

**METHODS AND APPARATUS FOR ARTIFICIAL INTELLIGENCE / MACHINE
LEARNING ENABLEMENT FUNCTION FOR APPLICATION DATA ANALYTICS
ENABLEMENT**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/518,905, filed August 11, 2023, which is hereby incorporated by reference in its entirety.

BACKGROUND

[0002] Machine learning (ML) is a branch of artificial intelligence (AI) directed towards building methods that leverage data to improve performance on a set of tasks. For example, one or more ML algorithms may be used to build a model based on sample data (e.g., training data) in order to make predictions or decisions without being explicitly programmed to do so. The process of training a ML model may include data collection, data preparation/processing, model building/training, model evaluation, model deployment, monitoring and update, etc. After a ML model is trained, it may be deployed and used for an intended purpose, such as performing an inferencing task and generating the inferencing results. Access to hardware resources (e.g., computing power, communication bandwidth, etc.) as well as inferencing data may be needed to support the deployed model and the inferencing task.

[0003] 3GPP defines a network data analytics function (NWDAF) and an application data analytics enablement (ADAE) service to provide analytics services to various consumers in the network.

[0004] The analytics services provided by the NWDAF (e.g., TS 23.288) aim to support network data analytics services in the 5G Core network. Such analytics may collect data from other NFs, AFs, or OAM and may be exposed to the 3rd party AF to provide statistics and predictions related to various types of analytics, such as slice load level, observed service experience, NF load, network performance, UE related analytics (mobility, communication), User data congestion, QoS sustainability, DN performance, etc. For example, 3GPP TS 23.288 sets forth architecture enhancements for 5G System (5GS) to support network data analytics services.

[0005] ADAE service (e.g., TS 23.436) provides functions to support the exposure of data analytics services from different 3GPP domains to the vertical ASP in a unified manner. For example, 3GPP TS 23.436 sets forth functional architecture and information flows for Application Data Analytics Enablement Service. ADAE service defines, at an overarching layer, value-add application data analytics services which cover statistics and predictions for the end-to-end application service. The major consumers of ADAE services may include vertical-specific applications and edge applications.

SUMMARY

[0006] Described herein are methods, apparatus, and systems to manage AI/ML resources and capabilities available to the application and service enablement layer and coordinate AI/ML-enabled analytics services. According to some examples, the ADAE service may be enhanced to support AI/ML-enabled analytics services.

[0007] According to some examples, an AI/ML resource and capability management is provided, where the ADAE service maintains profiles for AI/ML resources and capabilities and supports the registration and discovery procedures of the AI/ML resources and capabilities.

[0008] According to some examples, an AI/ML-enabled analytics service coordination is provided, where the ADAE service maintains and monitors status of the AI/ML service instances, coordinates AI/ML service requests received from different consumers, and dynamically determines the service action based on the information of the service instances and AI/ML resources/capabilities.

[0009] According to some examples, enhancements to the ADAE service may support AI/ML-enabled analytics services at the 3GPP application and service enablement layer. The enhanced ADAE service may support the management of AI/ML resources and capabilities and coordination of AI/ML-enabled analytics services.

[0010] According to some examples, a message from an AI/ML resource/capability provider may be received. The message may include descriptions of resources or capabilities that may be provided by the provider to support AI/ML-enabled analytics services. The message may be a registration request, or a message responsive to a query/request from the ADAE. The AI/ML resource or capability may include AI/ML data, AI/ML model, training capability, inferencing capability, etc. The AI/ML resource/capability provider may be an ADAE service, a VAL entity

(VAL server, VAL client), an application/service enablement layer entity (e.g., SEAL, EES, etc.), or a core network function (e.g., NWDAF, DCCF, ADRF). Descriptions of the AI/ML resource or capability may include: (1) AI/ML data description, data format, data purpose, data characteristics, data location, data source, access policy, etc.; (2) AI/ML model description, model requirements, model location, associated training entity and training data, access policy, etc.; or (3) AI/ML training/inferencing capability description, supported features, supporting resource, capability schedule, access policy, service KPIs, etc.

[0011] According to some examples, an AI/ML resource and capability profile may be generated based on the received message. The profile may be created by the ADAE service, or by the A-DCCF or A-ADRF function as requested by the ADAE service. In addition to the descriptions of the resources or capabilities, the profile may further include the identifier of the corresponding resource/capability provider, and information of the associated service instances that are using or will be using the resource/capability.

[0012] According to some examples, a message may be received from an AI/ML resource/capability consumer. The message may include filter criteria of AI/ML resources or capabilities that are of interest to the consumer. The message may be a discovery request, or a query/subscription request. The AI/ML resource/capability consumer may be an ADAE service, a VAL entity (VAL server, VAL client), an application/service enablement layer entity (e.g., SEAL, EES, EEC, etc.). The filter criteria may include information in the AI/ML resource/capability profile.

[0013] According to some examples, a response may be sent to the AI/ML resource/capability consumer. The response may include information of the AI/ML resources or capabilities that are of interest to the consumer. The response may include the profile(s) of the AI/ML resource/capability.

[0014] According to some examples, a request may be received from an AI/ML service consumer. The request may include identifier and context information of the consumer, notification conditions and targets, required analytics type, service objective (e.g., training or inferencing or both), service requirements, etc. The request may further include information of any of the training data, model, training entity, inferencing entity that is specified by the consumer.

[0015] According to some examples, an AI/ML service instance profile may be generated based on the received request. In addition to the information from the request, the profile may

further include instance status, context information of the consumer (e.g., if not provided by the consumer), information of the training data, model, training entity, inferencing entity (if not specified by the consumer).

[0016] According to some examples, the service action may be determined based on examining the information (e.g., profiles) of AI/ML resources/capabilities and existing AI/ML service instances. The service action may include selecting the data provider for training data and inferencing data, selecting the training entity, selecting the inferencing entity, selecting existing data as training data, selecting existing model, updating an existing training process, or updating an existing inferencing process. The service action of selecting the training entity may include sending a training request with training data information to the selected training entity. The service action of selecting the inferencing entity may include sending an inferencing request with trained model information to the selected inferencing entity. The service action of selecting the training/inferencing entity may further include selecting and configuring multiple entities to jointly perform the training/inferencing process. The service action of updating an existing training process may include re-selecting the training entity, or combining training processes associated with multiple service instances and re-configuring the training entity/data. The service action of updating an existing inferencing process may include re-selecting the inferencing entity, or combining the lists of notification targets associated with multiple service instance at the inferencing entity.

[0017] Moreover, the enhanced ADAE service may have alternative implementations including, but not limited to, a stand-alone service independent of ADAE, an AI/ML coordination service, an AI/ML enablement service, etc.

[0018] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Furthermore, the claimed subject matter is not limited to features that solve any or all disadvantages noted in any part of this disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] A more detailed understanding may be had from the following description, given by way of example in conjunction with the accompanying drawings.

- [0020] FIG. 1 shows an example architecture for ADAE;
- [0021] FIG. 2 shows an example ADAE internal function architecture;
- [0022] FIG. 3 shows an example ADAE general architecture;
- [0023] FIG. 4 shows an example ADAE internal architecture;
- [0024] FIG. 5 shows an example AI/ML-enabled ADAE service overview associated with native AI/ML capability;
- [0025] FIG. 6 shows an example AI/ML-enabled ADAE service overview associated with external AI/ML capability;
- [0026] FIG. 7 shows an example AI/ML data registration procedure;
- [0027] FIG. 8 shows an example AI/ML model registration procedure;
- [0028] FIG. 9 shows an example AI/ML training/inferencing capability registration procedure;
- [0029] FIG. 10 shows an example AI/ML data discovery procedure;
- [0030] FIG. 11 shows an example AI/ML model discovery procedure;
- [0031] FIG. 12 shows an example AI/ML capability discovery procedure;
- [0032] FIG. 13 shows an example AI/ML service coordination process;
- [0033] FIG. 14 shows an example AI/ML service procedure associated with training;
- [0034] FIG. 15 shows an example AI/ML service procedure associated with inferencing;
- [0035] FIG. 16 shows an example AI/ML-enabled ADAE support for application performance analytics;
- [0036] FIG. 17 shows an example AI/ML-enabled ADAE support for edge analytics;
- [0037] FIG. 18 shows an example GUI;
- [0038] FIG. 19A illustrates an example communications system;
- [0039] FIG. 19B shows a system diagram of an example RAN and core network;
- [0040] FIG. 19C shows a system diagram of an example RAN and core network;
- [0041] FIG. 19D shows a system diagram of an example RAN and core network;
- [0042] FIG. 19E illustrates another example communications system;
- [0043] FIG. 19F is a block diagram of an example apparatus or device, such as a WTRU;
- and
- [0044] FIG. 19G is a block diagram of an exemplary computing system.

DETAILED DESCRIPTION

[0045] Table 0.1 of the Appendix contains explanations of selected abbreviations used here.

[0046] FIG. 1 shows an example architecture for ADAE 50.

[0047] FIG. 2 shows an example ADAE internal function architecture 200.

[0048] Both NWDAF and ADAES support a data collection coordination function (DCCF) and an analytics data repository function (ADRF).

[0049] The DCCF coordinates the collection and distribution of data requested by the consumer or the analytics services. It prevents data sources from having to handle multiple subscriptions for the same data and send multiple notifications containing the same information due to uncoordinated requests from data consumers. The analytics services may send requests for data to the corresponding DCCF functions rather than directly to the data sources. DCCF may also perform data processing and data preparation based on the service requirements.

[0050] The ADRF stores historical data and/or analytics, e.g., data and/or analytics related to a past time period that has been obtained by the consumer or the analytics services. After data and/or analytics information is obtained, it may be stored in the ADRF either directly or via the DCCF.

[0051] In addition, NWDAF also defines the following functions that support AI/ML-enabled analytics services.

[0052] An NWDAF instance may contain analytics logical function (AnLF), which performs inferencing, derives analytics information based on the consumer's request and exposes analytics services.

[0053] An NWDAF instance may contain model training logical function (MTLF), which trains ML models and exposes new training services (e.g., providing trained ML model). The trained model may be stored in the ADRF.

[0054] Discovery and selection: The capabilities of a NWDAF instance are described in the NWDAF profile stored in the NRF. The NWDAF profile may include information such as supported analytics IDs/types, service area, capabilities (analytics aggregation, accuracy checking, federated learning, roaming, etc.), supported data source types or identifiers, etc. A NWDAF service consumer may select an NWDAF that supports requested analytics information and required analytics capabilities and/or requested ML model information by using the NWDAF

discovery procedure. The discovery filters may include information of the NWDAF profile, model information, etc.

[0055] Accuracy monitoring: A NWDAF has the accuracy checking capability of analytics and/or ML Model, where the NWDAF may provide the accuracy information to consumers when requested or use it for its internal processes. Analytics/ML model accuracy monitoring is to be achieved by comparing the predictions using the current trained ML model and its corresponding ground truth data e.g., the corresponding true observed events.

[0056] Federated learning: Federated Learning may be supported by multiple NWDAFs containing MTLF, where there is one NWDAF containing MTLF acting as a FL server and multiple NWDAFs containing MTLF acting as FL clients.

[0057] 3GPP SA6 has defined application data analytics enablement (ADAE) service to provide analytics service in the application and service enablement layer. Although NWDAF has defined several functions to support AI/ML-enabled analytics services in the core network, such functions are not yet available in the application and service enablement layer or the ADAE service. Functions to further support AI/ML-enabled analytics services are yet to be defined. Particularly, there is a lack of method for the ADAE service to manage the AI/ML resources and capabilities, or to coordinate the training and inferencing processes from multiple AI/ML service requests.

[0058] The ADAE service needs to be enhanced to support AI/ML-enabled analytics services, particularly, the training and inferencing functions. However, there are several challenges when integrating AI/ML enabling functions to the application and service enablement layer.

[0059] UE/Edge-based training and inferencing are described herein. For AI/ML services, usually it is beneficial to have the training and inferencing procedures (e.g., or part of the training and inferencing procedures) performed closer to the data sources to minimize data transmission overhead and response time. Many of the analytics services supported by ADAE rely on data collected from the UE and/or data collected from edge data networks (EDNs), such as application performance analytics, UE-to-UE application performance analytics, edge load analytics, etc. Therefore, it is desirable that the AI/ML-enabled analytics services (especially the training and inferencing procedures) can be supported on UE-based (e.g., ADAE client) or edge-based ADAE service (e.g., ADAE server implemented in the EDN).

[0060] Unlike analytics services that are provided by servers in the core network or the cloud, UE or edge-based ADAE service may have limited access to computing resources which are needed for performing ML training or inferencing process. For example, a UE-based ADAE service (e.g., an ADAE client) may not have access to computing resources at all times due to competing tasks running on the UE being given higher priority or the available battery level being low. The availability or capability of an edge-based ADAE service may be dependent on the status of the EDN or the corresponding edge enabler server (EES). Due to the dynamicity of resources and capabilities, the optimal configuration of the training/inferencing process may need to be changed adaptively, such as to determine which entity should be selected for training/inferencing and whether/when to switch to another entity so that the training/inferencing process may be performed successfully and efficiently.

[0061] Currently, there is a lack of a method for the ADAE service to manage the dynamic information of the resource and capabilities, and adaptively configure the training/inferencing procedures to support AI/ML-enabled analytics services in the application and service enablement layer.

[0062] Coordination of AI/ML procedures is described herein. Analytics service requests from different consumers may be requested for similar analytics results or services. In this case, coordination may be required to avoid unnecessary processing or messaging, especially for cases where the resource and capabilities supporting analytics services are limited.

[0063] Currently, a coordination function (e.g., DCCF) has been defined for the data collection process, which may support the coordination of analytics requests targeting only analytics data. However, AI/ML-enabled analytics service may further involve the training and inferencing stages. As a result, a coordination function is further needed for the training and inferencing procedures.

[0064] For example, the ADAE service may receive edge load prediction requests from different consumers targeting different sub areas of the edge data network. The ADAE service may first train an ML model with historical measurement data collected from the EDN. Then the trained model may be used to generate the prediction of edge load based on the current measurement data of the EDN. Although the prediction results for different sub-areas may vary, the ML model used to generate the prediction may be the same.

[0065] Without coordination, the ADAE service may process each request separately and configure a training process for each sub-area, resulting in repeated training procedures that are essentially training the same model. In existing 3GPP defined analytics services (e.g., NWDAF and ADAE), only data collection coordination function is supported, while coordination function beyond data collection (e.g., for training and inferencing process) has yet to be defined.

