

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2025 年 2 月 20 日 (20.02.2025)



(10) 国际公布号
WO 2025/035256 A1

(51) 国际专利分类号:

H04W 72/0446 (2023.01)

(21) 国际申请号:

PCT/CN2023/112522

(22) 国际申请日:

2023 年 8 月 11 日 (11.08.2023)

(25) 申请语言:

中文

(26) 公布语言:

中文

(71) 申请人: **OPPO** 广东移动通信有限公司 (**GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP., LTD.**) [CN/CN]; 中国广东省东莞市长安镇乌沙海滨路 18 号, Guangdong 523860 (CN)。

(72) 发明人: 陈文洪 (**CHEN, Wenhong**); 中国广东省东莞市长安镇乌沙海滨路 18 号, Guangdong 523860 (CN)。田文强 (**TIAN, Wenqiang**); 中国广东省东莞市长安镇乌沙海滨路 18 号, Guangdong 523860 (CN)。

(74) 代理人: 北京派特恩知识产权代理有限公司 (**CHINA PAT INTELLECTUAL PROPERTY OFFICE**); 中国北京市海淀区苏州街 3 号大恒科技大厦南座五层 503, Beijing 100080 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): 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, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (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)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) **Title:** WIRELESS COMMUNICATION METHOD AND APPARATUS, TERMINAL DEVICE, AND NETWORK DEVICE

(54) 发明名称: 无线通信方法及装置、终端设备、网络设备

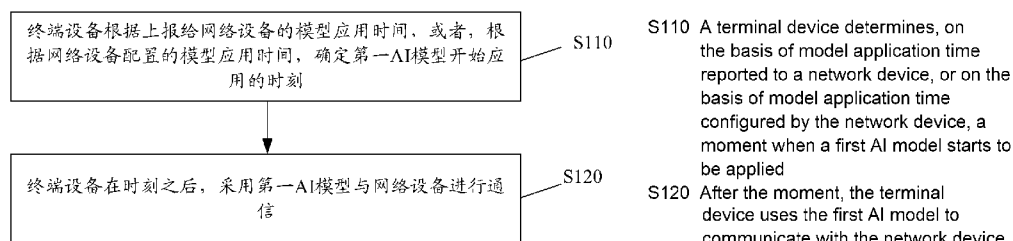


图 5

(57) **Abstract:** A wireless communication method and apparatus, a terminal device, and a network device. The method comprises: a terminal device determines, on the basis of model application time reported to a network device, or on the basis of model application time configured by the network device, a moment when a first AI model starts to be applied; and after the moment, the terminal device uses the first AI model to communicate with the network device.

(57) 摘要: 一种无线通信方法及装置、终端设备、网络设备, 该方法包括: 终端设备根据上报给网络设备的模型应用时间, 或者, 根据网络设备配置的模型应用时间, 确定第一AI模型开始应用的时刻; 终端设备在该时刻之后, 采用第一AI模型与网络设备进行通信。

WO 2025/035256 A1

无线通信方法及装置、终端设备、网络设备

技术领域

本申请实施例涉及移动通信技术领域，具体涉及一种无线通信方法及装置、终端设备、网络设备。

5 背景技术

鉴于人工智能 (Artificial Intelligence, AI) 技术在计算机视觉、自然语言处理等方面取得了巨大的成功，通信领域开始尝试利用 AI 技术来寻求新的技术思路来解决传统方法受限的技术难题。

AI 模型用于无线通信时，可以分为单端模型和双端模型。双端模型需要在终端设备和网络设备侧成对部署。示例性的，用于信道状态信息 (Channel State Information, CSI) 反馈的 AI 模型就是一种典型的双端模型，其中，终端设备可以将得到的信道信息 (例如特征向量，波束信息，时延信息等) 作为 AI 模型的输入，经 AI 模型推理出相应的 CSI 量化比特。相应的，网络设备侧对应的 AI 模型，可以将 CSI 量化比特作为输入，反向推理得到相应的信道信息。

对于双端模型，如果当前终端设备侧的 AI 模型性能较差，网络设备可以通过信令指示终端设备进行模型更新 (例如更换一个 AI 模型或者调整模型参数)。此时，网络设备也需要进行相应的模型更新，以使两侧的 AI 模型仍然是互相匹配的。如果两侧 AI 模型的更新不是完全同步的，会导致收端 AI 模型的输出是错误的，从而影响终端设备和网络设备之间的信息传输。如何保证网络设备与终端设备之间同步进行模型更新是需要解决的问题。

发明内容

本申请实施例提供一种信息传输方法及装置、终端设备、网络设备。

20 第一方面，本申请实施例提供的无线通信方法，包括：

终端设备根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

所述终端设备在所述时刻之后，采用所述第一 AI 模型与所述网络设备进行通信。

第二方面，本申请实施例提供的无线通信方法，包括：

25 网络设备根据终端设备上报的模型应用时间，或者，根据为所述终端设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

所述网络设备在所述时刻之后，采用所述第一 AI 模型对应的第二 AI 模型与所述终端设备进行通信。

第三方面，本申请实施例提供的无线通信装置，应用于终端设备，所述装置包括：

30 第一确定单元，配置为根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

第一通信单元，配置为在所述时刻之后，采用所述第一 AI 模型与所述网络设备进行通信。

第四方面，本申请实施例提供的无线通信装置，应用于网络设备，所述装置包括：

35 第二确定单元，配置为根据终端设备上报的模型应用时间，或者，根据为所述终端设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

第二通信单元，配置为在所述时刻之后，采用所述第一 AI 模型对应的第二 AI 模型与所述终端设备进行通信。

第五方面，本申请实施例提供的终端设备，该终端设备包括处理器和存储器。该存储器用于存储计算机程序，该处理器用于调用并运行该存储器中存储的计算机程序，执行上述的无线通信方法。

40 第六方面，本申请实施例提供的网络设备，该网络设备包括处理器和存储器。该存储器用于存储计算机程序，该处理器用于调用并运行该存储器中存储的计算机程序，执行上述的无线通信方法。

本申请实施例提供的芯片，用于实现上述的无线通信方法。

具体地，该芯片包括：处理器，用于从存储器中调用并运行计算机程序，使得安装有该芯片的

设备执行上述的无线通信方法。

本申请实施例提供的计算机可读存储介质，用于存储计算机程序，该计算机程序使得计算机执行上述的无线通信方法。

本申请实施例提供的计算机程序产品，包括计算机程序指令，该计算机程序指令使得计算机执行上述的无线通信方法。

本申请实施例提供的计算机程序，当其在计算机上运行时，使得计算机执行上述的无线通信方法。

本申请实施例提供的无线通信方法中，在终端设备侧和网络设备侧采用相对应的双端 AI 模型时，如果终端设备侧的 AI 模型发生更新或重配置，终端设备可以根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定更新或重配置后的 AI 模型开始应用的时刻。这样，可以确保两侧的模型开始应用的时刻是相同的，从而保证 AI 模型的输出是可用的，提高了通信系统的性能。

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

- 图 1 是本申请实施例提供的一种通信架构示意图；
- 图 2 是本申请实施例提供的一种神经元结构示意图；
- 图 3 是本申请实施例提供的一种神经网络结构示意图；
- 图 4 是本申请实施例提供的一种用于 CSI 反馈的神经网络示意图；
- 图 5 是本申请实施例提供的一种无线通信方法流程示意图；
- 图 6A 是本申请实施例提供的一种时隙结构示意图一；
- 图 6B 是本申请实施例提供的一种时隙结构示意图二；
- 图 7A 是本申请实施例提供的一种时隙结构示意图三；
- 图 7B 是本申请实施例提供的一种时隙结构示意图四；
- 图 8 是本申请实施例提供的一种时隙结构示意图五；
- 图 9A 是本申请实施例提供的一种时隙结构示意图六；
- 图 9B 是本申请实施例提供的一种时隙结构示意图七；
- 图 9C 是本申请实施例提供的一种时隙结构示意图八；
- 图 10 是本申请实施例提供的一种无线通信装置 1000 的结构示意图；
- 图 11 是本申请实施例提供的一种无线通信装置 1100 的结构示意图；
- 图 12 是本申请实施例提供的一种通信设备示意性结构图；
- 图 13 是本申请实施例的芯片的示意性结构图；
- 图 14 是本申请实施例提供的一种通信系统的示意性框图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

图 1 是本申请实施例的一个应用场景的示意图。

如图 1 所示，通信系统 100 可以包括终端设备 110 和网络设备 120。网络设备 120 可以通过空口与终端设备 110 通信。终端设备 110 和网络设备 120 之间支持多业务传输。

应理解，本申请实施例仅以通信系统 100 进行示例性说明，但本申请实施例不限于此。也就是说，本申请实施例的技术方案可以应用于各种通信系统，例如：长期演进（Long Term Evolution, LTE）系统、LTE 时分双工（Time Division Duplex, TDD）、通用移动通信系统（Universal Mobile Telecommunication System, UMTS）、物联网（Internet of Things, IoT）系统、窄带物联网（Narrow Band Internet of Things, NB-IoT）系统、增强的机器类型通信（enhanced Machine-Type Communications, eMTC）系统、5G 通信系统（也称为新无线（New Radio, NR）通信系统），或未来的通信系统等。

