



(51) International Patent Classification:

G06N 3/04 (2023.01) G06N 20/00 (2019.01)

(21) International Application Number:

PCT/CN2023/112733

(22) International Filing Date:

11 August 2023 (11.08.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant (for CN only): **NOKIA SHANGHAI BELL CO., LTD.** [CN/CN]; No. 388, Ningqiao Road, China (Shanghai) Pilot Free Trade Zone, Pudong New Area, Shanghai 201206 (CN).

(71) Applicant: **NOKIA SOLUTIONS AND NETWORKS OY** [FI/FI]; Karakaari 7, Espoo, 02610 (FI).

(72) Inventors: **FENG, Yijia**; Room# 502, Building# 48, 363 Taierzhuang Road, Pudong New Area, Shanghai 200129 (CN). **YE, Chen Hui**; Room 2402, Building 6, #388 Wan-

ping South Road, Xuhui District, Shanghai 200030 (CN). **HUANG, Yihang**; Room #1304, Building#13, 36 Bin-nan Road, Xuhui District, Shanghai 200235 (CN). **JIAO, Tianyu**; No. 800 Dongchuan Road, Minhang District, Shanghai 200240 (CN).

(74) Agent: **KING & WOOD MALLESONS**; 20th Floor, East Tower, World Financial Centre, No. 1 Dongsanhuan Zhonglu, Chaoyang District, Beijing 100020 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,

(54) Title: UNIVERSAL TASK-BASED MODEL

1500

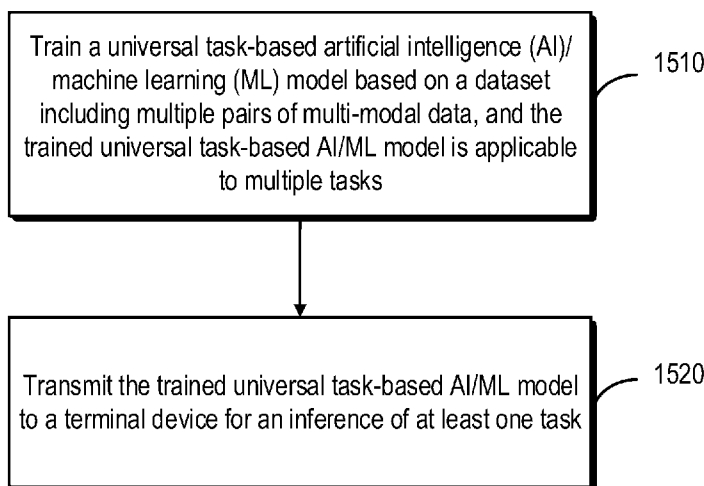


Fig.15

(57) Abstract: Embodiments of the present disclosure disclose devices, methods, apparatuses and a non-transitory computer readable medium for training a universal task-based artificial intelligence (AI)/machine learning (ML) model. In an aspect, a network device may train a universal task-based AI/ML model based on a dataset comprising multiple pairs of multi-modal data, and the trained universal task-based AI/ML model is applicable to multiple tasks. The network device may transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task. Embodiments of the present disclosure can save storage space and reduce the computation resources significantly.

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

UNIVERSAL TASK-BASED MODEL

FIELD

[0001] Various example embodiments relate to the field of communication, and in particular, to devices, methods, apparatuses and a computer readable storage medium for implementing a universal task-based model.

BACKGROUND

[0002] The rapid development of artificial intelligence (AI) and machine learning (ML) technology has provided a significant impact on various fields. For example, AI/ML technology has been deployed in various industries such as healthcare, business, automotive, etc., due to its advances in computing power and continuing breakthroughs in algorithms.

[0003] Meanwhile, for communication systems, in order to support different performance requirements in terms of data rates and reliability, the communication systems have been designed more and more sophisticated. The complexities, as well as desirability of intelligence and automation, in communication systems provide more and more challenges. It is expected for the AI/ML technology to play a crucial role in the communication system, due to its high capability of learning, reasoning, predicting, and perceiving.

SUMMARY

[0004] In general, example embodiments of the present disclosure provide a solution for implementing a universal task-based model.

[0005] In a first aspect, there is provided a network device. The network device may include at least one processor and at least one memory storing instructions that, when executed by the at least one processor, cause the network device to at least: train a universal task-based artificial intelligence (AI)/machine learning (ML) model based on a dataset comprising multiple pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to multiple tasks; and transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

[0006] In a second aspect, there is provided a terminal device. The terminal device may include at least one processor and at least one memory storing instructions that, when executed by the at least one processor, cause the terminal device at least to: receive, from a

network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, wherein the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset comprising multiple pairs of multi-modal data; and perform an inference of at least one task based on the trained universal task-based AI/ML model.

[0007] In a third aspect, there is provided a method. The method includes: training a universal task-based artificial intelligence (AI)/machine learning (ML) model based on a dataset comprising multiple pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to multiple tasks; and transmitting the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

[0008] In a fourth aspect, there is provided a method. The method includes: receiving, from a network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, wherein the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset comprising multiple pairs of multi-modal data; and performing an inference of at least one task based on the trained universal task-based AI/ML model.

[0009] In a fifth aspect, there is provided an apparatus. The apparatus includes: means for training a universal task-based artificial intelligence (AI)/machine learning (ML) model based on a dataset comprising multiple pairs of multi-modal data, wherein the universal task-based AI/ML model is applicable to multiple tasks; and means for transmitting the universal task-based AI/ML model to a terminal device for an inference of at least one task.

[0010] In a sixth aspect, there is provided an apparatus. The apparatus includes: means for receiving, from a network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, wherein the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset comprising multiple pairs of multi-modal data; and means for performing an inference of at least one task based on the trained universal task-based AI/ML model.

[0011] In a seventh aspect, there is provided a non-transitory computer readable medium comprising program instructions for causing an apparatus to perform at least the method in the third or fourth aspect.

[0012] In an eighth aspect, there is provided a computer program comprising instructions, which, when executed by an apparatus, cause the apparatus at least to: train a universal task-

based artificial intelligence (AI)/machine learning (ML) model based on a dataset comprising multiple pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to multiple tasks; and transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

[0013] In a ninth aspect, there is provided a computer program comprising instructions, which, when executed by an apparatus, cause the apparatus at least to: receive, from a network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, wherein the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset comprising multiple pairs of multi-modal data; and perform an inference of at least one task based on the trained universal task-based AI/ML model.

[0014] In a tenth aspect, there is provided a network device. The network device may include: training circuitry configured to train a universal task-based artificial intelligence (AI)/machine learning (ML) model based on a dataset comprising multiple pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to multiple tasks; and transmitting circuitry, configured to transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

[0015] In an eleventh aspect, there is provided a terminal device. The terminal device may include: receiving circuitry configured to receive, from a network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, wherein the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset comprising multiple pairs of multi-modal data; and performing circuitry, configured to perform an inference of at least one task based on the trained universal task-based AI/ML model.

[0016] It is to be understood that the summary section is not intended to identify key or essential features of embodiments of the present disclosure, nor is it intended to be used to limit the scope of the present disclosure. Other features of the present disclosure will become easily comprehensible through the following description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] Some example embodiments will now be described with reference to the accompanying drawings, in which:

[0018] Fig. 1 illustrates a schematic diagram of a communication environment in which a an artificial intelligence (AI)/machine learning (ML) related task may be implemented;

[0019] Fig. 2 illustrates an example of a ML-enabled feature (or task) using a functionality identification (ID) and a Model ID with associated information;

[0020] Fig. 3 illustrates a detailed schematic diagram of a communication system for AI/ML related tasks;

[0021] Fig. 4 illustrates an example signaling process for communicating a trained universal task-based AI/ML model in a communication system according to some embodiments of the present disclosure;

[0022] Fig. 5 illustrates an example signaling process for communicating an AI/ML model in a communication system according to some embodiments of the present disclosure;

[0023] Fig. 6 illustrates an example of a sample “text-image” pair for training the universal task-based AI/ML model according to some embodiments of the present disclosure;

[0024] Fig. 7 illustrates an example of a block diagram of a universal task-based AI/ML model being trained according to embodiments of the present disclosure;

[0025] Fig. 8 illustrates an example implementation of a text encoder in Fig. 7 according to some embodiments of the present disclosure;

[0026] Fig. 9 illustrates an example implementation of an image encoder in Fig. 7 according to some embodiments of the present disclosure;

[0027] Fig. 10 illustrates a process for training the universal task-based AI/ML model according to some embodiments of the present disclosure;

[0028] Fig. 11 illustrates a flowchart of a method for determining a compatible task for the universal task-based AI/ML model according to some embodiments of the present disclosure;

[0029] Fig. 12 illustrates a diagram for determining a compatible task according to some embodiments of the present disclosure;

[0030] Fig. 13 illustrates an example diagram of zero-shot learning for LOS/NLOS classification according to some embodiments of the present disclosure;

[0031] Fig. 14 illustrates an example of fine-tuning a task-oriented AI/ML model for a direct AI/ML positioning according to some embodiments of the present disclosure;

[0032] Fig. 15 illustrates a flowchart of a method implemented at a network device in accordance with some example embodiments of the present disclosure;

[0033] Fig. 16 illustrates a flowchart of a method implemented at a terminal device in accordance with some example embodiments of the present disclosure;

[0034] FIG. 17 illustrates a simplified block diagram of a device that is suitable for implementing some example embodiments of the present disclosure; and

[0035] Fig. 18 illustrates a block diagram of an example of a computer readable medium in accordance with some example embodiments of the present disclosure.

[0036] Throughout the drawings, the same or similar reference numerals represent the same or similar elements.

DETAILED DESCRIPTION

[0037] Principle of the present disclosure will now be described with reference to some example embodiments. It is to be understood that these embodiments are described only for the purpose of illustration and help those skilled in the art to understand and implement the present disclosure, without suggesting any limitation as to the scope of the disclosure. The disclosure described herein can be implemented in various manners other than the ones described below.

[0038] In the following description and claims, unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skills in the art to which this disclosure belongs.

[0039] References in the present disclosure to “one embodiment,” “an embodiment,” “an example embodiment,” and the like indicate that the embodiment described may include a particular feature, structure, or characteristic, but it is not necessary that every embodiment includes the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that it is within the knowledge of one skilled in the art to affect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described.

[0040] It shall be understood that although the terms “first” and “second” etc. may be used herein to describe various elements, these elements should not be limited by these terms.

These terms are only used to distinguish one element from another. For example, a first element could be termed a second element, and similarly, a second element could be termed a first element, without departing from the scope of example embodiments. As used herein, the term “and/or” includes any and all combinations of one or more of the listed terms.

[0041] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of example embodiments. As used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises”, “comprising”, “has”, “having”, “includes” and/or “including”, when used herein, specify the presence of stated features, elements, and/or components etc., but do not preclude the presence or addition of one or more other features, elements, components and/ or combinations thereof. As used herein, “at least one of the following: <a list of two or more elements>” and “at least one of <a list of two or more elements>” and similar wording, where the list of two or more elements are joined by “and” or “or”, mean at least any one of the elements, or at least any two or more of the elements, or at least all the elements.

[0042] As used in this application, the term “circuitry” may refer to one or more or all of the following:

- (a) hardware-only circuit implementations (such as implementations in only analog and/or digital circuitry) and

- (b) combinations of hardware circuits and software, such as (as applicable):

- (i) a combination of analog and/or digital hardware circuit(s) with software/firmware and

- (ii) any portions of hardware processor(s) with software (including digital signal processor(s)), software, and memory(ies) that work together to cause an apparatus, such as a mobile phone or server, to perform various functions) and

- (c) hardware circuit(s) and or processor(s), such as a microprocessor(s) or a portion of a microprocessor(s), that requires software (for example, firmware) for operation, but the software may not be present when it is not needed for operation.

[0043] This definition of circuitry applies to all uses of this term in this application, including in any claims. As a further example, as used in this application, the term circuitry also covers an implementation of merely a hardware circuit or processor (or multiple

processors) or portion of a hardware circuit or processor and its (or their) accompanying software and/or firmware. The term circuitry also covers, for example and if applicable to the particular claim element, a baseband integrated circuit or processor integrated circuit for a mobile device or a similar integrated circuit in server, a cellular network device, or other computing or network device.

[0044] As used herein, the term “communication network” refers to a network following any suitable communication standards, such as Long Term Evolution (LTE), LTE-Advanced (LTE-A), Wideband Code Division Multiple Access (WCDMA), High-Speed Packet Access (HSPA), Narrow Band Internet of Things (NB-IoT) and so on. Furthermore, the communications between a terminal device and a network device in the communication network may be performed according to any suitable generation communication protocols, including, but not limited to, the fourth generation (4G), 4.5G, the future fifth generation (5G) communication protocols, the future sixth generation (6G) communication protocols, and/or any other protocols either currently known or to be developed in the future. Embodiments of the present disclosure may be applied in various communication systems. Given the rapid development in communications, there will of course also be future type communication technologies and systems with which the present disclosure may be embodied. It should not be seen as limiting the scope of the present disclosure to only the aforementioned system.

[0045] As used herein, the term “network device” or “network node” refers to a node in a communication network via which a terminal device accesses the network and receives services therefrom. The network device may refer to a system simulator, a base station (BS) or an access point (AP), for example, a node B (NodeB or NB), an evolved NodeB (eNodeB or eNB), a NR NB (also referred to as a gNB), a Remote Radio Unit (RRU), a radio header (RH), a remote radio head (RRH), a relay, a low power node such as a femto, a pico, and so forth, depending on the applied terminology and technology.

[0046] The term “terminal device” refers to any end device that may be capable of wireless communication. By way of example rather than limitation, a terminal device may also be referred to as a communication device, user equipment (UE), a Subscriber Station (SS), a Portable Subscriber Station, a Mobile Station (MS), or an Access Terminal (AT). The terminal device may include, but not limited to, a mobile phone, a cellular phone, a smart phone, voice over IP (VoIP) phones, wireless local loop phones, a tablet, a wearable terminal device, a personal digital assistant (PDA), portable computers, desktop computer, image

capture terminal devices such as digital cameras, gaming terminal devices, music storage and playback appliances, vehicle-mounted wireless terminal devices, wireless endpoints, mobile stations, laptop-embedded equipment (LEE), laptop-mounted equipment (LME), USB dongles, smart devices, wireless customer-premises equipment (CPE), an Internet of Things (IoT) device, a watch or other wearable, a head-mounted display (HMD), a vehicle, a drone, a medical device and applications (for example, remote surgery), an industrial device and applications (for example, a robot and/or other wireless devices operating in an industrial and/or an automated processing chain contexts), a consumer electronics device, a device operating on commercial and/or industrial wireless networks, and the like. In the following description, the terms “terminal device”, “communication device”, “terminal”, “user equipment” and “UE” may be used interchangeably.

[0047] For artificial intelligence (AI)/machine learning (ML) related cases (e.g., positioning cases), task A (or feature A) (e.g., AI/ML direct positioning) may involve a first model with a first model identification (ID), and task-B (or feature B) (AI/ML-assisted positioning) may involve a second model with a second model ID. Fig. 1 illustrates a schematic diagram of a communication environment 100 in which a AI/ML related task may be implemented. As shown in Fig. 1, the communication environment 100, which may also be referred to as a communication network 100 or a communication system 100, may include a terminal device 110, a network (e.g. radio access network (RAN)) 120 and a network device 130.

[0048] The network 120 may implement any appropriate communication technology to provide access to the terminal device 110. The network device 130 may be, but not limited to, a Mobility Management Function (AMF), a Location Management Function (LMF), a Network Data Analytics Function (NWDAF), and so forth.

[0049] Although the network device 130 and the terminal device 110 are described in the communication environment 100 of Fig. 1, any other suitable communication devices in communication with one another may also be applied herein. In this regard, it is noted that although the network device 130 is schematically depicted as LMF and the terminal device 110 is schematically depicted as a mobile phone in Fig. 1, it is understood that these depictions are exemplary in nature without suggesting any limitation. In other words, the network device 130 and the terminal device 110 may be any other communication devices, for example, any other wireless communication devices.

[0050] In the following description, the network device 130 may be described with reference to as LMF, and the LMF 130 may be located in a base station or a RAN node or a core network of the communication system. However, it should be understood that, the network device 130 is not limited to the LMF, any other appropriate network device may be applied to be herein as the network device 130.

[0051] It is to be understood that the particular number of various communication devices, the particular number of various communication links, and the particular number of other elements as shown in Fig. 1 is for illustration purpose only without suggesting any limitations. The communication environment 100 may include any suitable number of communication devices, any suitable number of communication links, and any suitable number of other elements adapted for implementing communications. In addition, it should be appreciated that there may be various wireless as well as wireline communications (if needed) among all of the communication devices.

[0052] Communications among devices in the communication environment 100 may be implemented according to any appropriate communication protocol(s), including, but not limited to, cellular communication protocols of the third generation (3G), the fourth generation (4G) and the fifth generation (5G), the sixth generation (6G), and on the like, wireless local network communication protocols such as Institute for Electrical and Electronics Engineers (IEEE) 802.11 and the like, and/or any other protocols currently known or to be developed in the future. Moreover, the communication may utilize any appropriate wireless communication technology, comprising but not limited to: Code Division Multiple Access (CDMA), Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Frequency Division Duplex (FDD), Time Division Duplex (TDD), Multiple-Input Multiple-Output (MIMO), Orthogonal Frequency Division Multiple (OFDM), Discrete Fourier Transform spread OFDM (DFT-s-OFDM) and/or any other technologies currently known or to be developed in the future

[0053] Functionality-transferability in the artificial intelligence (AI)/machine learning (ML) field has been emerged. In a narrow sense, functionality-transferability may refer to domain adaptation, i.e., an AI/ML model which has been trained in scenario-A requires model fine-tuning to fit scenario-B. In a broad sense, AI/ML model functionality-transferability may refer to task adaptation, i.e., an AI/ML model which has been trained for task-A requires model fine-tuning to fit task B. Generally, task-adaptation is more challenging than domain-adaptation.

[0054] Taking AI-positioning for an example, there have been discussions in the third generation partnership project (3GPP). Some agreements have been reached in radio access network (RAN) workgroup (WG) 1 with respect to using different functionality identifications (IDs) and model IDs to discriminate different tasks. Fig. 1 illustrates an example of a ML-enabled feature (or task) using a functionality identification (ID) and a Model ID with associated information.

[0055] Fig. 2 illustrates an example of a ML-enabled feature (or task) using a functionality identification (ID) and a Model ID with associated information. As shown in Fig. 2, a block 200 may be related to a ML-enabled feature (or task) including, but not limited to, channel state information (CSI) compression with two-sided model, CSI prediction with user equipment (UE)-sided model, CSI prediction with two-sided model, etc. In block 100, three sub-blocks 210-230 are shown with each block having a corresponding functionality ID. For example, the sub-block 210 has a functionality ID of #1, the sub-block 220 has a functionality ID of #2, and the sub-block 230 has a functionality ID of #3. Each sub-block with a particular ID is unique within the feature with associated information and is optimized for a corresponding condition. For example, the sub-block 210 with functionality ID of #1 is optimized for an indoor condition, the sub-block 220 with functionality ID of #2 is optimized for an outdoor condition, and the sub-block 230 with functionality ID of #3 is optimized for a base station (BS) configuration.