[0066] Application Enabler Layer (Non-ADAE) resources and capabilities are described herein. Other than the ADAE service, entities in the application and service enablement layers (e.g., including the consumers of analytics services) may be able to provide local resources and capabilities to support a portion of the AI/ML process, such as entities from the EEL, SEAL, SEALDD, CAPIF, etc. For example, a SEAL server or an EES may be able to perform the inferencing process after the model is trained by the ADAE service. A consumer may initiate an AI/ML process locally and then request the ADAE service to complete the process. For example, a VAL server may train a ML model and request the ADAE service to assist in deploying the model to the UE and instructing the UE to perform the inferencing process with data collected from the application and service enablement layer. Currently, there is a lack of function to support and integrate non-ADAE AI/ML resources and capabilities to enable analytics services in the application and service enablement layer.

[0067] According to some examples, an AI/ML enablement function may enhance or interact with the ADAE service to support AI/ML-enabled analytics services in the application and service enablement layers.

[0068] FIG. 3 shows an example ADAE general architecture 300 of the enhanced ADAE service which may interact with the vertical application layer and the other application/service enablement layer services.

[0069] FIG. 4 shows an example ADAE internal architecture 400 of the enhanced ADAE service. According to some examples, the AI/ML enablement function may be implemented as an internal function within the ADAE service or as a service external to ADAE that may interact with the ADAE service (similar to A-DCCF or A-ADRF). Moreover, applications described herein may include all types of verticals, such as V2X, UAS, etc.

[0070] An AI/ML service consumer may request AI/ML-enabled analytics service from the ADAE service, or may request a particular AI/ML resource or capability that is managed by the ADAE service. The consumer may be an entity from the vertical application layer (e.g., VAL

server, VAL client), an entity from an application and service enablement layer (e.g., SEAL server, EES, ADAE client/server, A-DCCF, A-ADRF), or an entity from the core network (e.g., an NWDAF instance).

[0071] An AI/ML resource/capability provider may provide the major resources and/or capabilities that are needed to support the AI/ML-enabled analytics service. The AI/ML consumer may also be the resource/capability provider. A general AI/ML process may consist of procedures such as the collection of training and inferencing data, training AI/ML models with the training data, performing inferencing with the trained models, performance monitoring and model updates, etc. In this process, the AI/ML data, AI/ML model, training and inferencing entities may be managed by the ADAE service.

[0072] AI/ML data may be the data to be used for the training and/or inferencing procedures in an AI/ML process. The AI/ML data may be managed directly by the ADAE service or via the A-DCCF/A-ADRF functions. The data may be provided by an entity in the application and service enablement layer (e.g., SEAL, EEL, ADAE), by a VAL entity (VAL client or VAL server), or from the core network (e.g., NWDAF, OAM).

[0073] AI/ML model may be the model trained with AI/ML algorithms, which may be used for the inferencing procedure by an analytics service to generate inferencing/analytics results. The AI/ML models may be managed directly by the ADAE service or via the A-ADRF function. The model may be provided by an entity with training capability such as an analytics service entity (ADAE, NWDAF) or a VAL entity (VAL client or VAL server).

[0074] A training entity may be an entity with training function/capability that is capable of performing training tasks and generating or updating AI/ML models for AI/ML-enabled analytics services. The training entity may take training data as input and generate one or more trained/updated models as output. Entities that are capable of aggregating multiple partially trained models may also be viewed as a training entity, such as an FL server.

[0075] An inferencing entity may be an entity with inferencing function/capability that is capable of performing inferencing tasks and generating analytics results for AI/ML-enabled analytics services. The inferencing entity may take one or more trained models and inferencing data (if any) as input and generate inferencing/analytics results (e.g., predictions) as output.

[0076] Training and inferencing capabilities may be provided by ADAE-native entities or other entities from the application and service enablement layer, the vertical application layer,

or the core network. An ADAE server may function as a training entity, an inferencing entity, or both. Particularly, if the ADAE server is deployed to an EDN (e.g., as an EAS or edge-based ADAE service), there may be constraints on the availability of the capabilities due to factors such as EDN schedule, EDN load, limitation of service area/coverage, etc. An ADAE client at the UE (or UE-based ADAE service) may function as a training or inferencing entity. For example, a trained model may be deployed to the UE and the UE-based ADAE service may perform the inferencing task locally to reduce the latency in generating analytics results for the UE. There may be constraints on the availability of the capabilities due to factors such as UE's capability, UE's load, user's authorization/consent on the usage of its resources, etc. An application and service enablement layer entity (e.g., SEAL server, EES) may provide training or inferencing entity to support the ADAE service. For example, an EES may function as an inferencing entity to support edge-related analytics. A VAL server may provide training or inferencing capability to support the ADAE service. For example, an AI/ML application may provide trained model or training capability to the ADAE service. NWDAF may provide AI/ML resources and capabilities to the ADAE service, such as contributing training data (with DCCF or ADRF functions) or training capability (with NWDAF instance containing MTLF).

[0077] According to some examples, the AI/ML enablement function may enhance the ADAE service to support AI/ML-enabled analytics services. The enhanced ADAE service may provide management service of the AI/ML resources and capabilities in the application and service enablement layer. The enhanced ADAE service may further provide coordination service for the AI/ML processes, where the requests from different AI/ML service consumers are coordinated so that the AI/ML resources/capabilities or the training/inferencing procedures may be shared or re-used across different service instances. FIG. 5 and FIG. 6 show the overview of the AI/ML-enabled ADAE service with and without external AI/ML capability provider, respectively.

[0078] FIG. 5 shows an example AI/ML-enabled ADAE service overview associated with native AI/ML capability 500. As shown at step 1, the ADAE service may provide information of the available AI/ML resources and capabilities (e.g., such as AI/ML data and model at the A-ADRF, training and inferencing capabilities supported by an ADAE server/client) to a consumer via the discovery or exposure service. As shown at step 2, the ADAE service may provide data collection service (e.g., with A-DCCF and/or A-ADRF) to the consumer, which may generate a training dataset to be used for training a model. As shown at step 3, the ADAE service may provide

AI/ML training service to the consumer with the training capabilities available at an ADAE server/client, which may generate a trained model to be used for inferencing and generating inferencing results. As shown at step 4, the ADAE service may provide AI/ML inferencing service to the consumer with the inferencing capabilities available at an ADAE server/client, which may generate analytics results for the consumer. As shown at step 5, the ADAE service may provide coordination service to monitor and manage the AI/ML process and coordinate requests from multiple consumers.

[0079] FIG. 6 shows an example AI/ML-enabled ADAE service overview associated with external AI/ML capability 600. The ADAE service may further provide management service for external/non-ADAE AI/ML resources and capabilities. As shown at step 1, the ADAE service may obtain and maintain information for the AI/ML resource and capabilities from external providers via a registration service and may share/expose the information to the consumers (step 2). The ADAE service may support data collection (step 3), training (step 4) and inferencing (step 5) process that are performed by ADAE or non-ADAE entities. As shown at step 6, a coordination service may be provided.

[0080] AI/ML Resource and Capability Management is described herein. AI/ML resources may include AI/ML data and AI/ML models. AI/ML capabilities may include AI/ML inferencing capability and training capability. Such resources and capabilities may be used by the AI/ML-enabled analytics service to generate analytics results. The resources and capabilities may be provided by a resource/capability provider (e.g., SEAL entities, EEL entities, ADAE entities, VAL entities, NWDAF) and consumed by a resource/capability consumer (e.g., ADAE entities, VAL entities, NWDAF). The following management service and procedure are not limited to AI/ML-enabled analytics services. They may also be applied to managing other types of analytics resources and capabilities.

[0081] AI/ML resource and capability profiles are described herein. The ADAE service may maintain AI/ML data profiles of data resources that are available for AI/ML services. The ADAE service may use the AI/ML data profiles to determine what data may be used for the training process, instruct the inferencing entity to obtain inferencing data, or record inferencing/analytics results for performance monitoring. An AI/ML data profile may include information as shown in Table 1.

Table 1. AI/ML Data Profile

Information Element	Description
Profile Identifier	Identifier of the AI/ML data profile.
Data Identifier	Identifier of the AI/ML data instance (e.g., a dataset or a data stream) defined in this profile.
Data Description	Description of the data, such as “EAS/EES load,” “VAL session performance,” “UE-to-UE session performance,” “number of failure API invocations,” etc. The description may also include the associated analytics ID, such as “VAL session performance analytics,” “slice-specific application performance analytics,” “service API analytics,” etc.
Data Format	Information of the format of the data, which may be used for training data preparation and processing.
Data Purpose	Description of what the data is used for, such as “training data” (data used for training the model), “inferencing data” (input data of the inferencing process), “inferencing/analytics results,” etc. This information may be updated when the data purpose changes. For example, a data instance may be used as inferencing data first, and then used as training data. The data purpose may also be specified in Service Instance Information.
Data Characteristics	Information of the characteristics of the data.
Data Location	The location or address information of the data, such as where the data is stored or where the data can be accessed. For example, the data may be stored at the A-ADRF, or at the corresponding data sources such as a VAL server, an EAS, NWDAF, CAPIF core function, etc.
Data Source	Information of the data source(s) which generated or will generate the data.
> Data Source Identifier	Identifier of the data source.
> Data Source Profile	A profile with descriptions of the data source or an identifier of the profile (e.g., data producer profile).
Data Access Policy	Defines which entities are authorized to access the data and what actions can be performed by the authorized entities.
Service Instance Information	Information of the AI/ML service instances that are using or will be using this data instance.
> Service Instance Identifier	Identifier of the service instance.
> Data Purpose	Description of what the data is used for in this service instance, such as “training data,” “inferencing data,” “inferencing/analytics results,” etc.

[0082] The ADAE service may maintain AI/ML model profiles of trained models that are available for AI/ML services. The ADAE service may use the AI/ML model profiles to

determine if an existing model can be used for an AI/ML service request, if a model can be deployed to an inferencing entity (e.g., if the inferencing entity can support inferencing process with the model), or track the performance of the model to trigger update. An AI/ML model profile may include information as shown in Table 2.

Table 2. AI/ML Model Profile

Information Element	Description
Profile Identifier	Identifier of the AI/ML model profile.
Model Identifier	Identifier of the AI/ML model.
Model Description	Description of the model, such as “model for predicting VAL/edge server load,” “model for predicting UE-to-UE session performance,” “model for ACR (service continuity) detection in EDN,” “model for predicting API availability/failure,” “model for predicting slice usage pattern,” etc.
Model Requirements	Defines requirements of using the model to perform inferencing process, such as the required computing power, or a particular EDN or area of interest where the model is applicable.
Model Location	The location or address information of the model, such as where the model is stored or where the model can be accessed.
Training Information	Context information regarding how the model is trained. For example, if the model is trained with data from a particular EDN or area of interest and is only applicable to this EDN/area, then information of the EDN/area may be specified.
Training Entity Identifier	Identifier of the one or more entities where the model is trained/updated, such as the ADAE server or client ID. If the model is trained in a distributed manner (e.g., federated learning), the identifier of the aggregator or coordinator may be specified.
Training Data Information	Information of the training data with which the model is trained, such as the AI/ML data profile(s) (e.g., the identifier of the AI/ML data profile(s)) associated with the training data.
Training Performance	Information of the performance of the model during the training process, such as the precision, accuracy, processing time, etc.
Model Access Policy	Defines which entities are authorized to access the model and what actions can be performed by the authorized entities.
Service Instance Information	Information of the AI/ML service instances that are using or will be using this model for inferencing.
> Service Instance Identifier	Identifier of the service instance.
> Inferencing Performance	Information of the performance of the model during the inferencing process of this service instance, such as the precision, accuracy, processing time, etc.

[0083] The ADAE service may maintain AI/ML training/inferencing capability profiles of training/inferencing entities that are available for AI/ML services. The ADAE service may use the training/inferencing capability profiles to determine which entity is capable of performing training/inferencing process, monitor the training/inferencing process being performed on a training/inferencing entity, determine if a training/inferencing process can be shared or re-used by a different AI/ML service request, etc. An AI/ML training or inferencing profile may include information as shown in Table 3 and Table 4, respectively. For an ADAE server/client that contains AI/ML training and/or inferencing capabilities, an ADAE profile may be defined which contains information from the corresponding training and inferencing capability profiles.

Table 3. AI/ML Training Capability Profile (for Training Entity)

Information Element	Description
Profile Identifier	Identifier of the AI/ML training capability profile.
Training Entity	Identifier and/or other information of the training entity that is providing the training capabilities defined in this profile.
Endpoint Information	Endpoint information (e.g., URI, IP address) used to communicate with the training entity.
Entity Type	Description of the entity providing the training capability (training entity), such as an ADAE server, an ADAE client, an edge-based ADAE server (EAS), etc.
Capability Description	Description of the capabilities and supported functionalities/features. - Supported AI/ML models: limitations of the size/algorithm of the model to be trained - Supported features: distributed training, federated learning, model splitting, service continuity, fine-tuning, etc.
Supporting Resource	Resources available to support the training capabilities.
Capability Schedule	Schedule of when the capability is available.
Capability Access Policy	Defines which entities are authorized to access the capability and what actions can be performed by the authorized entities.
Service KPIs	Service characteristics provided by the training entity.
Service Instance Information	Information of the AI/ML service instances that are served by this training entity
> Service Instance Identifier	Identifier of the service instance.
> Notification Target	Information of the notification target(s) associated with this service instance, e.g., where to send the output (trained model) of the training process.
> Training Data Information	Information of the training data to be used for the training process, e.g., where to get the input data of the training process.

> Configuration Information	Information of the configurations of the training entity for this service instance.
-----------------------------	---

Table 4. AI/ML Inferencing Capability Profile (for Inferencing Entity)

Information Element	Description
Inferencing Entity	Identifier and/or other information of the inferencing entity that is providing the inferencing capabilities defined in this profile.
Endpoint Information	Endpoint information (e.g., URI, IP address) used to communicate with the inferencing entity.
Entity Type	Description of the entity providing the inferencing capability (inferencing entity), such as an ADAE server, an ADAE client, an edge-based ADAE server (EAS), etc.
Capability Description	Description of the capabilities and supported functionalities/features. - Supported AI/ML models: limitations of the size/algorithm of the model to be used for inferencing - Supported features: support for performance monitoring, drift detection, collaborative inferencing, service continuity, etc.
Supporting Resource	Resources available to support the inferencing capabilities.
Capability Schedule	Schedule of when the capability is available.
Capability Access Policy	Defines which entities are authorized to access the capability and what actions can be performed by the authorized entities.
Service KPIs	Service characteristics provided by the inferencing entity.
Service Instance Information	Information of the AI/ML service instances that are served by this inferencing entity
> Service Instance Identifier	Identifier of the service instance.
> Notification Target	Information of the notification target(s) associated with this service instance, e.g., where to send the output (inferencing/analytics results) of the inferencing process.
> Inferencing Data	Information of the inferencing data to be used for the inferencing process, e.g., where to obtain the input data of the inferencing process.
> Model Information	Information of the trained model to be used for the inferencing process, e.g., where to obtain the input model of the inferencing process.
> Configuration Information	Information of the configurations of the inferencing entity for this service instance.