在图 1 所示的通信系统 100 中，网络设备 120 可以是与终端设备 110 通信的接入网设备。接入

网络设备可以为特定的地理区域提供通信覆盖，并且可以与位于该覆盖区域内的终端设备 110（例如 UE）进行通信。

网络设备 120 可以是长期演进（Long Term Evolution, LTE）系统中的演进型基站（Evolutional Node B, eNB 或 eNodeB），或者是下一代无线接入网（Next Generation Radio Access Network, NG RAN）设备，或者是 NR 系统中的基站（gNB），或者是云无线接入网络（Cloud Radio Access Network, CRAN）中的无线控制器，或者该网络设备 120 可以为中继站、接入点、车载设备、可穿戴设备、集线器、交换机、网桥、路由器，或者未来演进的公共陆地移动网络（Public Land Mobile Network, PLMN）中的网络设备等。

终端设备 110 可以是任意终端设备，其包括但不限于与网络设备 120 或其它终端设备采用有线或者无线连接的终端设备。

例如，所述终端设备 110 可以指接入终端、用户设备（User Equipment, UE）、用户单元、用户站、移动站、移动台、远方站、远程终端、移动设备、用户终端、终端、无线通信设备、用户代理或用户装置。接入终端可以是蜂窝电话、无绳电话、会话启动协议（Session Initiation Protocol, SIP）电话、IoT 设备、卫星手持终端、无线本地环路（Wireless Local Loop, WLL）站、个人数字处理（Personal Digital Assistant, PDA）、具有无线通信功能的手持设备、计算设备或连接到无线调制解调器的其它处理设备、车载设备、可穿戴设备、5G 网络中的终端设备或者未来演进网络中的终端设备等。

终端设备 110 可以用于设备到设备（Device to Device, D2D）的通信。

无线通信系统 100 还可以包括与网络设备 120 进行通信的核心网设备 130，该核心网设备 130 可以是 5G 核心网（5G Core, 5GC）设备，例如，接入与移动性管理功能（Access and Mobility Management Function, AMF），又例如，认证服务器功能（Authentication Server Function, AUSF），又例如，用户面功能（User Plane Function, UPF），又例如，会话管理功能（Session Management Function, SMF）。可选地，核心网络设备 130 也可以是 LTE 网络的分组核心演进（Evolved Packet Core, EPC）设备，例如，会话管理功能+核心网的数据网关（Session Management Function + Core Packet Gateway, SMF+PGW-C）设备。应理解，SMF+PGW-C 可以同时实现 SMF 和 PGW-C 所能实现的功能。在网络演进过程中，上述核心网设备也有可能叫其它名字，或者通过对核心网的功能进行划分形成新的网络实体，对此本申请实施例不做限制。

通信系统 100 中的各个功能单元之间还可以通过下一代网络（next generation, NG）接口建立连接实现通信。

例如，终端设备通过 NR 接口与接入网设备建立空口连接，用于传输用户面数据和控制面信令；终端设备可以通过 NG 接口 1（简称 N1）与 AMF 建立控制面信令连接；接入网设备例如下一代无线接入基站（gNB），可以通过 NG 接口 3（简称 N3）与 UPF 建立用户面数据连接；接入网设备可以通过 NG 接口 2（简称 N2）与 AMF 建立控制面信令连接；UPF 可以通过 NG 接口 4（简称 N4）与 SMF 建立控制面信令连接；UPF 可以通过 NG 接口 6（简称 N6）与数据网络交互用户面数据；AMF 可以通过 NG 接口 11（简称 N11）与 SMF 建立控制面信令连接；SMF 可以通过 NG 接口 7（简称 N7）与 PCF 建立控制面信令连接。

图 1 示例性地示出了一个网络设备、一个核心网设备和两个终端设备，可选地，该无线通信系统 100 可以包括多个网络设备并且每个网络设备的覆盖范围内可以包括其它数量的终端设备，本申请实施例对此不做限定。

需要说明的是，图 1 只是以示例的形式示意本申请所适用的系统，当然，本申请实施例所示的方法还可以适用于其它系统。此外，本文中术语“系统”和“网络”在本文中常被可互换使用。本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。还应理解，在本申请的实施例中提到的“指示”可以是直接指示，也可以是间接指示，还可以是表示具有关联关系。举例说明，A 指示 B，可以表示 A 直接指示 B，例如 B 可以通过 A 获取；也可以表示 A 间接指示 B，例如 A 指示 C，B 可以通过 C 获取；还可以表示 A 和 B 之间具有关联关系。还应理解，在本申请的实施例中提到的“对应”可表示两者之间具有直接对应或间接对应的关系，也可以表示两者之间具有关联关系，也可以是指示与被指示、配置与被配置等关系。还应理解，在本申请的实施例中提到的“预定义”或“预定义规则”可以通过在设备（例如，包括终端设备和网络设备）中预先保存相应的代码、表格或其他可用于指示相关信息的方式来实现，本申请对于其具体的实现方式不做限定。比如预定义可以是指协议中定义的。还应理解，本申请实施例中，所述“协议”可以指通信领域的标准协议，例如可以包括 LTE 协议、

NR 协议以及应用于未来的通信系统中的相关协议，本申请对此不做限定。

为便于理解本申请实施例的技术方案，以下对本申请实施例的相关技术进行说明，以下相关技术作为可选方案与本申请实施例的技术方案可以进行任意结合，其均属于本申请实施例的保护范围。

AI 模型是一种可以胜任多种任务处理的模型，它们具备自我学习和自我适应的能力，可以根据环境的变化来进行动态调整和决策。AI 模型也可以称为机器学习（Machine Learning, ML）模型，两者等同或者可替换。

实际应用中，AI 模型可以由神经网络构成。神经网络是一种由多个神经元节点相互连接构成的运算模型，其中节点间的连接代表从输入信号到输出信号的加权值，称为权重；每个节点对不同的输入信号进行加权求和，并通过特定的激活函数输出。参考图 2 所示的神经元结构示意图，其中， a_1 、 a_2 、...、 a_n 和 1 为神经元的输入， w_1 、 w_2 、...、 w_n 和 b 表示权重，Sum 表示求和函数， f 表示激活函数， t 为输出结果。

一个简单的神经网络如图 3 所示，包含输入层、隐藏层和输出层，通过多个神经元不同的连接方式，权重和激活函数，可以产生不同的输出，进而拟合从输入到输出的映射关系。每一个上一级节点都与其全部的下一级节点相连。此全连接模型也可以叫做 DNN，即深度神经网络。

通过数据集的构建，训练，验证和测试等过程可以训练并得到一个 AI 模型。训练可以分为离线训练和在线训练。可以通过数据集离线训练的方式得到一个静态的训练结果，这里可以称之为离线训练。在网络设备或终端设备对 AI 模型的使用过程中，随着终端设备的进一步测量和/或上报，网络设备可以继续收集更多的数据，进行实时的在线训练来优化 AI 模型的参数，达到更好的推断和预测结果。在得到 AI 模型后，通过将当前得到的信息输入到 AI 模型中，可以推理得到相应的模型输出。

AI 用于无线通信时，可以分为单端模型和双端模型。其中，单端模型只在终端设备或网络设备一侧部署即可使用，AI 模型的训练也可以只在单侧进行即可；双端模型需要在终端设备和网络设备侧成对部署，两侧的模型需要一起进行训练，也就是说两侧部署的模型是对应的，不能单独使用或者更新。

示例性的，用于信道状态信息（Channel State Information, CSI）反馈的 AI 模型就是一种典型的双端模型。参考图 4 所示的一种用于 CSI 反馈的双端 AI 模型结构示意图。其中，在终端侧部署用于编码（encoder）的 AI 模型，在网络侧部署与之相对应的用于解码（decoder）的 AI 模型。终端设备基于 encoder 模型输出 CSI 量化比特，典型的如 PMI 比特，再通过上行信道（PUSCH/PUCCH）反馈给网络设备。网络设备将终端反馈的 CSI 量化比特（PMI）作为 decoder 模型的输入，从而输出与终端侧输入对应的信道信息，如各个子带的特征向量等，用于下行的预编码。同时，终端设备还需要进行模型性能监测，在模型性能不好时上报给网络设备，让网络设备进行模型的更新（例如模型结构的更新或者模型参数的更新）。由于两侧的 AI 模型是匹配的，如果网络设备进行模型更新，需要通过信令通知终端设备，令终端设备也更新成对应的 AI 模型，否则网络设备输出的结果就是不可用的。

需要说明的是，上述基于双端部署的 AI 模型不仅可以用于 CSI 反馈，也可以用于终端设备和网络设备侧具有对应操作/结构的其他处理方式，例如信道编码-信道解码，调制-解调，导频生成-信道估计，收发射频信号处理等，只需要训练出对应的终端设备侧和网络设备侧的一套 AI 模型即可。

可以理解的，针对双端模型，如果当前终端设备侧的模型性能不好，网络设备可以通过信令指示终端设备进行模型更新（例如更换一个 AI 模型或者调整模型参数）。此时，网络设备也需要进行相应的模型更新，以使两侧的 AI 模型仍然是互相匹配的。如果两侧 AI 模型的更新不是完全同步的，会导致收端 AI 模型的输出是错误的，从而影响终端设备和网络设备之间的信息传输。

