clauses 8.12 and 8.18. If AIMLE client selection criteria were provided in step 1, the AIMLE server performs AIMLE client selection as described in clauses 8.9 or 8.13 during the training of the candidate ML models.

6. The AIMLE server performs ML model information storage as described in clause 8.11 for each trained ML model.

7. The AIMLE server aggregates and determines the performance of each ML model with the given dataset.

8. The AIMLE server sends a notification to the AIMLE service consumer and include information as described in Table 8.23.3.3-1. The notification includes a list of trained candidate ML models with corresponding model information and performance. The AIMLE service consumer can then select the best performing ML models from the list provided in the notification.

8.23.3 Information flows

8.23.3.1 AIMLE assisted ML model selection subscription request

Table 8.23.3.1-1 shows the request sent by an AIMLE service consumer to an AIMLE server for the AIMLE assisted ML model selection subscription procedure.

Table 8.23.3.1-1: ML model selection subscription request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
AIML profile	M	Requirements for the ML model selection operation.
> Candidate ML models	M	A list of ML model identifiers (and initial model parameters) to train. The list provides candidate ML models to evaluate against the provided dataset.
> ML model requirements	O	ML model requirements for the AIMLE server to use for selecting additional candidate ML models for training with the provided datasets. The requirements can be any of the ML model information as described in Table 8.11.4.1-2.
> AIMLE client set identifiers	O (NOTE1)	A list of AIMLE client set identifiers to train the ML model.
> AIMLE client selection criteria	O (NOTE1)	Selection criteria for finding suitable AIMLE clients for training the ML model.
> Number of required AIMLE clients	O (NOTE2)	A minimal number of AIMLE clients required for training the ML model.
> Dataset identifiers	M	Dataset identifiers to use for training and evaluating model performance to obtain a list of ML model rankings.
> Training requirements	M	Training requirements as detailed in Table 8.23.3.1-2.
Notification target	O	Endpoint information for receiving notifications.
Notification settings	O	Notification settings for which the AIMLE server provides ML model status: after, after certain job percentage completion, periodically based on date and time, upon error events, etc.
NOTE1: At least one of the information elements shall be provided. NOTE2: Mandatory if AIMLE client selection criteria are present.		

Table 8.23.3.1-2: Training requirements

Information element	Status	Description
Performance metric	M	Identifies the performance metric to evaluate ML model training. Performance metric can be mean absolute error, mean squared error, accuracy, precision and recall, etc. The performance metric indicates the performance of the ML model.
Performance target	O	A target performance that indicates acceptable performance has been reached and training can be stopped.
Number of training rounds	M	A minimum number of training rounds for the ML training.
Number of data samples	M	A minimum number of data samples for the ML training.

8.23.3.2 ML model selection subscription response

Table 8.23.3.2-1 shows the response sent by the AIMLE server to the AIMLE service consumer for the AIMLE assisted ML model selection subscription procedure.

Table 8.23.3.2-1: ML model selection subscription response

Information element	Status	Description
Status	M	The status for the ML model selection operation
Subscription identifier	M	An identifier for the subscription.

8.23.3.3 ML model selection notification

Table 8.23.3.3-1 shows the notification sent by the AIMLE server to the AIMLE service consumer for the AIMLE assisted ML model selection subscription procedure.

Table 8.23.3.3-1: ML model selection notification

Information element	Status	Description
Subscription identifier	M	The identifier for the subscription that notification is associated with.
Operational status	M	The status for the ML model selection operation. The status can represent the estimate percentage completion or associated with the notification settings.
Trained ML models	M	The results of the ML model training.
> ML model information	M	Information about the ML model such as the ML model type as described in Table 8.11.4.1-2.
> Model performance	M	The performance metric for training the ML model.
Elapse time	O	The time that has elapsed for the ML model selection operation.
Timestamp	O	Timestamp of the notification.

8.24 AIMLE context transfer

8.24.1 General

This clause describes AIMLE context transfer procedure between AIMLE servers (over AIML-E reference point).

8.24.2 Procedure

Pre-conditions:

1. Each edge AIMLE server manages the AIMLE clients within its service area to perform AI/ML operations.
2. A UE associated with an AIMLE client moves from a source service area (managed by a source edge AIMLE server) to a target service area (managed by a target edge AIMLE server). The transition triggers application context relocation (ACR) procedure between the two edge AIMLE servers (as source EAS and target EAS, respectively) as specified in 3GPP TS 23.558 [12].

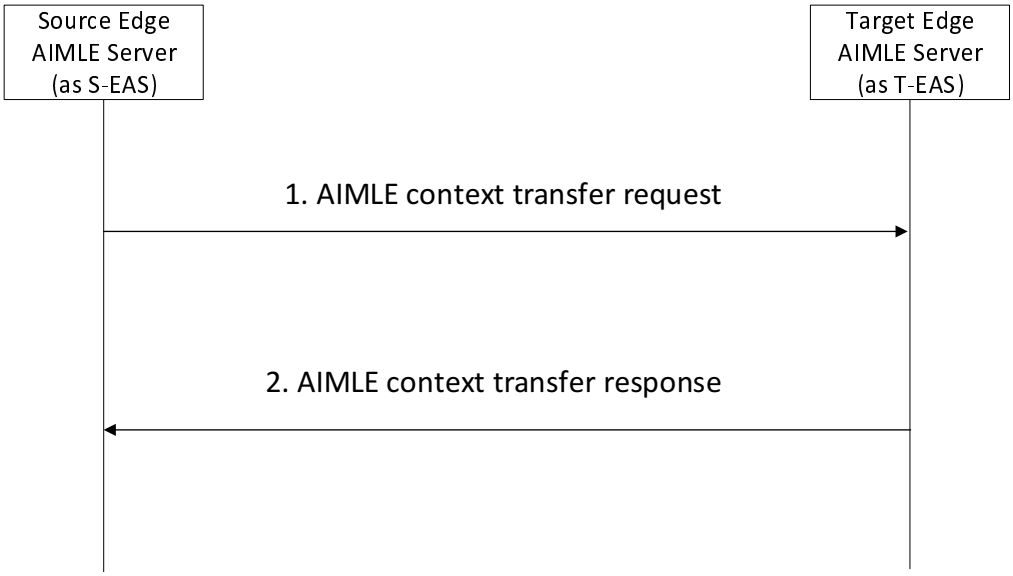

Figure 8.24.2-1. AIMLE context transfer

1. The source edge AIMLE server sends an AIMLE context transfer request to the target AIMLE server as described in Table 8.24.3.1-1. The request includes AIMLE context information which is generated based on the responses/notifications received from the transitioned AIMLE client (e.g. step 7 of clause 8.12.2 or step 3 of clause 8.20.1).

   The AIMLE context information is used by the target edge AIMLE server to determine whether the information (e.g. AI/ML operation output/result) received from the AIMLE client should be transferred to the source edge AIMLE server (or another edge AIMLE server that has been associated with the AIMLE client). For example, the target AIMLE server can forward the AIMLE service results received from the AIMLE client to the source AIMLE server if the results are only applicable to the source service area or the AIMLE client is part of a split operation pipeline formed in the source service area.

2. The target edge AIMLE server sends an AIMLE context transfer response to the source edge AIMLE server as described in Table 8.24.3.2-1.

8.24.3 Information flows

8.24.3.1 AIMLE context transfer request

Table 8.24.3.1-1 shows the AIMLE context transfer request that is sent by a source AIMLE server to a target AIMLE server.

Table 8.24.3.1-1: AIMLE context transfer request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor (e.g., AIMLE server).
AIMLE Context Information	M	The AIMLE context information as described in Table 8.24.3.1-2.

Table 8.24.3.1-2: AIMLE context information

Information element	Status	Description
AIMLE client ID	M	The identifier of the AIMLE client associated with the context.
Current managing AIMLE server	O	The identifier of the AIMLE server that is currently managing the AIMLE client, i.e. the AIMLE server associated with the service area that the AIMLE client is currently in.
Previous managing AIMLE server	O	List of identifiers of AIMLE servers that have been associated with the AIMLE client. The list is populated by adding the identifier of the source edge AIMLE server whenever the UE transitioned from a source edge area to a target edge area.
AIMLE service status	O	Status of the AIML operations (task) at the AIMLE client, e.g. “active”, “paused”, “completed”, percentage of completion.
AIMLE service results	O	Results of the AIML operations (task) performed by the AIMLE client.
AIMLE service applicability	O	Applicability information of the AIML operations performed by the AIMLE client, e.g. the operation results are applicable within a certain edge service area, the operations are applicable within a certain split operation pipeline.
>ML context information	O	Context information related to the ML operation that the AIMLE client is participating in or performing.
>> VAL service Information	O	Information related to the VAL service for which the AIMLE task is performed (e.g., the VAL service identifier for the AIMLE HFL training operation).
>> ML task	O	Type of ML task (model training, model testing, model inference, model transfer, model offload, model split, intermediate AI/ML operation/task) to be continued at the target AIMLE server.
>> ML task information	O	Information related to the ML task mentioned in “ML task” information element. The Model Training task Information may include training objective to be achieved, HFL training information, VFL training information, data set information for training, status of training operation at AIMLE client (e.g. “active”, “paused”, “completed”), training results, etc. The Model Inference task information may include Inference results, inference job id, etc. The Model split task information may include split operation profile as specified in table 8.14.3.3-2.
>> ML model information	M	Model information related to the ML task. This information may include, the model identifier, Information to fetch ML model information, address (e.g., a URL or an FQDN) of the ML model file or address of the model repository where the ML model resides, Model parameters from ML training, etc)

8.24.3.2 AIMLE context transfer response

Table 8.24.3.2-1 shows the AIMLE context transfer response that is sent by the target edge AIMLE server to the source edge AIMLE server.