[0056] In each sub-block, there are multiple models with respective model IDs. For example, a model with a model ID#1.1, a model with a model ID #1.2, and a model with a model ID #1.3 are shown in the sub-block 210. A model with a model ID#2.1, a model with a model ID #2.2, and a model with a model ID #2.3 are shown in the sub-block 220. A model with a model ID#3.1, a model with a model ID #3.2, and a model with a model ID #3.3 are shown in the sub-block 230. As shown in Fig. 2, the model with the model ID#1.1 is active for an ML-enabled feature, e.g., CSI prediction.

[0057] In addition, for a particular task or feature, different configurations may correspond to different functionality ID. Taking a direct positioning task for example, there may be multiple functionality IDs for different configurations. For example, Functionality 1-01 for the direct positioning task may correspond to a configuration of 64 antenna elements, 2 antenna ports, and 12 transmission and reception points (TRPs), Functionality 1-02 for the direct positioning task may correspond to a configuration of 128 antenna elements, 2 or 4 antenna ports, and N TRPs ($1 \leq N \leq 18$). Similarly, for an assisted positioning task, multiple

functionality IDs may be for different configurations, respectively. For example, Functionality 2-01 for the assisted positioning task may correspond to a configuration of intermediate feature being a time of arrival (TOA), 128 antenna elements, 1 antenna port, and 15 TRPs, and Functionality 2-02 for the assisted positioning task may correspond to a configuration of intermediate feature being a line of sight (LOS)/non line of sight (NLOS) indication, 128 antenna elements, 24 antenna ports, and N TRPs ($1 \leq N \leq 18$).

[0058] It is encouraged to propose in 3GPP about functionality and information elements of AI/ML functionality identification for AI/ML based positioning with UE-side model. Some agreements have been achieved as following, in RAN WG1#112:

Agreement (RAN WG1#112)

For UE-side models and UE-part of two-sided models:

- For AI/ML functionality identification
 - o Reuse legacy 3GPP framework of Features as a starting point for discussion.
 - o UE indicates supported functionalities/functionality for a given sub-use-case.
 - UE capability reporting is taken as starting point.
- For AI/ML model identification
 - o Models are identified by model ID at the Network. UE indicates supported AI/ML models.
- In functionality-based LCM
 - o Network indicates activation/deactivation/fallback/switching of AI/ML functionality via 3GPP signaling (e.g., RRC, MAC-CE, DCI).
 - o Models may not be identified at the Network, and UE may perform model-level LCM.
 - Study whether and how much awareness/interaction NW should have about model-level LCM
- In model-ID-based LCM, models are identified at the Network, and Network/UE may activate/deactivate/select/switch individual AI/ML models via model ID.

FFS: Relationship between functionality identification and model identification

FFS: Performance monitoring and RAN4 impact

FFS: detailed understanding on model

Agreement (RAN1#112)

Regarding AI/ML model inference, to study the potential specification impact (including the feasibility, and the necessity of specifying AI/ML model input and/or output) at least for the following aspects for AI/ML based positioning accuracy enhancement

- For direct AI/ML positioning (Case 2b and 3b), type of measurement(s) as model inference input considering performance impact and associated signaling overhead
 - o Potential new measurement: CIR/PDP
 - o existing measurement: e.g., RSRP/RSRPP/RSTD
 - o Note1: details of potential new measurement and/or potential enhancement to existing measurement is to be studied
 - o Note2: study the impact of model input for other cases are not precluded
- For AI/ML assisted positioning with UE-assisted (Case 2a) and NG-RAN node assisted positioning (Case 3a), measurement report to carry model output to LMF
 - o new measurement report: e.g., ToA, path phase
 - o existing measurement report: e.g., RSTD, LOS/NLOS indicator, RSRPP
 - o enhancement of existing measurement report: e.g., soft information/high resolution of RSTD
- Assistance signaling and procedure to facilitate model inference for both UE-side and Network-side model
 - o RS configurations
 - o Other assistance information is not precluded

Note: Companies are encouraged to report their assumption of functionality and their assumption of information element(s) of AI/ML functionality identification for AI/ML based positioning with UE-side model (Case 1 and 2a).

[0059] Fig. 3 illustrates a detailed schematic diagram of a communication system 300 for AI/ML related tasks. As shown in Fig. 3, the terminal device 110 may transmit a first task-oriented model (e.g., neural network (NN)) request related to a first task to the LMF 330. The LMF 330 may provide a NN1 (e.g., specific-NN-1 311) specific to the first task as requested by the terminal device 310. The terminal device 310 may pre-train the received specific-NN (e.g., specific-NN-1 311) by using a large volume of dataset, fine-tune the specific-NN on the first task-specific data (e.g., K1 L-volume data-1 312) with the first task-specific objectives, and perform the task inference by using the fine-tuned NN (e.g., specific-NN-1 311).

[0060] Similarly, the terminal device 110 may transmit a second task-oriented model (e.g., neural network (NN)) request related to a second task to the LMF 330. The LMF 330 may provide a NN2 (e.g., specific-NN-2 313) specific to the second task as requested by the terminal device 310. The terminal device 310 may pre-train the received specific-NN (e.g., specific-NN-2 312) by using a large dataset, fine-tune the specific-NN on the second task-specific data (e.g., K2 L-volume data-2 314) with the second task-specific objectives, and perform the task inference by using the fine-tuned NN (e.g., specific-NN-2 313).

[0061] Similarly, the terminal device 310 may transmit a third task-oriented model (e.g., neural network (NN)) request related to a third task to the LMF 330. The LMF 330 may provide a NN3 (e.g., specific-NN-3 315) specific to the third task as requested by the terminal device 310. The terminal device 310 may pre-train the received specific-NN (e.g., specific-NN-3 315) by using a large volume of dataset, fine-tune the specific-NN on the third task-specific data (e.g., K3 L-volume data-3 316) with the third task-specific objectives, and perform the task inference by using the fine-tuned NN (e.g., specific-NN-3 315).

[0062] However, one of the drawbacks in the framework 300 as shown in Fig. 3 lies in: each NN is specific to a task. Accordingly, for performing one task, a task-specific NN is downloaded to the terminal device 110. There have been difficulties in providing a unified solution to accommodate two or more of the first, second, or third tasks, and more other tasks.

[0063] In addition, as shown in Fig. 3, training a task-specific NN for a task may typically involves two-step processes: a pre-training process and a fine-tune process. The pre-training process may involve training the model on a large corpus of data to learn general information and capture contextual information. The fine-tune process may involve training the pre-trained model on the task-specific data with task-specific objectives. Fine-

tuning a pre-trained model to a specific task keeps the overall architecture, but needs to update the holistic network parameters with a task-specific objective. Accordingly, if various tasks are required, multiple models may be fine-tuned and stored, which would consume a large amount of storage and computation resources.

[0064] It's noteworthy that, for multiple AI/ML related tasks, e.g., AI/ML positioning cases, for a given environment, AI-assisted and AI-direct positioning tasks are intrinsically coherent. However, in legacy 3GPP, all the downstream tasks are considered independently, thus, training individually AI/ML models for different downstream tasks wastes the fundamental common features and leads to superfluous training-effort as well as model-storage. On the other hand, even though an AI/ML model may be generalized to multiple downstream tasks, using the conventional two-step training processes in model adaptation needs to update the holistic network parameters for each task with task-specific dataset and thus multiple models should be fine-tuned and stored, which would consume large amount of storage and computation resources.

[0065] It is desirable to provide an AI/ML model generalized to multiple downstream tasks as well as an efficient training implementation for the generalized AI/ML model, such that the storage space can be saved and the computation resources can be significantly reduced.

[0066] In view of the above discussions and analysis, example embodiments of the present disclosure provide a solution of providing an efficient approach for providing a signaling framework for a universal-task AI/M based-model. In some embodiment, a network device may train a universal task-based artificial intelligence (AI)/machine learning (ML) model based on a dataset comprising multiple pairs of multi-modal data, and the trained universal task-based AI/ML model is applicable to multiple tasks. The network device may transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task. By employing a method for training a universal-task AI/M based-model as well as a signaling framework, the storage space can be saved and the computation resources can be significantly reduced.

[0067] In some embodiments, a trained universal-task AI/M based-model may be deployed at a network device, and the network device may transmit the trained universal-task AI/M based-model to a terminal device in response to a request for an AI/ML model. Fig. 4 illustrates an example signaling process 400 for communicating a universal task-based AI/ML model in a communication system according to some embodiments of the present

disclosure. For the purpose of discussion, the process 400 will be described with reference to Fig. 1. The process 400 may involve a terminal device. The process 400 may further involve a network device. In some embodiments, the terminal device in Fig. 4 may be the terminal device 110 as shown in Fig. 1. In some other embodiments, the network device in Fig. 4 may be the LMF 130 as shown in Fig. 1. For ease of discussions, various embodiments in the following will be described in an example scenario that the network device is the LMF 130.

[0068] It may be understood that although the process 400 is described in combination with the communication network environment 100 of Fig. 1, the process 400 may be likewise applied to other scenarios than the communication system 100. Furthermore, in the process 400, it is possible to add, omit, modify one or more operations, or the operations may also be performed in any suitable order without departing from the scope of the present disclosure.

[0069] In the process 400, before the terminal device 110 request an AI/ML model, the LMF 130 may train (401) a universal task-based AI/ML model in advance. In some embodiments, the trained universal task-based AI/ML model may be applicable to multiple tasks, including, but not limited to, a task of direct AI/ML positioning, a task of AI/ML-assisted positioning, a task of AI/ML beam management, and so on. The universal task-based AI/ML model is trained to learn general knowledge about intrinsic characteristics of radio frequency (RF) propagation environment and may be generalized to multiple downstream tasks, such as direct positioning, assisted positioning, beam management, and so on. A more detailed description of training the universal task-based model will be described in combination with accompanying figures.

[0070] The LMF 130 may record (402) a model identification (ID) of the trained universal task-based AI/ML model and an identification of a compatible task. In some embodiments, the universal task-based AI/ML model may be applicable to multiple tasks. In other words, the trained universal task-based AI/ML model may be used for inferences of multiple tasks. A task that is applicable to or compatible with the trained universal task-based AI/ML model may be referred to be as a compatible task. The LMF 130 may associate respective task identifications of multiple tasks with the trained universal task-based AI/ML model based on the multiple tasks being compatible with the trained universal task-based AI/ML model. The LMF 130, may record the model identification of the trained universal task-based AI/ML model as well as identifier(s) of one or more compatible tasks. For example, a model identification of the trained universal task-based AI/ML model may be recorded as #1, and

the trained universal task-based AI/ML model may have three comparable tasks, with identifications of: Task-1, Task-2, and Task-3. Then the identifications of: Task-1, Task-2, and Task-3 are associated with the model identification #1 of the trained universal task-based AI/ML model.

[0071] It should be understood that, although, as shown in Fig. 4, the LMF 130 trains the universal task-based AI/ML model and records the model identifications and compatible task identifier, another entity may train the universal task-based AI/ML model and record related identifications, and transmit the trained universal task-based AI/ML model and related identifications to the LMF 130.

[0072] The terminal device 110 may transmit (403) capability information of the terminal device to the LMF 130. The capability information of the terminal device may include, but not limited to, at least one of operations per second (FLOPs), power constraints, a processor requirement (indicating whether a CPU or a GPU is required), computation capacity, storage capacity of the terminal device. For example, the terminal device 110 may report its maximum memory, maximum FLOPs the terminal device may provide, and other aspects including but not limited to, power constraints and device requirements indicating whether a CPU or a GPU is required. The capability information may be used by the LMF 130 to select the trained universal task-based AI/ML model or a specific-NN model from multiple AI/ML models, which may be described in detail below in combination with accompanying drawing.

[0073] The terminal device 110 may transmit (404) a request for an AI/ML model for one or more tasks, and the request may include an identification of each of the one or more tasks. For example, the terminal device 110 may request one or more AI/ML models for a cluster of tasks. The requested AI/ML models may be either a cluster of task-specific AI/ML models with each task-specific AI/ML model for each requested task, or the trained universal task-based AI/ML model for the cluster of tasks. In some embodiments, although shown separately, the capability information transmitted at 403 may be included in the request for an AI/ML model for one or more tasks. In other words, the request including the capability information and the identifications of the cluster of task may be transmitted (404) to the LMF 130.

[0074] In some embodiments, the request may include an identification of each of the one or more tasks. For example, assume a task of direct positioning has an identification of

Task-1, a task of assisted positioning has an identification of Task-2, and a task of beam management has an identification of Task-3, and the request for an AI/ML model for a cluster of tasks including direct positioning, assisted position, and beam management may include identifications of Task-1, Task-2, Task-3. The example is only for the purposes of illustration, and an identification of a task may include any appropriate formats or representations.

[0075] In some embodiments, the tasks, as indicated by the request, will be input to the requested AI/ML model and a result of the task will be output from the AI/ML model. For example, if the task is about direct positioning, the result of the task may output from the AI/ML model may include a coordinate of a location.

[0076] The LMF 130 may determine (405) one or more candidate AI/ML models that are applicable to the one or more tasks as indicated by the request among multiple AI/ML models, based on the identifications of the one or more tasks. In some embodiments, the determined candidate AI/ML model may include the trained universal task-based AI/ML model as trained at 401. In some embodiments, the determined candidate AI/ML model may not include the trained universal task-based AI/ML model.

[0077] In some embodiments, the LMF 130 may determine (405) the one or more candidate AI/ML models based on the identifications of the tasks as indicated by the request and the model identifications of the multiple AI/ML models. For example, the LMF 130 may compare a task identification of a task in the one or more tasks as indicated by the request to one or more task identifications of respective task identifications associated with the multiple AI/ML models. The LMF 130 may determine a candidate AI/ML model based on the task identification of the task matching with an identification of a task associated with the multiple AI/ML models. For example, if a task identification of a task indicated by the request is Task-1, an AI/ML model #1 have an associated identification Task-1, then the LMF 130 may determine the first AI/ML model #1 as a candidate AI/ML model.

[0078] The LMF 130 may select (406) the trained universal task-based AI/ML model or a specific-NN model from the candidate AI/ML models based on the capability information of the terminal device 110. Specifically, the LMF 130 may determine an operation indicator of each of the determined candidate AI/ML models, and the operation indicator may include one or more of FLOPs, a power constrain, a device requirement, or storage size of each corresponding candidate AI/ML model. The terminal device 110 may select the trained

universal task-based AI/ML model or a specific-NN model from the determined candidate AI/ML models based on an operation indicator of the trained universal task-based AI/ML model or the specific-NN model matching with the capability information of the terminal device. A detailed explanation of selecting the trained universal task-based AI/ML model will be described in combination with Fig. 5 below.

[0079] The LMF 130 may transmit (407), to the terminal device 110, an indication that indicates the trained universal task-based AI/ML model, for example, indicating the trained universal task-based AI/ML model is selected. In some embodiments, as described above, the trained universal task-based AI/ML model is selected from the multiple AI/ML models based on the identification of the one or more tasks as indicated by the request and the capability information of the terminal device 110.

[0080] The LMF 130 may transmit (408) the trained universal task-based AI/ML model to the terminal device 110. In some embodiments, the LMF 130 may transmit information about a model architecture of the trained universal task-based AI/ML model and parameter values for the parameters of the trained universal task-based AI/ML model to the terminal device 110.

[0081] In some embodiments, although shown separately, the indication at 407 and the trained universal task-based AI/ML model at 408 may be received in a response (e.g., a response message) to the request for the AI/ML model for the one or more tasks.

[0082] The terminal device 110 may perform (409) an inference of the one or more tasks indicated by the request at least based on the received trained universal task-based AI/ML model. In some embodiments, the terminal device 110 may perform an inference based on the trained universal task-based AI/ML model (i.e., zero-shot learning without a cascaded AI/ML model) or based on the trained universal task-based AI/ML model cascaded with a fine-tuned task-oriented AI/ML model. Methods of determining if a task-oriented AI/ML model is to be fine-tuned will be described in the following in combination with accompany figures.

[0083] Advantageously, as a terminal device generally handles multiple AI/ML tasks (like AI/ML-assisted positioning and direct AI/ML positioning) according to its wireless environment complexity, mobility, etc., providing the terminal device with the universal-task based-model can improve operation efficiency of model management as well as reducing storage space.

[0084] In some embodiments, the indication transmitted by the LMF 130 at 407 may also indicate that a task-oriented AI/ML model is to be fine-tuned and cascaded with the trained universal task-based AI/ML model to generate a cascaded model for a task in the one or more tasks as indicated by the request. The task-oriented AI/ML model may include a NN that may be specific to a corresponding task in the one or more tasks as indicated by the request. The task-oriented AI/ML model may be smaller in size compared with the trained universal task-based AI/ML model and may be trained or fine-tuned with a relatively small volume of task-specific dataset. For example, if a task-oriented AI/ML model needs to be trained for a task 1, the task-oriented AI/ML model may be fine-tuned on task 1-specific dataset. A detailed description with respect to the task-oriented AI/ML model will be described in combination with Fig. 5.

[0085] Fig. 5 illustrates an example signaling process 500 for communicating an AI/ML model in a communication system according to some embodiments of the present disclosure. For the purpose of discussion, the process 500 will be described with reference to Fig. 1. The process 500 may involve a terminal device. The process 500 may further involve a network device. In some embodiments, the terminal device in Fig. 5 may be the terminal device 110 as shown in Fig. 1. In some other embodiments, the network device in Fig. 5 may be the LMF 130 as shown in Fig. 1. For ease of discussions, various embodiments in the following will be described in an example scenario that the network device is the LMF 130.

[0086] The operations 501-506 are similar to these operations 401-406 as shown in Fig. 4 and may be understood with reference to description for 401-406, thus, the repetitive description of operations 501-506 is omitted here for the purposes of clarity and brevity.

[0087] At 507, if the LMF 130 determines that the trained universal task-based AI/ML model is selected, the LMF 130 may further determine, at 508, if a task-oriented AI/ML model is to be fine-tuned and cascaded with the universal task-based AI/ML model, so as to generate a cascaded model for a task in the one or more tasks as indicated by the request. In some embodiments, the LMF 130 may evaluate each task as indicated by the request with its one or more key performance indicators (KPIs) using a fine-tune detector to determine whether the task (e.g., task k) requires fine-tuning a corresponding task-oriented AI/ML model (e.g., task-oriented-NN-k) or may be directly inferred from Zero-Shot Learning (ZSL) without fine-tuning a corresponding task-oriented AI/ML model (e.g., task-oriented-NN-k).