另外，由于不同终端设备的处理能力不同，部署 AI 模型所需要的时间也不同，如何保证网络设备与不同终端设备之间能够同步进行模型更新是需要解决的问题。

基于此，本申请实施例提供一种无线通信方法，其中，在终端设备侧和网络设备侧采用相对应的双端 AI 模型时，如果终端设备侧的 AI 模型发生更新或重配置，终端设备可以根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定网络设备配置的 AI 模型开始应用的时刻。这样，可以确保两侧的模型应用时刻是相同的，从而保证 AI 模型的输出是可用的，提高了通信系统的性能。

为便于理解本申请实施例的技术方案，以下通过具体实施例详述本申请的技术方案。以上相关技术作为可选方案与本申请实施例的技术方案可以进行任意结合，其均属于本申请实施例的保护范围。本申请实施例包括以下内容中的至少部分内容。

图 5 示出了本申请实施例提供的一种无线通信方法，该方法可以包括步骤 S110 和步骤 S120。

S110、终端设备根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

S120、终端设备在时刻之后，采用第一 AI 模型与网络设备进行通信。

5 相应的，网络设备可以根据终端设备上报的模型应用时间，或者，根据为终端设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；以及，网络设备在时刻之后，采用第一 AI 模型对应的第二 AI 模型与所述终端设备进行通信。

需要说明的是，终端设备和网络设备采用相对应的双端 AI 模型进行通信。其中，第一 AI 模型为部署于终端设备侧的神经网络模型，第二 AI 模型为部署于网络设备侧的神经网络模型。

10 示例性的，第一 AI 模型可以用于以下任一过程：

下行 CSI 反馈；或者，

下行波束反馈；或者，

下行数据的信道解码；或者，

下行信号的解调；或者，

15 下行信道估计；或者，

下行射频信号处理；或者，

上行数据的信道编码；或者，

上行数据的调制；或者，

上行导频生成；或者，

20 上行射频信号处理。

其中，第一 AI 模型用于下行 CSI 反馈，可以是终端设备将得到的信道信息（例如特征向量，波束信息，时延信息等）作为第一 AI 模型的输入，经第一 AI 模型推理出相应的 CSI 量化比特。

第一 AI 模型用于下行数据的信道解码，可以是终端设备将接收到的编码后的数据作为第一 AI 模型的输入，输出解码后的信息比特。

25 第一 AI 模型用于下行信号的解调，可以是终端设备将接收到的下行调制信号作为第一 AI 模型的输入，输出解调后的信息比特。

第一 AI 模型用于下行信道估计，可以是终端设备将接收到的下行信道和导频信号作为第一 AI 模型的输入，输出信道估计结果。

30 第一 AI 模型用于下行射频信号处理，可以是终端设备将接收到的下行射频信号作为第一 AI 模型的输入，输出下行射频信号处理结果。

第一 AI 模型用于上行数据的信道编码，可以是终端设备将待发送的信息比特作为第一 AI 模型的输入，输出编码后的数据。

第一 AI 模型用于上行数据的调制，可以是终端设备将待发送的上行数据的信息比特作为第一 AI 模型的输入，输出上行调制信号。

35 第一 AI 模型用于上行导频生成，可以是终端设备将待发送的信号作为第一 AI 模型的输入，输出导频信号。

第一 AI 模型用于上行射频信号处理，可以是终端设备将待发送的信号作为第一 AI 模型的输入，输出上行射频信号。

可以理解的，第一 AI 模型可以通过数据集的构建、训练、验证和测试等过程训练得到的。

40 可以理解的，第一 AI 模型可以提前通过离线训练和/或在线训练的方式训练得到的。

示例性的，第二 AI 模型可以用于以下任一过程：

下行 CSI 反馈信息的解码；或者，

下行波束反馈信息的解码；或者，

上行数据的信道解码；或者，

45 上行信号的解调；或者，

上行信道估计；或者，

上行射频信号处理；或者，

下行数据的信道编码；或者，

下行数据的调制；或者，

50 下行导频生成；或者，

下行射频信号处理。

需要说明的是，第一 AI 模型和第二 AI 模型对应，可以是指第一 AI 模型和第二 AI 模型实现操作是一一对应的。示例性的，第一 AI 模型用于下行 CSI 反馈，则第二 AI 模型用于下行 CSI 反馈信息的解码；或者，第一 AI 模型用于下行波束信息的生成，则第二 AI 模型可以用于下行波束反馈信息的解码；第一 AI 模型用于上行数据的信道编码，则第二 AI 模型可以用于上行数据的信道解码等。

5 可以理解的，第二 AI 模型可以通过数据集的构建、训练、验证和测试等过程训练得到的。

可以理解的，第二 AI 模型可以提前通过离线训练和/或在线训练的方式训练得到的。

在一些实施例中，终端设备侧部署的第一 AI 模型可以是网络设备为终端设备配置的。需要说明的是，网络设备可以在终端设备初始接入时为其配置第一 AI 模型；或者，网络设备可以在终端设备监测到原 AI 模型性能较差时，为其配置第一 AI 模型，以指示终端设备进行模型更新。

10 在一些实施例中，网络设备可以向终端设备发送第一信息，相应的，终端设备接收网络设备发送的第一信息，第一信息用于配置第一 AI 模型。换句话说，第一信息为用于配置第一 AI 模型的下行信息。

在一些实施例中，第一信息可以通过广播信令发送给终端设备。示例性的，第一信息可以通过物理广播信道（Physical Broadcast Channel, PBCH），系统信息（System Information Bits, SIB），组公用下行控制信息（Group Common DCI）等承载，本申请实施例对此不做限制。需要说明的是，Group Common DCI 是通过公共搜索空间（Common Search Space, CSS）发给一组终端设备的 DCI。其中，Group Common DCI 采用公共的无线网临时标识（Radio Network Temporary Identifier, RNTI）进行加扰。在一些实施例中，第一信息可以是高层信息，通过高层信令发送。示例性的，第一信息可以通过无线资源控制（Radio Resource Control, RRC）信令发送，也可以通过媒体接入控制（Media Access Control, MAC）信令发送，如 MAC 控制单元（MAC Control Element, MAC CE）。在一些实施例中，第一信息也可以是物理层信息，通过物理层信令发送。示例性的，第一信息可以通过 DCI 承载。

25 在一种实施方式中，所述第一信息可以是网络设备发送的模型恢复信息，所述模型恢复信息中包含第一 AI 模型的配置信息。网络设备在接收到终端设备发送的模型失败上报信息后，可以发送相应的模型恢复信息，用于终端侧模型的更新。

在一些实施例中，第一信息可以包括以下中的一项或多项：

第一 AI 模型的标识信息；

第一 AI 模型的模型功能（functionality）；

第一 AI 模型的模型结构；以及，

30 第一 AI 模型的模型参数。

在一些实施例中，第一信息中可以携带第一 AI 模型的标识信息。应理解，终端设备可以预配置或者预定义一个或多个 AI 模型。终端设备可以根据第一信息中携带的第一 AI 模型的标识信息，从一个或多个 AI 模型中确定即将应用/更新的第一 AI 模型。其中，每个 AI 模型都可以有与之一一对应的一个标识信息，所述对应关系可以由网络设备与终端设备预先约定好，或者由网络设备配置给终端设备。

40 在一些实施例中，第一信息中可以指示第一 AI 模型的模型功能。其中，模型功能即通过 AI/ML 模型所实现的功能，不同的模型具有不同的模型功能，例如：PMI 上报，RI/PMI/CQI 上报，波束信息上报，RSRP 上报，CSI 压缩，信道编码解码，调制解调，CSI 预测，CSI 预测及压缩等。终端设备接收到第一信息后，可以根据 AI 模型所要实现的模型功能，确定所使用的第一 AI 模型。

需要说明的是，第一信息中携带的第一 AI 模型的模型结构可以是第一 AI 模型中的至少部分模型结构，或者说，第一信息可以携带全部或部分第一 AI 模型的模型结构。类似的，第一信息中携带的第一 AI 模型的模型参数可以是第一 AI 模型中的至少部分模型参数，或者说，第一信息可以携带全部或部分第一 AI 模型的模型参数。可以理解的，网络设备不一定要配置整个 AI 模型，可以只配置或更新第一 AI 模型中的部分结构或者部分模型参数。

45 在另一些实施例中，终端设备侧部署的第一 AI 模型也可以是终端设备从预配置或者预定义的一个或多个 AI 模型中自行选择并使用的模型。需要说明的是，终端设备可以在初始接入时，向网络设备上报其使用的第一 AI 模型；另外，终端设备也可以在监测到原 AI 模型性能较差时，进行模型更新（如通过在线训练），使用第一 AI 模型进行通信，同时向网络设备指示第一 AI 模型，以使网络设备对网络设备侧当前部署的对应的模型进行更新。

50 其中，终端设备可以向网络设备发送第二信息，相应的，网络设备接收终端设备发送的第二信息，第二信息用于向网络设备指示终端设备使用的第一 AI 模型。此时，第二信息可以通过物理层信

道发送给网络设备,典型的,通过 PUSCH 或者 PUCCH 上报给网络设备。如果所述第二信息是物理层信息,则可以通过 PUSCH 或者 PUCCH 上报给网络设备,例如作为 CSI 上报的一部分;如果所述第二信息是高层信息,则可以通过 PUSCH 上报给网络设备,例如 MAC 信令。