Table 8.24.3.2-1: AIMLE context transfer response

Information element	Status	Description
Successful response	O (NOTE)	Indicates that the request was successful.
Failure response	O (NOTE)	Indicates that the request failed.
> Cause	O	Indicates the cause of request failure
NOTE: One of the IEs shall be present.		

8.25 Support of AIML Services for Assisting Hierarchical Computing

This clause describes procedure for supporting AIML services for assisting hierarchical computing.

8.25.1 General

This clause describes the procedure for assisting hierarchical computing by the AIMLE server. A entity (e.g. CAS or EAS which are defined in 3GPP TS 23.558 [12]) can have different roles in hiearchitical computing (with one root node (e.g., CAS, EAS), the root has one or more children which are also known as sub-root node(s) (e.g., EAS), and multiple leaf nodes (e.g. EAS) with no children). Here, hierarchical computing represents a computation architecture with multiple computation entities involved and multiple levels of computations for a computation task.

8.25.2 Procedure for assisting hierarchical computing process

Figure 8.25.2-1 illustrates the procedures for AIMLE server to assist a hierarchical computing process.

Pre-conditions:

1. An AI/ML task be treated as a specical computing task being completed at consumer (e.g., CAS, EAS).

2.- The consumer decides its role in the hierarchical computing architecture for an AI/ML task (e.g. FL training) based on its local configuration.

3. The AIMLE server can assist a hierarchical computing process by providing time window(s) recommendation for computing task distribution if the consumer is a root node in a hierarchical computing process, providing time window(s) recommendation for intermediate output delivery if the consumer is a leaf node in a hierarchical computing process, and candidate execution node list provisioning or computing preparation status provisioning, etc.

4. The consumer decides that assistance from AIMLE server to support the hierarchical computing process (AI/ML task) is needed, due to lack of capability on e.g. execution node selection.

5. AIMLE Server is deployed following the hierarchical deployment model described in A.4.

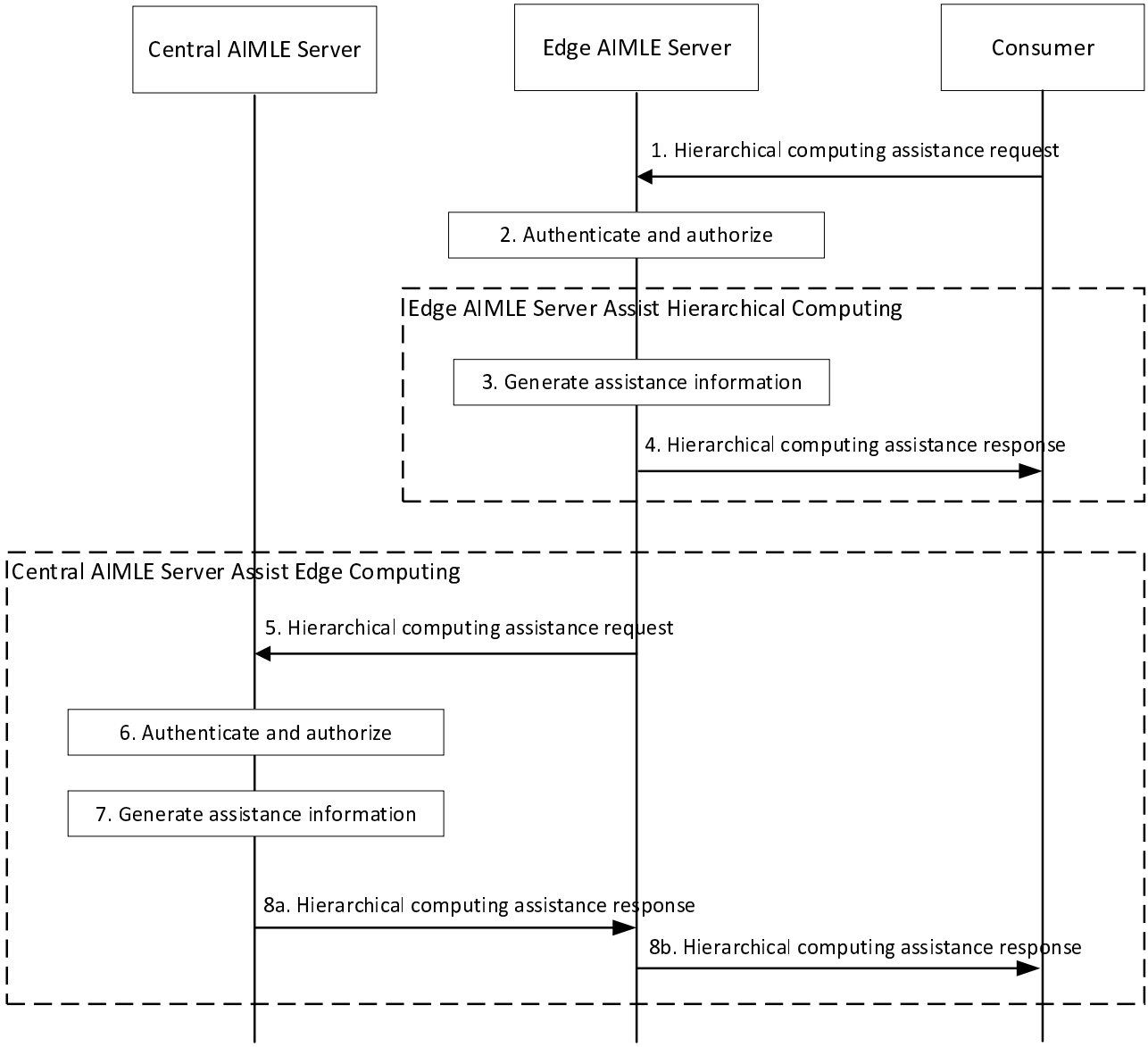

Figure 8.25.2-1: Procedure for assist a hiearchitical computing process

Figure 8.25.2-1 illustrates the procedure for AIMLE servers assist a hierarchical computing process. The corresponding procedure in detail is as follows:

1. The VAL server (e.g. CAS, EAS) sends hierarchical computing assistance request to the edge AIMLE server for a hierarchical computing process (AI/ML task). The request message includes information as described in Table 8.25.3.1-1.

2. The edge AIMLE server authenticates and authorizes the request from the consumer and checks its capability to provide the requested assistance information.

   If the request is authorized and the edge AIMLE server can generate the required assistance information (e.g. computing preparation status at an execution node which is registered to it), it performs step 3 and 4.

   If the request is authorized but the edge AIMLE server cannot generate the required assistance information (e.g. execution nodes are registered to different edge AIMLE servers), it performs step 5 and skips steps 3 and 4.

3. The edge AIMLE server performs operations to generate assistance information which is requested in step 1. For example, for computing preparation status at an execution node, the edge AIMLE server may subscribe/request analytics from ADAE server (e.g. edge load analytics, edge computing preparation analytics) and aggregate the information received to generate assistance information.

4. The edge AIMLE server sends hierarchical computing assistance response to the consumer (e.g. computing preparation status at the execution node). The response message contains the information as described in Table 8.25.3.2-1.

5. If the edge AIMLE server cannot generate the required assistance information, it sends hierarchical computing assistance request to central AIMLE server. The request message includes information as described in Table 8.25.3.1-1.

6. The central AIMLE server authenticates and authorizes the request from the edge AIMLE server.

7. If the request is authorized, the central AIMLE server performs operations to generate assistance information which is requested in step 1.

   For example, for execution node selection, according to the role of the VAL server in request message in step 1, the central AIMLE server derives the requirements on the execution node (e.g. high computation capability, high communication capability, or both). Then, the central AIMLE server may subscribe/request analytics from ADAE server (e.g. edge load analytics, edge computing preparation analytics), trigger generation of split operation assistance information, retrieve FL member information from ML Repository (the FL member with EAS ID). The existing services can be reused for execution node selection, e.g., the step 3 in clause 8.12.2.

8. The central AIMLE server sends hierarchical computing assistance response with the generated assistance information (e.g. a list of selected candidate execution nodes). The response message contains the information as described in Table 8.25.3.2-1.

   8a. The central AIMLE server sends the hierarchical computing assistance response to the edge AIMLE server.

   8b. The edge AIMLE server sends the hierarchical computing assistance response to the consumer.

The VAL server (e.g. CAS, EAS) uses the assistance information for decision making on its computing operations.

NOTE: If the VAL server (e.g. CAS, EAS) know the information of the central AIMLE server, it can send the hiearchitical computing assistance request to the central AIMLE server directly for the hiearchitical computing process (AI/ML task). The procedure is the same as the steps 5-8a in Figure 8.25.2-1 with replacing edge AIMLE server by consumer.

8.25.3 Information flows

8.25.3.1 Hiearchitical computing assistance request

Table 8.25.3.1-1 shows the request sent by VAL server (e.g. CAS, EAS) to the AIMLE server for assistance of a hiearchitical computing process (AI/ML task).

Table 8.25.3.1-1: Hiearchitical computing assistance request

Information element	Status	Description
Requestor identifier	M	The identifier of the requestor.
Original requestor identifier	O	The identifier of the original requestor, e.g. VAL server ID, EAS ID, CAS ID.
Role	M	Represents the role of the VAL server in a hiearchitical computing architecture (e.g. root node, sub-root node or leaf node of a hierarchical computing process).
Computing task type	M	The type of computing task (e.g. VFL, HFL).
Assistance information type	M	Represents the assistance information type, which is used to indicate the assistance information needed, e.g. candidate execution node list, computing preparation status at an execution node.
Execution node(s)	O	Represent one execution node or a list of candidate execution nodes.