[0084] AI/ML resource and capability registration is described herein. An AI/ML resource or capability provider may provide information of its resource/capability to the ADAE

with a registration procedure. The registration procedure allows an AI/ML resource or capability provider to provide information of its resource/capability that it is willing to contribute to the ADAE service in order to enable the discovery of the resources/capabilities. The registration procedure also allows the ADAE service to be aware of the AI/ML data even when the ADAE service itself does not possess the data. If there is a change in the information of the resource or capability, the provider may use the registration update procedure to update the ADAE service. The provider may use the de-registration procedure to remove information from the ADAE service.

[0085] Alternatively, the ADAE service may send a query/request to the provider to obtain information of its resource/capability, and then create/update the corresponding resource/capability profile. For example, the ADAE service may request information from a UE or EES which may function as a training or inferencing entity and create/update the training/inferencing capability profile to include the information.

[0086] If the provider is an ADAE service, the ADAE service may create the corresponding profiles to maintain information of the resources/capabilities and enable the discovery of the resources/capabilities.

[0087] FIG. 7 shows an example AI/ML data registration procedure 700 where an AI/ML data provider registers its data with the ADAE service. In this scenario, the ADAE service may be implemented as an ADAE server (as shown in the figure) or ADAE client, A-DCCF, or A-ADRF.

[0088] At step 1, an AI/ML data provider may send an AI/ML data registration request to the ADAE server. The request may include information of the data, such as the AI/ML data profile or information in Table 1. Note the registration request may be an initial registration or an update to or deletion of an existing registration.

[0089] At step 2, The ADAE server may perform an authorization check to verify whether the provider has the authorization to register its data. Upon successful authorization, the ADAE server may create/update/delete the AI/ML data profile based on the information provided in the registration request, or send a request to the A-DCCF or A-ADRF to create/update/delete the AI/ML data profile.

[0090] At step 3, if the ADAE server requests the A-DCCF or A-ADRF to create/update the AI/ML data profile, the ADAE server may receive a response from the A-DCCF or A-ADRF.

The response may include the identifier of the AI/ML data profile. For deletion requests, the A-DCCF or A-ADRF may simply acknowledge the deletion of the AI/ML data profile.

[0091] At step 4, the ADAE server may send a registration response to the AI/ML data provider. The response may include the identifier of the AI/ML data profile if it is created in step 2 or an acknowledgement of the update to or deletion of the AI/ML data profile.

[0092] At step 5, the AI/ML data may be transferred from the provider to the A-ADRF. The ADAE service may initiate the data transfer by sending a request to the provider and update the “data location” in the AI/ML data profile to be the address associated with the A-ADRF.

[0093] The provider may register multiple data instances to the ADAE service with one single registration or with separated registrations.

[0094] FIG. 8 shows an example AI/ML model registration procedure 800 where an AI/ML model provider registers its model with an ADAE service. In this scenario, the ADAE service may be implemented as an ADAE server (as shown in FIG. 8), as an ADAE client, or as an A-ADRF.

[0095] At step 1, the AI/ML model provider may send an AI/ML model registration request to the ADAE server. The request may include information of the model, such as the AI/ML model profile or information in Table 2. Note the registration request may be an initial registration or an update to or deletion of an existing registration.

[0096] At step 2, the ADAE server may perform an authorization check to verify whether the provider has the authorization to register its model. Upon successful authorization, the ADAE server may create/update/delete the AI/ML model profile based on the information provided in the registration request or send a request to the A-ADRF to create/update/delete the AI/ML model profile.

[0097] At step 3, if the ADAE server requests the A-ADRF to create/update the AI/ML model profile, the ADAE server may receive a response from the A-ADRF. The response may include the identifier of the AI/ML model profile. For deletion requests, the A-DCCF or A-ADRF may simply acknowledge the deletion of the AI/ML model profile.

[0098] At step 4, the ADAE server may send a registration response to the AI/ML model provider. The response may include the identifier of the AI/ML model profile if it is created in step 2 or an acknowledgement of the update to or deletion of the AI/ML model profile.

[0099] At step 5, the AI/ML model may be transferred from the provider to the A-ADRF. The ADAE service may initiate the model transfer by sending a request to the provider and update the “model location” in the AI/ML data profile to be the address associated with the A-ADRF.

[00100] The provider may register multiple models to the ADAE service with one single registration or with separated registrations.

[00101] FIG. 9 shows an example AI/ML training/inferencing capability registration procedure 900 where an AI/ML training/inferencing capability provider registers its capability with the ADAE service. In this scenario, the ADAE service may be implemented as an ADAE server (as shown in FIG. 9) or ADAE client.

[00102] At step 1, the AI/ML capability provider may send an AI/ML capability registration request to the ADAE server. The request may include information of the capability, such as the AI/ML training/inferencing capability profile or information in Table 4 and/or Table 3. Note the registration request may be an initial registration or an update to or deletion of an existing registration.

[00103] At step 2, the ADAE server may perform an authorization check to verify whether the provider has the authorization to register its capability. Upon successful authorization, the ADAE server may create/update/delete the AI/ML training/inferencing capability profile.

[00104] At step 3, the ADAE server may send a registration response to the AI/ML capability provider. The response may include the identifier of the AI/ML training/inferencing capability profile if it is created in step 2 or an acknowledgement of the update to or deletion of the AI/ML model profile.

[00105] AI/ML resource and capability discovery is described herein. An AI/ML resource and/or capability consumer may discover information of required/desired AI/ML resource/capability from the ADAE service with a discovery procedure. Discovery procedures enable AI/ML service consumers to obtain information about AI/ML resources or capabilities. The discovery is based on matching discovery filters provided in the discovery request.

[00106] Alternatively, the consumer may subscribe to the ADAE service for information of AI/ML resource/capability that is of interest, where the ADAE service may send a notification to the consumer when the required resource/capability is available or a change of the resource/capability is detected.

[00107] FIG. 10 shows an example AI/ML data discovery procedure 1000 where a consumer discovers AI/ML data from the ADAE service. In this scenario, the ADAE service may be implemented as an ADAE server (as shown in FIG. 10) or ADAE client, A-DCCF or A-ADRF.

[00108] At step 1, the AI/ML data consumer may send an AI/ML data discovery request to the ADAE server. The discovery request may include the requestor (consumer) identifier along with the security credentials and may include AI/ML data discovery filters, such as data identifier, data descriptions, data purpose, data characteristics, data source, etc.

[00109] At step 2, upon receiving the discovery request from the consumer, the ADAE server may check if the consumer is authorized to discover the requested AI/ML data. If the AI/ML data profiles are maintained at the ADAE server, the ADAE server may process the discovery request and apply discovery filters to identify the required AI/ML data. If the AI/ML data profiles are maintained at the A-DCCF or A-ADRF function, the ADAE server may send a request to the A-DCCF or A-ADRF to retrieve the profiles (of the AI/ML data that is of interest to the consumer by including the discovery filters in the request).

[00110] At step 3, if the ADAE server requests AI/ML data profiles from the A-DCCF or A-ADRF, the ADAE server may receive a response from the A-DCCF or A-ADRF with the requested AI/ML data profiles.

[00111] At step 4, if the processing of the request was successful, the ADAE server may send an AI/ML data discovery response to the consumer, which may include information about the discovered AI/ML data such as the profiles of the data.

[00112] At step 5, the discovered AI/ML data may be transferred from the A-ADRF to the consumer. The ADAE service may initiate the data transfer by sending a request to the A-ADRF and adding the address information associated the consumer to the “data location” in the AI/ML data profile.

[00113] FIG. 11 shows an example AI/ML model discovery procedure 1100 where a consumer discovers an AI/ML model from the ADAE service. In this scenario, the ADAE service may be implemented as an ADAE server (as shown in FIG. 11) or ADAE client, or A-ADRF.

[00114] At step 1, the AI/ML model consumer may send an AI/ML model discovery request to the ADAE server. The discovery request may include the requestor (consumer) identifier along with the security credentials and may include AI/ML model discovery filters, such as model identifier, model description, model requirements, training performance, etc.

[00115] At step 2, upon receiving the discovery request from the consumer, the ADAE server may check if the consumer is authorized to discover the requested AI/ML model. If the AI/ML model profiles are maintained at the ADAE server, the ADAE server may process the discovery request and apply discovery filters to identify the required AI/ML model(s). If the AI/ML model profiles are maintained at the A-ADRF function, the ADAE server may send a request to the A-ADRF to retrieve the profiles (of the AI/ML model that is of interest to the consumer by including the discovery filters in the request).

[00116] At step 3, if the ADAE server may request AI/ML model profiles from the A-ADRF, the ADAE server may receive a response from the A-ADRF with the requested AI/ML model profiles.

[00117] At step 4, if the processing of the request was successful, the ADAE server may send an AI/ML model discovery response to the consumer, which may include information about the discovered AI/ML model(s) such as the profiles of the model(s).

[00118] At step 5, the discovered AI/ML model(s) may be transferred from the A-ADRF to the consumer. The ADAE service may initiate the model transfer by sending a request to the A-ADRF and adding the address information associated the consumer to the “model location” in the AI/ML model profile.

[00119] FIG. 12 shows an example AI/ML capability discovery procedure 1200 where a consumer discovers AI/ML training/inferencing capability with the ADAE service. In this scenario, the ADAE service may be implemented as an ADAE server (as shown in FIG. 12) or ADAE client.

[00120] At step 1, the AI/ML capability consumer may send an AI/ML capability discovery request to the ADAE server. The discovery request may include the requestor (e.g., consumer) identifier along with the security credentials and may include AI/ML capability discovery filters, such as capability type (e.g., training or inferencing or both), entity type, capability description, capability schedule, service KPIs, etc.

[00121] At step 2, upon receiving the discovery request from the consumer, the ADAE server may check if the consumer is authorized to discovery the requested AI/ML capability. The ADAE server may process the discovery request and apply discovery filters to identify the required AI/ML capabilities (e.g., training entities or inferencing entities).

[00122] At step 3, if the processing of the request was successful, the ADAE server may send an AI/ML capability discovery response to the consumer, which includes information about the discovered AI/ML capabilities such as the profiles of the capabilities or information of the training/inferencing entities.

[00123] AI/ML-Enabled Analytics Service Coordination is described herein. After receiving an AI/ML-enabled analytics service request, the ADAE may create an AI/ML service instance to maintain information of the request, monitor the AI/ML process (e.g., including training and inferencing), and enable coordination among different instances or requests. Information of the profile may be provided by the consumer or configured by the ADAE service.

[00124] Information of the AI/ML service instance may be described in an AI/ML service instance profile as shown in Table 5.

Table 5. AI/ML Service Instance Profile

Information Element	Description
Profile Identifier	Identifier of the AI/ML service instance profile.
Instance Identifier	Identifier of the AI/ML service instance.
Consumer Identifier	Identifier of the service consumer who initiates the AI/ML service request.
Consumer Context Information	Context information of the service consumer, e.g., location, schedule, mobility, etc.
Request Identifier	Identifier of the AI/ML service request associated with this instance.
Notification Setting	Specify when and where to send a notification regarding events associated with the AI/ML process.
> Notification Event	Specify the event or condition of when to send a notification, such as “training data collection completed,” “model training completed,” “inferencing entity selected,” etc.
> Notification Target	Specify information of the notification target(s), such as the identifier(s) of the notification target(s) or an address of the notification target.
Analytics Type	Information of the analytics type or analytics ID of the requested AI/ML service, such as “VAL server performance analytics,” “VAL session performance analytics,” “edge load analytics,” etc.
Service Objective	Descriptions of the objective(s) of the AI/ML service, such as “training data collection only,” “model training only,” “inferencing only,” “complete AI/ML process,” etc.
Service Requirements	Specify requirements for the AI/ML resource/capability associated with this service request and/or the performance of the service, such as requirements of training data, requirements of

	model, requirements of training and inferencing entities, required performance indicators.
Service Schedule	Specify schedule information of the AI/ML service instance and the procedures, such as when to start/perform training or inferencing. The schedule may be specified by the consumer or be configured by the ADAE service according to the availability/schedule of the AI/ML resources and capabilities.
Instance Status	Control and/or indicator of the current status of the instance or service progress, such as: • Collecting training data • Training data collection completed or training data ready • Training model • Model training completed or model ready • Performing (online) inferencing
Training Data Information	Information of the training data associated with this instance, such as the corresponding AI/ML data profile(s) (e.g., the identifier of the AI/ML data profile(s)). The training data may be specified by the consumer or be determined by the ADAE service according to the service requirements and availability/schedule of the AI/ML resources and capabilities.
Inferencing Data Information	Information of the inferencing data associated with this instance, such as the corresponding AI/ML data profile(s) (e.g., the identifier of the corresponding AI/ML data profile(s)). The inferencing data may be specified by the consumer or be determined by the ADAE service according to the service requirements and availability/schedule of the AI/ML resources and capabilities.
Model Information	Information of the trained model(s) associated with this instance, such as the corresponding AI/ML model profile(s) (e.g., the identifier of the corresponding AI/ML model profile(s)). The model may be specified by the consumer or be determined by the ADAE service according to the service requirements and availability/schedule of the AI/ML resources and capabilities.
Training Entity Information	Information of the one or more training entities associated with this instance, such as the corresponding AI/ML training capability profile (e.g., the identifier of the corresponding AI/ML training capability profile), and the configuration information of each training entity. The training entity may be specified by the consumer or be determined by the ADAE service according to the service requirements and availability/schedule of the AI/ML resources and capabilities.
Inferencing Entity Information	Information of the one or more inferencing entities associated with this instance, such as the corresponding AI/ML inferencing capability profile (e.g., the identifier of the corresponding AI/ML

	inferencing capability profile), and the configuration information of each inferencing entity. The inferencing entity may be specified by the consumer or be determined by the ADAE service according to the service requirements and availability/schedule of the AI/ML resources and capabilities.
Analytics results and performance information	Information of the analytics results and/or the performance of the analytics service.

[00125] FIG. 13 shows an example AI/ML service coordination process 1300 illustrating the coordination process performed by the ADAE service when receiving a new AI/ML service request.

[00126] Start of FIG. 13: After receiving an AI/ML service request, the ADAE service may create a service instance for the request and start the coordination process. The ADAE service may check the existing AI/ML resources/capabilities and service instances by examining the corresponding profiles maintained at the ADAE service.

[00127] Decision 1 of FIG. 13: The ADAE service may first determine if there is any existing inferencing process (e.g., which may be associated with a service instance serving a previously received request) that may fulfil the request. This may be done by examining the profiles of the existing service instances that are in the status of “performing inferencing” and comparing the service objectives and service requirements. Depending on the result, the ADAE service may take the following actions:

[00128] If an existing inferencing process can be found, the ADAE service may associate the existing process with the new service instance by adding the new service instance to the capability profile of the inferencing entity and updating the corresponding notification target. Information of the inferencing process may be added to the profile of the new service instance, such as the inferencing entity information, model information, etc. If inferencing data is required, the ADAE service may instruct the inferencing entity to obtain the inferencing data by specifying the inferencing data in the service instance profile or inferencing capability profile. The ADAE service may then update the status of the new service instance to “performing inferencing.”