网络设备收到第二信息后,可以发送第二信息对应的响应信息,以使终端设备知道网络设备正确接收到了所述第二信息。例如,所述第二信息通过 PRACH 承载,则所述第二信息的响应信息可以是对应的 RAR (随机接入响应);所述第二信息通过 PUSCH 承载,则所述第二信息可以是所述 PUSCH 的 HARQ-ACK 信息。

在一些实施例中,第二信息可以包括以下中的一项或多项:

第一 AI 模型的标识信息;

第一 AI 模型的模型功能 (functionality);

第一 AI 模型的模型结构;以及,

第一 AI 模型的模型参数。

在一些实施例中,第二信息中可以携带第一 AI 模型的标识信息。应理解,网络设备可以根据第二信息中携带的第一 AI 模型的标识信息,从一个或多个 AI 模型中确定终端设备即将应用/更新的第一 AI 模型,从而确定与第一 AI 模型对应的第二 AI 模型。其中,每个 AI 模型都可以有与之一一对应的一个标识信息,所述对应关系可以由网络设备与终端设备预先约定好,或者由网络设备配置给终端设备。

在一些实施例中,第二信息中可以指示第一 AI 模型的模型功能。其中,模型的功能即通过 AI/ML 模型所实现的功能,不同的模型具有不同的模型功能,例如: PMI 上报, RI/PMI/CQI 上报, 波束信息上报, RSRP 上报, CSI 压缩, 信道编码解码, 调制解调, CSI 预测, CSI 预测及压缩等。网络设备接收到第二信息后,可以根据 AI 模型所要实现的模型功能,确定终端设备即将使用的/更新的第一 AI 模型,从而确定与第一 AI 模型对应的第二模型。

需要说明的是,第二信息中携带的第一 AI 模型的模型结构可以是第一 AI 模型中的至少部分模型结构,或者说,第二信息可以携带全部或部分第一 AI 模型的模型结构。类似的,第二信息中携带的第一 AI 模型的模型参数可以是第一 AI 模型中的至少部分模型参数,或者说,第二信息可以携带全部或部分第一 AI 模型的模型参数。可以理解的,终端设备不一定要指示整个第一 AI 模型,可以只配置或更新第一 AI 模型中的部分结构或者部分模型参数。

可以理解的,在终端设备接收到用于配置第一 AI 模型的第一信息,或者终端设备向网络设备发送指示第一 AI 模型的第二信息后,终端设备和网络设备均需要确定第一 AI 模型开始应用的时刻,从而在该时刻开始启用该第一 AI 模型与网络设备进行通信,保证终端设备侧和网络设备侧对应的模型的应用时间同步,避免两侧模型不匹配的问题。

在一种可能的实现方式中,终端设备可以根据其上报给网络设备的模型应用时间,确定网络设备配置的/终端设备指示的第一 AI 模型开始应用的时刻;相应的,网络设备可以根据终端设备上报的模型应用时间,确定网络设备为终端设备配置的/终端设备指示的第一 AI 模型开始应用的时刻。

在另一种可能的实现方式中,终端设备可以根据网络设备配置的模型应用时间,确定网络设备配置的/终端设备指示的第一 AI 模型开始应用的时刻;相应的,网络设备可以根据其为终端设备配置的模型应用时间,确定网络设备配置的/终端设备指示的第一 AI 模型开始应用的时刻。

下面通过方式#A 和方式#B 说明终端设备确定第一 AI 模型开始应用的时刻的详细流程。

方式#A,终端设备根据上报给网络设备的模型应用时间,确定网络设备配置的/终端设备指示的第一 AI 模型开始应用的时刻;相应的,网络设备可以根据终端设备上报的模型应用时间,确定其为终端设备配置的/终端设备指示的第一 AI 模型开始应用的时刻。

需要说明的是,在确定第一 AI 模型开始应用的时刻之前,终端设备可以向网络设备上报其支持的模型应用时间。

示例性的,终端设备通过 UE 能力信息 (例如 *UECapabilityInformation*),向网络设备上报该模型应用时间。

在一些实施例中,终端设备可以在接收网络设备发送的用于配置第一 AI 模型的第一信息之前,向网络设备上报模型应用时间。相应的,网络设备可以在发送第一信息之前,可以接收终端设备上报的模型应用时间。

在一些实施例中,终端设备可以在发送用于指示第一 AI 模型的第二信息之前,向网络设备上报模型应用时间;相应的,网络设备可以在接收第二信息之前,接收终端设备上报的模型应用时间。或者,终端设备可以在发送用于指示第一 AI 模型的第二信息中,携带模型应用时间,将使用的第一

AI 模型和模型应用时间一起上报给网络设备。这样，网络设备接收到第二信息后也能同时确定模型的应用时间，能够更快的应用模型。

在一些实施例中，对于网络设备通过第一信息为终端设备配置第一 AI 模型的场景，模型应用时间可以是以下中的任意一项：

- 5 终端设备接收到第一信息之后到应用第一 AI 模型的时刻之间的时间间隔；
- 终端设备发送第一信息的 HARQ-ACK 信息之后到应用第一 AI 模型的时刻之间的时间间隔；
- 终端设备接收到第一信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

- 10 终端设备发送第一信息的 HARQ-ACK 信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔。

- 在一种可能的实现方式中，对于模型的输出信息不需要上报网络设备的第一 AI 模型，例如用于下行数据的信道解码、下行信道的解调、下行信道估计、下行射频信号处理、上行数据的信道编码、上行数据的调制、上行导频生成、以及上行射频信号处理等操作的第一 AI 模型，其模型应用时间可以是终端设备接收到第一信息之后到应用第一 AI 模型的时刻之间的时间间隔，或者终端设备发送第一信息的 HARQ-ACK 信息之后到应用第一 AI 模型的时刻之间的时间间隔。

- 在一种可能的实现方式中，对于模型的输出信息需要上报给网络设备的第一 AI 模型，例如，用于下行 CSI 反馈或者下行波束信息反馈的第一 AI 模型，其模型应用时间可以是终端设备接收到第一信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔，或者，终端设备发送第一信息的 HARQ-ACK 信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔。

- 在一些实施例中，对于终端设备通过第二信息向网络设备指示第一 AI 模型的场景，模型应用时间可以为以下中的任意一项：

- 终端设备发送第二信息之后到应用第一 AI 模型的时刻之间的时间间隔；
- 终端设备发送第二信息之后到所述终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔；
- 终端设备接收第二信息的响应信息之后到应用第一 AI 模型的时刻之间的时间间隔；
- 终端设备接收第二信息的响应信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔。

- 其中，对于模型的输出信息不需要上报网络设备的第一 AI 模型，例如用于下行数据的信道解码、下行信道的解调、下行信道估计、下行射频信号处理、上行数据的信道编码、上行数据的调制、上行导频生成、以及上行射频信号处理等操作的第一 AI 模型，模型应用时间可以是终端设备发送第二信息之后到应用第一 AI 模型的时刻之间的时间间隔，或者，终端设备接收第二信息的响应信息之后到应用第一 AI 模型的时刻之间的时间间隔。

- 另外，对于模型的输出信息需要上报给网络设备的第一 AI 模型，例如，用于下行 CSI 反馈或者下行波束信息反馈的第一 AI 模型，模型应用时间可以是终端设备发送第二信息之后到所述终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔，或者，终端设备接收第二信息的响应信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔。

- 需要说明的是，上述两种场景中模型应用时间涉及的终端设备发送基于第一 AI 模型得到的反馈信息的时刻，可以是指终端设备第一次发送基于第一 AI 模型得到的反馈信息的时刻。

- 示例性的，对于用于下行 CSI 反馈的第一 AI 模型，模型应用时间可以是：终端设备接收到第一信息之后到终端设备（第一次）发送基于第一 AI 模型得到的 CSI 反馈信息的时刻之间的时间间隔，或者终端设备发送第一信息的 HARQ-ACK 信息之后到终端设备（第一次）发送基于第一 AI 模型得到的 CSI 反馈信息的时刻之间的时间间隔。模型应用时间也可以是：终端设备发送第二信息之后到终端设备（第一次）发送基于第一 AI 模型得到的 CSI 反馈信息的时刻之间的时间间隔，或者终端设备接收第二信息的响应信息之后到终端设备（第一次）发送基于第一 AI 模型得到的 CSI 反馈信息的时刻之间的时间间隔。

- 示例性的，对于用于下行波束信息反馈的第一 AI 模型，模型应用时间可以是：终端设备接收到第一信息之后到终端设备（第一次）发送基于第一 AI 模型得到的下行波束信息的时刻之间的时间间隔，或者终端设备发送第一信息的 HARQ-ACK 信息之后到终端设备（第一次）发送基于第一 AI 模型得到的下行波束信息的时刻之间的时间间隔。另外，模型应用时间也可以是终端设备发送第二信息之后到终端设备（第一次）发送基于第一 AI 模型得到的下行波束信息的时刻之间的时间间隔，或