8.25.3.2 Assist hierarchical computing response

Table 8.25.3.2-1 shows the response sent by AIMLE server to the VAL server (e.g. CAS, EAS) for assisting the hierarchical computing process.

Table 8.25.3.2-1: Hierarchical computing assistance response

Information element	Status	Description
Success response	O (NOTE 1)	Indicates that the assist hierarchical computing request was successful.
>Assistance information	M	The assistance information for assisting hierarchical computing process, e.g. candidate execution node list, computing preparation status at an execution node.
>>List of candidate execution nodes	O (NOTE 2)	A list of selected candidate execution nodes.
>>Preparation status	O (NOTE 2)	Computing preparation status at the execution node provided in request.
Failure response	O (NOTE 1)	Indicates that the assist hierarchical computing request was failure.
>Cause	M	Reason for the failure.
NOTE 1: One of the IEs shall be present. NOTE 2: At least one of the IEs shall be present.		

9 AIMLE APIs

9.1 General

The following AIMLE capabilities are offered as APIs:

- AIMLE server APIs;
- ML repository APIs;
- AIMLE client APIs;

9.2 AIMLE server APIs

9.2.1 AIMLE AI/ML Task Transfer API

9.2.1.1 General

Table 9.2.1.1-1 illustrates the APIs for AI/ML task transfer. This API enables the communication between VAL UE (via AIMLE Client) and the AIMLE server for AI/ML Task Transfer.

Table 9.2.1.1-1: Aimles_AIMLTaskTransferAssist API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AIMLTaskTransferAssist	Request	Request/Response	AIMLE Client

Table 9.2.1.1-2: Aimles_AIMLESControlledAIMLTaskTransfer API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AIMLESControlledAIMLTaskTransfer	Request	Request/Response	AIMLE Client

9.2.1.2 Aimles_AIMLTaskTransferAssist_Request operation

9.2.1.2.1 General

API description: This API enables source AI/ML member (AIMLE Client) to communicate with the AIMLE server for request assist AIML task transfer from the source AI/ML member (AIMLE Client) to target AI/ML member (AIMLE Client).

9.2.1.2.2 AIML task transfer assist request operation

API operation name: Aimles_AIMLTaskTransferAssist_Request

Description: The consumer requests for assist AI/ML task transfer.

Inputs: See clause 8.6.3.2.

Outputs: See clause 8.6.3.3.

See clause 8.6.2.1 for details of usage of this operation.

9.2.1.3 Aimles_AIMLESControlledAIMLTaskTransfer_Request operation

9.2.1.3.1 General

API description: This API enables the source AI/ML member (AIMLE Client) to communicate with the AIMLE server for request AIMLE server-controlled AIML task transfer from the source AI/ML member (AIMLE Client) to the target AI/ML member (AIMLE Client).

9.2.1.3.2 AIMLE server-controlled AIML task transfer request operation

API operation name: Aimles_AIMLESControlledAIMLTaskTransfer_Request

Description: The consumer requests for AIMLE server-controlled AI/ML task transfer.

Inputs: See clause 8.6.3.8.

Outputs: See clause 8.6.3.9.

9.2.2 ML model retrieval API

9.2.2.1 General

Table 9.2.2.1-1 illustrates the API for ML model retrieval.

Table 9.2.2.1-1: Aimles_MLModelRetrieval API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_MLModelRetrieval	Request	Request/Response	AIMLE Client, VAL Server
	Subscribe	Subscribe/Notify	AIMLE Client, VAL Server
	Notify		
	UpdateSubscription		
	Unsubscribe		

9.2.2.2 Aimles_MLModelRetrieval_Request operation

API operation name: Aimles_MLModelRetrieval_Request

Description: The consumer requests for one time ML Model retrieval.

Inputs: See clause 8.2.3.1.

Outputs: See clause 8.2.3.2.

See clause 8.2.2.2 for details of usage of this operation.

9.2.2.3 Aimles_MLModelRetrieval_Subscribe operation

API operation name: Aimles_MLModelRetrieval_Subscribe

Description: The consumer subscribe for ML Model retrieval.

Inputs: See clause 8.2.3.3.

Outputs: See clause 8.2.3.4.

See clause 8.2.2.3.2 for details of usage of this operation.

9.2.2.4 Aimles_MLModelRetrieval_Notify operation

API operation name: Aimles_MLModelRetrieval_Notify

Description: The consumer is notified with ML Model.

Inputs: See clause 8.2.3.5.

Outputs: None.

See clause 8.2.2.3.3 for details of usage of this operation.

9.2.2.5 Aimles_MLModelRetrieval_UpdateSubscription operation

API operation name: Aimles_MLModelRetrieval_UpdateSubscription

Description: The consumer updates an existing subscription for ML Model retrieval.

Inputs: See clause 8.2.3.6.

Outputs: See clause 8.2.3.7.

See clause 8.2.2.3.4 for details of usage of this operation.

9.2.2.6 Aimles_MLModelRetrieval_Unsubscribe operation

API operation name: Aimles_MLModelRetrieval_Unsubscribe

Description: The consumer cancels an existing subscription for ML Model retrieval.

Inputs: See clause 8.2.3.8.

Outputs: See clause 8.2.3.9.

See clause 8.2.2.3.5 for details of usage of this operation.

9.2.3 ML model training API

9.2.3.1 General

Table 9.2.3.1-1 illustrates the API for ML model training. This API enables the VAL server to communicate with the AIMLE server for ML model training.

Table 9.2.3.1-1: Aimles_MLModelTraining API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_MLModelTraining	Request	Request/Response	VAL Server

9.2.3.2 Aimles_MLModelTraining_Request operation

API operation name: Aimles_MLModelTraining_Request

Description: The consumer requests for ML Model training.

Inputs: See clause 8.3.3.1.

Outputs: See clause 8.3.3.2.

See clause 8.3.2 for details of usage of this operation.

9.2.4 AIMLE TL model selection assistance API

9.2.4.1 General

Table 9.2.4.1-1 illustrates the API for TL model selection assistance. This API enables the VAL server to communicate with the AIMLE server for requesting assistance on TL model selection.

Table 9.2.4.1-1: Aimles_TLModelSelectionAssistance API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_TLModelSelectionAssistance	Request	Request/Response	VAL Server

9.2.4.2 Aimles_TLModelSelectionAssistance_Request operation

API operation name: Aimles_TLModelSelectionAssistance_Request

Description: The consumer requests AIMLE server to assist in the TL model selection.

Inputs: See clause 8.16.4.1.

Outputs: See clause 8.16.4.2.

See clause 8.16.2 for details of usage of this operation.

9.2.5 FL member grouping support API

9.2.5.1 General

Table 9.2.5.1-1 illustrates the API for FL member grouping support. This API enables the VAL server to communicate with the AIMLE server for requesting support on FL member group management.

Table 9.2.5.1-1: Aimles_FLMemberGroupSupport API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_FLMemberGroupSupport	Request	Request/Response	VAL Server

9.2.5.2 Aimles_FLMemberGroupSupport_Request operation

API operation name: Aimles_FLMemberGroupSupport_Request

Description: The consumer requests AIMLE server to assist in the FL member grouping.

Inputs: See clause 8.17.3.1.

Outputs: See clause 8.17.3.2.

See clause 8.17.2 for details of usage of this operation.

9.2.6 AIMLE client Discovery API

9.2.6.1 General

Table 9.2.6.1-1 illustrates the APIs for AIMLE client discovery. This API enables the communication between the VAL server and the AIMLE server for AIMLE client discovery operation.

Table 9.2.6.1-1: Aimles_AIMLEClientDiscovery API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AIMLEClientDiscovery	Request	Request/Response	VAL Server

9.2.6.2 AIMLE client discovery request operation

API operation name: Aimles_AIMLEClientDiscovery_Request

Description: The consumer requests AIMLE server for AIMLE client discovery.

Inputs: See clause 8.8.3.1.

Outputs: See clause 8.8.3.2.

See clause 8.8.2.1 for details of usage of this operation.

9.2.7 AIMLE client Selection API

9.2.7.1 General

Table 9.2.7.1-1 illustrates the APIs for AIMLE client selection. This API enables the communication between the VAL server and the AIMLE server for AIMLE client selection operation.

Table 9.2.7.1-1: Aimles_AIMLEClientSelection API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AIMLEClientSelection	Request	Request/Response	VAL Server

9.2.7.2 AIMLE client selection request operation

API operation name: Aimles_AIMLEClientSelection_Request

Description: The consumer requests AIMLE server for AIMLE client selection.

Inputs: See clause 8.9.3.1.

Outputs: See clause 8.9.3.2.

See clause 8.9.2.1 for details of usage of this operation.

9.2.8 AIMLE client Selection Subscribe API

9.2.8.1 General

Table 9.2.8.1-1 illustrates the API for AIMLE client selection subscription for enabling a VAL server to monitor and select AIMLE clients for participation in AI/ML operations.

Table 9.2.8.1-1: Aimles_ClientSelection API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_ClientSelection	Subscribe	Subscribe/Notify	VAL Server
	Notify		
	Update		
	Unsubscribe		

9.2.8.2 Aimles_ClientSelection_subscribe operation

API operation name: Aimles_ClientSelection_subscribe

Description: The consumer subscribes for monitoring and selection of AIMLE clients for AI/ML operations.

Inputs: See clause 8.13.3.2.

Outputs: See clause 8.13.3.3.