[00129] If an existing inferencing process can be found, but it is unable to support the new service instance (e.g., due to insufficient computing power, unable to access the inferencing data, etc.) or the ADAE service may determine the current inferencing entity is not the optimal

choice, then the ADAE service may select a new inferencing entity or re-select an inferencing entity for the inferencing process.

[00130] For example, a UE may request for analytics information related to the EDN it is in. There is already an inferencing entity (e.g., edge-based ADAE server) that is generating analytics information for the EDN that the UE is currently in, but the UE is soon to leave the current EDN and move to another EDN. In this case, the ADAE service may determine to select an inferencing entity that is located in the new EDN or select a UE-based inferencing entity for the inferencing process.

[00131] In another example, a UE may request for analytics information related to the EDN it is in and the inferencing process is performed at the UE locally (e.g., with a UE-based inferencing entity). A second UE then requests for the same analytics information, but is unable to obtain the analytics results generated at the first UE. In this case, the ADAE service may re-select an edge-based ADAE server as the inferencing entity to serve both UEs.

[00132] If there is no existing inferencing process that can provide the required service, the ADAE service may continue to the next coordination action.

[00133] Decision 2 of FIG. 13: If no existing inferencing process can be found, the ADAE service may determine if there is any existing AI/ML model(s) that has been trained and may be used to generate the required analytics results for the new service instance by examining the AI/ML model profiles or querying the A-ADRF. Alternatively, the consumer may specify the model(s) to be used in the service request. Depending on the result, the ADAE service may take the following actions:

[00134] If an existing model(s) can be found, the ADAE service may associate the model(s) with the new service instance by updating the instance profile with the information of the found model(s). Then the ADAE service may proceed to identify an inferencing entity for performing the inferencing process with the model.

[00135] If there is no existing model that can be used for inferencing, the ADAE service may continue to the next coordination action.

[00136] Decision 3 of FIG. 13: If no trained model can be found or used to generate the required analytics results for the new service instance, the ADAE service may determine if there is any ongoing training process (e.g., which may be associated with a service instance serving a previously received request) that can generate the required model and fulfil the request. This may

be done by examining the profiles of the existing service instances that are in the status of “training model” and comparing the service objectives and service requirements. Depending on the result, the ADAE service may take the following actions:

[00137] If an existing training process can be found, the ADAE service may associate the existing process with the new service instance by adding the new service instance to the capability profile of the training entity and updating the corresponding notification target. Information of the training process may be added to the profile of the new service instance, such as the training entity information, training data information, etc. The ADAE service may then update the status of the new service instance to “training model.”

[00138] If an existing training process can be found, but needs to be updated (e.g., due to new training data that is introduced by the new service instance), then the ADAE service may update the configuration of the training process.

[00139] For example, a UE requests for analytics information related to the EDN it is in and an edge-based ADAE server located in the same EDN is training the model to be used for this request with training data originated from the local EDN. A second UE then requests for analytics information related to a second EDN where the second UE is in, where a model can be trained with data originated from the second EDN. The ADAE service may coordinate the two requests and re-configure the training process to have one common training entity performing the training process with aggregated training data from both EDNs. The trained model can be deployed to both EDNs for inferencing.

[00140] In the same example, the ADAE service may determine to utilize federated learning to train the model so that the training data from both EDNs may contribute to the model without having to be relocated (e.g., transferred to the common training entity). The ADAE may configure a training entity as the FL server and the local training entities in each EDN (e.g., edge-based ADAE server) as FL clients, if the training entities support federated learning.

[00141] If there is no existing training process that can provide the required service, the ADAE service may continue to the next coordination action.

[00142] Decision 4 of FIG. 13: If neither trained model nor ongoing training process can be found, the ADAE service may determine if there is any existing AI/ML data that can be used to train a model for the new service instance by examining the AI/ML data profiles or querying

the A-ADRF. Alternatively, the consumer may specify the training data to be used in the service request. Depending on the result, the ADAE service may take the following actions:

[00143] If existing training data can be found, the ADAE service may associate the data with the new service instance by updating the instance profile with the information of the found data. Then the ADAE service may proceed to identify a training entity for performing the training process with the data.

[00144] If there is no existing data that can be used for training, the ADAE service may identify where/how to collect the training data and initiate the data collection process to generate the required training data.

[00145] In an example, a system, apparatus, or method may be provided for providing an AI/ML enablement service. For example, a system may receive, from one or more AI/ML resource/capability providers, one or more provider messages comprising one or more descriptions of one or more AI/ML resources or capabilities of the one or more AI/ML resource/capability providers. The system may store, by the AI/ML enablement service and based on the one or more provider messages, the one or more descriptions of the one or more AI/ML resources or capabilities. The system may receive, from an AI/ML service consumer, a consumer request comprising an identifier of the AI/ML service consumer, one or more notification conditions and targets, a required analytics type, or one or more service requirements. The system may determine, based on processing information associated with the one or more descriptions of the one or more AI/ML resources or capabilities and the consumer request, a service action comprising selecting an AI/ML resource/capability provider of the one or more AI/ML resource/capability providers to perform AI/ML operations. The system may send, to the selected AI/ML resource/capability provider, a request to perform one or more AI/ML operations. The system may receive, from the selected AI/ML resource/capability provider, a response message indicating that the one or more AI/ML operations have been performed and comprising one or more results of the performed AI/ML operations. The system may send, to the AI/ML service consumer, a notification message comprising the one or more results of the performed AI/ML operations.

[00146] According to some examples, detailed service procedures are elaborated in FIG. 14 and FIG. 15. FIG. 14 shows an example AI/ML service procedure associated with training 1400, and FIG. 15 shows an example AI/ML service procedure associated with inferencing 1500 where a consumer requests for AI/ML-enabled analytics service, including the data collection,

training process, and inferencing process. Based on the service coordination results (e.g., step 2 in both FIG. 14 and FIG. 15), some of the steps may be skipped (e.g., detailed decision flow for skipping is elaborated in FIG. 13). Particularly, if the consumer only requests for training service (e.g., service objective is one or more trained models), the procedure in FIG. 15 may be skipped. If the consumer only requests for inferencing service (e.g., trained model is already available), the procedure in FIG. 14 may be skipped. Note that the AI/ML service consumer could also be the provider for part of the AI/ML resources and capabilities to be used for the requested service.

[00147] Referring to FIG. 14, at step 1, the ADAE service receives an AI/ML service request from a consumer. In the request, the consumer may specify the analytics type, service objective, service requirements, notification setting, etc. The consumer may also specify the resource/capability to be used for the request, such as the training data, trained model, training entity, inferencing entity. The ADAE service may create a service instance for the request and record information associated with the instance in an AI/ML service instance profile (e.g., as shown in Table 5). The request may be sent as a single message or multiple messages.

[00148] At step 2, the ADAE service may perform the service coordination process as described in FIG. 13. Based on the result of service coordination, the ADAE service may determine to skip to one or more of the following steps.

[00149] At step 3, if the ADAE service determines in step 2 that the training data is yet to be collected or needs to be updated/augmented, then the ADAE service may determine the data provider and initiate the training data collection process. The ADAE service may send a data collection request to the A-DCCF function to collect training data, where the A-DCCF may determine whether to use historical data (e.g., if available) or to collect new training data. Alternatively, the data collection process may be performed by the ADAE service and an AI/ML data profile may be created by the ADAE service for the collected data.

[00150] At step 4, the A-DCCF may first check whether the A-ADRF has historical training data that meets the requirement of the training request. If historical data is not available, the A-DCCF performs data collection process to collect the required training data (e.g., interactions between the A-DCCF and the data sources are not shown in FIG. 14). When data collection is completed, the A-DCCF may register the collected data to A-ADRF and create an AI/ML data profile. The A-DCCF may perform data processing as needed, such as data preparation, data cleaning, data transformation, data aggregation, etc.

[00151] At step 5, the A-DCCF may send a response to the ADAE service. The response may include the AI/ML data profile (e.g., the identifier of the AI/ML data profile) associated with the training data.

[00152] At step 6, after receiving the response about the data collection procedure, the ADAE service may update the status of the service instance to “training data collection completed,” and send a notification to the service consumer or the notification target according to the notification setting specified in the service request. The notification may include the AI/ML data profile (e.g., the identifier of the AI/ML data profile) associated with the training data.

[00153] At step 7, the ADAE service may determine the training entity and the corresponding configuration based on the service requirements in the request and the AI/ML training capability profiles. The ADAE service may select more than one training entity to perform the training process, such as in the scenarios of distributed training, federated learning, model splitting, or due to the schedule/availability of the training entity, etc. The ADAE service may update the selection of training entity when change of the training entity is detected (e.g., the capability profile is updated). If more than one model is to be trained, the ADAE service may determine the training entity for each model or a common training entity for all the models. Alternatively, the training entity may be determined by the consumer (e.g., via AI/ML training capability discovery procedure) and specified in the AI/ML service request.

[00154] At step 8, the ADAE service may send a model training request to the selected training entity and update the status of the service instance to “training model.” In the request, the ADAE service may include information of the training data (e.g., identifier of the AI/ML data profile, data location), information of the model to be trained (e.g., identifier of the AI/ML model profile, model location), and the configuration of the training entity as determined in step 7. The ADAE service may also request the training entity to register the trained model at the ADAE service or the A-ADRF function. Alternatively, the ADAE service may send a request to the determined training entity to update a model or update the training process of a model. In this case, the ADAE service may include information of the model to be updated (e.g., identifier of the AI/ML model profile, model location) in the request. For example, the ADAE service may request the training entity to update a trained model with newly collected training data that is obtained in step 5, or request the training entity to continue the training process that is migrated from another training entity.

[00155] At step 9, the training entity may retrieve the training data and AI/ML model from the A-ADRF by providing the identifiers of the AI/ML data profile and the AI/ML model profile provided by the ADAE service in step 8. Alternatively, the training entity may retrieve the training data/model based on the data/model location information provided by the ADAE service in step 8.

[00156] At step 10, after retrieving the training data and/or AI/ML model, the selected one or more training entities may perform the training process to train or update the AI/ML model.

[00157] At step 11, after the model is trained, the training entity may register the trained model at the ADAE service or A-ADRF function, and an AI/ML model profile may be created. If a model is updated, the training entity may update the AI/ML model profile with the ADAE service or A-ADRF function.

[00158] At step 12, the selected one or more training entities that trained the model may send a response to the ADAE service. The response may include the identifier of the created/updated AI/ML model profile.

[00159] At step 13, after receiving the response about the training procedure, the ADAE service may update the status of the service instance to “model training completed,” and send a notification to the service consumer or the notification target according to the notification setting specified in the service request. The notification may include the AI/ML model profile (e.g., the identifier of the AI/ML model profile) associated with the trained model and a status that the AI/ML training is complete.

[00160] If multiple models need to be trained for a service request, the ADAE service may create separate service instances for the training process of each model (e.g., each model requires different training data or training entity), or manage the training process of all the models jointly (e.g., if the training processes share the same training data or training entity).

[00161] FIG. 15 shows the example inferencing procedure 1500 for an AI/ML service request (e.g., step 1 and 2 may be skipped). Alternatively, the consumer may request for inferencing service only from the ADAE service (e.g., step 1 and 2 may be performed).

[00162] Pre-condition of FIG. 15: It may be assumed that the AI/ML model(s) needed for this service request is/are already available (e.g., either provided by the consumer, available and stored in the A-ADRF, or obtained by the ADAE service via the training procedure in FIG. 14).

[00163] Referring to FIG. 15, at step 1, the ADAE service receives an AI/ML service request from a consumer which is for inferencing service only. The ADAE service may create a service instance for the request and record information associated with the instance in an AI/ML service instance profile (e.g., as shown in Table 5).

[00164] At step 2, the ADAE service may perform the service coordination process as described in FIG. 13. Based on the result of service coordination, the ADAE service may determine to skip to one of the following steps.

[00165] At step 3, the ADAE service determines if inferencing data is required, if so, the ADAE service may initiate the inferencing data collection process. The ADAE service may send a data collection request to the A-DCCF function to collect inferencing data (interactions between the A-DCCF and the data sources are not shown in the figure). Alternatively, the data collection process may be performed by the ADAE service. An AI/ML data profile may be created for the inferencing data. The inferencing data collection process may be performed before the inferencing process (e.g., offline/batch inferencing) or simultaneously as the inferencing process (e.g., online inferencing).

[00166] At step 4, the ADAE service may determine the inferencing entity and the corresponding configuration based on the service requirements in the request and the AI/ML inferencing capability profiles. The ADAE service may select more than one inferencing entity to perform the inferencing process, such as in the scenarios of collaborative inferencing, model splitting, or due to the schedule/availability of the inferencing entity, etc. The ADAE service may select and configure multiple inferencing entities to perform the inferencing process in turn. The ADAE service may update the selection of inferencing entity when change of the inferencing entity is detected (e.g., the capability profile is updated). If more than one model is to be used for inferencing, the ADAE service may determine the configuration for the deployment of the models. Alternatively, the inferencing entity may be determined by the consumer (e.g., via AI/ML inferencing capability discovery procedure) and specified in the AI/ML service request.

[00167] At step 5, after determining the inferencing entity, the ADAE service may send a notification to the service consumer or the notification target according to the notification setting specified in the service request. The notification may include the AI/ML capability profile (e.g., the identifier of the AI/ML capability profile) associated with the inferencing entity. If the consumer is involved in the inferencing procedure as an inferencing entity, the ADAE service may

include configuration information in the notification. For example, the consumer may perform initial inferencing locally, and requests the ADAE service to perform the rest of the procedure. In this case, the ADAE service may instruct the consumer to retrieve the AI/ML model, perform inferencing, and send the intermediate results to the A-ADRF. Then the rest of the inferencing procedure may be performed on another inferencing entity determined by the ADAE service.

[00168] At step 6, the ADAE service may send an inferencing request to the determined inferencing entity, and update the status of the service instance to “performing inferencing.” In the request, the ADAE service may include information of the trained model (e.g., identifier of the AI/ML model profile, model location), information of the inferencing data (if any), and the configuration of the inferencing entity as determined in step 4. The ADAE service may also request the inferencing entity to register the inferencing/analytics results at the ADAE service or the A-ADRF function. Optionally, the ADAE service may request the inferencing entity to perform performance monitoring if it is supported by the inferencing entity.

[00169] At step 7, the inferencing entity may retrieve the trained model (and inferencing data) from the A-ADRF by performing discovery procedure with the identifier of the AI/ML model profile that is provided by the ADAE service in step 6. Alternatively, the inferencing entity may retrieve the trained model (e.g., and inferencing data) based on the model/data location information provided by the ADAE service in step 6.