[0088] In some embodiments, the LMF 130 may determine if a task-oriented AI/ML model is to be fine-tuned and cascaded with the trained universal task-based AI/ML model based on at least one of: a target of a corresponding task in the one or more tasks as indicated by the request received from the terminal device; or a key performance indicator (KPI) for the corresponding task. For example, for a corresponding task K, the LMF 130 may determine if a task-oriented AI/ML model (e.g., task-oriented-NN-K) is to be fine-tuned and cascaded with the trained universal task-based AI/ML model based on at least one of: a target of the corresponding task K; or a KPI for the corresponding task K. In some embodiments, the target for the corresponding task may include a result of the corresponding task or an output configuration of the corresponding task. The result may indicate an output of an AI/ML model processing the corresponding task. Different task may correspond to different results. In some embodiments, the result may include a coordinate indicating a position of an object or a classification result indicating a LOS or NLOS classification. It should be understood that, the result may include other formats or configurations, and not limited to the examples as described above.

[0089] As shown in Fig. 5, after the operation 508, the LMF 130 may transmit an indication to the request at 509 to the terminal device 110. The indication may indicate the trained universal task-based, AI/ML model, for example, that the trained universal task-based AI/ML model is selected. In some embodiments, the indication at 509 is similar to the indication at 407 in the signaling process 400 and may indicate the trained universal task-based AI/ML model. If the LMF 130 determines that a task-oriented AI/ML model is to be fine-tuned and cascaded with the universal task-based AI/ML model to generate a cascaded model for a task in the one or more tasks as indicated by the request, the indication may further indicate that a task-oriented AI/ML model is to be fine-tuned and cascaded with the trained universal task-based AI/ML model to generate a cascaded AI/ML model for a task in the at least one task. In other words, if the LMF 130 determines that a task-oriented AI/ML model is to be fine-tuned and cascaded with the universal task-based AI/ML model, the indication at 509 may indicate the trained universal task-based AI/ML model (for example, the trained universal task-based AI/ML model is selected) and that a task-oriented AI/ML model is to be fine-tuned and cascaded with the universal task-based AI/ML model.

[0090] At 510, if a task-oriented AI/ML model is to be fine-tuned for a corresponding task, the LMF 130 may transmit, to the terminal device 110, a request for a location for fine-tuning the task-oriented AI/ML model. The terminal device 110 may transmit, at 512, a response

(e.g., a response message in response to the request transmitted at 511) indicating the location where the task-oriented AI/ML model is to be fine-tuned. In some embodiments, the location may be the terminal device 110, that is, the terminal device 110 may fine-tune the task-oriented AI/ML model. Alternatively, the location may be the LMF 130, that is, the LMF 130 may fine-tune the task-oriented AI/ML model.

[0091] If the response from the terminal device 110 indicates the task-oriented AI/ML model is to be fine-tuned at the LMF 130, as case A shown in 513, the LMF 130 may fine-tune the task-oriented AI/ML model at 514, based on the response indicating the LMF 130 as the location for fine-tuning the task-oriented AI/ML model. In some embodiments, the LMF 130 may fine-tune the task-oriented AI/ML model using task-specific dataset maintained in the LMF 130.

[0092] The LMF 130 may transmit the fined-tuned task-oriented AI/ML model to the terminal device 110 as well as the trained universal task-based AI/ML model at 515. In some embodiments, the LMF 130 may transmit the trained universal task-based AI/ML model including information about a model architecture and/or one or more parameter values for parameters of the trained universal task-based AI/ML model to the terminal device. In some embodiments, the LMF 130 may transmit the fined-tuned task-oriented AI/ML model including information about a model architecture of the fined-tuned task-oriented AI/ML model and/or one or more parameter values for parameters of the fined-tuned task-oriented AI/ML model to the terminal device 110. In some embodiments, the LMF 130 may transmit the trained universal task-based AI/ML model cascaded with the fined-tuned task-oriented AI/ML model to the terminal device 110, for example, in a response to the request for the AI/ML model for one or more tasks. In some embodiments, the LMF 130 may transmit the trained universal task-based AI/ML model and the fined-tuned task-oriented AI/ML model separately to the terminal device 110, and the terminal device 110 may cascade the fined-tuned task-oriented AI/ML model with the trained universal task-based AI/ML model.

[0093] Alternatively, if the response from the terminal device 110 indicates the task-oriented AI/ML model is to be fine-tuned at the terminal device 110, as case B shown in 516, the LMF 130 may transmit, at 517, the trained universal task-based AI/ML model including information on a model architecture and/or one or more parameter values for parameters of the trained universal task-based AI/ML model to the terminal device 110. The terminal device 110 may fine-tune the task-oriented AI/ML model at 518. The fine-tuning process may include using a task-specific dataset to fine-tune the task-oriented AI/ML model. In

some embodiments, the terminal device 110 may fine-tune the task-oriented AI/ML model using task-specific dataset maintained in the terminal device 110. In some embodiments, if the task-oriented AI/ML model has not been deployed at the terminal device 110, the LMF 130 may transmit the task-oriented AI/ML model before the fine-tuning process. In some embodiments, the task-oriented AI/ML model and the trained universal task-based AI/ML model may be transmitted in a response to the request for the AI/ML model for one or more tasks.

[0094] In some embodiments, the terminal device 110 may perform an inference based on the trained universal task-based AI/ML model cascaded with the fine-tuned task-oriented AI/ML model. For example, in Case A, after receiving the fine-tuned task-oriented AI/ML model at 515, the terminal device 110 may perform an inference based on the trained universal task-based AI/ML model cascaded with the fine-tuned task-oriented AI/ML model. In case B, after fine-tuning the task-oriented AI/ML model at 518, the terminal device 110 may perform an inference based on the trained universal task-based AI/ML model cascaded with the fine-tuned task-oriented AI/ML model.

[0095] A detailed explanation of a signaling framework for a trained universal task-based AI/ML model has been described with reference to Fig. 4 and Fig. 5. According to the signaling processes 400 or 500, the network device 130 may train the universal task-based AI/ML model at 401 or 501. The network device 130 may transmit (for example, at 408 in the signaling process 400, or at 515 or 517 in the signaling process 500) the trained universal task-based AI/ML model to the terminal device 110 for an inference of one or more tasks. As described above, the trained universal task-based AI/ML model is applicable to multiple tasks and may be used for predicting or inferring one or more tasks. In some embodiments, the network device 130 may transmit the trained universal task-based AI/ML model directly to the terminal device 110 for an inference of one or more tasks. Alternatively, the network device 130 may transmit the trained universal task-based AI/ML model to the terminal device 110 via one or more intermediate devices. The present disclosure does not limit a specific transmission path of a transmission of the trained universal task-based.

[0096] In some embodiments, the network device 130 may transmit the trained universal task-based AI/ML model to the terminal device 110 in response to a request from the terminal device, such as the request transmitted by the terminal device 110 at 404 in the signaling process 400 or at 504 in the signaling process 500.

[0097] The terminal device 110 may perform an inference of at least one task based on the trained universal task-based AI/ML model, such as at the operation shown at 409 in the signaling process 400. In some embodiments, the terminal device 110 may perform an inference based on the trained universal task-based AI/ML model (i.e., zero-shot learning without a cascaded AI/ML model) or based on the trained universal task-based AI/ML model cascaded with a fine-tuned task-oriented AI/ML model. Methods of determining if a task-oriented AI/ML model is to be fine-tuned have been disclosed above, the detailed description is omitted here for the purposes of clarity and brevity.

[0098] A detailed description with respect to a method of training a universal task-based AI/ML model at block 401 in the signaling process 400 or in the block 501 at the signaling process 500 will be provided with accompanying figures. In other words, the training process as described below is a detailed implementation of operations in block 401 or in block 501. The trained universal task-based AI/ML model may be deployed at the network device 130.

[0099] In some embodiments, the network device 130 may train the universal task-based AI/ML model based on a dataset including multiple pairs of multi-modal data, and the trained universal task-based AI/ML model is applicable to multiple tasks, for example, multiple types of tasks, including, but not limited to, direct positioning, assisted positioning, or beam management, etc.

[00100] In some embodiments, the universal task-based AI/ML model is trained by using a dataset including multiple pairs of multi-modal data. For example, for each pair, there may include first data of a first modality, and second data of a second modality. In some embodiments, the first modality may be a text and the second modality may be an image, and this pair may be indicated as a “text-image” pair. The text in a “text-image” pair may include text information with descriptions of physical environment (e.g., LOS/NLOS classification, BS location, terminal device location, environment classification, etc.). The image in a “text-image” pair may include image information of the wireless channel features corresponding to the text description. In some embodiments, the text information is corresponding to the image information in a “text-image” pair, and in other words, the image is another format for representing the text information. In some embodiments, text information may be matrixed and represented in an image format.

[00101] Fig. 6 illustrates an example of a sample “text-image” pair for training the universal task-based AI/ML model according to some embodiments of the present disclosure. As shown in Fig. 6, the text-image pair includes a text 610 with descriptions of physical environment (e.g., LOS/NLOS classification, BS location, terminal device location, environment classification, etc.) in a text form, and the image 620 represent wireless channel features in an image form. The wireless channel features may include, but not limited to, channel state information (CSI), channel impulse response (CIR), power delay profile (PDP), etc. The text 610 and the image 620 are corresponding to each other, and may be referred as a “positive pair”. While a text and an image not associated with or correspond to the text may be referred to be as “negative pair”. By leveraging the “positive pair” and “negative pair” as labels, the ground truth relationship between each text description and wireless channel image may be derived. The training objective is to maximize the similarity between positive pairs and minimize the similarity between negative pairs in the shared embedding space.

[00102] It should be understood that, the example “text-image” pair as shown in Fig. 6 is only for the purposes of illustration, and the dataset for training the universal task-based AI/ML model may include any appropriate format and/or modality with any appropriate descriptions on communication systems.

[00103] Advantageously, using a “text-image” data pair, a specific wireless channel may be described from two modalities’ perspectives to derive more comprehensive wireless environment knowledge. This dataset may cover a diverse range of text-image pairs to ensure a comprehensive representation of the data.

[00104] Now an explanation of training the universal task-based AI/ML model is provided below. In some embodiments, the universal task-based AI/ML model may include a first modal encoder and a second modal encoder. During the training process, the network device may obtain a first output vector from the first modal encoder, and obtain a second output vector from the second encoder. The first output vector and the second output vector are obtained based on a set of pairs of multi-modal data. The network device 130 may train the universal task-based AI/ML model by adjusting parameter values of the universal task-based AI/ML model based on the first output vector and the second output vector.

[00105] In some embodiments, the network device 130 may determine a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector,

and the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data (i.e., a positive pair of data). The network device may train the universal task-based AI/ML model by maximizing a similarity between the first sub-vector and the second sub-vector.

[00106] In some embodiments, the network device 130 may determine a similarity between a third sub-vector in the first output vector and a fourth sub-vector in the second output vector, and the third sub-vector and the fourth sub-vector are generated based on different pairs of multi-modal data (i.e., negative pairs of data). The network device may train the universal task-based AI/ML model by minimizing a similarity between the third sub-vector and the fourth sub-vector.

[00107] In some embodiments, the network device 130 may train the universal task-based AI/ML model by maximizing a similarity between the first sub-vector and the second sub-vector. In some embodiments, the network device 130 may train the universal task-based AI/ML model by minimizing a similarity between the third sub-vector and the fourth sub-vector. In some embodiments, the network device 130 may train the universal task-based AI/ML model by maximizing a similarity between the first sub-vector and the second sub-vector and by minimizing a similarity between the third sub-vector and the fourth sub-vector. A more detailed explanation with respect to the minimization and maximization will be described with accompanying figures.

[00108] A detailed description of training the universal AI/ML model will be described in combination with accompanying figures. Fig. 7 illustrates an example of a block diagram of a universal task-based AI/ML model being trained according to embodiments of the present disclosure.

[00109] As shown in the example of Fig. 7, the universal task-based model 730 is trained by using a text-image pair including a text 710 indicating descriptions of physical enjoinment, and an image 720 indicating wireless channel measurements. The text 710 and the image 720 in the text-image pair are corresponding to each other and constitutes a positive pair.

[00110] The universal task-based model 730 may include a first modal encoder, e.g., a text encoder 732 as shown in Fig. 7, and a second model encoder, e.g., an image encoder 734. In some embodiments, the first model encoder e.g., a text encoder 732, may receive the text 710 with physical environment descriptions, embed environment semantic information from the text 710, and output a vector (e.g., vector 1 as shown in Fig. 7). In some embodiments,

the vector 1 may be a m -dimension vector, and m is an integer. The second model encoder, e.g., an image encoder 734, may receive the image 720 indicating wireless channel measurements, extract wireless channel features, and output a vector (e.g., vector 2 as shown in Fig. 7). In some embodiments, the vector 2 may be a m -dimension vector, and m is an integer. In some embodiments, the vector 1 may be in a format of a matrix, and the vector 2 may be in a format of a matrix.

[00111] The universal task-based model 730 may also include a cross-modal semantic alignment module 736 for implementing a cross-modal semantic information alignment in a common feature space. Specifically, the cross-modal semantic alignment module 736 may maximize a similarity between the vector 1 and the vector 2, in which the vector 1 and the vector 2 are generated based on a positive pair of data 710 and 720. In other words, the cross-modal semantic alignment module 736 may maximize a similarity between the two output m -dimension vectors from text encoder 732 and image encoder 734, such that cross-modal semantic information alignment in a common feature space may be achieved. The universal task-based AI/ML model may be trained on the basis of cross-modal semantic information alignment in a common feature space.

[00112] Generally, the universal task-based AI/ML model is trained for multiple epochs, with each epoch having multiple iterations. In each iteration, a batch of dataset is used for training the universal task-based AI/ML model. The following description will be described with reference to an iteration in an epoch, in which a batch size N (N is an integer) of dataset is used for training the universal task-based AI/ML model. In other words, for each iteration, there are N samples, which may be N pairs of data.

[00113] Fig. 8 illustrates an example implementation of a text encoder 732 in Fig. 7 according to some embodiments of the present disclosure. As shown in Fig. 8, the text encoder 732 may receive multiple samples (i.e., text inputs) including a sample text input 710. The number of the received sample text inputs is N , which is the batch size for training the universal task-based AI/ML model in one iteration. The text encoder 732 may convert natural language text into embedding vectors T 840.

[00114] For example, as shown in Fig. 8, a text input 710 among N text inputs includes text descriptions of a base station location, a user location and a LOS status, etc. For example, the text input 710 may include the text “The base station location is -24.8875, 11.0972, 5. The user location is -21.9427, 12.4355, 1. The status is NLOS.” The text encoder 732

may convert natural language text into embedding vectors T 840, which may be a $N \times m$ matrix, with N being the batch size and m being the number of dimensions for an output of each sample input, which is an text input according to embodiments of the present disclosure.

[00115] In some embodiments, the text encoder 732 may be implemented with various AI models or neural networks. The present disclosure does not limit the detailed implementation of the text encoder, and any existing or future-developed technology may be applied herein for an implementation of the text encoder.

[00116] Fig. 9 illustrates an example implementation of an image encoder 734 in Fig. 7 according to some embodiments of the present disclosure. As shown in Fig. 9, the text encoder 734 may receive multiple samples (i.e., image inputs) including a sample image input 720. The number of the received image inputs is N , which is the batch size for training the universal task-based AI/ML model in one iteration. The image encoder 734 may convert a sample image input into an embedding vector I .

[00117] For example, as shown in Fig. 9, an image input 720 among N image inputs may represent wireless channel features and correspond to the text input 710. In some embodiments, the input image may be a $2 \times 32 \times 64$ channel matrix, where 2 represents the real and imaginary parts of the complex channel, 32 is the number of transmission antennas (including 1 receiving antenna), and 64 is the number of subcarriers. The image encoder 734 may convert the sample image 720 into an embedding vector 2. For a batch size N , the image encoder 734 may convert N images into embedding vectors I 950, which may be a $N \times m$ matrix, with N being the batch size and m being the number of dimensions for an output of each sample, which is an image input according to embodiments of the present disclosure.

[00118] In some embodiments, the image encoder 734 may be implemented with various AI models or neural networks, for example, the image encoder 734 may be implemented with a Vision Transformer (ViT), etc. However, the present disclosure does not limit the detailed implementation of the image encoder, and any existing or future-developed technology may be applied herein for an implementation of the image encoder.

[00119] Now refer back to Fig. 7, as described above, the cross-modal semantic alignment module 736 is employed for implementing a cross-modal semantic information alignment in a common feature space. The cross-modal semantic alignment module 736 may maximum a similarity between vectors generated based on a positive pair and/or minimize a similarity between vectors generated based on a negative pair, such that cross-modal semantic

information alignment in a common feature space may be achieved. A more detailed description of the cross-modal semantic alignment module 736 is described in combination with Fig. 10.

[00120] Fig. 10 illustrates a process for training the universal task-based AI/ML model according to some embodiments of the present disclosure. As shown in Fig. 10, for an iteration for training the universal task-based AI/ML model, the text encoder 732 and the image encoder 734 may receive N pairs of sample data, respectively. For example, for each pair of data, the text encoder 732 may receive the text (such as text 710 as shown in Fig. 10), and the image encoder 1434 may receive the image (such as an image 720 as shown in Fig. 10). Each pair may be a positive pair, because the text in the pair and the image in the pair are corresponding to each other. For example, for the text 710 as shown in Fig. 10, the image 720 may be the wireless channel feature corresponding to the text 710.

[00121] The text encoder 732 may receive each text input $Text\ i$ and extract each vector T_i ($1 \leq i \leq N$) accordingly. In some embodiments, vector T_i is a m -dimensions matrix. For example, vector T_1 is the vector extracted from a first text input $Text\ 1$, vector T_2 is the vector extracted from a second text input $Text\ 2$, and so on. Similarly, the image encoder 1434 may receive each image input $Image\ i$ and extract each vector I_i ($1 \leq i \leq N$) accordingly. In some embodiments, vector I_i is a m -dimensions matrix. For example, vector I_1 is the vector extracted from a first image input $Image\ 1$, vector I_2 is the vector extracted from a second image input $Image\ 2$, and so on. In some embodiments, each vector T_i or I_i may be a m -dimensional matrix. Accordingly, an output vector 1020 from the text encoder 732 is a $N \times m$ matrix, and an output vector 1040 from the image encoder 734 is a $N \times m$ matrix, in which N is the batch size and m is the number of dimensions for a sub-vector generated based on a sample input. And each vector T_i in the output vector 1020 may be referred as a sub-vector in the following description, and each vector I_i in the output vector 1040 may also be referred as a sub-vector in the following description.

[00122] The cross-modal semantic alignment module 736 may compare a similarity between the output vector 1020 and the output vector 1040. Specifically, the cross-modal semantic alignment module 736 may determine a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, and the first sub-vector and the second sub-vector are generated based on a same pair of multi-modal data. For example, as shown in Fig. 10, the cross-modal semantic alignment module 736 may determine a similarity between a first sub-vector T_i in the first output vector 1020 and a

second sub-vector I_i in the second output vector 1040, the first sub-vector T_i , and the second sub-vector I_i are generated based on a positive pair including a text input $Text\ i$ and a corresponding image input $image\ i$. A text input $Text\ i$ and a corresponding image input $image\ I$ are from a positive pair. The similarity between the first sub-vector T_i in the first output vector 1020 and a second sub-vector I_i in the second output vector 1040 are shown with a block filled with patterns. The cross-modal semantic alignment module 736 may further determine a similarity between a third sub-vector T_j in the first output vector 1020 and a fourth sub-vector I_k in the second output vector 1040, in which $1 \leq j \leq N$, $1 \leq k \leq N$, $j \neq k$. The third sub-vector T_j and fourth sub-vector I_k are generated based on a negative pair including a text input $Text\ j$ and an image input $image\ k$ that are not corresponding to each other. The similarity between the third sub-vector T_j and fourth sub-vector I_k is shown with a block without any pattern filled, as shown in Fig. 10.