See clause 8.13.2.2 for details of usage of this operation.

9.2.8.3 Aimles_ClientSelection_notify operation

API operation name: Aimles_ClientSelection_notify

Description: The consumer is notified of selected AIMLE client status update events.

Inputs: See clause 8.13.3.4.

Outputs: None.

See clause 8.13.2.3 for details of usage of this operation.

9.2.8.4 Aimles_ClientSelection_update operation

API operation name: Aimles_ClientSelection_update

Description: The consumer updates selection of AIMLE clients for AI/ML operations.

Inputs: See clause 8.13.3.5.

Outputs: See clause 8.13.3.6.

See clause 8.13.2.2 for details of usage of this operation.

9.2.8.5 Aimles_ClientSelection_unsubscribe operation

API operation name: Aimles_ClientSelection_ unsubscribe

Description: The consumer unsubscribes of selected AIMLE clients for AI/ML operations.

Inputs: See clause 8.13.3.7.

Outputs: See clause 8.13.3.8.

See clause 8.13.2.3 for details of usage of this operation.

9.2.9 AIMLE Data Management API

9.2.9.1 General

Table 9.2.9.1-1 illustrates the API for AIMLE data management for enabling a VAL server to request assistance with data management operations.

Table 9.2.9.1-1: Aimles_DataManagement API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_DataManagement	Subscribe	Subscribe/Notify	VAL server
	Notify		

9.2.9.2 Aimles_DataManagement_Subscribe operation

API operation name: Aimles_DataManagementAssistance_Subscribe

Description: The consumer subscribes for assistance with data management operations.

Inputs: See clause 8.15.3.1.

Outputs: See clause 8.15.3.2.

See clause 8.15.2 for details of usage of this operation.

9.2.9.3 Aimles_DataManagement_Notify operation

API operation name: Aimles_DataManagementAssistance_Notify

Description: The consumer received notifications for data management operations.

Inputs: See clause 8.15.3.3.

Outputs: None.

See clause 8.15.2 for details of usage of this operation.

9.2.10 AIMLE Service Operations Management API

9.2.10.1 General

Table 9.2.10.1-1 illustrates the API for AIMLE Service Operations Management.

Table 9.2.10.1-1: Aimles_AIMLEServiceOperationsManagement APIs

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AIMLEServiceOperationsManagement	Request	Request/Response	VAL Server

9.2.10.2 Aimles_AIMLEServiceOperationsManagement_Request operation

API operation name: Aimles_AIMLEServiceOperationsManagement_Request

Description: The consumer requests for AIMLE Service operations management

Inputs: See clause 8.20.3.1

Outputs: See clause 8.20.3.2

See clause 8.20.2.2 for details of usage of this operation.

9.2.11 AIMLE Client Registration APIs

9.2.11.1 General

Table 9.2.11.1-1 illustrates the APIs for AIMLE client registration, update and de-registration. This API enables the communication between the AIMLE client and the AIMLE server for AIMLE client registration operations.

Table 9.2.11.1-1: Aimles_AIMLEClientRegistration API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AIMLEClientRegistration	Request	Request/Response	AIMLE Client
	Update		AIMLE Client
	Delete		AIMLE Client

9.2.11.2 AIMLE client registration request operation

API operation name: Aimles_AIMLEClientRegistration_Request

Description: The consumer requests AIMLE server for AIMLE client registration.

Inputs: See clause 8.7.3.2-1.

Outputs: See clause 8.7.3.3.

See clause 8.7.2.2 for details of usage of this operation.

9.2.11.3 AIMLE client registration update request operation

API operation name: Aimles_AIMLEClientRegistration_Update Request

Description: The consumer requests AIMLE server for update AIMLE client registration.

Inputs: See clause 8.7.3.4.

Outputs: See clause 8.7.3.5.

See clause 8.7.2.3 for details of usage of this operation.

9.2.11.4 AIMLE client registration delete request operation

API operation name: Aimles_AIMLEClientRegistration_Delete Request

Description: The consumer requests AIMLE server for delete AIMLE client registration.

Inputs: See clause 8.7.3.6.

Outputs: See clause 8.7.3.7.

See clause 8.7.2.4 for details of usage of this operation.

9.2.12 Split AI/ML Operation API

9.2.12.1 General

Table 9.2.12.1-1 illustrates the API for Split AI/ML Operation. This API enables the AIMLE Client or VAL server to communicate with the AIMLE server for split AI/ML operation pipeline discovery or creation, for node registration and for split operation event subscription.

Table 9.2.12.1-1: Aimles_SplitOperation API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_SplitOpPipeline	Discover	Request/Response	AIMLE Client
	Create		
	Update		
	Delete		
Aimles_SplitOpNodeRegistration	Request	Request/Response	VAL Server
	Update		
	Deregister		
Aimles_SplitOpEvent	Subscribe	Subscribe/Notify	AIMLE Client, VAL Server
	Notify	Subscribe/Notify	AIMLE Client, VAL Server
	UpdateSubscription		
	Unsubscribe		

9.2.12.2 Aimles_SplitOpPipeline_Discover operation

API operation name: Aimles_SplitOpPipeline_Discover_Request

Description: The consumer requests for one time split AI/ML operation pipeline discovery.

Inputs: See clause 8.14.3.2.

Outputs: See clause 8.14.3.3.

See clause 8.14.2.2 for details of usage of this operation.

9.2.12.3 Aimles_SplitOpPipeline_Create operation

API operation name: Aimles_SplitOpPipeline_Create_Request

Description: The consumer creates a split AI/ML operation pipeline.

Inputs: See clause 8.14.3.4.

Outputs: See clause 8.14.3.5.

See clause 8.14.2.3 for details of usage of this operation.

9.2.12.4 Aimles_SplitOpPipeline_Update operation

API operation name: Aimles_SplitOpPipeline_Update_Request

Description: The consumer updates a split AI/ML operation pipeline.

Inputs: See clause 8.14.3.19.

Outputs: See clause 8.14.3.20.

See clause 8.14.2.6 for details of usage of this operation.

9.2.12.5 Aimles_SplitOpPipeline_Delete operation

API operation name: Aimles_SplitOpPipeline_Delete_Request

Description: The consumer deletes a split AI/ML operation pipeline.

Inputs: See clause 8.14.3.21.

Outputs: See clause 8.14.3.22.

See clause 8.14.2.7 for details of usage of this operation.

9.2.12.6 Aimles_SplitOpNodeRegistration_Request operation

API operation name: Aimles_SplitOpNodeRegistration_Request

Description: The consumer requests for one time split AI/ML operation node registration.

Inputs: See clause 8.14.3.6.

Outputs: See clause 8.14.3.7.

See clause 8.14.2.4.2 for details of usage of this operation.

9.2.12.7 Aimles_SplitOpNodeRegistration_Update operation

API operation name: Aimles_SplitOpNodeRegistration_Update

Description: The consumer updates a split AI/ML operation node registration.

Inputs: See clause 8.14.3.8.

Outputs: See clause 8.14.3.9.

See clause 8.14.2.4.3 for details of usage of this operation.

9.2.12.8 Aimles_SplitOpNodeRegistration_Deregister operation

API operation name: Aimles_SplitOpNodeRegistration_Deregister

Description: The consumer deregisters a split AI/ML operation node registration.

Inputs: See clause 8.14.3.10.

Outputs: See clause 8.14.3.11.

See clause 8.14.2.4.4 for details of usage of this operation.

9.2.12.9 Aimles_SplitOpEvent_Subscribe operation

API operation name: Aimles_SplitOpEvent_Subscribe

Description: The consumer subscribes for split AI/ML operation events.

Inputs: See clause 8.14.3.12.

Outputs: See clause 8.14.3.13.

See clause 8.14.2.5.2 for details of usage of this operation.

9.2.12.10 Aimles_SplitOpEvent_Notify operation

API operation name: Aimles_SplitOpEvent_Notify

Description: The consumer is notified of split AI/ML operation events.

Inputs: See clause 8.14.3.14.

Outputs: None.

See clause 8.14.2.5.3 for details of usage of this operation.

9.2.12.11 Aimles_SplitOpEvent_UpdateSubscription operation

API operation name: Aimles_SplitOpEvent_UpdateSubscription

Description: The consumer updates an existing subscription for split AI/ML operation events.

Inputs: See clause 8.14.3.15.

Outputs: See clause 8.14.3.16.

See clause 8.14.2.5.4 for details of usage of this operation.

9.2.12.12 Aimles_SplitOpEvent_Unsubscribe operation

API operation name: Aimles_SplitOpEvent_Unsubscribe

Description: The consumer cancels an existing subscription for split AI/ML operation events.

Inputs: See clause 8.14.3.17.

Outputs: See clause 8.14.3.18.

See clause 8.14.2.5.5 for details of usage of this operation.

9.2.13 ML model update API

9.2.13.1 General

Table 9.2.13.1-1 illustrates the API for ML model update.

Table 9.2.13.1-1: Aimles_MLModelUpdate API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_MLModelUpdate	Request	Request/Response	VAL server, ADAE server, AIMLE client

9.2.13.2 Aimles_MLModelUpdate_Request operation

API operation name: Aimles_MLModelUpdate_Request

Description: The AIMLE consumer requests ML Model update.

Inputs: See clause 8.21.3.1.

Outputs: See clause 8.21.3.2.

See clause 8.21.2.1 for details of usage of this operation.

9.2.14 ML model performance monitoring API

9.2.14.1 General

Table 9.2.14.1-1 illustrates the API for ML model performance monitoring subscription and notification. This API enables the VAL server to communicate with the AIMLE server for subscribing for model performance monitoring.