[00170] At step 8, after retrieving the trained model(s), the selected one or more inferencing entities may perform the inferencing process and generate analytics results.

[00171] At step 9, the ADAE service may configure the inferencing entity to store the generated inferencing results or analytics results at the A-ADRF for performance evaluation or monitoring purpose. An AI/ML data profile may be created for the stored inferencing/analytics results.

[00172] At step 10, the selected one or more inferencing entities may send a response to the ADAE service. The response may include the identifier of the AI/ML data profile associated with the inferencing results.

[00173] At step 11, during the inferencing process, a model update may be triggered (e.g., due to detection of performance drift) by the inferencing entity or by the ADAE service. The ADAE service may perform a training service procedure as in FIG. 14 and then inform the inferencing entity of the updated model.

[00174] At step 12, after receiving the response about the inferencing procedure, the ADAE service may send a notification to the service consumer or the notification target according to the notification setting specified in the service request. The notification may include the AI/ML data profile (e.g., the identifier of the AI/ML data profile) associated with the inferencing/analytics results. Alternatively, the ADAE service may send the information of the inferencing entity to the consumer so that the consumer may request the inferencing entity directly for the inferencing/analytics results.

[00175] With the proposed function, the enhanced ADAE service may support AI/ML-enabled application performance related analytics, such as VAL server performance analytics, VAL session performance analytics, UE-to-UE application performance analytics, etc. A consumer (e.g., a VAL server) may request from the ADAE service AI/ML-enabled analytics results, which may be generated based on the QoS/performance data measured with a UE.

[00176] FIG. 16 shows an example procedure of AI/ML-enabled ADAE support for application performance analytics 1600. Not all entities or steps involved in the procedure are shown in the figure.

[00177] At step 0, an ADAE server may expose information of its services and capabilities to a consumer (e.g., VAL server), such as the training and inferencing capabilities available at the ADAE server and an ADAE client (e.g., via AI/ML resource and capability discovery procedure). For example, inferencing capability at the ADAE client may indicate that AI/ML-analytics results may be generated locally at the UE without having to upload the real-time measurement data.

[00178] At step 1, the consumer may send an analytics (e.g., subscription) request to the ADAE server. In the request, the consumer may specify that AI/ML-enabled analytics results are required, which may be achieved by using a specific analytics ID (e.g., “AI/ML-enabled VAL session performance analytics”) or including an indicator for AI/ML enablement in the request. In addition, the consumer may specify other requirements of the analytics service (e.g., data sources, model, training or inferencing entity) as described in step 1 of FIG. 14 (and Table 5). The consumer may further specify that the analytics results should be generated at the UE locally whenever it is supported by specifying the ADAE client to be the inferencing entity in the request. If the consumer does not specify where the analytics results should be generated, the ADAE server may adaptively determine the inferencing entity based on the status of the AI/ML process.

[00179] At step 2, the ADAE server may perform service coordination as described in FIG. 13 and determine the data providers, the training entity (e.g., step 7 of FIG. 14), and the inferencing entity (e.g., step 4 of FIG. 15). For example, the ADAE server may evaluate the AI/ML capability profile of the ADAE client to determine if it is capable of performing the training and/or inferencing process.

[00180] At step 3, the ADAE server may send an analytics request to the ADAE client for collecting training data as described in step 3 of FIG. 14. If the ADAE client is to perform training, the ADAE server may provide information of the model to be trained (e.g., a URI to the ML model) and the corresponding configuration information to the ADAE client (e.g., as described in step 8 of FIG. 14).

[00181] At step 4, the VAL client may establish connection with a VAL server, which may be different from the VAL server making the request in step 1, for VAL server/session performance analytics or with a VAL client of another UE for UE-to-UE VAL session performance analytics. Application QoS measurements about the connection may be collected for use as training data.

[00182] At step 5, the ADAE client may send the collected training data to the ADAE server (e.g., directly or via the A-DCCF/A-ADRF) if the model is to be trained at the ADAE server. If the model is to be trained locally at the ADAE client, then the ADAE client may send a notification to the ADAE server indicating the completion of the training data collection and the start of training process.

[00183] At step 6, the AI/ML model may be trained by the ADAE server (or the ADAE client) with the training data. If the training is performed at the ADAE client, then the ADAE client may send a notification to the ADAE server when the training is completed.

[00184] At step 7, the ADAE server may send an analytics request to the ADAE client for inferencing process. The request may include information of the trained model, information of the inferencing data (e.g., to be collected), and the corresponding configuration information as described in step 6 of FIG. 15.

[00185] At step 8, real-time QoS measurements about the connection may be collected, which may be used as the inferencing data.

[00186] At step 9, the ADAE client may derive analytics results with the trained model and the collected inferencing data. The ADAE server may adaptively change the inferencing entity.

For example, if computing power is not available for the ADAE client during a certain period of time, the ADAE server may temporarily take over the inferencing process while requesting the inferencing data from the ADAE client.

[00187] At step 10, the ADAE client may send an analytics notification to the ADAE server including the analytics results. The notification may also include analytics performance monitoring results which may trigger a model update.

[00188] At step 11, the ADAE server may send an analytics notification to the consumer with the generated analytics results.

[00189] The procedure described in FIG. 16 may also be applied to other analytics IDs where the analytics results are requested to be generated locally at the UE.

[00190] With the proposed function, the enhanced ADAE service may support AI/ML-enabled edge related analytics, such as EDN load analytics, edge server load analytics, edge service/application performance analytics, etc. A consumer (e.g., an ECS, an EES, an EAS, a VAL server) may request from the ADAE service AI/ML-enabled analytics results, which may be generated based on the data collected from the EDN(s).

[00191] FIG. 17 shows an example procedure of AI/ML-enabled ADAE support for edge analytics 1700. Not all entities or steps involved in the procedure are shown in FIG. 17.

[00192] At step 0, an ADAE server may expose information of its services and capabilities to consumers (e.g., EEL entity or VAL server), such as the training and inferencing capabilities available at ADAE servers that are deployed to EDNs (e.g., via AI/ML resource and capability discovery procedure). For example, training capability at the edge-based ADAE server may indicate that AI/ML models may be trained/updated locally at the EDN with the data collected from the EDN without having to upload the data.

[00193] At step 1, a consumer may send an analytics (e.g., subscription) request to the ADAE server. In the request, the consumer may specify that AI/ML-enabled edge analytics results for EDN-1 are required, which may be achieved by using a specific analytics ID (e.g., “AI/ML-enabled edge load analytics”) or including an indicator for AI/ML enablement in the request. In addition, the consumer may specify other requirements of the analytics service (e.g., data sources, model, training or inferencing entity) as described in step 1 of FIG. 14 (and Table 5). The consumer may further specify that the AI/ML models should be trained at EDN-1 locally whenever it is supported by specifying the ADAE server in EDN-1 to be the training entity in the request.

[00194] At step 2, the ADAE server may perform service coordination as described in FIG. 13. For example, the ADAE server may use the Training Information and Model Description IEs of the AI/ML model profile described in Table 2 to identify that an AI/ML model associated with the same analytics ID is already available, but the model is trained for EDN-2 (e.g., with training data collected from EDN-2), which may or may not apply to EDN-1.

[00195] At step 3, the ADAE server may send an analytics request to an edge-based ADAE server in EDN-1 for inferencing process. The request may include the model trained for EDN-2 or a URI to the trained model that the ADAE server may download.

[00196] At step 4, real-time edge related data of EDN-1 is collected, which may be used as the inferencing data.

[00197] At step 5, the ADAE server in EDN-1 may derive analytics results with the model of EDN-2 and inferencing data collected from EDN-1. The ADAE server in EDN-1 may also monitor the performance of the inferencing process (e.g., accuracy) and record the information.

[00198] At step 6, the ADAE server in EDN-1 may send an analytics notification including the analytics results to the ADAE server from which the analytics request was received or to the consumer if the consumer is also in EDN-1 (e.g., the EES in EDN-1). The notification may also include analytics performance monitoring results which may trigger a model update. For example, the performance monitoring results may indicate that the model for EDN-2 does not fit EDN-1, which may trigger the update or re-training of the model to better fit EDN-1.

[00199] At step 7, based on the received analytics performance information, the ADAE server may determine that the AI/ML model of EDN-2 needs to be updated/re-trained for EDN-1. The ADAE server may check the AI/ML capability profile of the edge-based ADAE server in EDN-1 to determine its configuration (e.g., training schedule, support for FL). The ADAE server may send an analytics request with the determined configuration information to the edge-based ADAE server in EDN-1 to update/re-train the model with data collected from EDN-1. The ADAE server may determine to apply distributed learning or federated learning to train the model. For example, the ADAE server in the core network (e.g., as the FL server) may send the model update requests to ADAE servers at different EDNs (e.g., as the FL clients) and aggregate the updated models.

[00200] At step 8, the ADAE server in EDN-1 may use data collected from EDN-1 to update the model of EDN-2 or re-train a model. The previously collected inferencing data and performance monitoring data may be used as training data for updating/re-training the model. The “data purpose” in the AI/ML data profile may be updated accordingly. If needed, the ADAE server in EDN-1 may initiate additional data collection process to collect more training data.

[00201] At step 9, the ADAE server in EDN-1 derives analytics results with the updated/re-trained model. The ADAE server in EDN-1 may send an analytics notification to the ADAE server in the core network with information of the updated/re-trained model.

[00202] At step 10, the ADAE server in EDN-1 sends an analytics notification to the ADAE server and/or the consumer. The notification may include analytics results generated with the updated/re-trained model.

[00203] The procedure described in FIG. 17 may also be applied to other analytics IDs where the AI/ML models are requested to be trained/updated locally at the EDN.

[00204] FIG. 18 shows an example graphical user interface (GUI) 1800 where the consumer may request services from the AI/ML-enabled analytics service. The consumer may discover any existing AI/ML resources or capabilities that are available, or request to create a new AI/ML service instance which may perform a training and/or inferencing task.

[00205] The 3rd Generation Partnership Project (3GPP) develops technical standards for cellular telecommunications network technologies, including radio access, the core transport network, and service capabilities - including work on codecs, security, and quality of service. Recent radio access technology (RAT) standards include WCDMA (commonly referred as 3G), LTE (commonly referred as 4G), LTE-Advanced standards, and New Radio (NR), which is also referred to as “5G”. 3GPP NR standards development is expected to continue and include the definition of next generation radio access technology (new RAT), which is expected to include the provision of new flexible radio access below 7 GHz, and the provision of new ultra-mobile broadband radio access above 7 GHz. The flexible radio access is expected to consist of a new, non-backwards compatible radio access in new spectrum below 7 GHz, and it is expected to include different operating modes that may be multiplexed together in the same spectrum to address a broad set of 3GPP NR use cases with diverging requirements. The ultra-mobile broadband is expected to include cmWave and mmWave spectrum that may provide the opportunity for ultra-mobile broadband access for, e.g., indoor applications and hotspots. In

particular, the ultra-mobile broadband is expected to share a common design framework with the flexible radio access below 7 GHz, with cmWave and mmWave specific design optimizations.

[00206] 3GPP has identified a variety of use cases that NR is expected to support, resulting in a wide variety of user experience requirements for data rate, latency, and mobility. The use cases include the following general categories: enhanced mobile broadband (eMBB) ultra-reliable low-latency Communication (URLLC), massive machine type communications (mMTC), network operation (e.g., network slicing, routing, migration and interworking, energy savings), and enhanced vehicle-to-everything (eV2X) communications, which may include any of Vehicle-to-Vehicle Communication (V2V), Vehicle-to-Infrastructure Communication (V2I), Vehicle-to-Network Communication (V2N), Vehicle-to-Pedestrian Communication (V2P), and vehicle communications with other entities. Specific service and applications in these categories include, e.g., monitoring and sensor networks, device remote controlling, bi-directional remote controlling, personal cloud computing, video streaming, wireless cloud-based office, first responder connectivity, automotive ecall, disaster alerts, real-time gaming, multi-person video calls, autonomous driving, augmented reality, tactile internet, virtual reality, home automation, robotics, and aerial drones to name a few. All of these use cases and others are contemplated herein.

[00207] FIG. 19A illustrates an example communications system 100 in which the systems, methods, and apparatuses described and claimed herein may be used. The communications system 100 may include wireless transmit/receive units (WTRUs) 102a, 102b, 102c, 102d, 102e, 102f, and/or 102g, which generally or collectively may be referred to as WTRU 102 or WTRUs 102. The communications system 100 may include, a radio access network (RAN) 103/104/105/103b/104b/105b, a core network 106/107/109, a public switched telephone network (PSTN) 108, the Internet 110, other networks 112, and Network Services 113. 113. Network Services 113 may include, for example, a V2X server, V2X functions, a ProSe server, ProSe functions, IoT services, video streaming, and/or edge computing, etc.

[00208] It may be appreciated that the concepts disclosed herein may be used with any number of WTRUs, base stations, networks, and/or network elements. Each of the WTRUs 102 may be any type of apparatus or device configured to operate and/or communicate in a wireless environment. In the example of FIG. 19A, each of the WTRUs 102 is depicted in Figures 8A-8E as a hand-held wireless communications apparatus. It is understood that with the wide variety of use cases contemplated for wireless communications, each WTRU may comprise or be included

in any type of apparatus or device configured to transmit and/or receive wireless signals, including, by way of example only, user equipment (UE), a mobile station, a fixed or mobile subscriber unit, a pager, a cellular telephone, a personal digital assistant (PDA), a smartphone, a laptop, a tablet, a netbook, a notebook computer, a personal computer, a wireless sensor, consumer electronics, a wearable device such as a smart watch or smart clothing, a medical or eHealth device, a robot, industrial equipment, a drone, a vehicle such as a car, bus or truck, a train, or an airplane, and the like.

[00209] The communications system 100 may also include a base station 114a and a base station 114b. In the example of FIG. 19A, each base stations 114a and 114b is depicted as a single element. In practice, the base stations 114a and 114b may include any number of interconnected base stations and/or network elements. Base stations 114a may be any type of device configured to wirelessly interface with at least one of the WTRUs 102a, 102b, and 102c to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, Network Services 113, and/or the other networks 112. Similarly, base station 114b may be any type of device configured to wirelessly interface with at least one of the Remote Radio Heads (RRHs) 118a, 118b, Transmission and Reception Points (TRPs) 119a, 119b, and/or Roadside Units (RSUs) 120a and 120b to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, other networks 112, and/or Network Services 113. RRHs 118a, 118b may be any type of device configured to wirelessly interface with at least one of the WTRUs 102, e.g., WTRU 102c, to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, Network Services 113, and/or other networks 112.

[00210] TRPs 119a, 119b may be any type of device configured to wirelessly interface with at least one of the WTRU 102d, to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, Network Services 113, and/or other networks 112. RSUs 120a and 120b may be any type of device configured to wirelessly interface with at least one of the WTRU 102e or 102f, to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, other networks 112, and/or Network Services 113. By way of example, the base stations 114a, 114b may be a Base Transceiver Station (BTS), a Node-B, an eNode B, a Home Node B, a Home eNode B, a Next Generation

Node-B (gNode B), a satellite, a site controller, an access point (AP), a wireless router, and the like.