[00123] In some embodiments, the cross-modal semantic alignment module 736 may calculate cosine similarity matrixes M_{it} and M_{ti} via $M_{it} = I \cdot T^T$ and $M_{ti} = T \cdot I^T$. The cross-modal semantic alignment module 736 may further compare the cosine similarity matrixes M_{it} and M_{ti} with a ground truth matrix G , which indicates “positive-pair” and “negative pair” label between each text data and image data, two loss terms are calculated using cross entropy as:

$$L_{it} = cross\ entropy(M_{it}, G) \quad (1)$$

$$L_{ti} = cross\ entropy(M_{ti}, G) \quad (2)$$

[00124] The final loss L for training the universal task-based AI/ML model may be calculated by averaging the above two loss terms, as shown in equation (3)

$$L = (L_{it} + L_{ti})/2 \quad (3)$$

[00125] Although in equation (3), the weight for each loss term is 0.5, it should be understood that, other weight may be applied for each loss term.

[00126] By maximizing the similarity between output vectors from data in a first modality and data in a second modality in a positive pair of data, the text encoder and image encoder may be trained jointly to align the multi-modal information in a shared embedding space. By back propagating the final loss L to the text encoder and image encoder, the text and image information may be aligned in a common embedding space, thus the universal task-based AI/ML model may learn the intrinsic representations in the semantic space.

[00127] In some embodiments, the trained universal task-based AI/ML model may be associated with multiple tasks. In other words, the trained universal task-based AI/ML model is applicable to multiple tasks, e.g., multiple types of tasks including, but not limited to, direct positioning, assisted positioning, or beam management, etc., and may be associated with these applicable tasks, which may be referred to be as compatible tasks. In some embodiments, the network device 130 may determine one or more compatible tasks which are compatible with the trained universal task-based AI/ML model.

[00128] In some embodiments, the network device 130 may include a task-model compatibility analyzer to analyze which type(s) of task (e.g., downstream tasks) are suitable for such a trained universal task-based AI/ML model and determine the compatible task to the trained universal task-based AI/ML model. In some embodiments, the network device may receive an input of a task input, and determine if the input of the task is a subset of an input of the trained universal task-based AI/ML model or if the input of the task may be inferred from the input of the trained universal task-based AI/ML model. If the network device determines that the input of the task is a subset of an input of the trained universal task-based AI/ML model or may be inferred from the input of the trained universal task-based AI/ML model, the network device may determine that the trained universal task-based AI/ML model may accommodate the task, and the task is compatible with the trained universal task-based AI/ML model. The network device may further associate an identification of the task with the trained universal task-based AI/ML model based on the determination that the input of the task is a subset of an input of the trained universal task-based AI/ML model or may be inferred from the input of the trained universal task-based AI/ML model.

[00129] In some embodiments, the network device may record the identification of a compatible task for the trained universal task-based AI/ML model to associate the task identification of the compatible task with the trained universal task-based AI/ML model. For example, if the network device 130 determines that a task with an identification of Task-1 is a compatible task, the network device may record the identification of Task-1 for the trained universal task-based AI/ML model and associate the identification of Task-1 with the trained universal task-based AI/ML model.

[00130] A detailed explanation for determining a compatible task will be provided in combination with Fig. 11. Fig. 11 illustrates a flowchart of a method for determining a compatible task for the trained universal task-based AI/ML model according to some

embodiments of the present disclosure. The process as shown in Fig. 11 may be performed by a task-model compatibility analyzer in the network device to analyze which types of tasks are suitable or applicable for the trained universal task-based AI/ML model and determine the identifications of compatible tasks to such a trained universal task-based AI/ML model.

[00131] Generally, an input of each task may be compared with an input of the trained universal task-based AI/ML model. If an input of a task is the subset of or may be inferred from an input of the trained universal task-based AI/ML model, the trained universal task-based AI/ML model may accommodate this task and match with this task's identification. Otherwise, the trained universal task-based AI/ML model may not accommodate this task and may not match with this task's identification.

[00132] As shown in Fig. 11, a task K is taken for an example. The network device may receive an input of the task K. The task-model compatibility analyzer may determine if the input of the task K is a subset of an input of the trained universal task-based AI/ML model. In some embodiments, an input of the trained universal task-based AI/ML model may indicate an input configuration or setting of the trained universal task-based AI/ML model. If the task-model compatibility analyzer determines that the input of the task K is a subset of an input of the trained universal task-based AI/ML model, the flowchart proceeds to block 1830, in which the task-model compatibility analyzer may determine that the task K is a compatible task for the trained universal task-based AI/ML model, and the universal task-based AI/ML model may accommodate the task-K.

[00133] If the task-model compatibility analyzer determines that the input of the task K is not a subset of an input of the trained universal task-based AI/ML model, the flowchart proceeds to block 1120, in which the task-model compatibility analyzer may further determine if the input of the task K may be inferred from the input of the trained universal task-based AI/ML model. If the task-model compatibility analyzer determines that the input of the task K may be inferred from the input of the trained universal task-based AI/ML model, the flowchart proceeds to the block 1140, in which the task-model compatibility analyzer may determine that the task K is a compatible task for the trained universal task-based AI/ML model, and the trained universal task-based AI/ML model may accommodate the task K. Otherwise, if the task-model compatibility analyzer determines that the input of the task K may not be inferred from the input of the trained universal task-based AI/ML model, the flowchart proceeds to the block 1150, in which the task-model compatibility analyzer may determine that the task K is not a compatible task for the trained universal task-

based AI/ML model, and the trained universal task-based AI/ML model may not accommodate the task K.

[00134] An example of determining a compatible task will be described in combination with Fig. 12. Fig. 12 illustrates a diagram for determining a compatible task according to some embodiments of the present disclosure. In some embodiments, an input of the trained universal task-based AI/ML mode model and an input of a downstream task are compared to check whether the trained universal task-based AI/ML mode is compatible with the downstream task.

[00135] As shown in Fig. 12, the inputs of the trained universal task-based AI/ML mode are shown as a text input 710 with text descriptions of physical environment, and an image input 720 with wireless channel measurements. It should be understood that, the inputs 710 and 720 are only for the purposes of illustrations, any suitable input configuration may be applied and input to the trained universal task-based AI/ML mode.

[00136] There are four tasks shown in Fig. 12, in which the first task 1220 (e.g., with an identification of “Task-1”) is for LOS/NLOS classification, the second task 1240 (e.g., with an identification of “Task-2”) is for direct AI/ML positioning, the third task 1260 (e.g., with an identification of “Task-3”) is for AI/ML assisted positioning, and the fourth task 1280 (e.g., with an identification of “Task-4”) is for obstacle motion prediction.

[00137] The first task 1220 has an input of an image 1222 with wireless channel measurements and an optional input 1224. The task-model compatibility analyzer may determine that the input of the first task 1220 is a subset of the input of the trained universal task-based AI/ML model, and may determine that the trained universal task-based AI/ML model is compatible with the first task 1220, or in other words, the first task 1220 is a compatible task for the trained universal task-based AI/ML model.

[00138] The second task 1240 has an input of an image 1242 with wireless channel measurements. The task-model compatibility analyzer may determine that the input of the second task 1240 is a subset of the input of the trained universal task-based AI/ML model, and may determine that the trained universal task-based AI/ML model is compatible with the second task 1240, or in other words, the second task 1240 is a compatible task for the trained universal task-based AI/ML model.

[00139] The third task 1260 has an input of TOA/received signal strength indicator (RSSI)/other intermediate features. The task-model compatibility analyzer may determine

that the input of the third task 1260 may be inferred from an input of the trained universal task-based AI/ML model, and may determine that the trained universal task-based AI/ML model is compatible with the third task 1260, or in other words, the third task 1260 is a compatible task for the trained universal task-based AI/ML model.

[00140] The fourth task 1280 has an input of historical motions of an obstacle. The task-model compatibility analyzer may determine that the input of the fourth task 1280 is not a subset of the input of the trained universal task-based AI/ML model, and may not be inferred from an input of the trained universal task-based AI/ML model. The task-model compatibility analyzer may determine that the trained universal task-based AI/ML model is not compatible with the second task 1280.

[00141] Accordingly, the network device may further associate the identifications of tasks 1220, 1240, and 1260 with the trained universal task-based AI/ML model by recording these task identifications for the trained universal task-based AI/ML model, accordingly, the trained universal task-based AI/ML model may have associated identifications of tasks: Task-1, Task-2, and Task-3, as shown in block 1290 in Fig. 12.

[00142] In some embodiments, by checking each potential task with task-model compatibility analyzer, the compatible task identifications may be derived for the trained universal task-based AI/ML model at network side.

[00143] In some embodiments, after the trained universal task-based AI/ML model has been trained at the network device, the trained universal task-based AI/ML model may be transmitted by the network device, for example, in response to a request for an AI/ML model as transmitted in 404, to the terminal device for a cluster of tasks. As described above, the terminal device may perform an inference based on the trained universal task-based AI/ML model (i.e., zero-shot learning without a cascaded AI/ML model) or the trained universal task-based AI/ML model cascaded with a fine-tuned task-oriented AI/ML model. Methods of determining if a task-oriented AI/ML model is to be fine-tuned have been disclosed above, the detailed description is omitted here for the purposes of clarity and brevity.

[00144] An example of zero-shot learning is provided herein for better understanding embodiments of the present disclosure. In some embodiments, the universal task-based AI/ML model is trained using dataset including text samples of the BS location, the terminal device location, and LOS/NLOS classification, and including image samples of corresponding channel state information (CSI) data. For a classification task which has

been learned in training process, such as the LOS/NLOS classification task, the trained universal task-based AI/ML model, once being determined to infer the LOS/NLOS classification task according the signaling process 400 or 500, may be directed inferred for the task without any task-oriented AI/ML model, which may be regarded as Zero-Shot Learning (ZSL).

[00145] Fig. 13 illustrates an example diagram of zero-shot learning for LOS/NLOS classification according to some embodiments of the present disclosure. As shown in Fig. 13, if the network device determines that the trained universal task-based AI/ML model may be directed inferred for the classification task without any task-oriented AI/ML model, for example, according to the operation 508 as shown in Fig. 5, the text encoder 732 and the image encoder 734 are frozen in zero-shot learning state. As shown in Fig. 13, an input 1330 for the image encoder 734 is the CSI data, which has the same configuration as that in the training process. Input data 1310, 1320 for the text encoder 732 is the text description to indicate LOS/NLOS classification, which is the sub-set of an input of the universal task-based AI/ML model during the training process. By calculating similarities between an output image embedding I_1 and class embeddings T_1 and T_2 obtained from the textual descriptions in each sample 1310, 1320, respectively, the most similar class of the image embedding may be derived to indicate LOS/NLOS classification. For example, a similarity of the image embedding I_1 with the first class embedding T_1 is 0.7, and a similarity of the image embedding I_1 with the second class embedding T_2 is 0.3. Assuming the first class embedding T_1 is obtained from the text description of the sample 1310, then the output of the trained universal task-based AI/ML model may be “the status is NLOS”, as shown in Fig. 13.

[00146] In some embodiments, for a task not learned in the training stage, such as the direct AI/ML positioning task, the trained universal task-based AI/ML model may not be direct inferred for this task using zero-shot learning, because only classification tasks have been learned in the training stage. The network device determines that a task-oriented AI/ML model is to be fine-tuned for this task, for example, according to the operation 508 as shown in Fig. 5. In some embodiments, for a task which has been learned in training stage, fine-tuning may also be conducted for performance enhancement.

[00147] Fig. 14 illustrates an example of fine-tuning a task-oriented AI/ML model for a direct AI/ML positioning according to some embodiments of the present disclosure. As shown in Fig. 14, when generalizing the trained universal task-based AI/ML model to direct AI/ML positioning task, the image encoder 734 is frozen and a task-oriented AI/ML model

1420 is fine-tuned. In some embodiments, the task-oriented AI/ML model 1420 may include N-layer Multi-layer Perceptron (MLP), and may include a linear layer 1422 and ReLU 1424. The detailed architecture of the task-oriented AI/ML model 1420 is not limited. The task-oriented AI/ML model 1420 may be fine-tuned at the terminal device or at the network device on a task-specific dataset.

[00148] In some embodiments, if the task-oriented AI/ML model 1420 is to be fine-tuned at the network device, the network device may fine-tune the task-oriented AI/ML model 1420 (for example, according to the operation 514 in the signaling process 500) based on the response from terminal device, for example, received at 512 in signaling process 500, and transmit the fine-tuned task-oriented AI/ML model (for example, according to the operation 515 in the signaling process 500). In some embodiments, the fine-tuned task-oriented AI/ML model may be cascaded with the trained universal task-based AI/ML model, and transmitted to the terminal device. Alternatively, the fine-tuned task-oriented AI/ML model may be transmitted separately from the trained universal task-based AI/ML model, and the terminal device may cascade the fine-tuned task-oriented AI/ML model with the trained universal task-based AI/ML model.

[00149] In some embodiments, the task-oriented AI/ML model 1420 is to be fine-tuned at the terminal device, for example, according to the operation 518 in the signaling process 500. In some embodiments, if the task-oriented AI/ML model 1420 has not been deployed at the terminal device, the network device may transmit the task-oriented AI/ML model 1420 to the terminal device, and the terminal device may fine-tune the task-oriented AI/ML model 1420 and cascade the fine-tuned task-oriented AI/ML model 1420 with the trained universal task-based AI/ML model. In some embodiments, if the task-oriented AI/ML model 1420 has been deployed at the terminal device, the terminal device may fine-tune the task-oriented AI/ML model 1420 at the terminal device, and cascade the fine-tuned task-oriented AI/ML model 1420 with the trained universal task-based AI/ML model. The terminal device may use the trained universal task-based AI/ML model cascaded with the fine-tuned task-oriented AI/ML model 1420 for inferring the positioning task and may obtain an output (x, y) indicating a position of an object, as shown in Fig. 14.

[00150] Since the trained universal task-based AI/ML model has learned useful representations and semantic relationships between images and text during the raining stage, the trained universal task-based AI/ML model may provide a strong initial starting point. Thus, inheriting and building upon the trained model's capabilities, the fine-tuned task-

oriented AI/ML model 1420 may better align with the specific task's objectives and data characteristics. In this example, a mean square error (MSE) loss may be adopted for supervised fine-tuning the task-oriented AI/ML model 1420.

[00151] In some embodiments, the sample dataset used to train the universal task-based AI/ML model is generated by DeepMIMO, which is a framework for generating large-scale MIMO datasets based on accurate Remcom 3D ray-tracing. The configurations of DeepMIMO are given in Table 1.

Table 1. DeepMIMO configurations

Configurations	Value
Scenario	O1
Carrier Frequency	28GHz
Bandwidth	50MHz
Subcarrier	64
Antenna Configuration	32 transmission ports: (1,32,1) 1 receiving ports: (1,1,1)
Antenna Spacing	0.5λ
Channel Model	Ray-tracing
BS Number	18
BS height	6m
UE Number	1184923
UE height	2m

[00152] For evaluation effects of the solution according to embodiments of the present disclosure, two downstream tasks have been tested LOS/NLOS classification and direct AIML positioning.

[00153] For the LOS/NLOS classification task, following three schemes with the dataset not included in the training dataset have been evaluated. Scheme (1) is training and zero-shot learning (this disclosure): 0 data samples for fine-tuning, 284957 data samples for testing. Scheme (2) is training and fine-tuning (this disclosure): 2850 data samples for fine-tuning,

the remaining 282107 data samples for testing. Scheme (3) is baseline ViT: 2850 data samples for fine-tuning, the remaining 282107 data samples for testing.

[00154] Table 2 compares the trainable parameters and LOS/NLOS classification accuracy of the 3 schemes. As shown in Table 3, zero-shot learning scheme does not need to train any NN parameters in the trained universal task-based model and presents a low LOS/NLOS classification accuracy. Meanwhile, the fine-tuning scheme only needs to fine-tune 258 NN parameters but can improve the LOS/NLOS classification accuracy to 98.92%, which is less than 1% degradation but 99.999% fine-tuning complexity reduction (42M to 258) compared the baseline.

Table 2. Performance comparison of LOS classification task

Downstream task	Trainable parameters	LOS/NLOS classification accuracy
training and zero-shot learning (this disclosure)	0	77.0%
training and fine-tuning (this disclosure)	258	98.92%
Baseline ViT	42 M	99.74%

[00155] For the direct AI/ML positioning task, the following two schemes with the datasets not included in the training dataset have been evaluated. Scheme (1) is training and fine-tuning (this disclosure): 196646 data samples for fine-tuning, the remaining data samples for testing. Scheme (2) is baseline ViT: 196.646 data samples for fine-tuning, the remaining data samples for testing.

[00156] Table 3 compares the trainable parameters and positioning accuracy at CDF90 between the above 2 schemes. As shown in Table 3, the proposed scheme can achieve 1.40m positioning accuracy at CDF90 while reducing 99.86% fine-tuning complexity (42M to 60K) compared to baseline scheme.

Table 3. Performance comparison of positioning task

Downstream tasks	Trainable parameters	Positioning accuracy at CDF90
training and fine-tuning (this disclosure)	60 K	1.40m

Baseline ViT	42 M	0.826m
--------------	------	--------

[00157] From above two tables, it can be observed that the proposed universal task-based AI/ML model can achieve comparable performance as the baseline scheme while reducing more than 99.5% fine-tuning complexity, indicating the proposed embodiments of the present disclosure is very promising in future AI/ML wireless air interface applications.

[00158] Fig. 15 illustrates a flowchart of a method 1500 implemented at a network device in accordance with some example embodiments of the present disclosure. For the purpose of discussion, the method 1500 will be described from the perspective of the network device 130 with reference to Fig. 1.

[00159] At 1510, the network device may train a universal task-based artificial intelligence, AI/machine learning, ML, model based on a dataset comprising multiple pairs of multi-modal data, and the trained universal task-based AI/ML model is applicable to multiple tasks. At 1520, the network device may transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

[00160] In some embodiments, a pair of multi-modal data may include data of a first modality and data of a second modality.

[00161] In some embodiments, the universal task-based AI/ML model may include a first modal encoder and a second modal encoder, and the network device may train the universal task-based AI/ML model by: obtaining a first output vector from the first modal encoder; obtaining a second output vector from the second encoder, wherein the first output vector and the second output vector are obtained based on a set of pairs of multi-modal data; and adjusting a parameter value of the universal task-based AI/ML model based on the first output vector and the second output vector.

[00162] In some embodiments, the network device may adjust the parameter value of the universal task-based AI/ML model by: determining a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, wherein the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data; and maximizing the similarity between the first sub-vector and the second sub-vector.