Table 9.2.14.1-1: Aimles_MLModelPerfMonitor API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_MLModelPerfMonitor	Subscribe	Subscribe/Notify	VAL Server
	Notify		

9.2.14.2 Subscribe

API operation name: MLModelPerfMonitor_Subscribe

Description: The consumer subscribes for ML model performance monitoring.

Inputs: See clause 8.22.3.1.

Outputs: See clause 8.22.3.2.

See clause 8.22.2 for details of usage of this operation.

9.2.14.3 Notify

API operation name: MLModelPerfMonitor_Notify

Description: The consumer is notified by ML repository on the ML model performance monitoring.

Inputs: -

Outputs: See clause 8.22.3.3.

See clause 8.22.2 for details of usage of this operation.

9.2.15 AIMLE assisted ML model selection API

9.2.15.1 General

Table 9.2.15.1-1 illustrates the API for AIMLE assisted ML model selection. This API enables an AIMLE service consumer to request assistance with ML model selection.

Table 9.2.15.1-1: Aimles_AssistedMLModelSelection API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_AssistedMLModelSelection	Subscribe	Subscribe/Notify	VAL Server
	Notify		

9.2.15.2 Aimles_AssistedMLModelSelection_Subscribe operation

API operation name: Aimles_AssistedMLModelSelection_Subscribe

Description: The consumer subscribes to AIMLE assisted ML model selection from an AIMLE server.

Inputs: See clause 8.23.3.1.

Outputs: See clause 8.23.3.2.

See clause 8.23.2 for details of usage of this operation.

9.2.15.3 Aimles_AssistedMLModelSelection_Notify operation

API operation name: Aimles_AssistedMLModelSelection_Notify

Description: The consumer receives notifications for AIMLE assisted ML model selection.

Inputs: See clause 8.23.3.3.

Outputs: None.

See clause 8.23.2 for details of usage of this operation.

9.2.16 AIMLE context transfer API

9.2.16.1 General

Table 9.2.16.1-1 illustrates the API for AIMLE context transfer. This API enables an AIMLE server to communicate AIMLE context to another AIMLE server in edge scenarios.

Table 9.2.16.1-1: Aimles_ContextTransfer API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_ContextTransfer	Request	Request/Response	AIMLE Server

9.2.16.2 Aimles_ContextTransfer_Request operation

API operation name: Aimles_ContextTransfer_Request

Description: The consumer requests AIMLE context transfer to an AIMLE server.

Inputs: See clause 8.24.3.1.

Outputs: See clause 8.24.3.2.

See clause 8.24.2 for details of usage of this operation.

9.2.17 AIMLE Assistance of Hierarchical Computing API

9.2.17.1 General

Table 9.2.17.1-1 illustrates the API for assist hierarchical computing. This API enables the communication between VAL server (e.g. CAS, EAS) and the AIMLE server for assistance of a hierarchical computing process.

Table 9.2.17.1-1: Aimles_HierarchicalComputingAssist API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimles_HierarchicalComputingAssist	Request	Request/Response	VAL Server (e.g. CAS, EAS)

9.2.17.2 Aimles_HierarchicalComputingAssist_Request operation

API operation name: Aimles_HierarchicalComputingAssist_Request

Description: The consumer requests for assistance of a hierarchical computing process.

Inputs: See clause 8.25.3.1.

Outputs: See clause 8.25.3.2.

See clause 8.25.2 for details of usage of this operation.

9.3 ML repository APIs

9.3.1 MLR FL member Registration API

9.3.1.1 General

Table 9.3.1.1-1 illustrates the API for FL member registration. This API enables the VAL server (candidate FL member) to communicate with the ML repository for registering as candidate FL member.

Table 9.3.1.1-1: MLR_FLMemberRegister API

API Name	API Operations	Operation Semantics	Consumer(s)
MLR_FLMemberRegistration	Request	Request/Response	VAL Server
MLR_FLMemberRegistrationUpdate	Request	Request/Response	VAL Server
MLR_FLMemberRegistrationFetch	Request	Request/Response	VAL Server
MLR_FLMemberDeregistration	Request	Request/Response	VAL Server

9.3.1.2 MLR_FLMemberRegistration_Request operation

API operation name: MLR_FLMemberRegistration_Request

Description: The consumer requests for registering as FL member.

Inputs: See clause 8.4.4.1.

Outputs: See clause 8.4.4.2.

See clause 8.4.2 for details of usage of this operation.

9.3.1.3 MLR_FLMemberRegistrationUpdate_Request operation

API operation name: MLR_FLMemberRegistrationUpdate_Request

Description: The consumer requests for updating the registration as FL member.

Inputs: See clause 8.4.4.3.

Outputs: See clause 8.4.4.4.

See clause 8.4.3 for details of usage of this operation.

9.3.1.4 MLR_FLMemberRegistrationFetch_Request operation

API operation name: MLR_FLMemberRegistrationFetch_Request

Description: The consumer requests for Fetching the registering as FL member.

Inputs: See clause 8.17.3.1, Table 8.17.3.1-2.

Outputs: See clause 8.17.3.2, Table 8.17.3.2-2.

See clause 8.17.2 for details of usage of this operation.

9.3.1.5 MLR_FLMemberDeregistration_Request operation

API operation name: MLR_FLMemberDeregistration_Request

Description: The consumer requests for deregistering a candidate registered FL member.

Inputs: See clause 8.4.4.6.

Outputs: See clause 8.4.4.7.

See clause 8.3a for details of usage of this operation.

9.3.2 MLR FL Event API

9.3.2.1 General

Table 9.3.2.1-1 illustrates the API for FL-related event subscription and notification. This API enables the AIMLE server to communicate with the ML repository for subscribing for FL related events.

Table 9.3.2.1-1: MLR_FLEvents API

API Name	API Operations	Operation Semantics	Consumer(s)
MLR_FLEvents	Subscribe	Subscribe/Notify	AIMLE Server
	Notify		

9.3.2.2 Subscribe

API operation name: _FLEvents_Subscribe

Description: The consumer subscribes for FL-related events.

Inputs: See clause 8.5.5.2.

Outputs: See clause 8.5.5.3.

See clause 8.5.2 for details of usage of this operation.

9.3.2.3 Notify

API operation name: FLEvents_Notify

Description: The consumer is notified by ML repository on the FL-related events.

Inputs: -

Outputs: See clause 8.5.5.4.

See clause 8.5.3 for details of usage of this operation.

9.3.3 MLR model management API

9.3.3.1 General

Table 9.3.3.1-1 illustrates the APIs for MLR model management to enable an AIMLE server to discover and store ML models in an ML repository.

Table 9.3.3.1-1: MLR_ModelManagement API

API Name	API Operations	Operation Semantics	Consumer(s)
MLR_ModelInformationStorage	Request	Request/Response	AIMLE server
MLR_ModelInformationDiscovery	Request	Request/Response	AIMLE server

9.3.3.2 MLR_ModelInformationStorage_Request operation

API operation name: MLR_MLModelInformationStorage_Request

Description: The consumer requests to perform storage of ML models.

Inputs: See clause 8.11.4.1.

Outputs: See clause 8.11.4.2.

See clause 8.11.2 for details of usage of this operation.

9.3.3.3 MLR_ModelInformationDiscovery_Request operation

API operation name: MLR_MLModelInformationDiscovery_Request

Description: The consumer requests to perform discovery of ML models.

Inputs: See clause 8.11.4.3.

Outputs: See clause 8.11.4.4.

See clause 8.11.3 for details of usage of this operation.

9.4 AIMLE client APIs

9.4.1 ML model training capability evaluation API

9.4.1.1 General

Table 9.4.1.1-1 illustrates the API for ML model training capability evaluation. This API enables the AIMLE Server to communicate with the AIMLE client(s) for ML model training capability evaluation.

Table 9.4.1.1-1: Aimlec_MLModelTrainingCapabilityEva API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_MLModelTrainingCapabilityEva	Request	Request/Response	AIMLE Server
	Response		AIMLE Client

9.4.1.2 Aimlec_MLModelTrainingCapabilityEva_Request operation

API operation name: Aimlec_MLModelTrainingCapabilityEva_Request

Description: The consumer requests for ML Model training capability evaluation.

Inputs: See clause 8.19.3.1.

Outputs: None.

See clause 8.19.2 for details of usage of this operation.

9.4.1.3 Aimlec_MLModelTrainingCapabilityEva_Response operation

API operation name: Aimlec_MLModelTrainingCapabilityEva_Response

Description: The consumer responses for ML Model training capability evaluation.

Inputs: See clause 8.19.3.2.

Outputs: None.

See clause 8.19.2 for details of usage of this operation.

9.4.2 HFL training API

9.4.2.1 General

Table 9.4.2.1-1 illustrates the API for HFL training to enable an AIMLE server to subscribe to HFL training.

Table 9.4.2.1-1: Aimlec_HFLTraining API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_HFLTraining	Subscribe	Subscribe/Notify	AIMLE server
	Notify		

9.4.2.2 Aimlec_HFLTraining_Subscribe operation

API operation name: Aimlec_HFLTraining_Subscribe

Description: The consumer subscribes to HFL training.

Inputs: See clause 8.12.3.1.

Outputs: See clause 8.12.3.2.

See clause 8.12.2 for details of usage of this operation.

9.4.2.3 Aimlec_HFLTraining_Notify operation

API operation name: Aimlec_HFLTraining_Subscribe

Description: The consumer receives notifications from HFL training.

Inputs: See clause 8.12.3.3.

Outputs: None.

See clause 8.12.2 for details of usage of this operation.