者终端设备接收到第二信息的响应信息之后到终端设备（第一次）发送基于第一 AI 模型得到的下行波束信息的时刻之间的时间间隔。

还需要说明的是，上述时间间隔可以以时隙或 OFDM 符号为单位。此外，上述时间间隔还可以以微时隙、OFDM 符号集合为单位，或者以绝对时间（例如 N 毫秒，N 秒， $N>0$ ）为单位，本申请实施例对此不做限制。

在一些实施例中，终端设备上报给网络设备的模型应用时间可以包括以下中的一项或多项：

终端设备支持的一个或多个 AI 模型分别对应的模型应用时间；其中，终端设备支持的一个或多个 AI 模型包括第一 AI 模型；

终端设备支持的一种或多种模型更新方式分别对应的模型应用时间；其中，终端设备支持的一种或多种模型更新方式包括第一 AI 模型的更新方式；

终端设备支持的一个或多个 AI 模型功能分别对应的模型应用时间；其中，终端设备支持的一个或多个模型功能包括第一 AI 模型的功能；

终端设备支持的一个或多个模型类型分别对应的模型应用时间；其中，终端设备支持的一个或多个模型类型包括第一 AI 模型的类型。

在一种可能的实现方式中，终端设备可以支持一个或多个 AI 模型，该一个或多个 AI 模型可以是预配置或者预定义的。可以理解的，不同的 AI 模型的模型结构和模型参数不同，因此不同的 AI 模型的模型应用时间也就不同。终端设备可以为支持一个或多个 AI 模型分别上报各自的模型应用时间。可以理解的，终端设备可以为不同的模型上报不同的模型应用时间。

在一种可能的实现方式中，终端设备可以支持一种或多个模型更新方式。示例性的，模型更新方式包括：更新模型参数和/或更新模型结构。

其中，终端设备可以为支持的一种或多种 AI 模型更新方式上报各自对应的模型应用时间。可以理解的，终端设备可以为不同的模型更新方式上报不同的模型应用时间。示例性的，只更新模型参数需要的应用时间为 t_1 ，更新模型结构和模型参数需要的应用时间为 t_2 ，且 $t_1 < t_2$ 。

需要说明的是，同时更新模型结构和模型参数可以认为是更新了整个 AI 模型。

在一种可能的实现方式中，终端设备可以支持一个或多个模型类型。示例性的，终端设备支持的模型类型可以包括高精度模型和低精度模型；或者，已知模型和未知模型。

其中，高精度模型的复杂度高于低精度模型，所以需要的模型应用时间也较长。这里，相对于低精度模型，高精度模型可以是输出比特数更多的模型，或者是神经网络层数更多的模型，或者是神经网络结构更复杂的模型。

另外，已知模型可以为终端设备已经部署过的模型，未知模型为终端设备尚未部署过的模型。其中，已知模型需要的应用时间短于未知模型需要的应用时间。

需要说明的是，终端设备上报的模型应用时间可以根据终端设备的处理能力或硬件配置，以及不同 AI 模型/不同 AI 模型类型推理需要的时间来确定。示例性的，终端设备采用 CPU 用于 AI 模型的模型应用时间，可以长于采用 GPU 用于 AI 模型的模型应用时间。

可以理解的，终端设备和网络设备根据上报的模型应用时间，确定第一 AI 模型（第一 AI 模型可以是网络设备通过第一信息配置的，也可以是终端设备通过第二信息指示的）开始应用的时刻，可以通过以下中的一项或多项实现：

根据网络设备配置的/终端设备指示的第一 AI 模型，以及终端设备上报的一个或多个 AI 模型分别对应的模型应用时间，确定第一 AI 模型开始应用的时刻；

根据网络设备配置的/终端设备指示的第一 AI 模型的功能，以及终端设备上报的一个或多个模型功能分别对应的模型应用时间，确定第一 AI 模型开始应用的时刻；

根据网络设备配置的/终端设备指示的第一 AI 模型的模型更新方式，以及终端设备上报的一种或多种模型更新方式分别对应的模型应用时间，确定该第一 AI 模型开始应用的时刻；

根据网络设备配置的/终端设备指示的第一 AI 模型的模型类型，以及终端设备上报的一个或多个模型类型分别对应的模型应用时间，确定第一 AI 模型开始应用的时刻。

示例性的，终端设备可以上报 AI 模型 1 对应的模型应用时间为 T_1 ，AI 模型 2 对应的模型应用时间为 T_2 。若网络设备通过第一信息为终端设备配置模型 1，则终端设备和网络设备可以根据 AI 模型 1 对应的模型应用时间 T_1 ，确定第一信息配置的 AI 模型 1 开始应用的时刻。或者，若终端设备通过第二信息指示模型 2，则终端设备和网络设备可以根据 AI 模型 2 对应的模型应用时间 T_2 ，确定第二信息指示的 AI 模型 2 的开始应用的时刻。

示例性的，终端设备可以上报模型功能 1（例如 CSI 压缩功能）对应的模型应用时间为 T_3 ，模

型功能 2（例如信道编码解码功能）对应的模型应用时间为 T4。若网络设备通过第一信息为终端设备配置模型功能 1，则终端设备和网络设备可以根据模型功能 1 对应的模型应用时间 T3，确定第一信息配置的 AI 模型开始应用的时刻。或者，若终端设备通过第二信息指示模型功能 2，则终端设备和网络设备可以根据模型功能 2 对应的模型应用时间 T4，确定第二信息指示的 AI 模型开始应用的时刻。

示例性的，终端设备可以上报高精度模型对应的模型应用时间 T5，低精度模型对应的模型应用时间 T6。若网络设备通过第一信息为终端设备配置的 AI 模型为低精度模型，则终端设备和网络设备可以基于低精度模型的模型应用时间 T6，确定第一信息配置的 AI 模型开始应用的时刻。若终端设备通过第二信息指示 AI 模型的类型为高精度模型，则终端设备和网络设备可以基于高精度模型的应用时间 T5，确定第二信息指示的 AI 模型开始应用的时刻。

示例性的，终端设备上报已知模型对应的模型应用时间 T7，未知模型对应的模型应用时间 T8。若网络设备通过第一信息为终端设备配置的 AI 模型 1 为未知模型（例如 AI 模型 1 为终端设备预配置或者预定义的候选模型之外的模型），则终端设备和网络设备可以基于未知模型的模型应用时间 T8，确定第一信息配置的 AI 模型 1 开始应用的时刻。

需要说明的是，在终端设备上报的第一 AI 模型对应的模型应用时间，第一 AI 模型的模型功能对应的模型应用时间，第一 AI 模型的模型类型对应的模型应用时间，以及第一 AI 模型的模型更新方式对应的模型应用时间中的两个或两个以上不同的情况下，终端设备可以根据最长（或最短）的模型应用时间，来确定第一 AI 模型的开始应用时刻。

方式#B，终端设备根据网络设备配置的模型应用时间，确定网络设备配置的/终端设备指示的第一 AI 模型开始应用的时刻；相应的，网络设备可以根据其为终端设备配置的模型应用时间，确定网络设备配置的/终端设备指示的第一 AI 模型开始应用的时刻。

需要说明的是，模型应用时间的相关解释说明可以参考上述实施例中的描述，为了简洁，此处不再赘述。

在一些实施例中，网络设备可以为终端设备直接配置模型应用时间。其中，网络设备可以向终端设备发送第三信息，相应的，终端设备可以接收网络设备发送的第三信息，其中，第三信息用于配置模型应用时间。也就是说，网络设备通过第三信息为终端设备配置模型应用时间。

在一些实施例中，第三信息可以通过广播信令发送给终端设备。示例性的，第三信息可以通过 PBCH，SIB，Group Common DCI 等承载，本申请实施例对此不做限制。在一些实施例中，第三信息可以是高层信息，通过高层信令发送。示例性的，第三信息可以通过 RRC 信令，或者 MAC CE 信令发送。在一些实施例中，第三信息也可以是物理层信息，通过物理层信令发送。示例性的，第三信息可以通过 DCI 信令承载。

针对网络设备通过第一信息为终端设备配置第一 AI 模型的场景，第三信息可以包括第一信息中所配置的第一 AI 模型的模型应用时间。

需要说明的是，所述第三信息可以与第一信息一起指示给终端设备，这样终端设备收到模型的配置信息后也能同时确定模型的应用时间，能够更快的应用模型。

在一种可能的实现方式中，网络设备可以根据预设规则或者预设的映射关系，为终端设备配置第一 AI 模型的模型应用时间。

示例性的，网络设备可以根据为终端设备配置的第一 AI 模型的类型、第一 AI 模型的更新方式等因素，为终端设备配置第一 AI 模型的应用时间。

示例性的，网络设备可以预先存储每种 AI 模型对应的模型应用时间（AI 模型和模型应用时间之间的映射关系）。当网络设备为终端设备配置第一 AI 模型时，可以基于该映射关系，为终端设备配置第一 AI 模型对应的模型应用时间。

在另一种可能的实现方式中，网络设备可以接收终端设备上报的模型应用时间，根据终端设备上报的模型应用时间来为终端设备配置实际的模型应用时间。

需要说明的是，网络设备可以根据覆盖范围内多个终端设备分别上报的模型应用时间，来确定为终端设备配置的模型应用时间。

在一些实施例中，网络设备可以采用多个终端设备分别上报的模型应用时间中的最大值，作为覆盖范围内各个终端设备实际使用的模型应用时间，并将该模型应用时间配置给这些终端设备。也就是说，网络设备为终端设备配置的模型应用时间的时长，大于或等于终端设备上报的模型应用时间的时长。