[00211] The base station 114a may be part of the RAN 103/104/105, which may also include other base stations and/or network elements (not shown), such as a Base Station Controller (BSC), a Radio Network Controller (RNC), relay nodes, etc. Similarly, the base station 114b may be part of the RAN 103b/104b/105b, which may also include other base stations and/or network elements (not shown), such as a BSC, a RNC, relay nodes, etc. The base station 114a may be configured to transmit and/or receive wireless signals within a particular geographic region, which may be referred to as a cell (not shown). Similarly, the base station 114b may be configured to transmit and/or receive wired and/or wireless signals within a particular geographic region, which may be referred to as a cell (not shown). The cell may further be divided into cell sectors. For example, the cell associated with the base station 114a may be divided into three sectors. Thus, for example, the base station 114a may include three transceivers, e.g., one for each sector of the cell. The base station 114a may employ Multiple-Input Multiple Output (MIMO) technology and, therefore, may utilize multiple transceivers for each sector of the cell, for instance.

[00212] The base station 114a may communicate with one or more of the WTRUs 102a, 102b, 102c, and 102g over an air interface 115/116/117, which may be any suitable wireless communication link (e.g., Radio Frequency (RF), microwave, infrared (IR), ultraviolet (UV), visible light, cmWave, mmWave, etc.). The air interface 115/116/117 may be established using any suitable Radio Access Technology (RAT).

[00213] The base station 114b may communicate with one or more of the RRHs 118a and 118b, TRPs 119a and 119b, and/or RSUs 120a and 120b, over a wired or air interface 115b/116b/117b, which may be any suitable wired (e.g., cable, optical fiber, etc.) or wireless communication link (e.g., RF, microwave, IR, UV, visible light, cmWave, mmWave, etc.). The air interface 115b/116b/117b may be established using any suitable RAT.

[00214] The RRHs 118a, 118b, TRPs 119a, 119b and/or RSUs 120a, 120b, may communicate with one or more of the WTRUs 102c, 102d, 102e, 102f over an air interface 115c/116c/117c, which may be any suitable wireless communication link (e.g., RF, microwave, IR, ultraviolet UV, visible light, cmWave, mmWave, etc.) The air interface 115c/116c/117c may be established using any suitable RAT.

[00215] The WTRUs 102 may communicate with one another over a direct air interface 115d/116d/117d, such as Sidelink communication which may be any suitable wireless communication link (e.g., RF, microwave, IR, ultraviolet UV, visible light, cmWave, mmWave, etc.) The air interface 115d/116d/117d may be established using any suitable RAT.

[00216] The communications system 100 may be a multiple access system and may employ one or more channel access schemes, such as CDMA, TDMA, FDMA, OFDMA, SC-FDMA, and the like. For example, the base station 114a in the RAN 103/104/105 and the WTRUs 102a, 102b, 102c, or RRHs 118a, 118b, TRPs 119a, 119b and/or RSUs 120a and 120b in the RAN 103b/104b/105b and the WTRUs 102c, 102d, 102e, and 102f, may implement a radio technology such as Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access (UTRA), which may establish the air interface 115/116/117 and/or 115c/116c/117c respectively using Wideband CDMA (WCDMA). WCDMA may include communication protocols such as High-Speed Packet Access (HSPA) and/or Evolved HSPA (HSPA+). HSPA may include High-Speed Downlink Packet Access (HSDPA) and/or High-Speed Uplink Packet Access (HSUPA).

[00217] The base station 114a in the RAN 103/104/105 and the WTRUs 102a, 102b, 102c, and 102g, or RRHs 118a and 118b, TRPs 119a and 119b, and/or RSUs 120a and 120b in the RAN 103b/104b/105b and the WTRUs 102c, 102d, may implement a radio technology such as Evolved UMTS Terrestrial Radio Access (E-UTRA), which may establish the air interface 115/116/117 or 115c/116c/117c respectively using Long Term Evolution (LTE) and/or LTE-Advanced (LTE-A), for example. The air interface 115/116/117 or 115c/116c/117c may implement 3GPP NR technology. The LTE and LTE-A technology may include LTE D2D and/or V2X technologies and interfaces (such as Sidelink communications, etc.) Similarly, the 3GPP NR technology may include NR V2X technologies and interfaces (such as Sidelink communications, etc.)

[00218] The base station 114a in the RAN 103/104/105 and the WTRUs 102a, 102b, 102c, and 102g or RRHs 118a and 118b, TRPs 119a and 119b, and/or RSUs 120a and 120b in the RAN 103b/104b/105b and the WTRUs 102c, 102d, 102e, and 102f may implement radio technologies such as IEEE 802.16 (e.g., Worldwide Interoperability for Microwave Access (WiMAX)), CDMA2000, CDMA2000 1X, CDMA2000 EV-DO, Interim Standard 2000 (IS-2000), Interim Standard 95 (IS-95), Interim Standard 856 (IS-856), Global System for Mobile

communications (GSM), Enhanced Data rates for GSM Evolution (EDGE), GSM EDGE (GERAN), and the like.

[00219] The base station 114c in FIG. 19A may be a wireless router, Home Node B, Home eNode B, or access point, for example, and may utilize any suitable RAT for facilitating wireless connectivity in a localized area, such as a place of business, a home, a vehicle, a train, an aerial, a satellite, a manufactory, a campus, and the like. The base station 114c and the WTRUs 102, e.g., WTRU 102e, may implement a radio technology such as IEEE 802.11 to establish a Wireless Local Area Network (WLAN). Similarly, the base station 114c and the WTRUs 102, e.g., WTRU 102d, may implement a radio technology such as IEEE 802.15 to establish a wireless personal area network (WPAN). The base station 114c and the WTRUs 102, e.g., WTRU 102e, may utilize a cellular-based RAT (e.g., WCDMA, CDMA2000, GSM, LTE, LTE-A, NR, etc.) to establish a picocell or femtocell. As shown in FIG. 19A, the base station 114c may have a direct connection to the Internet 110. Thus, the base station 114c may not be required to access the Internet 110 via the core network 106/107/109.

[00220] The RAN 103/104/105 and/or RAN 103b/104b/105b may be in communication with the core network 106/107/109, which may be any type of network configured to provide voice, data, messaging, authorization and authentication, applications, and/or Voice Over Internet Protocol (VoIP) services to one or more of the WTRUs 102. For example, the core network 106/107/109 may provide call control, billing services, mobile location-based services, pre-paid calling, Internet connectivity, packet data network connectivity, Ethernet connectivity, video distribution, etc., and/or perform high-level security functions, such as user authentication.

[00221] Although not shown in FIG. 19A, it may be appreciated that the RAN 103/104/105 and/or RAN 103b/104b/105b and/or the core network 106/107/109 may be in direct or indirect communication with other RANs that employ the same RAT as the RAN 103/104/105 and/or RAN 103b/104b/105b or a different RAT. For example, in addition to being connected to the RAN 103/104/105 and/or RAN 103b/104b/105b, which may be utilizing an E-UTRA radio technology, the core network 106/107/109 may also be in communication with another RAN (not shown) employing a GSM or NR radio technology.

[00222] The core network 106/107/109 may also serve as a gateway for the WTRUs 102 to access the PSTN 108, the Internet 110, and/or other networks 112. The PSTN 108 may include circuit-switched telephone networks that provide Plain Old Telephone Service (POTS). The

Internet 110 may include a global system of interconnected computer networks and devices that use common communication protocols, such as the Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and the internet protocol (IP) in the TCP/IP internet protocol suite. The other networks 112 may include wired or wireless communications networks owned and/or operated by other service providers. For example, the networks 112 may include any type of packet data network (e.g., an IEEE 802.3 Ethernet network) or another core network connected to one or more RANs, which may employ the same RAT as the RAN 103/104/105 and/or RAN 103b/104b/105b or a different RAT.

[00223] Some or all of the WTRUs 102a, 102b, 102c, 102d, 102e, and 102f in the communications system 100 may include multi-mode capabilities, e.g., the WTRUs 102a, 102b, 102c, 102d, 102e, and 102f may include multiple transceivers for communicating with different wireless networks over different wireless links. For example, the WTRU 102g shown in FIG. 19A may be configured to communicate with the base station 114a, which may employ a cellular-based radio technology, and with the base station 114c, which may employ an IEEE 802 radio technology.

[00224] Although not shown in FIG. 19A, it may be appreciated that a User Equipment may make a wired connection to a gateway. The gateway maybe a Residential Gateway (RG). The RG may provide connectivity to a Core Network 106/107/109. It may be appreciated that many of the ideas contained herein may equally apply to UEs that are WTRUs and UEs that use a wired connection to connect to a network. For example, the ideas that apply to the wireless interfaces 115, 116, 117 and 115c/116c/117c may equally apply to a wired connection.

[00225] FIG. 19B is a system diagram of an example RAN 103 and core network 106. As noted above, the RAN 103 may employ a UTRA radio technology to communicate with the WTRUs 102a, 102b, and 102c over the air interface 115. The RAN 103 may also be in communication with the core network 106. As shown in FIG. 19B, the RAN 103 may include Node-Bs 140a, 140b, and 140c, which may each include one or more transceivers for communicating with the WTRUs 102a, 102b, and 102c over the air interface 115. The Node-Bs 140a, 140b, and 140c may each be associated with a particular cell (not shown) within the RAN 103. The RAN 103 may also include RNCs 142a, 142b. It may be appreciated that the RAN 103 may include any number of Node-Bs and Radio Network Controllers (RNCs.)

[00226] As shown in FIG. 19B, the Node-Bs 140a, 140b may be in communication with the RNC 142a. Additionally, the Node-B 140c may be in communication with the RNC 142b. The Node-Bs 140a, 140b, and 140c may communicate with the respective RNCs 142a and 142b via an Iub interface. The RNCs 142a and 142b may be in communication with one another via an Iur interface. Each of the RNCs 142a and 142b may be configured to control the respective Node-Bs 140a, 140b, and 140c to which it is connected. In addition, each of the RNCs 142a and 142b may be configured to carry out or support other functionality, such as outer loop power control, load control, admission control, packet scheduling, handover control, macro-diversity, security functions, data encryption, and the like.

[00227] The core network 106 shown in FIG. 19B may include a media gateway (MGW) 144, a Mobile Switching Center (MSC) 146, a Serving GPRS Support Node (SGSN) 148, and/or a Gateway GPRS Support Node (GGSN) 150. While each of the foregoing elements are depicted as part of the core network 106, it may be appreciated that any one of these elements may be owned and/or operated by an entity other than the core network operator.

[00228] The RNC 142a in the RAN 103 may be connected to the MSC 146 in the core network 106 via an IuCS interface. The MSC 146 may be connected to the MGW 144. The MSC 146 and the MGW 144 may provide the WTRUs 102a, 102b, and 102c with access to circuit-switched networks, such as the PSTN 108, to facilitate communications between the WTRUs 102a, 102b, and 102c, and traditional land-line communications devices.

[00229] The RNC 142a in the RAN 103 may also be connected to the SGSN 148 in the core network 106 via an IuPS interface. The SGSN 148 may be connected to the GGSN 150. The SGSN 148 and the GGSN 150 may provide the WTRUs 102a, 102b, and 102c with access to packet-switched networks, such as the Internet 110, to facilitate communications between and the WTRUs 102a, 102b, and 102c, and IP-enabled devices.

[00230] The core network 106 may also be connected to the other networks 112, which may include other wired or wireless networks that are owned and/or operated by other service providers.

[00231] FIG. 19C is a system diagram of an example RAN 104 and core network 107. As noted above, the RAN 104 may employ an E-UTRA radio technology to communicate with the WTRUs 102a, 102b, and 102c over the air interface 116. The RAN 104 may also be in communication with the core network 107.

[00232] The RAN 104 may include eNode-Bs 160a, 160b, and 160c, though it may be appreciated that the RAN 104 may include any number of eNode-Bs. The eNode-Bs 160a, 160b, and 160c may each include one or more transceivers for communicating with the WTRUs 102a, 102b, and 102c over the air interface 116. For example, the eNode-Bs 160a, 160b, and 160c may implement MIMO technology. Thus, the eNode-B 160a, for example, may use multiple antennas to transmit wireless signals to, and receive wireless signals from, the WTRU 102a.

[00233] Each of the eNode-Bs 160a, 160b, and 160c may be associated with a particular cell (not shown) and may be configured to handle radio resource management decisions, handover decisions, scheduling of users in the uplink and/or downlink, and the like. As shown in FIG. 19C, the eNode-Bs 160a, 160b, and 160c may communicate with one another over an X2 interface.

[00234] The core network 107 shown in FIG. 19C may include a Mobility Management Gateway (MME) 162, a serving gateway 164, and a Packet Data Network (PDN) gateway 166. While each of the foregoing elements are depicted as part of the core network 107, it may be appreciated that any one of these elements may be owned and/or operated by an entity other than the core network operator.

[00235] The MME 162 may be connected to each of the eNode-Bs 160a, 160b, and 160c in the RAN 104 via an S1 interface and may serve as a control node. For example, the MME 162 may be responsible for authenticating users of the WTRUs 102a, 102b, and 102c, bearer activation/deactivation, selecting a particular serving gateway during an initial attach of the WTRUs 102a, 102b, and 102c, and the like. The MME 162 may also provide a control plane function for switching between the RAN 104 and other RANs (not shown) that employ other radio technologies, such as GSM or WCDMA.

[00236] The serving gateway 164 may be connected to each of the eNode-Bs 160a, 160b, and 160c in the RAN 104 via the S1 interface. The serving gateway 164 may generally route and forward user data packets to/from the WTRUs 102a, 102b, and 102c. The serving gateway 164 may also perform other functions, such as anchoring user planes during inter-eNode B handovers, triggering paging when downlink data is available for the WTRUs 102a, 102b, and 102c, managing and storing contexts of the WTRUs 102a, 102b, and 102c, and the like.

[00237] The serving gateway 164 may also be connected to the PDN gateway 166, which may provide the WTRUs 102a, 102b, and 102c with access to packet-switched networks, such as

the Internet 110, to facilitate communications between the WTRUs 102a, 102b, 102c, and IP-enabled devices.

[00238] The core network 107 may facilitate communications with other networks. For example, the core network 107 may provide the WTRUs 102a, 102b, and 102c with access to circuit-switched networks, such as the PSTN 108, to facilitate communications between the WTRUs 102a, 102b, and 102c and traditional land-line communications devices. For example, the core network 107 may include, or may communicate with, an IP gateway (e.g., an IP Multimedia Subsystem (IMS) server) that serves as an interface between the core network 107 and the PSTN 108. In addition, the core network 107 may provide the WTRUs 102a, 102b, and 102c with access to the networks 112, which may include other wired or wireless networks that are owned and/or operated by other service providers.