9.4.3 Client data processing API

9.4.3.1 General

Table 9.4.3.1-1 illustrates the API for Client data processing to enable an AIMLE server to request data processing to be performed on AIMLE clients.

Table 9.4.3.1-1: Aimlec_ClientDataProcessing API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_ClientDataProcessing	Request	Request/Response	AIMLE server

9.4.3.2 Aimlec_ClientDataProcessing_Request operation

API operation name: Aimlec_ClientDataProcessing_Request

Description: The consumer requests data processing to be performed on AIMLE clients.

Inputs: See clause 8.15.3.4.

Outputs: See clause 8.15.3.5.

See clause 8.15.2 for details of usage of this operation.

9.4.4 AIMLE Client Service Operations API

9.4.4.1 General

Table 9.4.4.1-1 illustrates the API for AIMLE Client Service Operations.

Table 9.4.4.1-1: Aimlec_AIMLEClientServiceOperations APIs

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_AIMLEClientServiceOperations	Request	Request/Response	AIMLE Server
	Response		AIMLE Client

9.4.4.2 Aimlec_ AIMLEClientServiceOperations_Request operation

API operation name: Aimlec_AIMLEClientServiceOperations_Request

Description: The consumer requests for AIMLE client service operations.

Inputs: See clause 8.20.3.3.

Outputs: None.

See clause 8.20.2.2 for details of usage of this operation.

9.4.4.3 Aimlec_ AIMLEClientServiceOperations_Response operation

API operation name: Aimlec_AIMLEClientServiceOperations_Response

Description: The consumer responses for AIMLE client service operations

Inputs: See clause 8.20.3.4.

Outputs: None.

See clause 8.20.2.2 for details of usage of this operation.

9.4.5 AIMLE Client Participation API

9.4.5.1 General

Table 9.4.5.1-1 illustrates the APIs for AIMLE client participation. This API enables the communication between the AIMLE client and the AIMLE server for AIMLE client participation.

Table 9.4.5.1-1: Aimlec_AIMLEClientParticipation API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_AIMLEClientParticipation	Request	Request/Response	AIMLE Server

9.4.5.2 Aimlec_AIMLEClientParticipation_Request operation

API operation name: Aimlec_AIMLEClientParticipation_Request

Description: The consumer requests AIMLE server for AIMLE client participation.

Inputs: See clause 8.10.3.1.

Outputs: See clause 8.10.3.2.

See clause 8.10.2.1 for details of usage of this operation.

9.4.6 AIMLE AI/ML Task Transfer APIs

9.4.6.1 General

Table 9.4.6.1-1 illustrates the APIs for AI/ML task transfer. This API enables the communication between VAL UE (via AIMLE Client) and the AIMLE server for AI/ML Task Transfer.

Table 9.4.6.1-1: Aimlec_AIMLTaskTransfer API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_AIMLTaskTransfer	Request	Request/Response	AIMLE Server

Table 9.4.6.1-2: Aimlec_DirectAIMLTaskTransfer API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_DirectAIMLTaskTransfer	Request	Request/Response	AIMLE Client

9.4.6.2 Aimlec_AIMLTaskTransfer_Request operation

9.4.6.2.1 General

API description: This API enables the AIMLE server to communicate with the target AI/ML member (AIMLE Client) for request AIML task transfer from source AI/ML member (AIMLE Client) to the target AI/ML member (AIMLE Client).

9.4.6.2.2 AIML task transfer request operation

API operation name: Aimlec_AIMLTaskTransfer_Request

Description: The consumer requests for AI/ML task transfer.

Inputs: See clause 8.6.3.4.

Outputs: See clause 8.6.3.5.

See clause 8.6.2.2 for details of usage of this operation.

9.4.6.3 Aimlec_DirectAIMLTaskTransfer_Request operation

9.4.6.3.1 General

API description: This API enables the AIMLE client to communicate with the target AI/ML member (AIMLE Client) for request direct AIML task transfer from the source AI/ML member (AIMLE Client) to the target AI/ML member (AIMLE Client).

9.4.6.3.2 Direct AIML task transfer request operation

API operation name: Aimlec_DirectAIMLTaskTransfer_Request

Description: The consumer requests for direct AI/ML task transfer.

Inputs: See clause 8.6.3.6.

Outputs: See clause 8.6.3.7.

See clause 8.6.2.3 for details of usage of this operation.

9.4.7 FL grouping indication API

9.4.7.1 General

Table 9.4.7.1-1 illustrates the APIs for FL grouping indication. This API enables the communication between the AIMLE server and the FL member for indicating the FL grouping procedure.

Table 9.4.7.1-1: Aimlec_FLGroupIndication API

API Name	API Operations	Operation Semantics	Consumer(s)
Aimlec_FLGroupIndication	Request	Request/Response	AIMLE Server

9.4.7.2 Aimlec_FLGroupIndication operation

API operation name: Aimlec_FLGroupIndication

Description: The consumer indicates to FL member (AIMLE clients which are deployed on UEs) for FL grouping procedure.

Inputs: See clause 8.17.3.3.

Outputs: See clause 8.17.3.4.

See clause 8.17.2 for details of usage of this operation.

Annex A (informative): Deployment scenarios

A.1 General

This Annex provides the different deployment models for AIMLE services. There could be three deployment options:

- AIMLE server can be deployed at a centralized cloud platform and collects data from multiple EDNs.
- AIMLE server can be deployed at the edge platform.
- Hierarchical AIMLE server deployment, where multiple AIML enablement services are deployed in edge or central clouds (e.g., in hierarchical architecture). Such deployment allows for local-global analytics for system wide optimization.

A.2 Deployment model #1: Cloud-deployed AIMLE server

In this deployment, , the AIMLE server is centrally located and can provide support for AIML operations to the application and edge services (EAS/EES, VAL server). An example deployment option for AIMLE server at the cloud is shown in Figure A.2-1

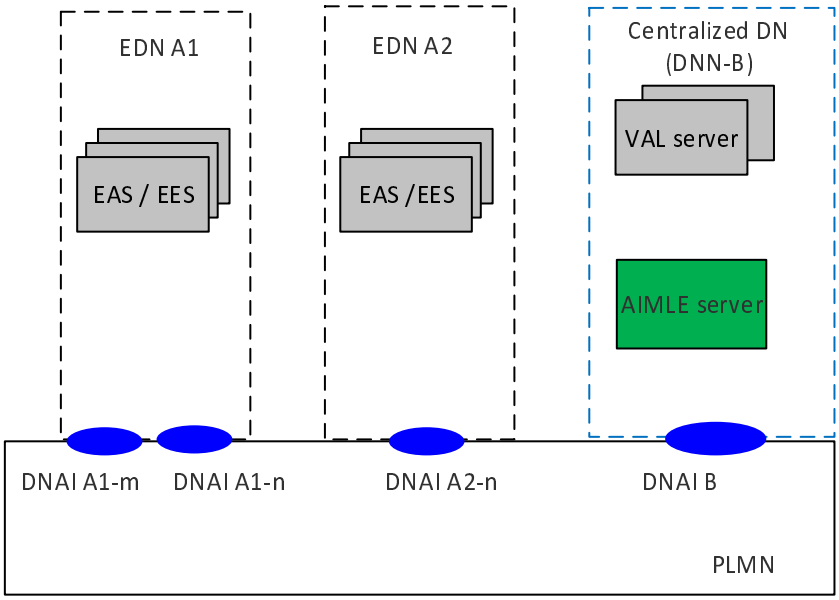

Figure A.2-1: Example deployment for AIMLE at the cloud

A.3 Deployment model #2: Edge-deployed AIMLE server

In this deployment, the AIMLE server deployed as EAS is located at the EDN and provides AIML enablement services to the other EAS(s) or other edge native applications at the edge platform. AIMLE services can be deployed by the ECSP or the MNO to provide value-add services related to AI/ML operations.

The ML support operations, that the edge deployed AIMLE Server provides, are applicable to the AIMLE service areas (as shown in the example deployment scenario in Figure A.3-1), which are equivalent to the EDN service areas.

NOTE: AIMLE server deployed as EAS can provide application enablement service to EES.

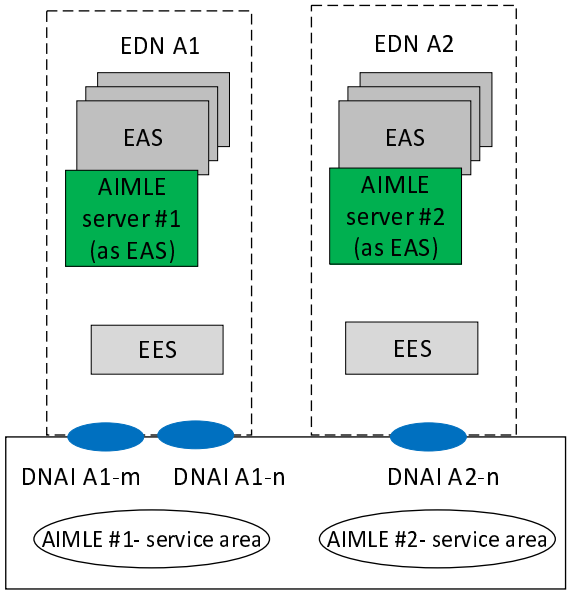

Figure A.3-1: Example deployment for AIMLE at the cloud

A.4 Deployment model #3: Hierarchical AIMLE server deployment

In this deployment, multiple AIMLE servers can be located at different EDNs (deployed as EASs)/DNs and can be deployed by the same provider. Such hierarchical deployments allow the local – global ML operations (e.g., federated learning across domains).