由此可见，网络设备可以为同时其覆盖范围内的多个不同的终端设备配置模型应用时间，这样，

网络设备与不同终端设备能够同步进行模型更新,从而保证接收端的模型输出是可用的,提高通信性能。

针对终端设备通过第二信息指示第一 AI 模型的场景,在一种可能的实现方式中,网络设备可以在接收到第二信息之后,确定第二信息指示的第一 AI 模型的模型应用时间,并通过第三信息向终端设备指示该模型应用时间。也就是说,网络设备可以在接收到第二信息后,向终端设备发送第三信息,以为终端设备配置第二信息中所指示的第一 AI 模型的模型应用时间。

需要说明的是,网络设备可以根据预设规则或者预设的映射关系,为终端设备配置第一 AI 模型的模型应用时间。或者,网络设备可以根据终端设备上报的模型应用时间来为终端设备配置实际的模型应用时间。例如,网络设备可以采用多个终端设备分别上报的模型应用时间中的最大值,作为覆盖范围内各个终端设备实际使用的模型应用时间,并将该模型应用时间配置给这些终端设备。

针对终端设备通过第二信息指示第一 AI 模型的场景,在另一种可能的实现方式中,网络设备还可以在第二信息之前,向终端设备发送第三信息。该实现方式中第三信息可以包括以下中的一项或多项:

一个或多个 AI 模型分别对应的模型应用时间;其中,所述一个或多个 AI 模型包括所述第一 AI 模型;

一个或多个模型功能分别对应的模型应用时间;其中,所述一个或多个模型功能包括所述第一 AI 模型的功能;

一种或多种模型更新方式分别对应的模型应用时间;其中,所述一种或多种模型更新方式包括所述第一 AI 模型的更新方式;

一个或多个模型类型分别对应的模型应用时间;其中,所述一个或多个模型类型包括所述第一 AI 模型的类型。

可以理解的,终端设备接收到第三信息后,可以结合第二信息指示的第一 AI 模型,以及第三信息配置的每种 AI 模型、每种模型功能、每种模型更新方式、或者每个模型类型对应的模型应用时间,确定第二信息所指示的第一 AI 模型实际的模型应用时间。

需要说明的是,第三信息中配置的模型应用时间可以是网络设备根据终端设备上报的模型应用时间确定的。可以理解的,网络设备可以根据覆盖范围内多个终端设备分别上报的模型应用时间,来确定为终端设备配置的模型应用时间。在一些实施例中,网络设备可以采用多个终端设备分别上报的模型应用时间中的最大值,作为覆盖范围内各个终端设备实际使用的模型应用时间,并将该模型应用时间配置给这些终端设备。

示例性的,第三信息可以配置 AI 模型 1 (或者模型功能 1/模型更新方式 1/模型类型 1) 和 AI 模型 2 (或者模型功能 2/模型更新方式 2/模型类型 2) 分别对应的模型应用时间。其中,第三信息中配置的 AI 模型 1 (或者模型功能 1/模型更新方式 1/模型类型 1) 的模型应用时间,可以是多个终端设备上报的 AI 模型 1 (或者模型功能 1/模型更新方式 1/模型类型 1) 对应的模型应用时间中的最大值,第三信息中配置的 AI 模型 2 (或者模型功能 2/模型更新方式 2/模型类型 2) 的模型应用时间,可以是多个终端设备上报的 AI 模型 2 (或者模型功能 2/模型更新方式 2/模型类型 2) 对应的模型应用时间中的最大值。

这样,网络设备可以为同时其覆盖范围内的多个不同的终端设备配置模型应用时间。如此,网络设备与不同终端设备能够同步进行模型更新,从而保证接收端的模型输出是可用的,提高通信性能。

本申请实施例中,终端设备基于上述方式确定了第一 AI 模型开始应用的时刻之后,进一步地,终端设备可以在该时刻后,采用第一 AI 模型进行与网络设备之间的信息传输。相应的,网络设备可以在该时刻之后,采用与第一 AI 模型对应的第二模型进行与终端设备之间的信息传输。

在一些实施例中,在终端设备应用第一 AI 模型之前,终端设备可以采用第三 AI 模型进行与网络设备之间的信息传输,或者,终端设备不基于 AI 模型进行与网络设备之间的信息传输。相应的,在终端设备应用第一 AI 模型之前,网络设备可以采用与第三 AI 模型对应的第四 AI 模型进行与终端设备之间的信息传输,或者网络设备不基于 AI 模型进行与终端设备之间的信息传输,例如,网络设备基于传统的非 AI 方式来接收终端设备传输的信息。

需要说明的是,第三 AI 模型可以是预定义的 AI 模型,或者历史 AI 模型。其中,预定义的 AI 模型即终端设备和网络设备预先约定好的一个 AI 模型。历史 AI 模型可以是网络设备在第一 AI 模型开始应用的时刻之前,网络设备为终端设备配置的 AI 模型或者终端设备指示的 AI 模型,这里的历史 AI 模型可以是与第一 AI 模型的模型结构和模型参数完全不同的 AI 模型,也可以是模型结构相同

但模型参数不同的 AI 模型,也可以是模型功能不同的 AI 模型,本申请实施例对此不做限制。另外,第四 AI 模型可以是与第三 AI 模型对应的网络设备侧部署的 AI 模型。

示例性的,如果网络设备通过第一信息配置的第一 AI 模型/终端设备通过第二信息指示的第一 AI 模型为新的模型(即第一 AI 模型的模型 ID 与第三 AI 模型的模型 ID 不同),则在第一 AI 模型开始应用的时刻之前,终端设备可以使用之前使用的第三 AI 模型进行与网络设备之间的信息传输。

示例性的,如果网络设备通过第一信息配置了一个新的模型的功能/终端设备通过第二信息指示了一个新的模型的功能(即第一 AI 模型的 functionality 与第三 AI 模型的 functionality 不同),则在第一 AI 模型开始应用的时刻之前,终端设备可以使用之前的 functionality 对应的第三模型进行与网络设备之间的信息传输。这里的功能也可以是功能 ID。

示例性的,如果网络设备通过第一信息更新了模型结构和模型参数/终端设备通过第二信息更新了模型结构和模型参数,在终端设备确定的第一 AI 模型开始应用的时刻之前,终端设备可以使用之前采用的模型结构和模型参数进行与网络设备之间的信息传输。

示例性的,如果网络设备通过第一信息只更新了模型参数/终端设备通过第二信息只更新了模型参数,在终端设备确定的第一 AI 模型开始应用的时刻之前,终端设备可以使用之前采用的模型参数进行与网络设备之间的信息传输。

还需要说明的是,终端设备不基于 AI 模型进行与网络设备之间的信息传输,也就是指终端设备采用传统的非 AI 的方式进行与网络设备之间的信息传输。

综上所述,本申请实施例中,在终端设备侧和网络设备侧采用相对应的双端 AI 模型时,如果终端设备侧的 AI 模型发生更新或重配置,终端设备可以根据上报给网络设备的模型应用时间,或者,根据网络设备配置的模型应用时间,确定网络设备配置的 AI 模型开始应用的时刻。这样,可以确保两侧的模型应用时刻是相同的,从而保证 AI 模型的输出是可用的,提高了通信系统的性能。

针对网络设备通过第一信息为终端设备配置第一 AI 模型的场景,终端设备和网络设备确定的第一 AI 模型开始应用的时刻与第一信息有关。下面通过方式#1~方式#3 说明,网络设备通过第一信息为终端设备配置第一 AI 模型的场景中,第一 AI 模型开始应用的时刻。

方式#1,第一 AI 模型开始应用的时刻可以为:

承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一 OFDM 符号;或者,

承载第一信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一 OFDM 符号;其中, k 为模型应用时间。

需要说明的是,方式#1 中模型应用时间所表征的时间间隔可以以 OFDM 符号为单位。模型应用时间可以是 k 个 OFDM 符号,或者说,模型应用时间表征的时间间隔为 k 个 OFDM 符号。其中, k 为大于或等于 1 的整数。

可以理解的,如上文所述, k 可以是终端设备上报给网络设备的,也可以是网络设备配置的。

在一种可能的实现方式中,终端设备可以将承载第一信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一 OFDM 符号作为第一 AI 模型开始应用的时刻。或者说,终端设备可以将承载第一信息的下行信道的最后一个 OFDM 符号之后间隔 k 个 OFDM 符号的第一个时隙或者第一个 OFDM 符号作为第一 AI 模型开始应用的时刻。

需要说明的是,这里的下行信道可以是 PBCH, PDSCH, PDCCH 等。示例性的,第一信息为广播信息时,下行信道可以是 PBCH;第一信息为高层信息时,下行信道可以是 PDSCH;第一信息为物理层信息时,下行信道可以是 PDCCH。

还需要说明的是,本申请实施例中的 k 个 OFDM 符号也可以是 k-1 或 k+1 个 OFDM 符号,本申请实施例对此不做限制。

示例性的,参考图 6A 所示,假设承载第一信息的下行信道的最后一个 OFDM 符号为符号 m,则终端设备可以确定第一 AI 模型开始应用的时刻为符号 m+k 所在符号之后的第一个时隙。