[00239] FIG. 19D is a system diagram of an example RAN 105 and core network 109. The RAN 105 may employ an NR radio technology to communicate with the WTRUs 102a and 102b over the air interface 117. The RAN 105 may also be in communication with the core network 109. A Non-3GPP Interworking Function (N3IWF) 199 may employ a non-3GPP radio technology to communicate with the WTRU 102c over the air interface 198. The N3IWF 199 may also be in communication with the core network 109.

[00240] The RAN 105 may include gNode-Bs 180a and 180b. It may be appreciated that the RAN 105 may include any number of gNode-Bs. The gNode-Bs 180a and 180b may each include one or more transceivers for communicating with the WTRUs 102a and 102b over the air interface 117. When integrated access and backhaul connection are used, the same air interface may be used between the WTRUs and gNode-Bs, which may be the core network 109 via one or multiple gNBs. The gNode-Bs 180a and 180b may implement MIMO, MU-MIMO, and/or digital beamforming technology. Thus, the gNode-B 180a, for example, may use multiple antennas to transmit wireless signals to, and receive wireless signals from, the WTRU 102a. It should be appreciated that the RAN 105 may employ of other types of base stations such as an eNode-B. It may also be appreciated the RAN 105 may employ more than one type of base station. For example, the RAN may employ eNode-Bs and gNode-Bs.

[00241] The N3IWF 199 may include a non-3GPP Access Point 180c. It may be appreciated that the N3IWF 199 may include any number of non-3GPP Access Points. The non-3GPP Access Point 180c may include one or more transceivers for communicating with the

WTRUs 102c over the air interface 198. The non-3GPP Access Point 180c may use the 802.11 protocol to communicate with the WTRU 102c over the air interface 198.

[00242] Each of the gNode-Bs 180a and 180b may be associated with a particular cell (not shown) and may be configured to handle radio resource management decisions, handover decisions, scheduling of users in the uplink and/or downlink, and the like. As shown in FIG. 19D, the gNode-Bs 180a and 180b may communicate with one another over an Xn interface, for example.

[00243] The core network 109 shown in FIG. 19D may be a 5G core network (5GC). The core network 109 may offer numerous communication services to customers who are interconnected by the radio access network. The core network 109 comprises a number of entities that perform the functionality of the core network. As used herein, the term “core network entity” or “network function” refers to any entity that performs one or more functionalities of a core network. It is understood that such core network entities may be logical entities that are implemented in the form of computer-executable instructions (software) stored in a memory of, and executing on a processor of, an apparatus configured for wireless and/or network communications or a computer system, such as system 90 illustrated in FIG. 19G.

[00244] In the example of FIG. 19D, the 5G Core Network 109 may include an access and mobility management function (AMF) 172, a Session Management Function (SMF) 174, User Plane Functions (UPFs) 176a and 176b, a User Data Management Function (UDM) 197, an Authentication Server Function (AUSF) 190, a Network Exposure Function (NEF) 196, a Policy Control Function (PCF) 184, a Non-3GPP Interworking Function (N3IWF) 199, a User Data Repository (UDR) 178. While each of the foregoing elements are depicted as part of the 5G core network 109, it may be appreciated that any one of these elements may be owned and/or operated by an entity other than the core network operator. It may also be appreciated that a 5G core network may not consist of all of these elements, may consist of additional elements, and may consist of multiple instances of each of these elements. FIG. 19D shows that network functions directly connect to one another, however, it should be appreciated that they may communicate via routing agents such as a diameter routing agent or message buses.

[00245] In the example of FIG. 19D, connectivity between network functions is achieved via a set of interfaces, or reference points. It may be appreciated that network functions may be modeled, described, or implemented as a set of services that are invoked, or called, by other

network functions or services. Invocation of a Network Function service may be achieved via a direct connection between network functions, an exchange of messaging on a message bus, calling a software function, etc.

[00246] The AMF 172 may be connected to the RAN 105 via an N2 interface and may serve as a control node. For example, the AMF 172 may be responsible for registration management, connection management, reachability management, access authentication, access authorization. The AMF may be responsible forwarding user plane tunnel configuration information to the RAN 105 via the N2 interface. The AMF 172 may receive the user plane tunnel configuration information from the SMF via an N11 interface. The AMF 172 may generally route and forward NAS packets to/from the WTRUs 102a, 102b, and 102c via an N1 interface. The N1 interface is not shown in FIG. 19D.

[00247] The SMF 174 may be connected to the AMF 172 via an N11 interface. Similarly the SMF may be connected to the PCF 184 via an N7 interface, and to the UPFs 176a and 176b via an N4 interface. The SMF 174 may serve as a control node. For example, the SMF 174 may be responsible for Session Management, IP address allocation for the WTRUs 102a, 102b, and 102c, management and configuration of traffic steering rules in the UPF 176a and UPF 176b, and generation of downlink data notifications to the AMF 172.

[00248] The UPF 176a and UPF 176b may provide the WTRUs 102a, 102b, and 102c with access to a Packet Data Network (PDN), such as the Internet 110, to facilitate communications between the WTRUs 102a, 102b, and 102c and other devices. The UPF 176a and UPF 176b may also provide the WTRUs 102a, 102b, and 102c with access to other types of packet data networks. For example, Other Networks 112 may be Ethernet Networks or any type of network that exchanges packets of data. The UPF 176a and UPF 176b may receive traffic steering rules from the SMF 174 via the N4 interface. The UPF 176a and UPF 176b may provide access to a packet data network by connecting a packet data network with an N6 interface or by connecting to each other and to other UPFs via an N9 interface. In addition to providing access to packet data networks, the UPF 176 may be responsible packet routing and forwarding, policy rule enforcement, quality of service handling for user plane traffic, downlink packet buffering.

[00249] The AMF 172 may also be connected to the N3IWF 199, for example, via an N2 interface. The N3IWF facilitates a connection between the WTRU 102c and the 5G core network

170, for example, via radio interface technologies that are not defined by 3GPP. The AMF may interact with the N3IWF 199 in the same, or similar, manner that it interacts with the RAN 105.

[00250] The PCF 184 may be connected to the SMF 174 via an N7 interface, connected to the AMF 172 via an N15 interface, and to an Application Function (AF) 188 via an N5 interface. The N15 and N5 interfaces are not shown in FIG. 19D. The PCF 184 may provide policy rules to control plane nodes such as the AMF 172 and SMF 174, allowing the control plane nodes to enforce these rules. The PCF 184, may send policies to the AMF 172 for the WTRUs 102a, 102b, and 102c so that the AMF may deliver the policies to the WTRUs 102a, 102b, and 102c via an N1 interface. Policies may then be enforced, or applied, at the WTRUs 102a, 102b, and 102c.

[00251] The UDR 178 may act as a repository for authentication credentials and subscription information. The UDR may connect to network functions, so that network function may add to, read from, and modify the data that is in the repository. For example, the UDR 178 may connect to the PCF 184 via an N36 interface. Similarly, the UDR 178 may connect to the NEF 196 via an N37 interface, and the UDR 178 may connect to the UDM 197 via an N35 interface.

[00252] The UDM 197 may serve as an interface between the UDR 178 and other network functions. The UDM 197 may authorize network functions to access of the UDR 178. For example, the UDM 197 may connect to the AMF 172 via an N8 interface, the UDM 197 may connect to the SMF 174 via an N10 interface. Similarly, the UDM 197 may connect to the AUSF 190 via an N13 interface. The UDR 178 and UDM 197 may be tightly integrated.

[00253] The AUSF 190 performs authentication related operations and connects to the UDM 178 via an N13 interface and to the AMF 172 via an N12 interface.

[00254] The NEF 196 exposes capabilities and services in the 5G core network 109 to Application Functions (AF) 188. Exposure may occur on the N33 API interface. The NEF may connect to an AF 188 via an N33 interface and it may connect to other network functions in order to expose the capabilities and services of the 5G core network 109.

[00255] Application Functions 188 may interact with network functions in the 5G Core Network 109. Interaction between the Application Functions 188 and network functions may be via a direct interface or may occur via the NEF 196. The Application Functions 188 may be considered part of the 5G Core Network 109 or may be external to the 5G Core Network 109 and deployed by enterprises that have a business relationship with the mobile network operator.

[00256] Network Slicing is a mechanism that may be used by mobile network operators to support one or more ‘virtual’ core networks behind the operator’s air interface. This involves ‘slicing’ the core network into one or more virtual networks to support different RANs or different service types running across a single RAN. Network slicing enables the operator to create networks customized to provide optimized solutions for different market scenarios which demands diverse requirements, e.g., in the areas of functionality, performance and isolation.

[00257] 3GPP has designed the 5G core network to support Network Slicing. Network Slicing is a good tool that network operators may use to support the diverse set of 5G use cases (e.g., massive IoT, critical communications, V2X, and enhanced mobile broadband) which demand very diverse and sometimes extreme requirements. Without the use of network slicing techniques, it is likely that the network architecture would not be flexible and scalable enough to efficiently support a wider range of use cases need when each use case has its own specific set of performance, scalability, and availability requirements. Furthermore, introduction of new network services should be made more efficient.

[00258] Referring again to FIG. 19D, in a network slicing scenario, a WTRU 102a, 102b, or 102c may connect to an AMF 172, via an N1 interface. The AMF may be logically part of one or more slices. The AMF may coordinate the connection or communication of WTRU 102a, 102b, or 102c with one or more UPF 176a and 176b, SMF 174, and other network functions. Each of the UPFs 176a and 176b, SMF 174, and other network functions may be part of the same slice or different slices. When they are part of different slices, they may be isolated from each other in the sense that they may utilize different computing resources, security credentials, etc.

[00259] The core network 109 may facilitate communications with other networks. For example, the core network 109 may include, or may communicate with, an IP gateway, such as an IP Multimedia Subsystem (IMS) server, that serves as an interface between the 5G core network 109 and a PSTN 108. For example, the core network 109 may include, or communicate with a short message service (SMS) service center that facilitates communication via the short message service. For example, the 5G core network 109 may facilitate the exchange of non-IP data packets between the WTRUs 102a, 102b, and 102c and servers or applications functions 188. In addition, the core network 170 may provide the WTRUs 102a, 102b, and 102c with access to the networks 112, which may include other wired or wireless networks that are owned and/or operated by other service providers.

[00260] The core network entities described herein and illustrated in Figures 8A, 8C, 8D, and 8E are identified by the names given to those entities in certain existing 3GPP specifications, but it is understood that in the future those entities and functionalities may be identified by other names and certain entities or functions may be combined in future specifications published by 3GPP, including future 3GPP NR specifications. Thus, the particular network entities and functionalities described and illustrated in Figures 8A, 8B, 8C, 8D, and 8E are provided by way of example only, and it is understood that the subject matter disclosed and claimed herein may be embodied or implemented in any similar communication system, whether presently defined or defined in the future.

[00261] FIG. 19E illustrates an example communications system 111 in which the systems, methods, apparatuses described herein may be used. Communications system 111 may include Wireless Transmit/Receive Units (WTRUs) A, B, C, D, E, F, a base station gNB 121, a V2X server 124, and Road Side Units (RSUs) 123a and 123b. In practice, the concepts presented herein may be applied to any number of WTRUs, base station gNBs, V2X networks, and/or other network elements. One or several or all WTRUs A, B, C, D, E, and F may be out of range of the access network coverage 131. WTRUs A, B, and C form a V2X group, among which WTRU A is the group lead and WTRUs B and C are group members.

[00262] WTRUs A, B, C, D, E, and F may communicate with each other over a Uu interface 129 via the gNB 121 if they are within the access network coverage 131. In the example of FIG. 19E, WTRUs B and F are shown within access network coverage 131. WTRUs A, B, C, D, E, and F may communicate with each other directly via a Sidelink interface (e.g., PC5 or NR PC5) such as interface 125a, 125b, or 128, whether they are under the access network coverage 131 or out of the access network coverage 131. For instance, in the example of FIG. 19E, WTRU D, which is outside of the access network coverage 131, communicates with WTRU F, which is inside the coverage 131.

[00263] WTRUs A, B, C, D, E, and F may communicate with RSU 123a or 123b via a Vehicle-to-Network (V2N) 133 or Sidelink interface 125b. WTRUs A, B, C, D, E, and F may communicate to a V2X Server 124 via a Vehicle-to-Infrastructure (V2I) interface 127. WTRUs A, B, C, D, E, and F may communicate to another UE via a Vehicle-to-Person (V2P) interface 128.

[00264] FIG. 19F is a block diagram of an example apparatus or device WTRU 102 that may be configured for wireless communications and operations in accordance with the systems,

methods, and apparatuses described herein, such as a WTRU 102 of FIG. 19A, 8B, 8C, 8D, or 8E. As shown in FIG. 19F, the example WTRU 102 may include a processor 118, a transceiver 120, a transmit/receive element 122, a speaker/microphone 124, a keypad 126, a display/touchpad/indicators 128, non-removable memory 130, removable memory 132, a power source 134, a global positioning system (GPS) chipset 136, and other peripherals 138. It may be appreciated that the WTRU 102 may include any sub-combination of the foregoing elements. Also, the base stations 114a and 114b, and/or the nodes that base stations 114a and 114b may represent, such as but not limited to transceiver station (BTS), a Node-B, a site controller, an access point (AP), a home node-B, an evolved home node-B (eNodeB), a home evolved node-B (HeNB), a home evolved node-B gateway, a next generation node-B (gNode-B), and proxy nodes, among others, may include some or all of the elements depicted in FIG. 19F and described herein.

[00265] The processor 118 may be a general purpose processor, a special purpose processor, a conventional processor, a digital signal processor (DSP), a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits (ASICs), Field Programmable Gate Array (FPGAs) circuits, any other type of integrated circuit (IC), a state machine, and the like. The processor 118 may perform signal coding, data processing, power control, input/output processing, and/or any other functionality that enables the WTRU 102 to operate in a wireless environment. The processor 118 may be coupled to the transceiver 120, which may be coupled to the transmit/receive element 122. While FIG. 19F depicts the processor 118 and the transceiver 120 as separate components, it may be appreciated that the processor 118 and the transceiver 120 may be integrated together in an electronic package or chip.

[00266] The transmit/receive element 122 of a UE may be configured to transmit signals to, or receive signals from, a base station (e.g., the base station 114a of FIG. 19A) over the air interface 115/116/117 or another UE over the air interface 115d/116d/117d. For example, the transmit/receive element 122 may be an antenna configured to transmit and/or receive RF signals. The transmit/receive element 122 may be an emitter/detector configured to transmit and/or receive IR, UV, or visible light signals, for example. The transmit/receive element 122 may be configured to transmit and receive both RF and light signals. It may be appreciated that the transmit/receive element 122 may be configured to transmit and/or receive any combination of wireless or wired signals.