The ML support services that the edge deployed AIMLE server correspond to the AIMLE service areas (as shown in the example in Figure A.4-1), which is equivalent to the EDN service areas. The central AIMLE server covers all PLMN area and is used to coordinate the ML related operations (e.g., FL server / aggregator) with the distributed AIMLE servers.

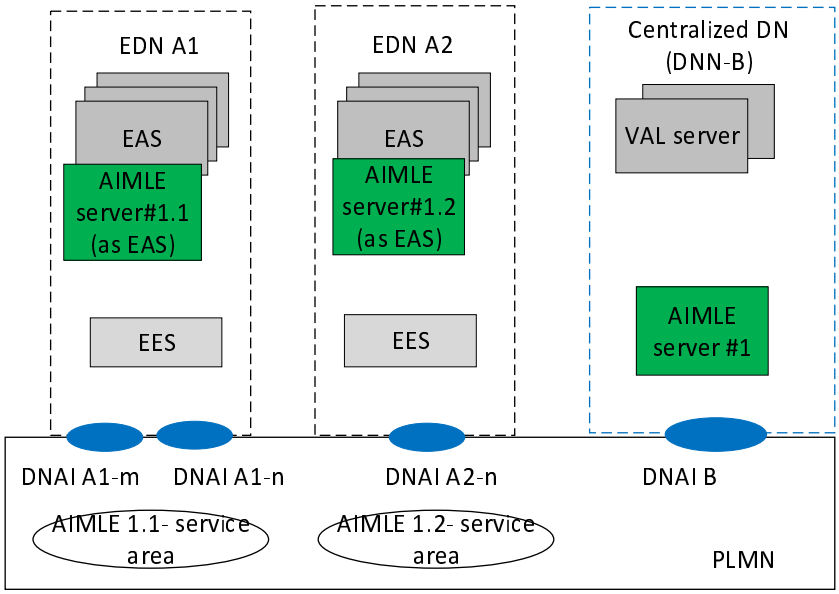

Figure A.4-1: Example hierarchical deployment of AIMLE

Annex B (informative): Business Scenarios

Figure B-1 shows the business relationships that exist for the AIMLE functionality and that are needed to support a single VAL user.

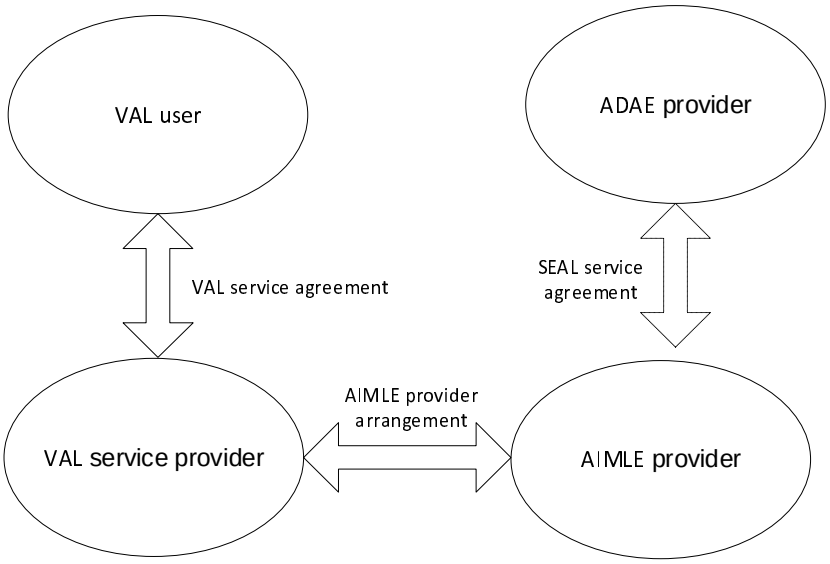

Figure B-1: Business relationships for VAL services

The VAL user belongs to a VAL service provider based on a VAL service agreement between the VAL user and the VAL service provider. The VAL service provider can have VAL service agreements with several VAL users. The VAL user can have VAL service agreements with several VAL service providers.

The VAL service provider can have AIMLE provider arrangements with multiple AIMLE providers.

The AIMLE server is part of the AIMLE provider. The AIMLE provider can have SEAL service agreements with other SEAL service providers and in particular ADAE provider if needed for utilizing AIMLE services for ADAE analytics. Such arrangements allow the ADAE provider to utilize the AIMLE provider services.

The AIMLE and ADAE providers can either be part of the PLMN or have service arrangements with PLMN operators.

NOTE: The ADAE provider can have further arrangements with PLMN operator and VAL service provider; however, this is not shown in the figure.

Annex C (informative): Role of AIMLE in ML Model Lifecycle

C.1 General

ML model lifecycle (aka ML model operation workflow) contains several parts including ML model training, ML model testing, AI/ML inference emulation, ML model deployment, AI/ML inference, Data Management, intermediate model aggregation, Trained AI/ML model delivery, and so on. The consumer can delegate all or part of ML model lifecycle to the AIML enablement layer. In doing so the AIML enablement layer can take over some AI/ML related work for the consumer and reduce the complexity of implementation for consumers. In the below sub-clauses, some options for the role of AIMLE in the ML model lifecycle are provided.

C.2 Role#1 of AIMLE in ML Model Lifecycle

In this scenario, the consumer completely rely on the AIML enablement layer for managing the ML operational workflow. That is, the consumer send the requirements of the AI/ML application, then the AIML enablement layer can perform ML operational workflows and determine the required AI/ML model to send to the consumer. Such role of AIMLE is captured in the capability related to AIML service operations control and management as described in clause 8.20.

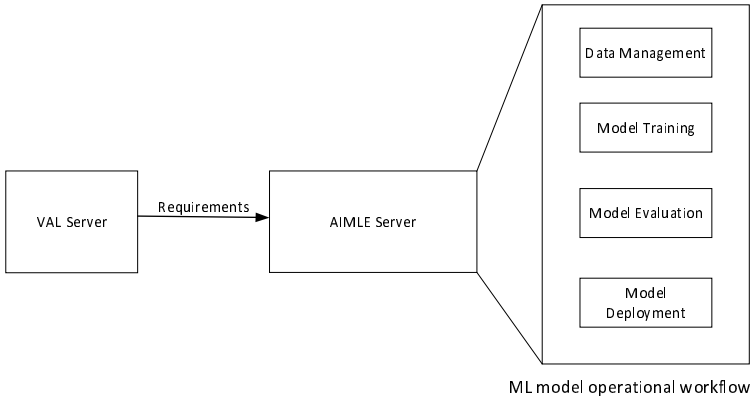

Figure C.2-1: Example of Role#1

C.3 Role#2 of AIMLE in ML Model Lifecycle

In this scenario, consumer would like to perform part of the ML operational workflow, then consumer partially rely on the AIML enablement layer to assist with ML operational workflowss. Due to different ML operational workflow division, the AIML enablement layer can provide different levels of assistance on ML workflows for consumers.

Such capability is captured for example for ML model training (as in clause 8.3), HFL training (as in clause 8.12), model evaluation (which is captured in clause 8.19 in model capability evaluation, and 8.22 in model performance monitoring procedures). Figure C.3-1 illustrates an example where ML model training is handled by the AIMLE where the other operations are performed by the VAL / ASP.

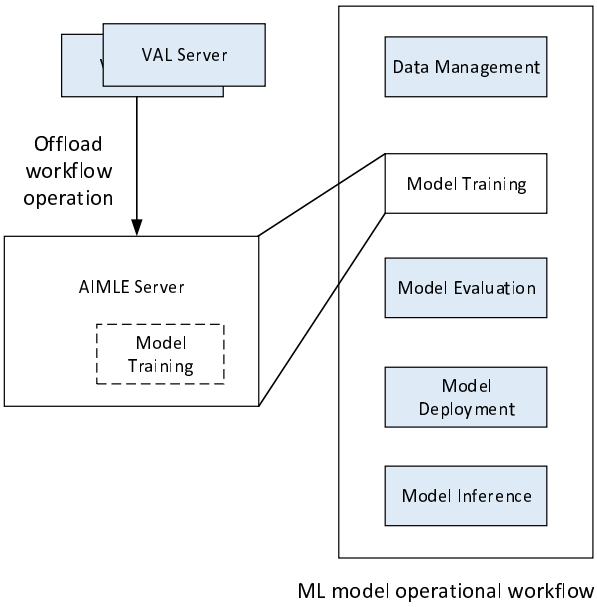

Figure C.3-1: Example of Role#2

C.4 Role#3 of AIMLE in ML Model Lifecycle

In this scenario, the AIMLE is not performing part of the ML operational workflow; however it serves as a platform to enable the AI/ML apps to utilize ML operational workflows which are provided by VAL. In this role, the AIMLE supports tasks like the discovery, registration, storage, grouping and selection of entities to be performing the ML operations in the lifecycle. Such role is more applicable to ML model lifecycle enablement which provides assistance for use cases where an ASP/VAL wants to find other application entities to perform some ML operations (e.g. ML model inference) and AIMLE server as a mediator to accomplish this.

An example including some capabilities is illustrated in Figure C.4-1. In this figure, the support capabilities are based on AIMLE capabilities identified in this specification. In particular, AIMLE is undertaking:

- ML model related support capabilities such as model retrieval, discovery and storage (as covered in procedures in clauses 8.2 and 8.11)
- ML operation related support capabilities such as VFL/ HFL and TL enablement, Split AI/ML Operation support, Data management assistance, AI/ML task transfer, FL assistance in member grouping, registration and event notification (as covered in procedures in clauses 8.4, 8.6, 8.12, 8.14, 8.15-8.18).
- AIMLE client related support capabilities, including AIMLE client registration, discovery, participation, monitoring, selection (as covered in procedures in clauses 8.7-8.10, 8.13).