[00163] In some embodiments, the network device may adjust the parameter value of the universal task-based AI/ML model by: determining a similarity between a third sub-vector

in the first output vector and a fourth sub-vector in the second output vector, wherein the third sub-vector and the fourth sub-vector are obtained based on different pairs of multi-modal data; and minimizing the similarity between the third sub-vector and the fourth sub-vector.

[00164] In some embodiments, the network device may determine at least one compatible task which is compatible with the trained universal task-based AI/ML model.

[00165] In some embodiments, the network device may determine the at least one compatible task by: receiving an input of a task; determining that the input of the task is a subset of an input of the trained universal task-based AI/ML model or is inferred from the input of the trained universal task-based AI/ML model; and determining the task is a compatible task based on determining that the input of the task is a subset of the input of the trained universal task-based AI/ML model or is inferred from the input of the trained universal task-based AI/ML model.

[00166] In some embodiments, the network device may associate an identification of the compatible task with the trained universal task-based AI/ML model.

[00167] In some embodiments, the network device may determine a task-oriented AI/ML model is to be fine-tuned and cascaded with the trained universal task-based AI/ML model based on at least one of: a target of a task in the at least one task; or a key performance indicator (KPI) for the task.

[00168] In some embodiments, the network device may transmit, to the terminal device, a task-oriented AI/ML model, wherein the task-oriented AI/ML model is to be fine-tuned at the terminal device.

[00169] In some embodiments, the network device may fine-tune a task-oriented AI/ML model; and transmit, to the terminal device, the fine-tuned task-oriented AI/ML model.

[00170] Fig. 16 illustrates a flowchart of a method 1600 implemented at a terminal device in accordance with some example embodiments of the present disclosure. For the purpose of discussion, the method 1600 will be described from the perspective of the terminal device 110 with reference to Fig. 1.

[00171] At 1610, the terminal device may receive, from a network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, and the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset

including multiple pairs of multi-modal data. At 1620, the terminal device may perform an inference of at least one task based on the trained universal task-based AI/ML model.

[00172] In some embodiments, a pair of multi-modal data may include data of a first modality and data of a second modality.

[00173] In some embodiments, the trained universal task-based AI/ML model is obtained by training a universal task-based AI/ML model, and the universal task-based AI/ML model may include a first modal encoder and a second modal encoder, and the universal task-based AI/ML model is trained by: obtaining a first output vector from the first modal encoder; obtaining a second output vector from the second encoder, wherein the first output vector and the second output vector are obtained based on a set of pairs of multi-modal data; adjusting a parameter value of the universal task-based AI/ML model based on the first output vector and the second output vector.

[00174] In some embodiments, the parameter value of the universal task-based AI/ML model is adjusted by: determining a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, wherein the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data; and maximizing the similarity between the first sub-vector and the second sub-vector.

[00175] In some embodiments, the parameter value of the universal task-based AI/ML model is adjusted by: determining a similarity between a third sub-vector in the first output vector and a fourth sub-vector in the second output vector, wherein the third sub-vector and the fourth sub-vector are obtained based on different pairs of multi-modal data; and minimizing the similarity between the third sub-vector and the fourth sub-vector.

[00176] In some embodiments, the terminal device may perform the inference by: performing the inference based on the trained universal task-based AI/ML model or the trained universal task-based AI/ML model cascaded with a fine-tuned task-oriented AI/ML model.

[00177] In some embodiments, the terminal device may receive the fine-tuned task-oriented AI/ML model from the network device.

[00178] In some embodiments, the terminal device may receive a task-oriented AI/ML model from the network device; fine-tune the received task-oriented AI/ML model to generate the fine-tuned task-oriented AI/ML model; and cascade the fine-tuned task-oriented AI/ML model with the trained universal task-based AI/ML model.

[00179] In some example embodiments, an apparatus capable of performing the method 1500 (for example, the network device) may comprise means for performing the respective steps of the method 1500. The means may be implemented in any suitable form. For example, the means may be implemented in a circuitry or software module.

[00180] In some embodiments, the apparatus may include means for training a universal task-based artificial intelligence (AI)/machine learning (ML) model based on a dataset including multiple multi-modal data, wherein the trained universal task-based AI/ML model is applicable to multiple tasks. The apparatus may include means for transmitting the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

[00181] In some embodiments, a pair of multi-modal data may include data of a first modality and data of a second modality.

[00182] In some embodiments, the universal task-based AI/ML model may include a first modal encoder and a second modal encoder, and the network device may train the universal task-based AI/ML model by: obtaining a first output vector from the first modal encoder; obtaining a second output vector from the second encoder, wherein the first output vector and the second output vector are obtained based on a set of pairs of multi-modal data; and adjusting a parameter value of the universal task-based AI/ML model based on the first output vector and the second output vector.

[00183] In some embodiments, the apparatus may adjust the parameter value of the universal task-based AI/ML model by: determining a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, wherein the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data; and maximizing the similarity between the first sub-vector and the second sub-vector.

[00184] In some embodiments, the apparatus may adjust the parameter value of the universal task-based AI/ML model by: determining a similarity between a third sub-vector in the first output vector and a fourth sub-vector in the second output vector, wherein the third sub-vector and the fourth sub-vector are obtained based on different pairs of multi-modal data; and minimizing the similarity between the third sub-vector and the fourth sub-vector.

[00185] In some embodiments, the apparatus may include means for determining at least one compatible task which is compatible with the trained universal task-based AI/ML model.

[00186] In some embodiments, the apparatus may determine the at least one compatible task by: receiving an input of a task; determining that the input of the task is a subset of an input of the trained universal task-based AI/ML model or is inferred from the input of the trained universal task-based AI/ML model; and determining the task is a compatible task based on determining that the input of the task is a subset of an input of the trained universal task-based AI/ML model or is inferred from the input of the universal task-based AI/ML model.

[00187] In some embodiments, the apparatus may include means for associating an identification of the compatible task with the trained universal task-based AI/ML model.

[00188] In some embodiments, the apparatus may include means for determining a task-oriented AI/ML model is to be fine-tuned and cascaded with the trained universal task-based AI/ML model based on at least one of: a target of a task in the at least one task; or a key performance indicator (KPI) for the task.

[00189] In some embodiments, the apparatus may include means for transmitting, to the terminal device, a task-oriented AI/ML model, wherein the task-oriented AI/ML model is to be fine-tuned at the terminal device.

[00190] In some embodiments, the apparatus may include means for fine-tuning a task-oriented AI/ML model; and means for transmitting, to the terminal device, the fine-tuned task-oriented AI/ML model.

[00191] In some embodiments, the apparatus further comprises means for performing other steps in some embodiments of the method 1500. In some embodiments, the means comprises at least one processor and at least one memory including computer program code, the at least one memory and computer program code configured to, with the at least one processor, cause the performance of the apparatus.

[00192] In some example embodiments, an apparatus capable of performing the method 1600 (for example, the terminal device) may include means for performing the respective steps of the method 1600. The means may be implemented in any suitable form. For example, the means may be implemented in a circuitry or software module.

[00193] In some embodiments, the apparatus may include means for receiving, from a network device, a trained universal task-based artificial intelligence (AI)/machine learning (ML) model, and the trained universal task-based AI/ML model is applicable to multiple tasks and is trained based on a dataset including multiple pairs of multi-modal data. The

apparatus may include means for performing an inference of at least one task based on the trained universal task-based AI/ML model.

[00194] In some embodiments, a pair of multi-modal data may include data of a first modality and data of a second modality.

[00195] In some embodiments, the trained universal task-based AI/ML model is obtained by training a universal task-based AI/ML model, and the universal task-based AI/ML model may include a first modal encoder and a second modal encoder, and the universal task-based AI/ML model is trained by: obtaining a first output vector from the first modal encoder; obtaining a second output vector from the second encoder, wherein the first output vector and the second output vector are obtained based on a set of pairs of multi-modal data; adjusting a parameter value of the universal task-based AI/ML model based on the first output vector and the second output vector.

[00196] In some embodiments, the parameter value of the universal task-based AI/ML model is adjusted by: determining a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, wherein the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data; and maximizing the similarity between the first sub-vector and the second sub-vector.

[00197] In some embodiments, the parameter value of the universal task-based AI/ML model is adjusted by: determining a similarity between a third sub-vector in the first output vector and a fourth sub-vector in the second output vector, wherein the third sub-vector and the fourth sub-vector are obtained based on different pairs of multi-modal data; and minimizing the similarity between the third sub-vector and the fourth sub-vector.

[00198] In some embodiments, the apparatus may perform the inference by: performing the inference based on the trained universal task-based AI/ML model or the trained universal task-based AI/ML model cascaded with a fine-tuned task-oriented AI/ML model.

[00199] In some embodiments, the apparatus may include means for receiving the fine-tuned task-oriented AI/ML model from the network device.

[00200] In some embodiments, the apparatus may include means for receiving a task-oriented AI/ML model from the network device; fine-tune the received task-oriented AI/ML model to generate the fine-tuned task-oriented AI/ML model; and means for cascading the fine-tuned task-oriented AI/ML model with the trained universal task-based AI/ML model.

[00201] In some embodiments, the apparatus further comprises means for performing other steps in some embodiments of the method 1600. In some embodiments, the means comprises at least one processor and at least one memory including computer program code, the at least one memory and computer program code configured to, with the at least one processor, cause the performance of the apparatus.

[00202] Fig. 17 illustrates a simplified block diagram of a device 1700 that is suitable for implementing some example embodiments of the present disclosure. The device 1700 may be provided to implement a device, for example, the terminal device or the network device as shown in Fig. 1. As shown, the device 1700 includes one or more processors 1710, one or more memories 1720 coupled to the processor 1710, and one or more communication modules 1740 coupled to the processor 1710.

[00203] The communication module 1740 is for bidirectional communications. The communication module 1740 has at least one antenna to facilitate communication. The communication interface may represent any interface that is necessary for communication with other network elements.

[00204] The processor 1710 may be of any type suitable to the local technical network and may include one or more of the following: general purpose computers, special purpose computers, microprocessors, digital signal processors (DSPs) and processors based on multicore processor architecture, as non-limiting examples. The device 2400 may have multiple processors, such as an application specific integrated circuit chip that is slaved in time to a clock which synchronizes the main processor.

[00205] The memory 1720 may include one or more non-volatile memories and one or more volatile memories. Examples of the non-volatile memories include, but are not limited to, a Read Only Memory (ROM) 1724, an electrically programmable read only memory (EPROM), a flash memory, a hard disk, a compact disc (CD), a digital video disk (DVD), and other magnetic storage and/or optical storage. Examples of the volatile memories include, but are not limited to, a random access memory (RAM) 1722 and other volatile memories that will not last in the power-down duration.

[00206] A computer program 1730 includes computer executable instructions that are executed by the associated processor 1710. The program 1730 may be stored in the ROM 1724. The processor 1710 may perform any suitable actions and processing by loading the program 1730 into the RAM 1722.

[00207] The embodiments of the present disclosure may be implemented by means of the program 1730 so that the device 1700 may perform any process of the disclosure as discussed with reference to Figs. 4 to 16. The embodiments of the present disclosure may also be implemented by hardware or by a combination of software and hardware.

[00208] In some example embodiments, the program 1730 may be tangibly contained in a computer readable medium which may be included in the device 1700 (such as in the memory 1720) or other storage devices that are accessible by the device 1700. The device 1700 may load the program 1730 from the computer readable medium to the RAM 1722 for execution. The computer readable medium may include any types of tangible non-volatile storage, such as ROM, EPROM, a flash memory, a hard disk, CD, DVD, and the like.

[00209] Fig. 18 illustrates a block diagram of an example of a computer readable medium 1800 in accordance with some example embodiments of the present disclosure. The computer readable medium 1800 has the program 1830 stored thereon. It is noted that although the computer readable medium 1800 is depicted in form of CD or DVD in Fig. 18, the computer readable medium 1800 may be in any other form suitable for carry or hold the program 1730.

[00210] Generally, various embodiments of the present disclosure may be implemented in hardware or special purpose circuits, software, logic or any combination thereof. Some aspects may be implemented in hardware, while other aspects may be implemented in firmware or software which may be executed by a controller, microprocessor or other computing device. While various aspects of embodiments of the present disclosure are illustrated and described as block diagrams, flowcharts, or using some other pictorial representations, it is to be understood that the block, apparatus, system, technique or method described herein may be implemented in, as non-limiting examples, hardware, software, firmware, special purpose circuits or logic, general purpose hardware or controller or other computing devices, or some combination thereof.

[00211] The present disclosure also provides at least one computer program product tangibly stored on a non-transitory computer readable storage medium. The computer program product includes computer-executable instructions, such as those included in program modules, being executed in a device on a target real or virtual processor, to carry out the method 1500 or 1600 as described above with reference to Fig. 15 or Fig. 16. Generally, program modules include routines, programs, libraries, objects, classes, components, data

structures, or the like that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or split between program modules as desired in various embodiments. Machine-executable instructions for program modules may be executed within a local or distributed device. In a distributed device, program modules may be located in both local and remote storage media.

[00212] Program code for carrying out methods of the present disclosure may be written in any combination of one or more programming languages. These program codes may be provided to a processor or controller of a general purpose computer, special purpose computer, or other programmable data processing apparatus, such that the program codes, when executed by the processor or controller, cause the functions/operations specified in the flowcharts and/or block diagrams to be implemented. The program code may execute entirely on a machine, partly on the machine, as a stand-alone software package, partly on the machine and partly on a remote machine or entirely on the remote machine or server.

[00213] In the context of the present disclosure, the computer program codes or related data may be carried by any suitable carrier to enable the device, apparatus or processor to perform various processes and operations as described above. Examples of the carrier include a signal, computer readable medium, and the like.

[00214] The computer readable medium may be a computer readable signal medium or a computer readable storage medium. A computer readable medium may include but not limited to an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples of the computer readable storage medium would include an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. The term “non-transitory,” as used herein, is a limitation of the medium itself (i.e., tangible, not a signal) as opposed to a limitation on data storage persistency (e.g., RAM vs. ROM).

[00215] Further, while operations are depicted in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results.

In certain circumstances, multitasking and parallel processing may be advantageous. Likewise, while several specific implementation details are contained in the above discussions, these should not be construed as limitations on the scope of the present disclosure, but rather as descriptions of features that may be specific to particular embodiments. Certain features that are described in the context of separate embodiments may also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment may also be implemented in multiple embodiments separately or in any suitable sub-combination.

[00216] Although the present disclosure has been described in languages specific to structural features and/or methodological acts, it is to be understood that the present disclosure defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

WHAT IS CLAIMED IS:

1. A network device comprising:
at least one processor; and
at least one memory storing instructions that, when executed by the at least one processor, cause the network device at least to:
train a universal task-based artificial intelligence, AI/machine learning, ML, model based on a dataset comprising a plurality of pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to a plurality of tasks; and
transmit the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.
2. The network device of claim 1, wherein a pair of multi-modal data comprises data of a first modality and data of a second modality.
3. The network device of claim 1 or 2, wherein the universal task-based AI/ML model comprises a first modal encoder and a second modal encoder, and the network device is caused to train the universal task-based AI/ML model by:
obtaining a first output vector from the first modal encoder;
obtaining a second output vector from the second encoder, wherein the first output vector and the second output vector are obtained based on a set of pairs of multi-modal data; and
adjusting a parameter value of the universal task-based AI/ML model based on the first output vector and the second output vector.
4. The network device of claim 3, wherein the network device is caused to adjust the parameter value of the universal task-based AI/ML model by:
determining a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, wherein the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data; and
maximizing the similarity between the first sub-vector and the second sub-vector.
5. The network device of claim 3 or 4, wherein the network device is caused to adjust the parameter value of the universal task-based AI/ML model by:

determining a similarity between a third sub-vector in the first output vector and a fourth sub-vector in the second output vector, wherein the third sub-vector and the fourth sub-vector are obtained based on different pairs of multi-modal data; and

minimizing the similarity between the third sub-vector and the fourth sub-vector.

6. The network device of any of claims 1 to 5, wherein the network device is further caused to:

determine at least one compatible task which is compatible with the trained universal task-based AI/ML model.

7. The network device of claim 6, wherein the network device is caused to determine the at least one compatible task by:

receiving an input of a task;

determining that the input of the task is a subset of an input of the trained universal task-based AI/ML model or is inferred from the input of the trained universal task-based AI/ML model; and

determining the task is a compatible task based on determining that the input of the task is a subset of the input of the trained universal task-based AI/ML model or is inferred from the input of the universal task-based AI/ML model.

8. The network device of claim 6 or 7, wherein the network device is further caused to:

associate an identification of the compatible task with the trained universal task-based AI/ML model.

9. The network device of claim 1, wherein the network device is further caused to:

determine a task-oriented AI/ML model is to be fine-tuned and cascaded with the trained universal task-based AI/ML model based on at least one of: a target of a task in the at least one task; or a key performance indicator, KPI, for the task.

10. The network device of claim 9, wherein the network device is further caused to:

transmit, to the terminal device, a task-oriented AI/ML model, wherein the task-oriented AI/ML model is to be fine-tuned at the terminal device.

11. The network device of claim 9, wherein the network device is further caused to:
fine-tune a task-oriented AI/ML model; and
transmit, to the terminal device, the fine-tuned task-oriented AI/ML model.

12. A terminal device comprising:
at least one processor; and
at least one memory storing instructions that, when executed by the at least one processor, cause the terminal device at least to:
receive, from a network device, a trained universal task-based artificial intelligence, AI/machine learning, ML, model, wherein the trained universal task-based AI/ML model is applicable to a plurality of tasks and is trained based on a dataset comprising a plurality of pairs of multi-modal data; and
perform an inference of at least one task based on the trained universal task-based AI/ML model.

13. The terminal device of claim 12, wherein a pair of multi-modal data comprises data of a first modality and data of a second modality.

14. The terminal device of claim 12 or 13, wherein the trained universal task-based AI/ML model is obtained by training a universal task-based AI/ML model, wherein the universal task-based AI/ML model comprise a first modal encoder and a second modal encoder, wherein the universal task-based AI/ML model is trained by:
obtaining a first output vector from the first modal encoder;
obtaining a second output vector from the second encoder, wherein the first output vector and the second output vector are obtained based on a set of pairs of multi-modal data;
and
adjusting a parameter value of the universal task-based AI/ML model based on the first output vector and the second output vector.

15. The terminal device of claim 14, wherein the parameter value of the universal task-based AI/ML model is adjusted by:
determining a similarity between a first sub-vector in the first output vector and a second sub-vector in the second output vector, wherein the first sub-vector and the second sub-vector are obtained based on a same pair of multi-modal data; and

maximizing the similarity between the first sub-vector and the second sub-vector.

16. The terminal device of claim 14 or 15, wherein the parameter value of the universal task-based AI/ML model is adjusted by:

determining a similarity between a third sub-vector in the first output vector and a fourth sub-vector in the second output vector, wherein the third sub-vector and the fourth sub-vector are obtained based on different pairs of multi-modal data; and

minimizing the similarity between the third sub-vector and the fourth sub-vector.