[00267] In addition, although the transmit/receive element 122 is depicted in FIG. 19F as a single element, the WTRU 102 may include any number of transmit/receive elements 122. More specifically, the WTRU 102 may employ MIMO technology. Thus, the WTRU 102 may include two or more transmit/receive elements 122 (e.g., multiple antennas) for transmitting and receiving wireless signals over the air interface 115/116/117.

[00268] The transceiver 120 may be configured to modulate the signals that are to be transmitted by the transmit/receive element 122 and to demodulate the signals that are received by the transmit/receive element 122. As noted above, the WTRU 102 may have multi-mode capabilities. Thus, the transceiver 120 may include multiple transceivers for enabling the WTRU 102 to communicate via multiple RATs, for example NR and IEEE 802.11 or NR and E-UTRA, or to communicate with the same RAT via multiple beams to different RRHs, TRPs, RSUs, or nodes.

[00269] The processor 118 of the WTRU 102 may be coupled to, and may receive user input data from, the speaker/microphone 124, the keypad 126, and/or the display/touchpad/indicators 128 (e.g., a liquid crystal display (LCD) display unit or organic light-emitting diode (OLED) display unit). The processor 118 may also output user data to the speaker/microphone 124, the keypad 126, and/or the display/touchpad/indicators 128. In addition, the processor 118 may access information from, and store data in, any type of suitable memory, such as the non-removable memory 130 and/or the removable memory 132. The non-removable memory 130 may include random-access memory (RAM), read-only memory (ROM), a hard disk, or any other type of memory storage device. The removable memory 132 may include a subscriber identity module (SIM) card, a memory stick, a secure digital (SD) memory card, and the like. The processor 118 may access information from, and store data in, memory that is not physically located on the WTRU 102, such as on a server that is hosted in the cloud or in an edge computing platform or in a home computer (not shown).

[00270] The processor 118 may receive power from the power source 134, and may be configured to distribute and/or control the power to the other components in the WTRU 102. The power source 134 may be any suitable device for powering the WTRU 102. For example, the power source 134 may include one or more dry cell batteries, solar cells, fuel cells, and the like.

[00271] The processor 118 may also be coupled to the GPS chipset 136, which may be configured to provide location information (e.g., longitude and latitude) regarding the current

location of the WTRU 102. In addition to, or in lieu of, the information from the GPS chipset 136, the WTRU 102 may receive location information over the air interface 115/116/117 from a base station (e.g., base stations 114a, 114b) and/or determine its location based on the timing of the signals being received from two or more nearby base stations. It may be appreciated that the WTRU 102 may acquire location information by way of any suitable location-determination method.

[00272] The processor 118 may further be coupled to other peripherals 138, which may include one or more software and/or hardware modules that provide additional features, functionality, and/or wired or wireless connectivity. For example, the peripherals 138 may include various sensors such as an accelerometer, biometrics (e.g., finger print) sensors, an e-compass, a satellite transceiver, a digital camera (for photographs or video), a universal serial bus (USB) port or other interconnect interfaces, a vibration device, a television transceiver, a hands free headset, a Bluetooth® module, a frequency modulated (FM) radio unit, a digital music player, a media player, a video game player module, an Internet browser, and the like.

[00273] The WTRU 102 may be included in other apparatuses or devices, such as a sensor, consumer electronics, a wearable device such as a smart watch or smart clothing, a medical or eHealth device, a robot, industrial equipment, a drone, a vehicle such as a car, truck, train, or an airplane. The WTRU 102 may connect to other components, modules, or systems of such apparatuses or devices via one or more interconnect interfaces, such as an interconnect interface that may comprise one of the peripherals 138.

[00274] FIG. 19G is a block diagram of an exemplary computing system 90 in which one or more apparatuses of the communications networks illustrated in Figures 8A, 8C, 8D and 8E may be embodied, such as certain nodes or functional entities in the RAN 103/104/105, Core Network 106/107/109, PSTN 108, Internet 110, Other Networks 112, or Network Services 113. Computing system 90 may comprise a computer or server and may be controlled primarily by computer readable instructions, which may be in the form of software, wherever, or by whatever means such software is stored or accessed. Such computer readable instructions may be executed within a processor 91, to cause computing system 90 to do work. The processor 91 may be a general purpose processor, a special purpose processor, a conventional processor, a digital signal processor (DSP), a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits (ASICs), Field

Programmable Gate Array (FPGAs) circuits, any other type of integrated circuit (IC), a state machine, and the like. The processor 91 may perform signal coding, data processing, power control, input/output processing, and/or any other functionality that enables the computing system 90 to operate in a communications network. Coprocessor 81 is an optional processor, distinct from main processor 91, that may perform additional functions or assist processor 91. Processor 91 and/or coprocessor 81 may receive, generate, and process data related to the methods and apparatuses disclosed herein.

[00275] In operation, processor 91 fetches, decodes, and executes instructions, and transfers information to and from other resources via the computing system's main data-transfer path, system bus 80. Such a system bus connects the components in computing system 90 and defines the medium for data exchange. System bus 80 typically includes data lines for sending data, address lines for sending addresses, and control lines for sending interrupts and for operating the system bus. An example of such a system bus 80 is the PCI (Peripheral Component Interconnect) bus.

[00276] Memories coupled to system bus 80 include random access memory (RAM) 82 and read only memory (ROM) 93. Such memories include circuitry that allows information to be stored and retrieved. ROMs 93 generally contain stored data that may not easily be modified. Data stored in RAM 82 may be read or changed by processor 91 or other hardware devices. Access to RAM 82 and/or ROM 93 may be controlled by memory controller 92. Memory controller 92 may provide an address translation function that translates virtual addresses into physical addresses as instructions are executed. Memory controller 92 may also provide a memory protection function that isolates processes within the system and isolates system processes from user processes. Thus, a program running in a first mode may access only memory mapped by its own process virtual address space; it may not access memory within another process's virtual address space unless memory sharing between the processes has been set up.

[00277] In addition, computing system 90 may contain peripherals controller 83 responsible for communicating instructions from processor 91 to peripherals, such as printer 94, keyboard 84, mouse 95, and disk drive 85.

[00278] Display 86, which is controlled by display controller 96, is used to display visual output generated by computing system 90. Such visual output may include text, graphics, animated graphics, and video. The visual output may be provided in the form of a graphical user interface

(GUI). Display 86 may be implemented with a CRT-based video display, an LCD-based flat-panel display, gas plasma-based flat-panel display, or a touch-panel. Display controller 96 includes electronic components required to generate a video signal that is sent to display 86.

[00279] Further, computing system 90 may contain communication circuitry, such as for example a wireless or wired network adapter 97, that may be used to connect computing system 90 to an external communications network or devices, such as the RAN 103/104/105, Core Network 106/107/109, PSTN 108, Internet 110, WTRUs 102, or Other Networks 112 of Figures 8A, 8B, 8C, 8D, and 8E, to enable the computing system 90 to communicate with other nodes or functional entities of those networks. The communication circuitry, alone or in combination with the processor 91, may be used to perform the transmitting and receiving steps of certain apparatuses, nodes, or functional entities described herein.

[00280] It is understood that any or all of the apparatuses, systems, methods and processes described herein may be embodied in the form of computer executable instructions (e.g., program code) stored on a computer-readable storage medium which instructions, when executed by a processor, such as processors 118 or 91, cause the processor to perform and/or implement the systems, methods and processes described herein. Specifically, any of the steps, operations, or functions described herein may be implemented in the form of such computer executable instructions, executing on the processor of an apparatus or computing system configured for wireless and/or wired network communications. Computer readable storage media includes volatile and nonvolatile, removable and non-removable media implemented in any non-transitory (e.g., tangible or physical) method or technology for storage of information, but such computer readable storage media do not include signals. Computer readable storage media include, but are not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other tangible or physical medium which may be used to store the desired information and which may be accessed by a computing system.

APPENDIX

Table 0.1 - Abbreviations

3GPP	3 rd Generation Partnership Program
AI/ML	Artificial Intelligence / Machine Learning
ACR	Application Context Relocation
ADAE	Application Data Analytics Enablement
(A-) ADRF	(Application layer) Analytical Data Repository Function
AF	Application Function
CAPIF	Common API Framework
(A-) DCCF	(Application layer) Data Collection Coordination Function
EAS	Edge Application Server
ECS	Edge Configuration Server
EDN	Edge Data Network
EEC	Edge Enabler Client
EEL	Edge Enabler Layer
EES	Edge Enabler Server
FL	Federated Learning
NRF	Network Repository Function
NWDAF	Network Data Analytics Function
OAM	Operations, Administration, and Management
QoS	Quality of Service
SEAL	Service Enabler Architecture Layer
SEALDD	Service Enabler Architecture Layer Data Delivery
UAS	Uncrewed Aerial Systems
UE	User Equipment
URI	Uniform Resource Identifier
VAL	Vertical Application Layer
V2X	Vehicle-to-Everything

CLAIMS

What is claimed is:

1. A method for providing an artificial intelligence/machine learning (AI/ML) enablement service, the method comprising:

receiving, from one or more AI/ML resource/capability providers, one or more provider messages comprising one or more descriptions of one or more AI/ML resources or capabilities of the one or more AI/ML resource/capability providers;

storing, by the AI/ML enablement service and based on the one or more provider messages, the one or more descriptions of the one or more AI/ML resources or capabilities;

receiving, from an AI/ML service consumer, a consumer request comprising an identifier of the AI/ML service consumer, one or more notification conditions and targets, a required analytics type, or one or more service requirements;

determining, based on processing information associated with the one or more descriptions of the one or more AI/ML resources or capabilities and the consumer request, a service action comprising selecting an AI/ML resource/capability provider of the one or more AI/ML resource/capability providers to perform AI/ML operations;

sending, to the selected AI/ML resource/capability provider, a request to perform one or more AI/ML operations;

receiving, from the selected AI/ML resource/capability provider, a response message indicating that the one or more AI/ML operations have been performed and comprising one or more results of the performed AI/ML operations; and

sending, to the AI/ML service consumer, a notification message comprising the one or more results of the performed AI/ML operations.

2. The method of claim 1, wherein the provider message is a registration request or a message responsive to a query/request from the AI/ML enablement service.

3. The method of claim 1, wherein the one or more AI/ML resource/capability providers comprise at least one of an AI/ML enablement service, a Vertical Application Layer (VAL) entity, an application/service enablement layer entity, or a core network function.

4. The method of claim 1, wherein the one or more descriptions of the one or more AI/ML resources or capabilities comprise at least of an AI/ML data description, an AI/ML model description, or an AI/ML training/inferencing capability description.

5. The method of claim 1, further comprising:
receiving, from AI/ML service consumer, a consumer message comprising filter criteria of AI/ML resources or capabilities that are of interest to the AI/ML service consumer, wherein the consumer message is a discovery request or a query/subscription request;
sending, to the AI/ML service consumer, a response comprising information associated with the AI/ML resources or capabilities that are of interest to the AI/ML service consumer, wherein the response comprises the one or more descriptions of one or more AI/ML resources or capabilities of the one or more AI/ML resource/capability providers.

6. The method of claim 1, wherein the AI/ AI/ML service consumer is an AI/ML enablement service, a Vertical Application Layer (VAL) entity, or an application/service enablement layer entity.

7. The method of claim 1, wherein the service action comprises re-selecting a training entity or combining training processes associated with multiple service instances and re-configuring the training entity or training data.

8. The method of claim 1, wherein the service action comprises re-selecting an inferencing entity or combining lists of notification targets associated with multiple service instances at the inferencing entity.

9. The method of claim 1, wherein the one or more AI/ML operations comprise one or more of collecting AI/ML data, training an AI/ML model, updating the AI/ML model, or using the AI/ML model for inferencing.

10. The method of claim 1, wherein the storing further comprises storing information associated with service instances that are using or will use the one or more AI/ML resources or capabilities of the one or more AI/ML resource/capability providers.

11. The method of claim 1, wherein the consumer request further comprises information associated with training data, a model, or an AI/ML resource/capability provider, of the one or more AI/ML resource/capability providers, that is specified by the AI/ML service consumer.

12. The method of claim 1, further comprising:

generating, based on the consumer request, an AI/ML service instance profile comprising information associated with the consumer request, an instance status, context information of the AI/ML service consumer, or information associated with training data, an AI/ML model, or an AI/ML resource/capability provider of the one or more AI/ML resource/capability providers.

13. An apparatus for providing an artificial intelligence/machine learning (AI/ML) enablement service, the apparatus comprising one or more processors and memory storing instructions which, when executed by the one or more processors, cause the apparatus to:

receive, from one or more AI/ML resource/capability providers, one or more provider messages comprising one or more descriptions of one or more AI/ML resources or capabilities of the one or more AI/ML resource/capability providers;

store, by the AI/ML enablement service and based on the one or more provider messages, the one or more descriptions of the one or more AI/ML resources or capabilities;

receive, from an AI/ML service consumer, a consumer request comprising an identifier of the AI/ML service consumer, one or more notification conditions and targets, a required analytics type, or one or more service requirements;

determine, based on processing information associated with the one or more descriptions of the one or more AI/ML resources or capabilities and the consumer request, a service action

comprising selecting an AI/ML resource/capability provider of the one or more AI/ML resource/capability providers to perform AI/ML operations;

send, to the selected AI/ML resource/capability provider, a request to perform one or more AI/ML operations;

receive, from the selected AI/ML resource/capability provider, a response message indicating that the one or more AI/ML operations have been performed and comprising one or more results of the performed AI/ML operations; and

send, to the AI/ML service consumer, a notification message comprising the one or more results of the performed AI/ML operations.

14. The apparatus of claim 13, wherein the provider message is a registration request or a message responsive to a query/request from the AI/ML enablement service.

15. The apparatus of claim 13, wherein the one or more AI/ML resource/capability providers comprise at least one of an AI/ML enablement service, a Vertical Application Layer (VAL) entity, an application/service enablement layer entity, or a core network function.

16. The apparatus of claim 13, wherein the one or more descriptions of the one or more AI/ML resources or capabilities comprise at least of an AI/ML data description, an AI/ML model description, or an AI/ML training/inferencing capability description.

17. The apparatus of claim 13, wherein the AI/ AI/ML service consumer is an AI/ML enablement service, a Vertical Application Layer (VAL) entity, or an application/service enablement layer entity.

18. The apparatus of claim 13, wherein the service action comprises re-selecting:
a training entity or combining training processes associated with multiple service instances and re-configuring the training entity or training data, or
re-selecting an inferencing entity or combining lists of notification targets associated with multiple service instances at the inferencing entity.

19. The apparatus of claim 13, wherein the one or more AI/ML operations comprise one or more of collecting AI/ML data, training an AI/ML model, updating the AI/ML model, or using the AI/ML model for inferencing.

20. The apparatus of claim 13, wherein the storing further comprises storing information associated with service instances that are using or will use the one or more AI/ML resources or capabilities of the one or more AI/ML resource/capability providers.