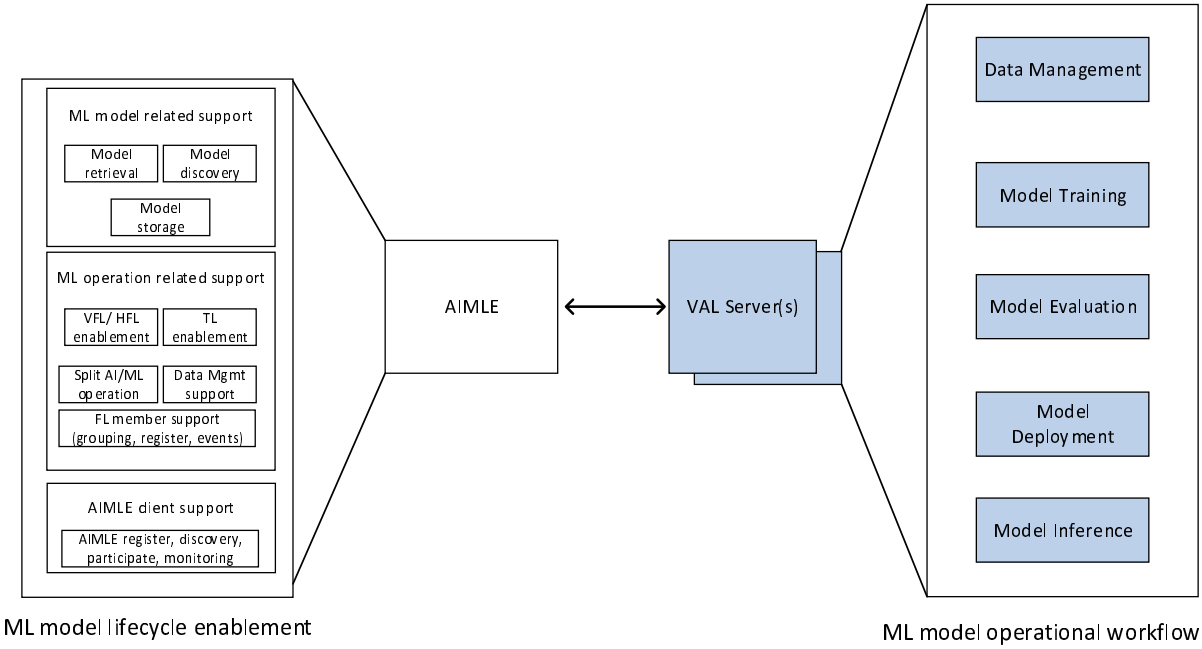

Figure C.4-1: Example of Role#3

Annex D:
Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	Rev	**Cat**	**Subject/Comment**	**New version**
2024-07	SA6#62 ad hoc-e					TS skeleton	0.0.0
2024-07	SA6#62 ad hoc-e					Implementation of the following pCRs approved by SA6: S6a240288, S6a240289.	0.1.0
2024-08	SA6#62					Implementation of the following pCRs approved by SA6: S6-243211, S6-243580, S6-243699, S6-243589, S6-243714, S6-243591, S6-243734, S6-243727, S6-243728, S6-243729, S6-243751, S6-243752, S6-243702, S6-243659, S6-243660, S6-243752, S6-243642, S6-243721, S6-243644, S6-243645.	0.2.0
2024-10	SA6#63					Implementation of the following pCRs approved by SA6: S6-244583, S6-244584, S6-244585, S6-244586, S6-244677, S6-244587, S6-244707, S6-244590, S6-244591, S6-244068, S6-244592, S6-244713, S6-244071, S6-244595, S6-244678, S6-244162, S6-244597, S6-244662, S6-244736, S6-244164, S6-244601, S6-244663, S6-244708, S6-244605, S6-244684, S6-244607, S6-244705, S6-244727.	0.3.0
2024-11	SA6#64					Implementation of the following pCRs approved by SA6: S6-245637, S6-245666, S6-245557, S6-245685, S6-245560, S6-245324, S6-245562, S6-245561,S6-245667, S6-245564, S6-245565, S6-245638, S6-245567,S6-245668, S6-245694, S6-245571, S6-245572, S6-245573, S6-245669, S6-245672, S6-245670,S6-245671, S6-245578, S6-245695, S6-245580, S6-245701, S6-245582, S6-245583, S6-245321, S6-245587, S6-245588, S6-245729, S6-245586, S6-245673, S6-245591,S6-245675, S6-245593, S6-245731.	0.4.0
2024-12	SA#106	SP-241703				Submitted to SA#106 for information and approval	1.0.0
2024-12	SA#106	SP-241703				MCC Editorial update for publication after TSG SA approval (SA#106)	19.0.0
2025-03	SA#107	SP-250206	0001	1	F	Delete term of AI/ML server	19.1.0
2025-03	SA#107	SP-250206	0002	1	F	Unified use of AIML enablement	19.1.0
2025-03	SA#107	SP-250206	0003	1	F	Resolving Editor's Notes in clause 8.8 and 8.13	19.1.0
2025-03	SA#107	SP-250206	0004	1	F	Aligning AIMLE client discovery with the AIMLE client registration IEs	19.1.0
2025-03	SA#107	SP-250206	0005	1	F	AIMLE client selection subscription update and subscription cancel	19.1.0
2025-03	SA#107	SP-250206	0006	1	F	Corrections related to ML model data types	19.1.0
2025-03	SA#107	SP-250206	0007	1	F	Consistent VAL Service ID usage	19.1.0
2025-03	SA#107	SP-250206	0008	2	F	Corrections to AIMLE client registration	19.1.0
2025-03	SA#107	SP-250206	0009	1	F	Update to AIMLE client selection	19.1.0
2025-03	SA#107	SP-250206	0010	1	F	Additional AIMLE identifiers	19.1.0
2025-03	SA#107	SP-250206	0013	1	F	EN resolutions in TS 23.482	19.1.0
2025-03	SA#107	SP-250206	0018	1	F	Definitions of terms and abbreviations	19.1.0
2025-03	SA#107	SP-250206	0022	1	F	Updates to the Response of ML Model Information Storage and Discovery	19.1.0
2025-03	SA#107	SP-250206	0023	1	F	Updates to ML Model Training Notification	19.1.0
2025-03	SA#107	SP-250206	0024	1	F	Updates to AI/ML Task Transfer	19.1.0
2025-03	SA#107	SP-250206	0025	1	F	Corrections to AIML Service Operations	19.1.0
2025-03	SA#107	SP-250206	0026	2	F	Updates to FL Member Grouping	19.1.0
2025-03	SA#107	SP-250206	0027	1	F	Adding AIMLE Client to FL Member Registration Procedure and Information Flows	19.1.0
2025-03	SA#107	SP-250206	0028		F	Adding Missing Service Operations to API Clauses	19.1.0
2025-03	SA#107	SP-250206	0029	1	F	Add Functional Description for the Hierarchical Computing	19.1.0
2025-03	SA#107	SP-250206	0030		F	Align Termination for ADAE Analytics ID	19.1.0
2025-03	SA#107	SP-250206	0031	1	F	Corrections to Procedures and Information Flows	19.1.0
2025-06	SA#108	SP-250593	0032	1	F	Adding Reference and General Description for Security Aspects	19.2.0
2025-06	SA#108	SP-250593	0033	1	F	Updates to Information Flows for Responses	19.2.0
2025-06	SA#108	SP-250593	0034		F	Clarification on AIMLE Server Deployment for Supporting Hierarchical Computing	19.2.0
2025-06	SA#108	SP-250593	0035	2	F	Modification of the business relationship in TS23.482	19.2.0
2025-06	SA#108	SP-250593	0036	1	F	EN resolutions and minor fixes	19.2.0
2025-06	SA#108	SP-250593	0038	4	F	Addition of missing operations in FL Member Grouping	19.2.0
2025-06	SA#108	SP-250593	0041	2	F	Deregistration of an FL Member	19.2.0
2025-06	SA#108	SP-250593	0042	2	F	Correction on AIMLE Server to assist AI/ML task transfer	19.2.0
2025-06	SA#108	SP-250593	0043	1	F	Correction on ML model information storage and discovery	19.2.0
2025-09	SA#109	SP-251051	0044	3	F	ML model training procedure alignment with client selection and HFL/VFL procedures	19.3.0
2025-09	SA#109	SP-251051	0051	1	F	Corrections to Deployment Models for Edge AIMLE Server	19.3.0
2025-09	SA#109	SP-251051	0052	1	F	Correction on ML model information discovery	19.3.0
2025-09	SA#109	SP-251051	0053	2	F	AIMLE context transfer correction	19.3.0
2025-09	SA#109	SP-251051	0054	1	F	AIMLE ML model training request correction	19.3.0
2026-01	SA#110	SP-251470	0056	1	F	HFL training completion	19.4.0

2026-01	SA#110	SP-251470	0057	1	F	Clarifications to clause 8.15.2	19.4.0
2026-03	SA#111	SP-260191	0059	1	F	AIMLE client selection corrections	19.5.0
2026-03	SA#111	SP-260191	0078	1	F	VAL server registration as FL member	19.5.0
2026-03	SA#111	SP-260191	0082		F	Registration and discovery of VAL client as AI/ML member	19.5.0

History

Version	Date	Status
V19.3.0	January 2026	Publication
V19.4.0	February 2026	Publication
V19.5.0	April 2026	Publication