17. The terminal device of any of claims 12 to 16, wherein the terminal device is caused to perform the inference by:

performing the inference based on the trained universal task-based AI/ML model or the trained universal task-based AI/ML model cascaded with a fine-tuned task-oriented AI/ML model.

18. The terminal device of claim 17, wherein the terminal device is further caused to: receive the fine-tuned task-oriented AI/ML model from the network device.

19. The terminal device of claim 17, wherein the terminal device is further caused to: receive a task-oriented AI/ML model from the network device;

fine-tune the received task-oriented AI/ML model to generate the fine-tuned task-oriented AI/ML model; and

cascade the fine-tuned task-oriented AI/ML model with the trained universal task-based AI/ML model.

20. A method, comprising:

training a universal task-based artificial intelligence, AI/machine learning, ML, model based on a dataset comprising a plurality of pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to a plurality of tasks; and

transmitting the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

21. A method comprising:

receiving, from a network device, a trained universal task-based artificial intelligence, AI/machine learning, ML, model, wherein the trained universal task-based AI/ML model is applicable to a plurality of tasks and is trained based on a dataset comprising a plurality of pairs of multi-modal data; and

performing an inference of at least one task based on the trained universal task-based AI/ML model.

22. An apparatus comprising:

means for training a universal task-based artificial intelligence, AI/machine learning, ML, model based on a dataset comprising a plurality of pairs of multi-modal data, wherein the trained universal task-based AI/ML model is applicable to a plurality of tasks; and

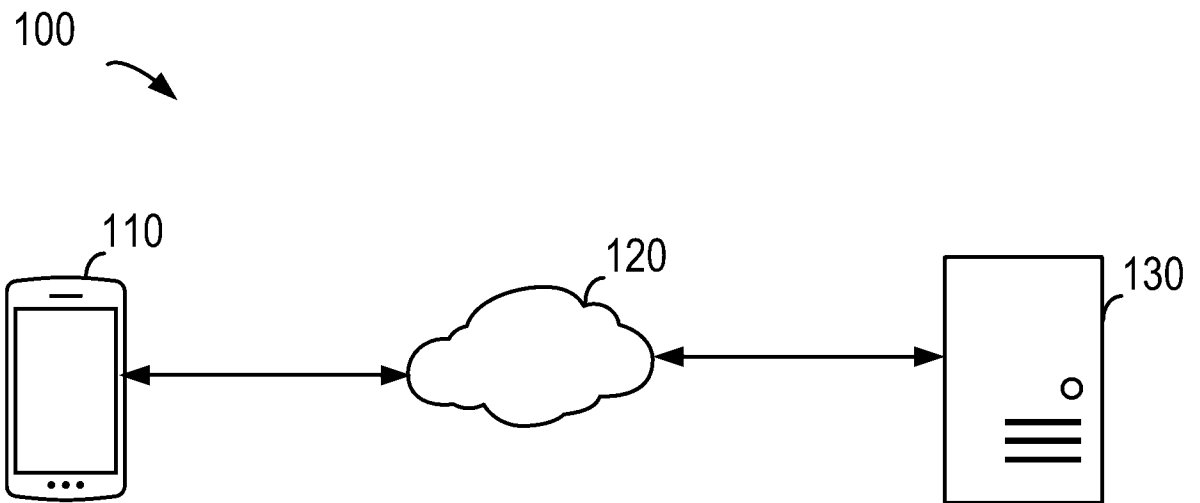
means for transmitting the trained universal task-based AI/ML model to a terminal device for an inference of at least one task.

23. An apparatus comprising:

means for receiving, from a network device, a trained universal task-based artificial intelligence, AI/machine learning, ML, model, wherein the trained universal task-based AI/ML model is applicable to a plurality of tasks and is trained based on a dataset comprising a plurality of pairs of multi-modal data; and

means for performing an inference of at least one task based on the trained universal task-based AI/ML model.

24. A non-transitory computer readable medium comprising program instructions for causing an apparatus to perform at least the method of claim 20 or 21.

**Fig.1**

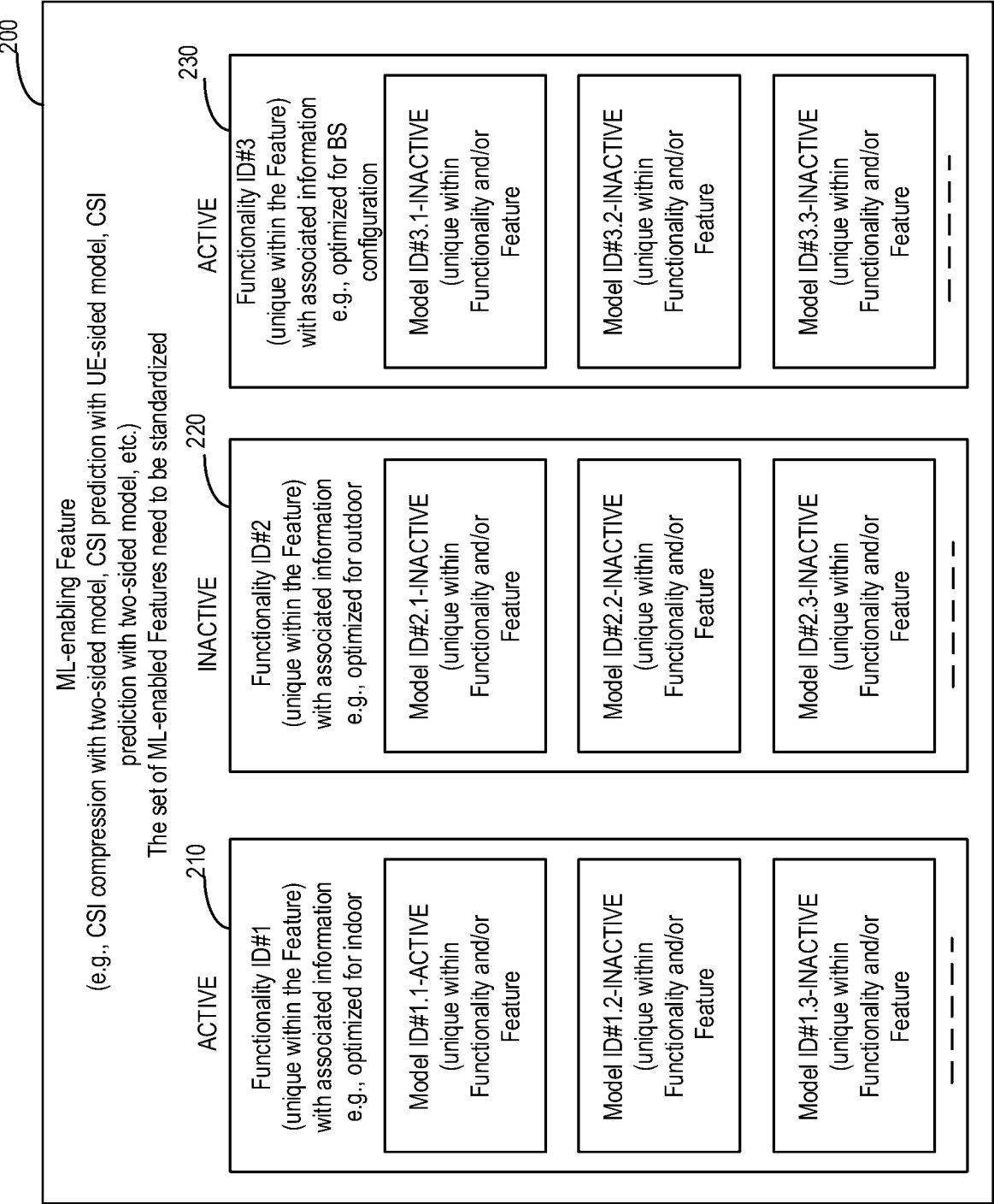


Fig.2

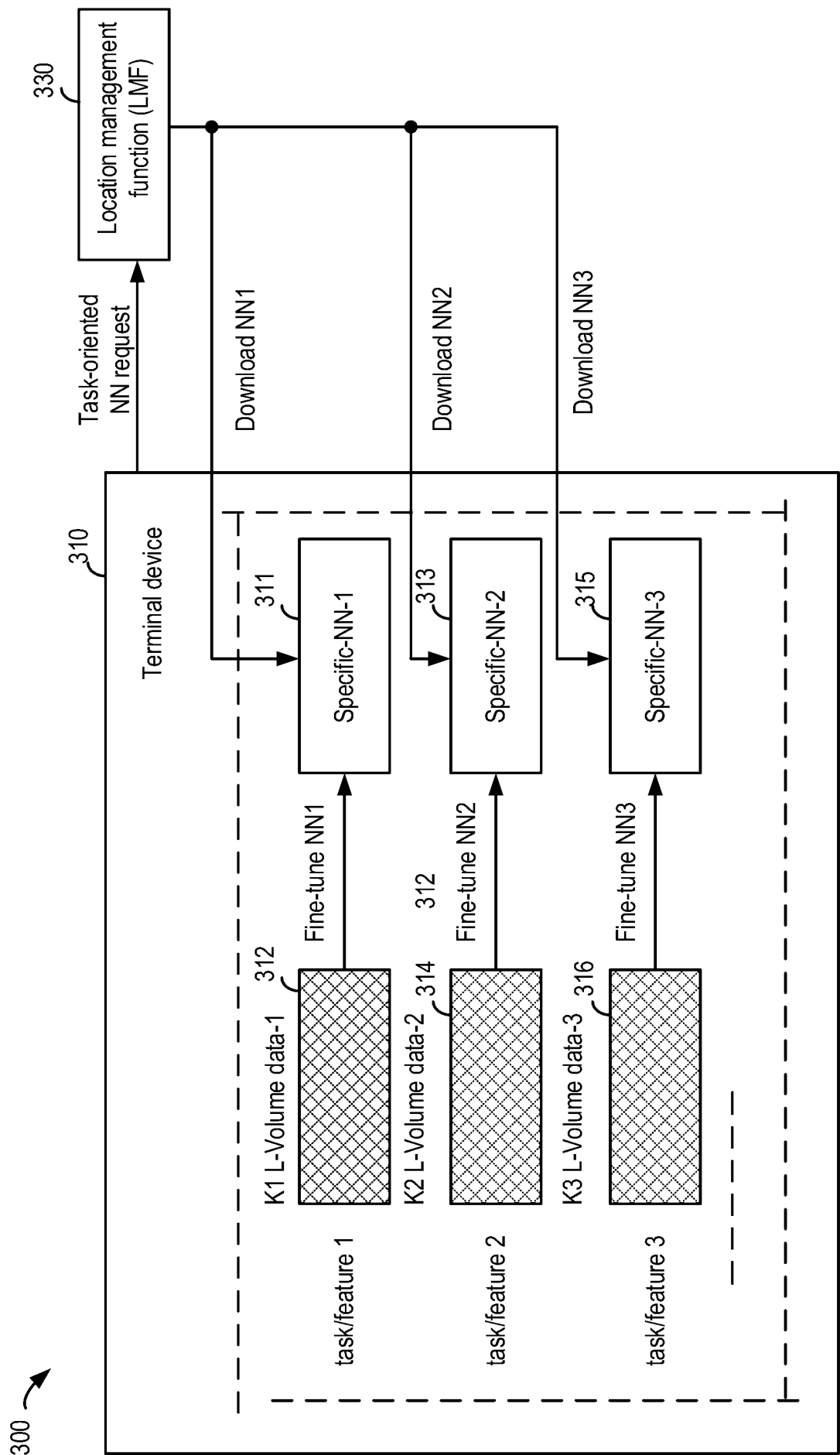
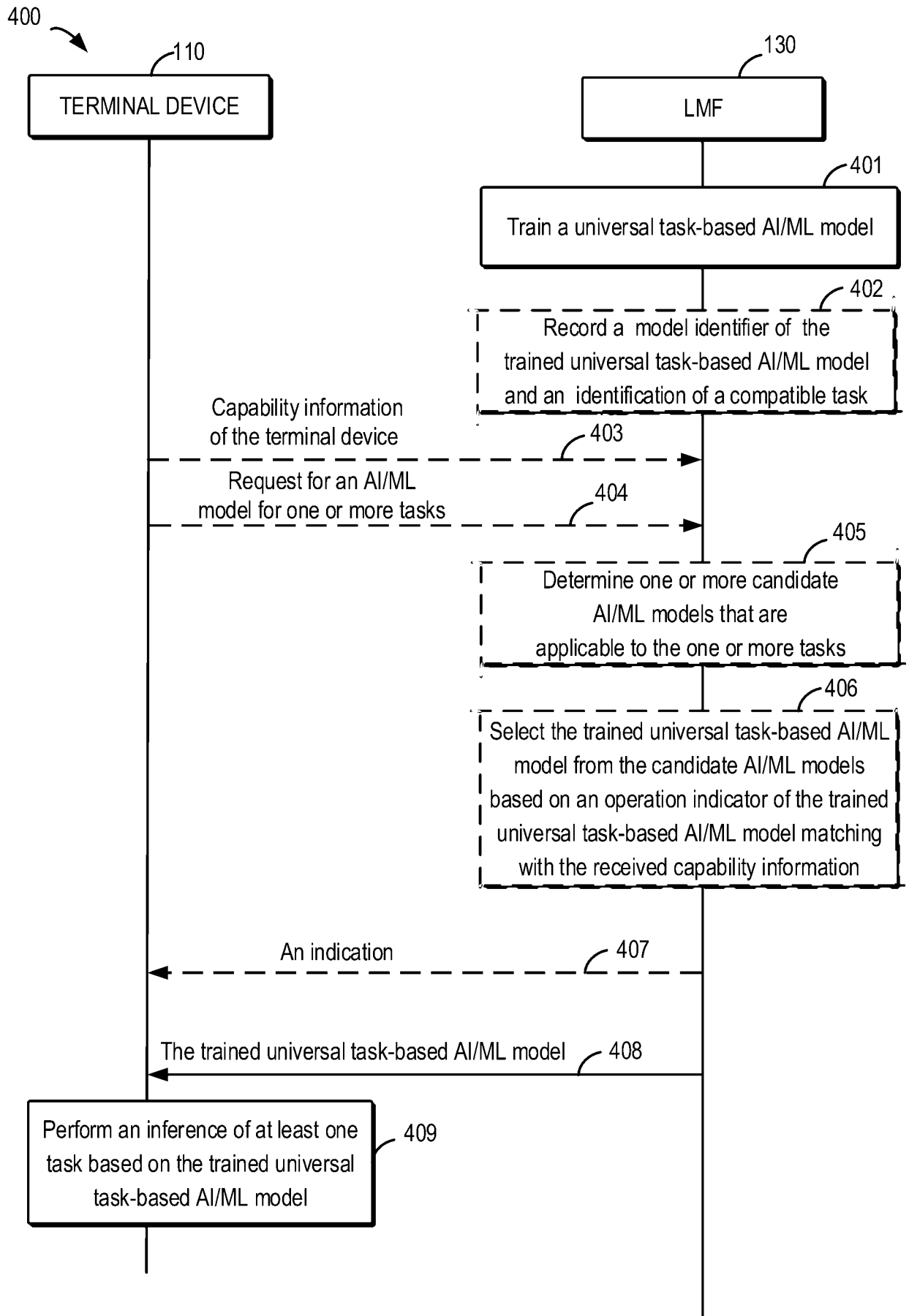


Fig.3

**Fig.4**

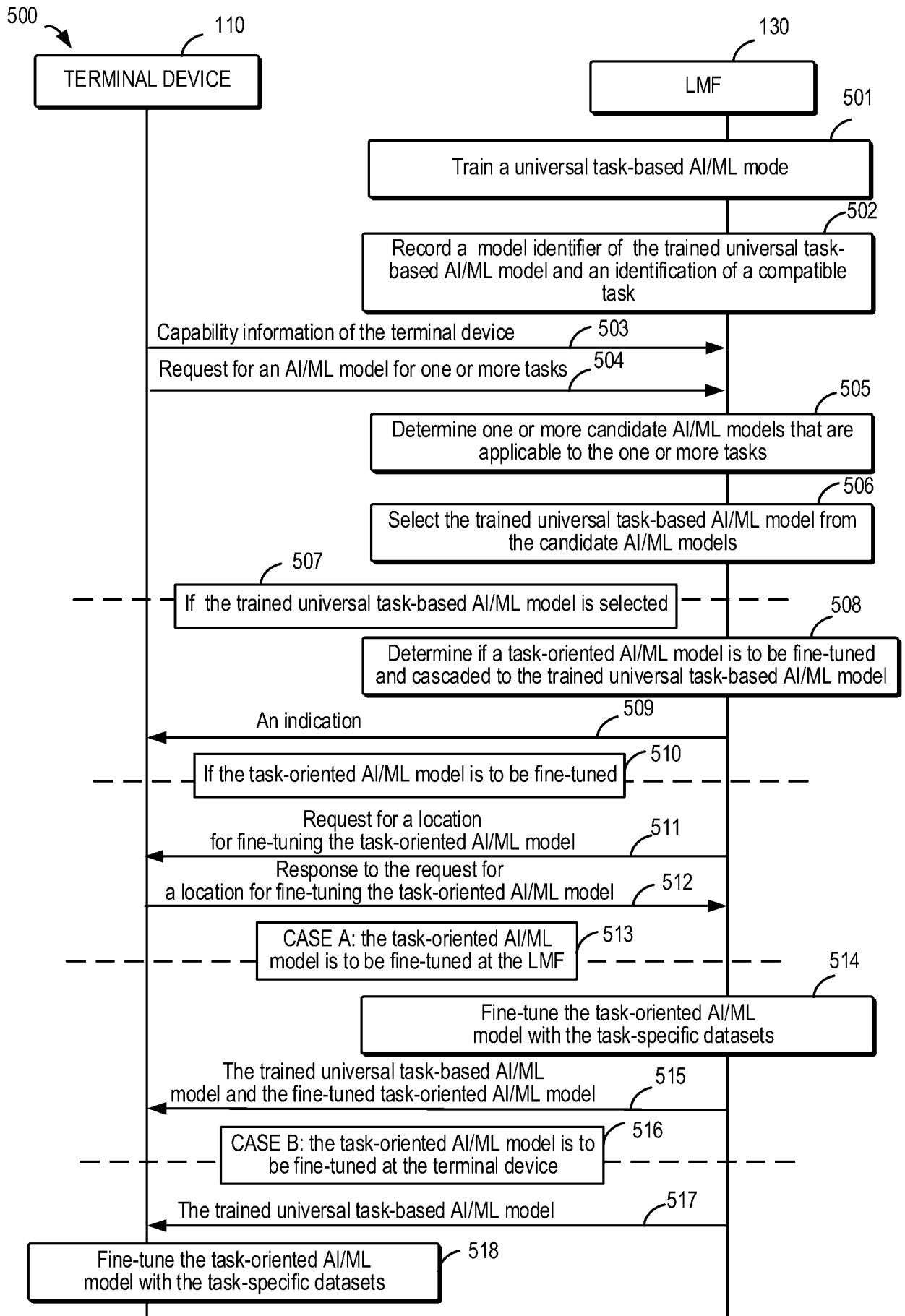


Fig.5

600

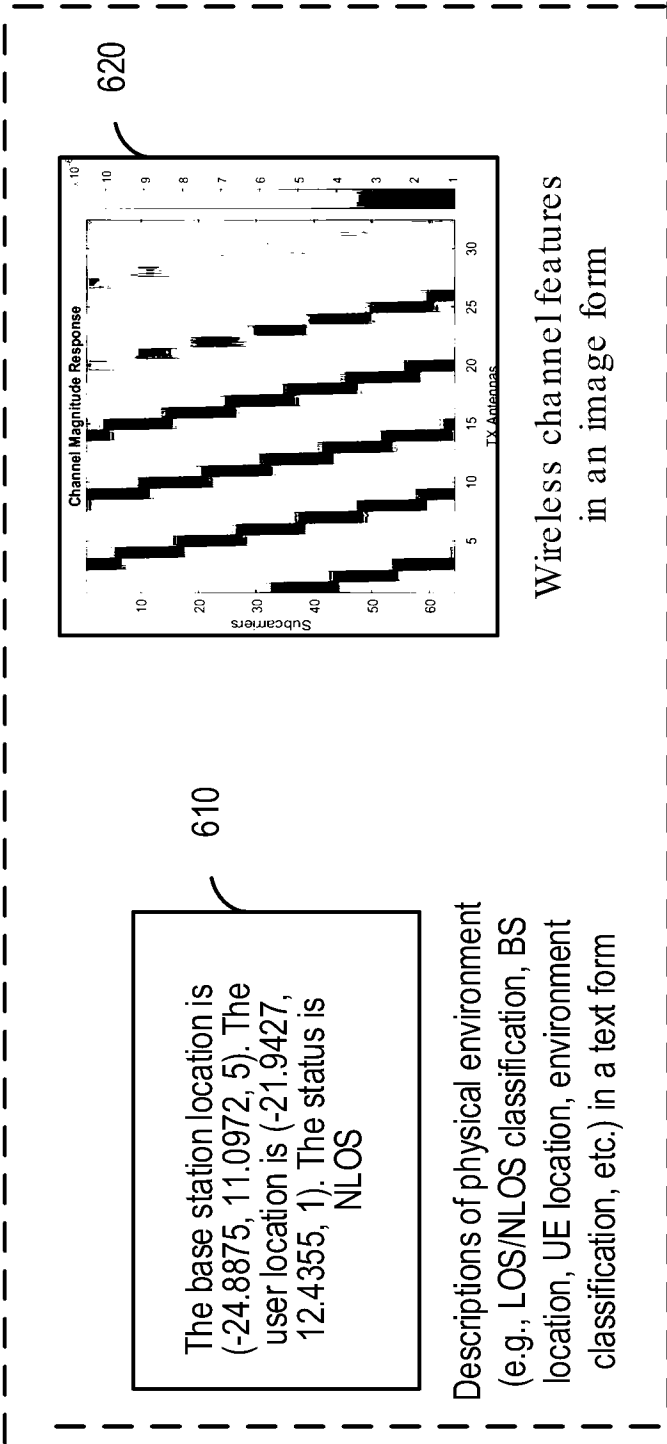


Fig.6

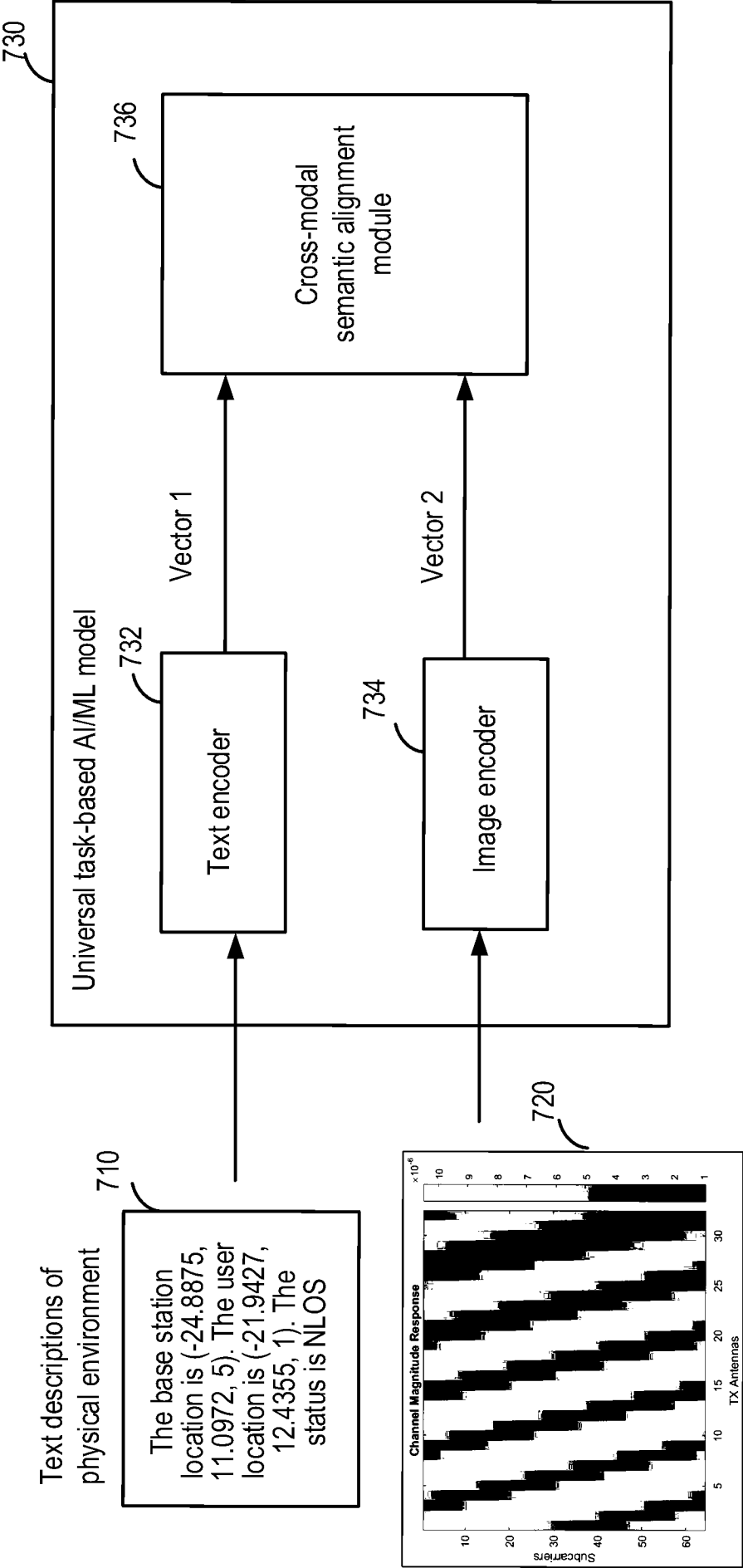


Fig.7

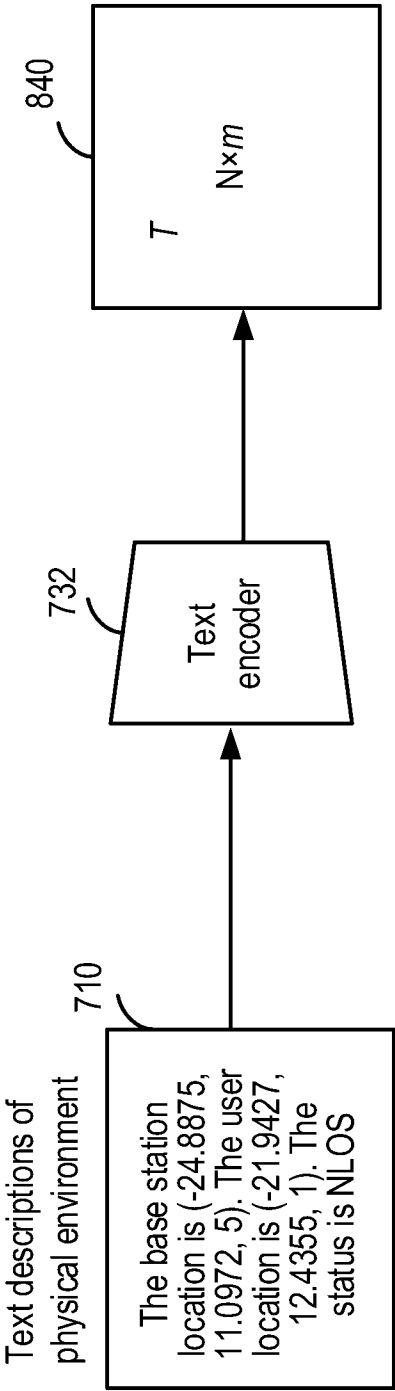
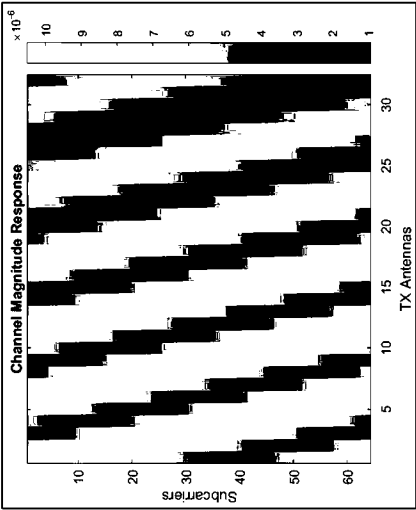
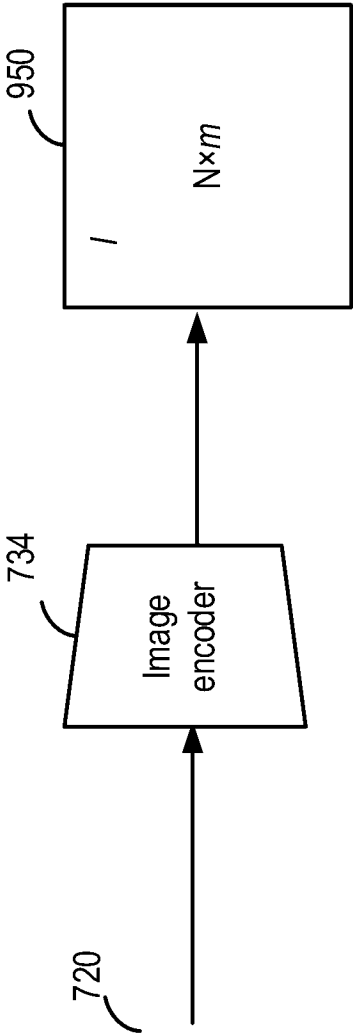


Fig.8

Fig.9



Wireless channel measurements



m is the output dimension of each sample

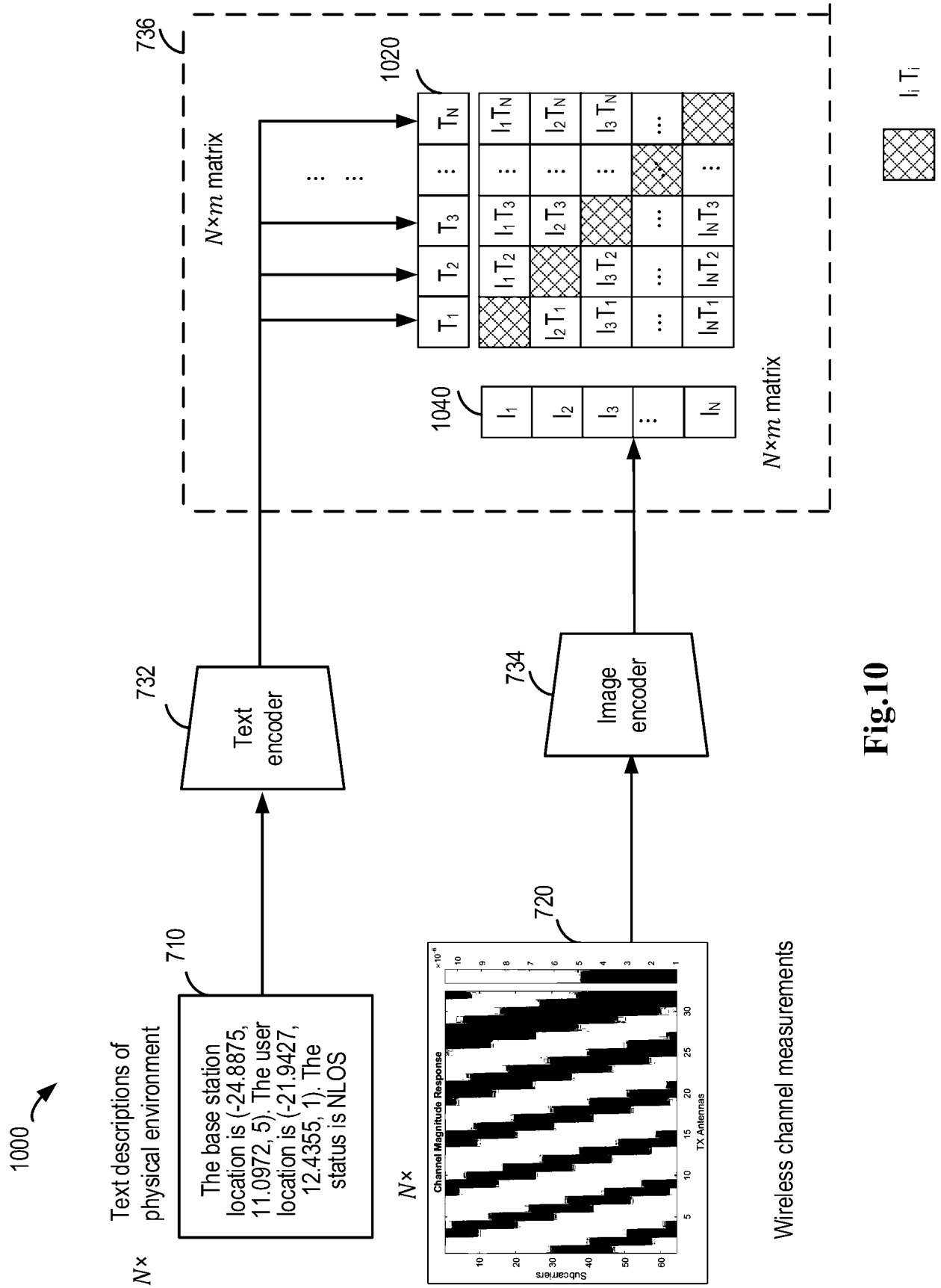
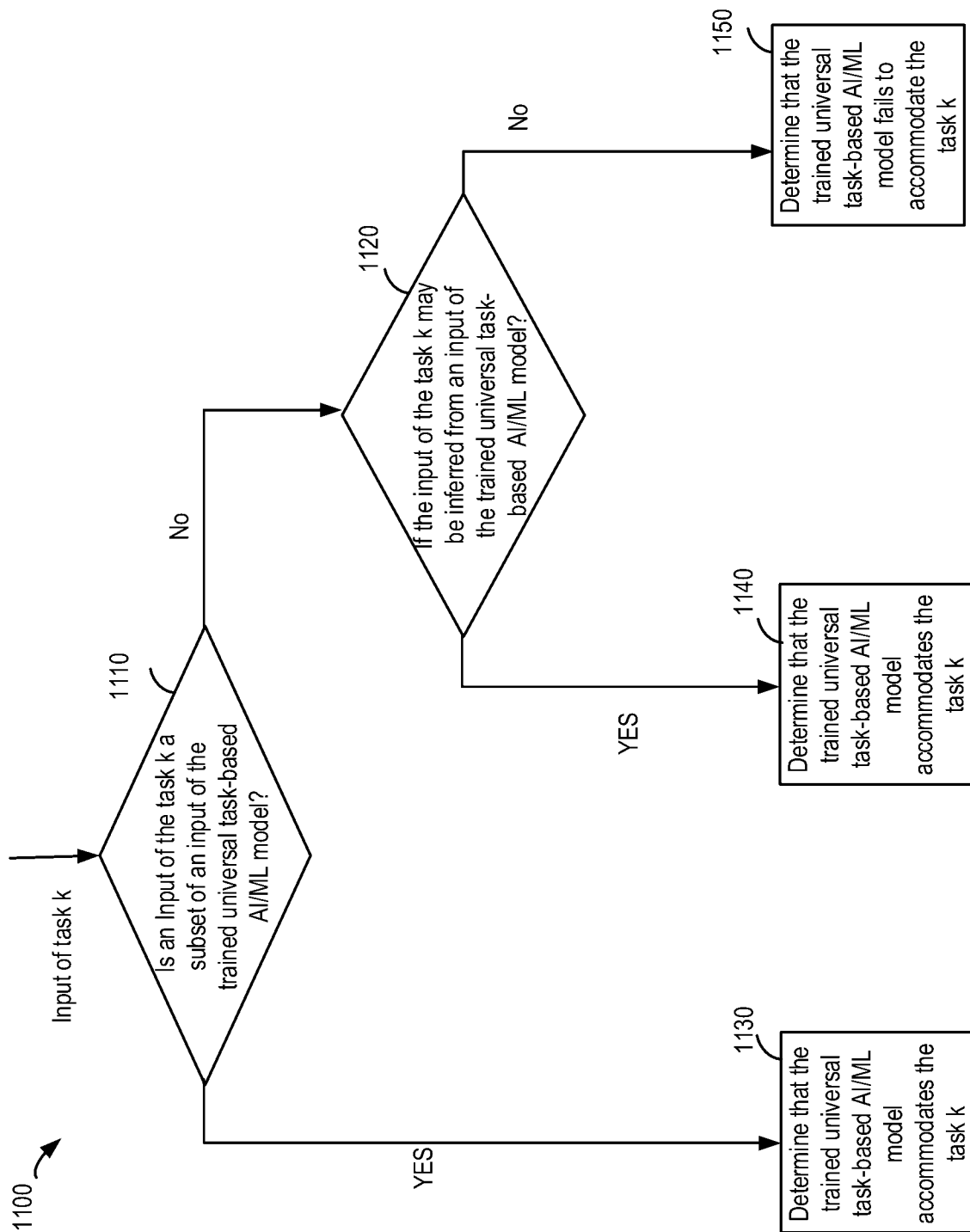


Fig. 10

**Fig.11**

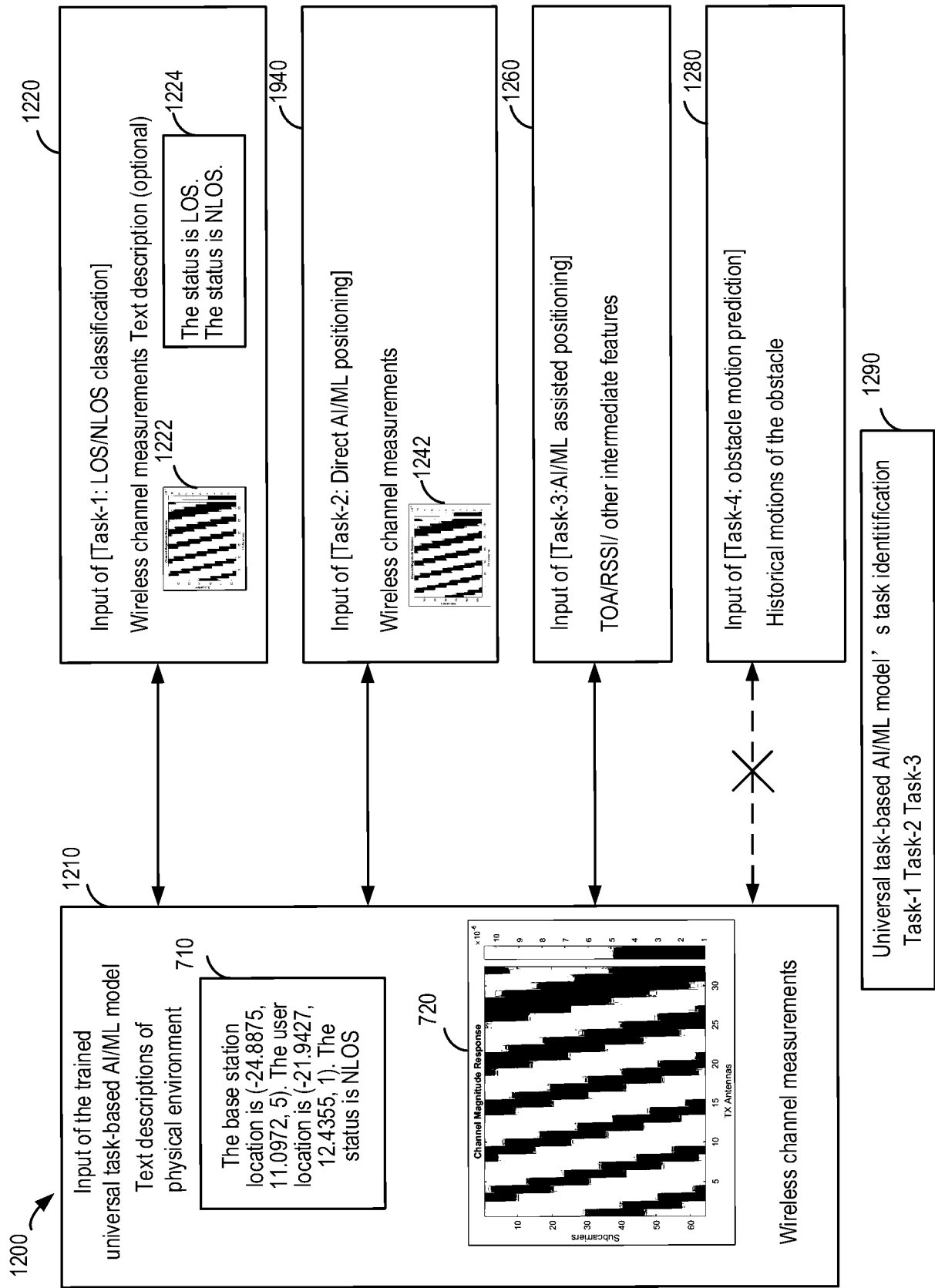


Fig.12

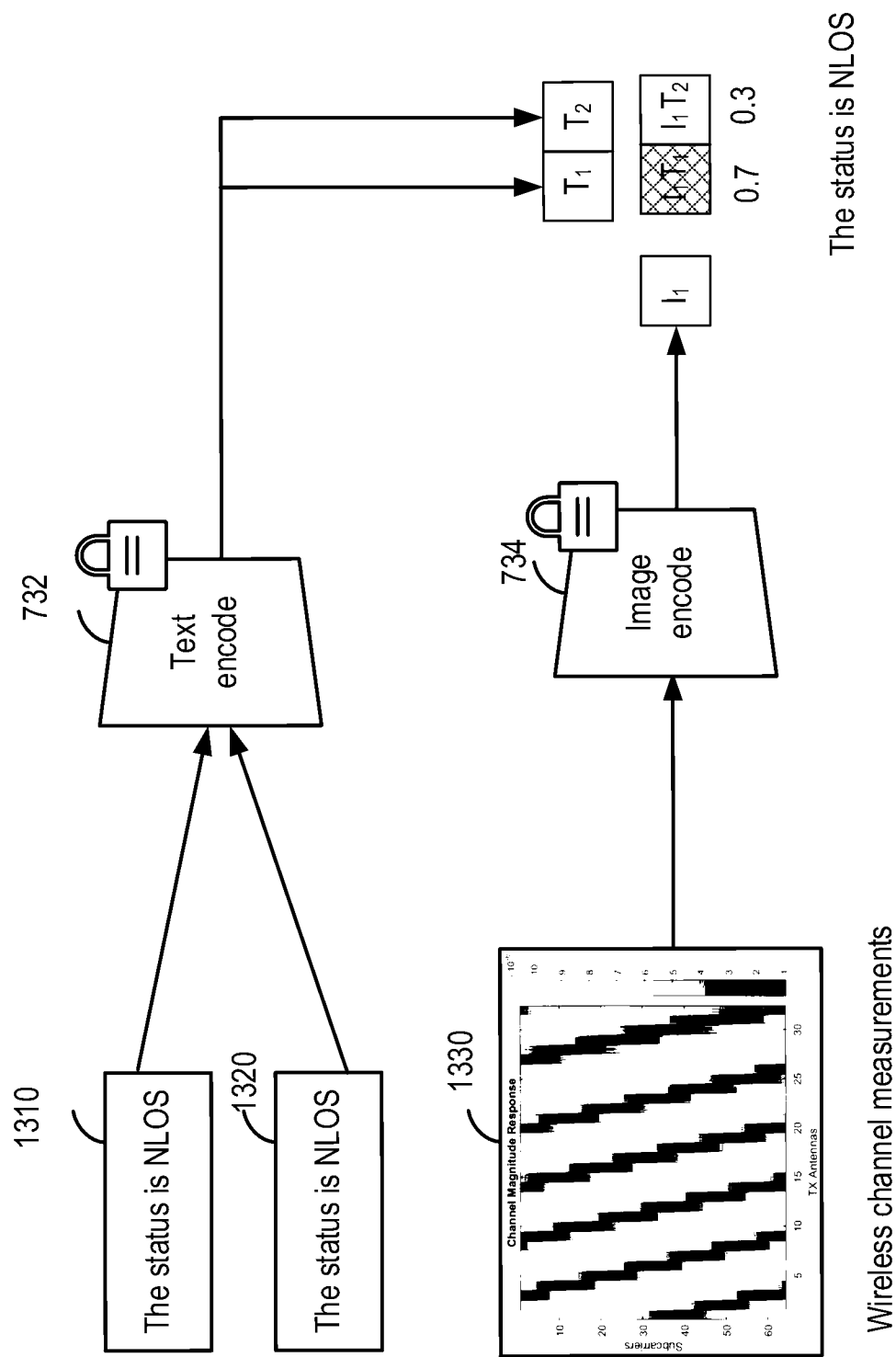


Fig.13

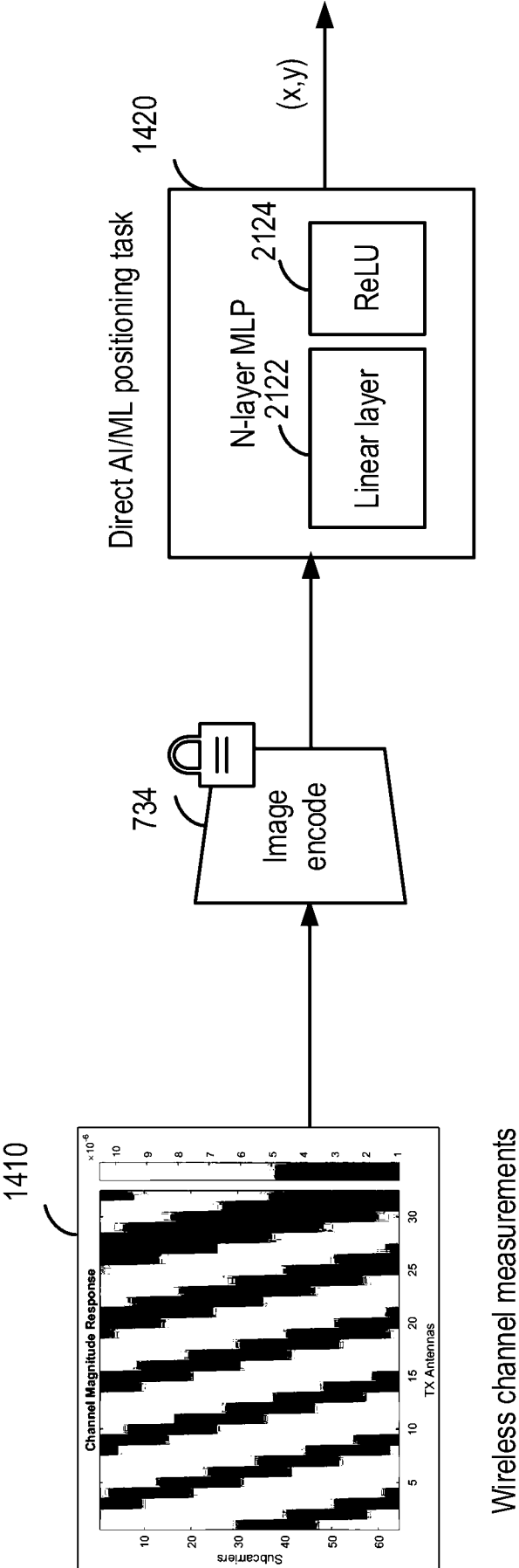
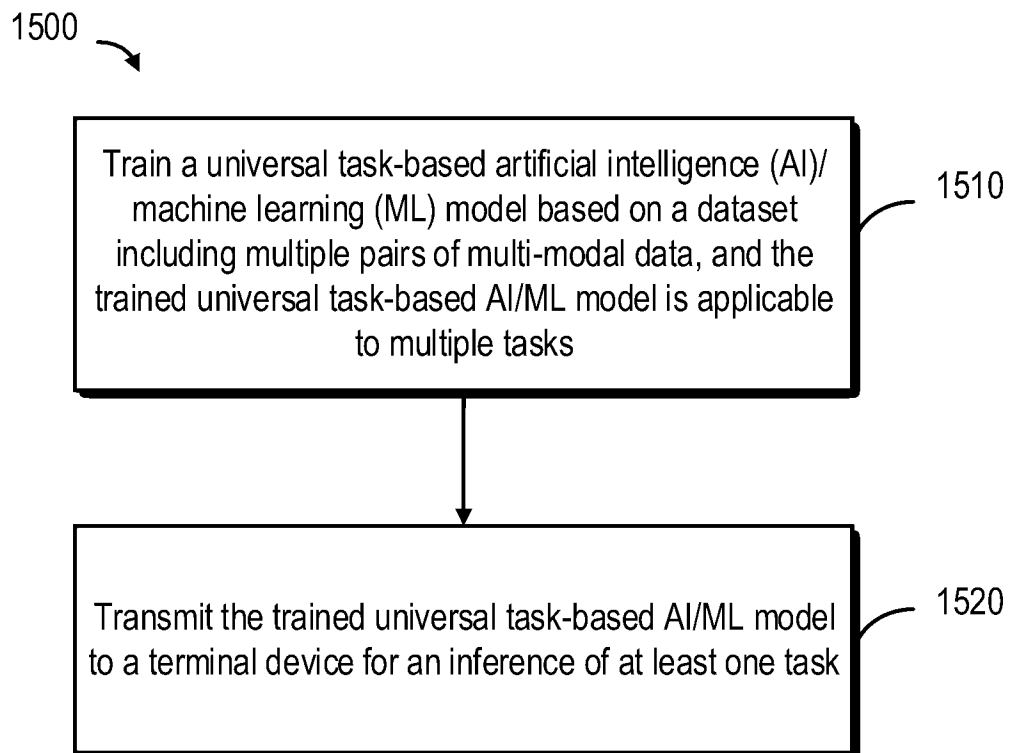
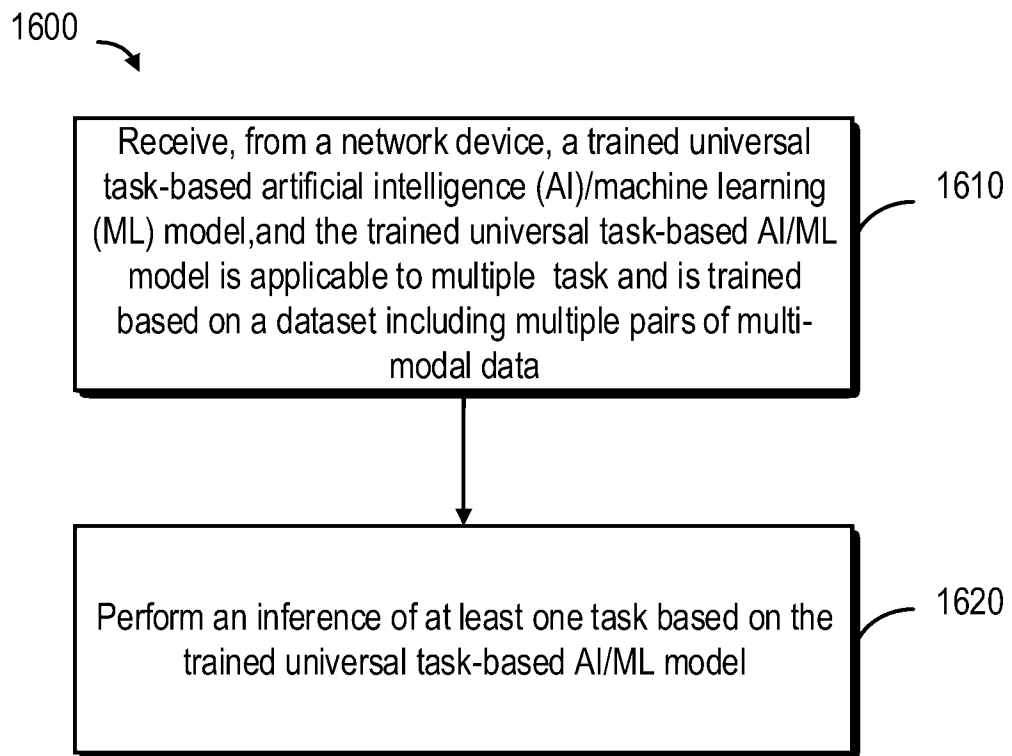
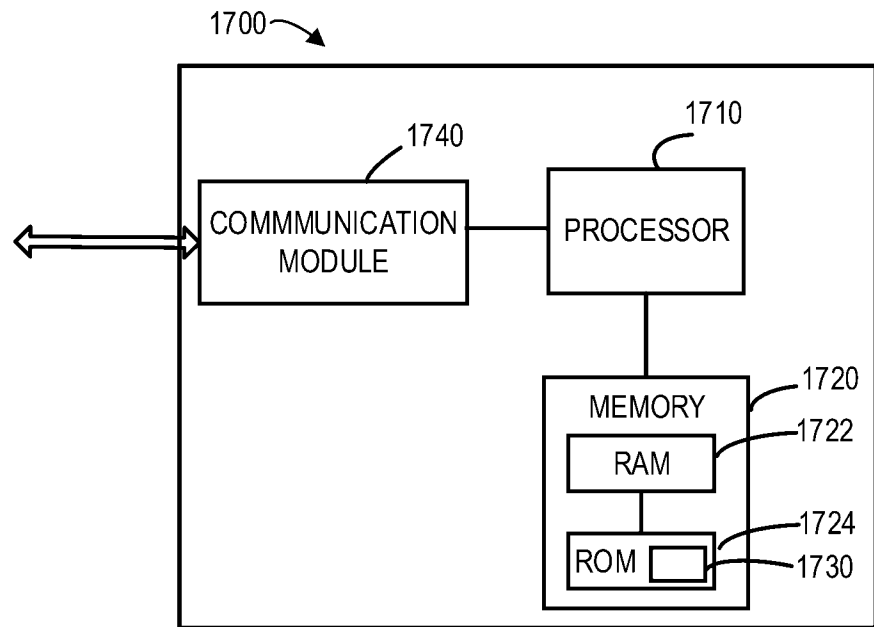


Fig.14

**Fig.15**

**Fig.16**

**Fig.17**

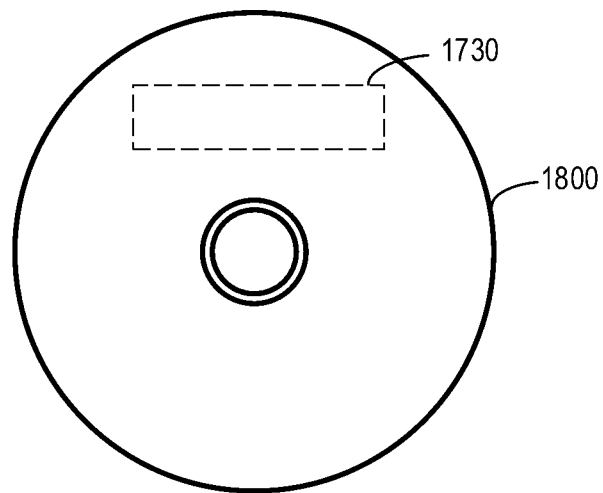


Fig.18

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/112733

A. CLASSIFICATION OF SUBJECT MATTER

G06N3/04(2023.01)i; G06N20/00(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN, CNABS, CNTXT, USTXT, CNKI, IEEE: artificial intelligence, AI, machine learning, ML, multi modal, train, task, inference, fine-tune, encoder, text, image, similarity, maximize, minimize

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2022292269 A1 (BEIJING BAIDU NETCOM SCIENCE TECHNOLOGY CO., LTD.) 15 September 2022 (2022-09-15) claims 1-20; description, paragraphs [0003]-[0010]	1-24
X	CN 112990297 A (BEIJING ACADEMY OF ARTIFICIAL INTELLIGENCE et al.) 18 June 2021 (2021-06-18) description, paragraphs [0005]-[0045]	1-24
A	US 2021397941 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 23 December 2021 (2021-12-23) the whole document	1-24
A	CN 114912540 A (SHANGHAI SHANGTANG INTELLIGENT TECHNOLOGY CO., LTD.) 16 August 2022 (2022-08-16) the whole document	1-24
A	CN 115909374 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 04 April 2023 (2023-04-04) the whole document	1-24



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

08 January 2024

Date of mailing of the international search report

28 March 2024

Name and mailing address of the ISA/CN

**CHINA NATIONAL INTELLECTUAL PROPERTY
ADMINISTRATION**
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088, China

Authorized officer

KONG,Xin

Telephone No. (+86) 010-53961371

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/112733

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2022292269	A1	15 September 2022	EP	4060565	A1	21 September 2022
				JP	2022141587	A	29 September 2022
				KR	20220128935	A	22 September 2022
				CN	112668671	A	24 December 2021
CN	112990297	A	18 June 2021	None			
US	2021397941	A1	23 December 2021	None			
CN	114912540	A	16 August 2022	None			
CN	115909374	A	04 April 2023	None			