示例性的,参考图 6B 所示,假设承载第一信息的下行信道的最后一个 OFDM 符号为符号 m,则终端设备可以确定第一 AI 模型开始应用的时刻为符号 m+k 所在符号之后的第一个 OFDM 符号,即图 6B 所示的符号 m+k+1。

在另一种可能的实现方式中,终端设备可以将承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一 OFDM 符号作为第一 AI 模型开始应用的时刻。或者说,终端设备可以将承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号之后间隔 k 个 OFDM 符号的第一个时隙或者第一个 OFDM 符号作为第一 AI 模型

开始应用的时刻。

可以理解的,终端设备接收到第一信息后,可以向网络设备发送第一信息的 HARQ-ACK 信息。示例性的,第一信息为高层信息或者物理层信息时,终端设备接收到第一信息后,可以向网络设备反馈第一信息的 HARQ-ACK 信息。

5 需要说明的是,承载第一信息的 HARQ-ACK 信息的上行信道可以是 PUCCH。

还需要说明的是,本申请实施例中的 k 个 OFDM 符号也可以是 $k-1$ 或 $k+1$ 个 OFDM 符号,本申请实施例对此不做限制。

10 示例性的,参考图 7A 所示,假设承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号为符号 m ,则终端设备可以确定第一 AI 模型开始应用的时刻为符号 $m+k$ 所在符号之后的第一个时隙。

示例性的,参考图 7B 所示,假设承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号为符号 m ,则终端设备可以确定第一 AI 模型开始应用的时刻为符号 $m+k$ 所在符号之后的第一个 OFDM 符号,即图 7B 所示的符号 $m+k+1$ 。

15 可以理解的,将 k 个 OFDM 符号之后的第一 OFDM 符号作为第一 AI 模型开始应用的时刻,相比于将 k 个 OFDM 符号之后的第一时隙作为第一 AI 模型开始应用的时刻,可以缩短第一 AI 模型应用的时间间隔,同时还可以提高模型应用时刻的准确度。

方式#2,第一 AI 模型开始应用的时刻可以为:

终端设备接收到第一信息的时隙开始 k 个时隙之后的第一个时隙, k 为所述模型应用时间。

20 需要说明的是,方式#2 中模型应用时间所表征的时间间隔可以以时隙为单位。模型应用时间可以是 k 个时隙,或者说,模型应用时间表征的时间间隔为 k 个时隙。其中, k 为大于或等于 1 的整数。

可以理解的,如上文所述, k 可以是终端设备上报给网络设备的,也可以是网络设备配置的。

25 在该方式#2 中,终端设备可以将接收第一信息的时隙开始 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说,终端设备可以将接收第一信息的时隙开始间隔 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是,本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙,本申请实施例对此不做限制。

示例性的,参考图 8 所示,假设终端设备在时隙 n 接收到第一信息,终端设备可以确定第一 AI 模型开始应用的时刻为时隙 $n+k$ 之后的第一个时隙,即时隙 $n+k+1$ 。

30 方式#3,第一 AI 模型开始应用的时刻可以为:

承载第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 k 个时隙,或 $(k+3N)$ 个时隙,或 $\max(k, 3N)$ 个时隙之后的第一个时隙,其中, k 为所述模型应用时间, N 为 1ms 包含的时隙数量。

35 需要说明的是,方式#3 中模型应用时间所表征的时间间隔可以以时隙为单位。模型应用时间可以是 k 个时隙,或者说,模型应用时间表征的时间间隔为 k 个时隙。其中, k 为大于或等于 1 的整数。

可以理解的,如上文所述, k 可以是终端设备上报给网络设备的,也可以是网络设备配置的。

40 在一种实施方式中,终端设备可以将携带第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说,终端设备可以将携带第一信息的 HARQ-ACK 信息的上行信道所在时隙开始,间隔 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是,该实施方式中,第一信息为高层信息,典型地,第一信息通过 RRC 信令发送。

还需要说明的是,本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙,本申请实施例对此不做限制。

45 示例性的,参考图 9A 所示,假设终端设备在时隙 n 发送携带第一信息的 PDSCH 的 HARQ-ACK 信息,则第一 AI 模型开始应用的时刻为时隙 $n+k$ 之后的第一个时隙,即时隙 $n+k+1$ 。

在一种实施方式中,终端设备可以将携带第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 $(k+3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说,终端设备可以将携带第一信息的 HARQ-ACK 信息的上行信道所在时隙开始,间隔 $(k+3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

50 需要说明的是,该实施方式中,第一信息为高层信息,典型地,第一信息通过 MAC 层信令发送,例如 3GPP TS38.321 中的定义的信令。

可以理解的, $3N$ 个时隙 (即 3ms) 为 MAC 层信令生效至少需要的时间, 通过间隔 $k+3N$ 个时隙从而保证 MAC 层信令有足够的生效时间。

还需要说明的是, 本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙, 本申请实施例对此不做限制。

示例性的, 参考图 9B 所示, 假设终端设备在时隙 n 发送携带第一信息的 PDSCH 的 HARQ-ACK 信息, 则第一 AI 模型开始应用的时刻为时隙 $n+k+3N$ 之后的第一个时隙, 即时隙 $n+k+3N+1$ 。

在一种实施方式中, 终端设备可以将携带第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 $\max(k, 3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说, 终端设备可以将携带第一信息的 HARQ-ACK 信息的上行信道所在时隙开始, 间隔 $\max(k, 3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是, 该实施方式中, 第一信息为高层信息, 典型地, 第一信息通过 MAC 层信令, 例如 3GPP TS38.321 中的定义的信令。

可以理解的, $3N$ 个时隙 (即 3ms) 为 MAC 层信令生效至少需要的时间, 通过取 k 和 $3N$ 中的较大值从而保证 MAC 层信令有足够的生效时间。

还需要说明的是, 本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙, 本申请实施例对此不做限制。

示例性的, 参考图 9C 所示, 假设终端设备在时隙 n 发送携带第一信息的 PDSCH 的 HARQ-ACK 信息且 $k < 3N$, 则第一 AI 模型开始应用的时刻为时隙 $n+3N$ 之后的第一个时隙, 即时隙 $n+3N+1$ 。

需要说明的是, 终端设备和网络设备可以约定方式#1~方式#3 中的一种方式, 或者网络设备为终端设备指定方式#1~方式#3 中的一种方式, 来确定第一 AI 模型开始应用的时刻, 从而确保两侧相对应的 AI 模型开始应用的时刻是相同的。

另外, 针对终端设备通过第二信息向网络设备指示第一 AI 模型的场景, 终端设备和网络设备确定的第一 AI 模型开始应用的时刻可以与第二信息相关。下面通过方式#1'~方式#3' 说明, 终端设备通过第二信息向网络设备指示第一 AI 模型的场景中, 第一 AI 模型开始应用的时刻。

方式#1', 第一 AI 模型开始应用的时刻为:

承载第二信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号; 或者,

承载第二信息的上行信道的响应信息的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号; 其中, k 为所述模型应用时间。

需要说明的是, 方式#1' 中模型应用时间所表征的时间间隔可以以 OFDM 符号为单位。模型应用时间可以是 k 个 OFDM 符号, 或者说, 模型应用时间表征的时间间隔为 k 个 OFDM 符号。其中, k 为大于或等于 1 的整数。

在一种可能的实现方式中, 终端设备可以将承载第二信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一 OFDM 符号作为第一 AI 模型开始应用的时刻。或者说, 终端设备可以将承载第二信息的上行信道的最后一个 OFDM 符号之后间隔 k 个 OFDM 符号的第一个时隙或者第一个 OFDM 符号作为第一 AI 模型开始应用的时刻。

需要说明的是, 这里的上行信道可以是 PUSCH 或者 PUCCH。示例性的, 若第二信息为高层信息, 则上行信道可以是 PUSCH; 若第二信息为物理层信息, 则上行信道可以是 PUCCH。

还需要说明的是, 本申请实施例中的 k 个 OFDM 符号也可以是 $k-1$ 或 $k+1$ 个 OFDM 符号, 本申请实施例对此不做限制。

在另一种可能的实现方式中, 终端设备可以将承载第二信息的响应信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一 OFDM 符号作为第一 AI 模型开始应用的时刻。或者说, 终端设备可以将承载第二信息的响应信息的下行信道的最后一个 OFDM 符号之后间隔 k 个 OFDM 符号的第一个时隙或者第一个 OFDM 符号作为第一 AI 模型开始应用的时刻。

可以理解的, 网络设备收到第二信息后, 可以发送第二信息对应的响应信息, 以使终端设备知道网络设备正确接收到了所述第二信息。例如, 所述第二信息通过 PRACH 承载, 则所述第二信息的响应信息可以是对应的 RAR; 所述第二信息通过 PUSCH 承载, 则所述第二信息可以是所述 PUSCH 的 HARQ-ACK 信息。

需要说明的是, 上述的 k 个 OFDM 符号也可以是 $k-1$ 或 $k+1$ 个 OFDM 符号, 本申请实施例对此不做限制。

方式#2', 第一 AI 模型开始应用的时刻为:

终端设备发送第二信息的时隙开始 k 个时隙之后的第一个时隙；第二信息为用于指示所述第一 AI 模型的上行信息， k 为所述模型应用时间。

需要说明的是，方式#2'中模型应用时间所表征的时间间隔可以以时隙为单位。模型应用时间可以是 k 个时隙，或者说，模型应用时间表征的时间间隔为 k 个时隙。其中， k 为大于或等于 1 的整数。

可以理解的，如上文所述， k 可以是终端设备上报给网络设备的，也可以是网络设备配置的。

在该方式#2'中，终端设备可以将发送第二信息的时隙开始 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说，终端设备可以将发送第二信息的时隙开始间隔 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是，本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙，本申请实施例对此不做限制。

方式#3'，第一 AI 模型开始应用的时刻为：

承载第二信息的响应信息的下行信道所在时隙开始 k 个时隙，或 $(k+3N)$ 个时隙，或 $\max(k, 3N)$ 个时隙之后的第一个时隙，其中，所述第二信息为用于指示所述第一 AI 模型的上行信息， k 为所述模型应用时间， N 为 1ms 包含的时隙数量。

需要说明的是，方式#3'中模型应用时间所表征的时间间隔可以以时隙为单位。模型应用时间可以是 k 个时隙，或者说，模型应用时间表征的时间间隔为 k 个时隙。其中， k 为大于或等于 1 的整数。

可以理解的，如上文所述， k 可以是终端设备上报给网络设备的，也可以是网络设备配置的。

在一种实施方式中，终端设备可以将携带第二信息的响应信息的下行信道所在时隙开始 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说，终端设备可以将携带第二信息的响应信息的下行信道所在时隙开始，间隔 k 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是，该实施方式中，第二信息为高层信息，典型地，第二信息通过 RRC 信令发送。

还需要说明的是，本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙，本申请实施例对此不做限制。

在一种实施方式中，终端设备可以将携带第二信息的响应信息的下行信道所在时隙开始 $(k+3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说，终端设备可以将携带第二信息的响应信息的下行信道所在时隙开始，间隔 $(k+3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是，该实施方式中，第二信息为高层信息，典型地，第二信息通过 MAC 层信令发送，例如 3GPP TS38.321 中的定义的信令。

可以理解的， $3N$ 个时隙（即 3ms）为 MAC 层信令生效至少需要的时间，通过间隔 $k+3N$ 个时隙从而保证 MAC 层信令有足够的生效时间。

还需要说明的是，本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙，本申请实施例对此不做限制。

在一种实施方式中，终端设备可以将携带第二信息的响应信息的下行信道所在时隙开始 $\max(k, 3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。或者说，终端设备可以将携带第二信息的响应信息的下行信道所在时隙开始，间隔 $\max(k, 3N)$ 个时隙之后的第一个时隙作为第一 AI 模型开始应用的时刻。

需要说明的是，该实施方式中，第二信息为高层信息，典型地，第二信息通过 MAC 层信令发送，例如 3GPP TS38.321 中的定义的信令。

可以理解的， $3N$ 个时隙（即 3ms）为 MAC 层信令生效至少需要的时间，通过取 k 和 $3N$ 中的较大值从而保证 MAC 层信令有足够的生效时间。

还需要说明的是，本申请实施例中的 k 个时隙也可以是 $k-1$ 或 $k+1$ 个时隙，本申请实施例对此不做限制。

需要说明的是，终端设备和网络设备可以约定方式#1'~方式#3'中的一种方式，或者网络设备为终端设备指定方式#1'~方式#3'中的一种方式，来确定第一 AI 模型开始应用的时刻，从而确保两侧相对应的 AI 模型开始应用的时刻是相同的。

在一些实施例中，所述第一 AI 模型开始应用的时刻和/或所述模型应用时间，可以根据以下任意一项确定：

承载第一信息的下行信道的子载波间隔；

承载承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔;
承载第二信息的上行信道的子载波间隔;
承载第二信息的响应信息的下行信道的子载波间隔;
应用第一 AI 模型的载波上所激活的 BWP 的子载波间隔。

- 5 需要说明的是，第一 AI 模型开始应用的时刻和模型应用时间是以时隙或者 OFDM 符号为单位的参数，需要根据子载波间隔确定实际对应的时间长度/时刻（或者称为绝对时间长度/时隙），不同子载波间隔对应的时间长度不同。
- 示例性的，参考表 1 所示所示的子载波间隔与时隙个数和时隙长度之间的关系。

表 1

子载波间隔（kHz）	每个子帧的时隙个数	每个时隙时长（毫秒 ms）
15	1	1
30	2	0.5
60	4	0.25
120	8	0.125
240	16	0.0625

- 10 参考表 1 所示，子载波间隔为 30kHz，对应每个时隙的长度为 0.5ms。若模型应用时间为 k 个时隙，则 k 个时隙的长度为 0.5k ms。以此类推，子载波间隔为 60kHz，对应每个时隙的长度为 0.25ms。若模型应用时间为 k 个时隙，则 k 个时隙的长度为 0.25k ms。
- 可以理解的，第一 AI 模型开始应用的时刻和/或模型应用时间，可以根据与第一 AI 模型相关的子载波间隔确定。

- 15 在一些实施例中，针对网络设备通过第一信息为终端设备配置第一 AI 模型的场景，第一 AI 模型开始应用的时刻和/或模型应用时间，可以根据以下中的任意一项确定：
- 承载第一信息的下行信道的子载波间隔；
承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔；
应用所述第一 AI 模型的载波上所激活的 BWP 的子载波间隔。

- 20 在一种可能的实现方式中，第一 AI 模型开始应用的时刻和/或模型应用时间，可以根据承载第一信息的下行信道的子载波间隔确定。
- 示例性的，若第一信息通过 PBCH 承载，则第一 AI 模型开始应用的时刻和/或模型应用时间根据 PBCH 的子载波间隔确定。若第一信息通过 Group Common DCI 承载，则第一 AI 模型开始应用的时刻和/或模型应用时间根据承载所述 DCI 的 PDCCH 的子载波间隔确定。

- 25 示例性的，结合图 8 所示的时隙结构示意图，模型开始应用的时刻为时隙 n+k+1，终端设备可以根据承载第一信息的子载波间隔，得到时隙 n+k+1 与时隙 n 之间间隔的时间长度，从而确定模型应用的时间。
- 在另一种可能的实现方式中，第一 AI 模型开始应用的时刻和/或模型应用时间，可以根据承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔确定。

- 30 示例性的，若第一信息通过高层信令承载，则第一 AI 模型开始应用的时刻和/或模型应用时间可以根据携带该高层信令的 PDSCH 的 HARQ-ACK 信息所在的上行信道的子载波间隔确定。
- 示例性的，结合图 9B 所示的时隙结构示意图，如果模型开始应用的时刻为时隙 n+k+3N+1，终端设备可以根据所确定的子载波间隔，得到时隙 n+k+3N+1 与时隙 n 之间间隔的时间长度，从而确定模型应用的时间。

- 35 在又一种可能的实现方式中，第一 AI 模型开始应用的时刻和/或模型应用时间，可以根据应用第一 AI 模型的载波上所激活的 BWP 的子载波间隔。
- 在一些实施例中，当应用第一 AI 模型的载波的数量为多个时，第一 AI 模型开始应用的时刻和/或模型应用时间可以根据第一子载波间隔确定。其中，第一子载波间隔为应用第一 AI 模型的多个载波上分别激活的 BWP 中任一个子载波间隔。第一子载波间隔可以是网络设备与终端设备约定好的。
- 40 示例性的，第一子载波间隔可以为应用第一 AI 模型的多个载波上分别激活的 BWP 中最小的子载波间隔。例如，应用第一 AI 模型的载波分别为 CC1，CC2 和 CC3，其中三个 CC 上激活的 BWP 分别为 BWP1，BWP2 和 BWP3，且 BWP1 配置的子载波间隔<BWP2 配置的子载波间隔<BWP3 配置的子载波间隔，此时，第一 AI 模型开始应用的时刻和/或模型应用时间可以根据 BWP1 配置的子载波间隔确定。

- 45 在另一些实施例中，针对终端设备通过第二信息向网络设备指示第一 AI 模型的场景，第一 AI

模型开始应用的时刻和/或模型应用时间,可以根据以下中的任意一项确定:

承载第二信息上行信道的子载波间隔;

承载第二信息的响应信息的下行信道的子载波间隔;

应用所述第一 AI 模型的载波上所激活的 BWP 的子载波间隔。

- 5 在一种可能的实现方式中,第一 AI 模型开始应用的时刻和/或模型应用时间,可以根据承载第二信息上行信道的子载波间隔确定。

示例性的,若第二信息通过 UCI 携带,则第一 AI 模型开始应用的时刻和/或模型应用时间根据该 UCI 的 PUCCH 的子载波间隔确定。若第二信息通过高层信令携带,则第一 AI 模型开始应用的时刻和/或模型应用时间根据承载该高层信令的 PUSCH 的子载波间隔确定。若第二信息通过 PRACH 携带,则第一 AI 模型开始应用的时刻和/或模型应用时间根据该 PRACH 的子载波间隔确定。

10 在另一种可能的实现方式中,第一 AI 模型开始应用的时刻和/或模型应用时间,可以根据承载第二信息的响应信息的下行信道的子载波间隔确定。

示例性的,若第二信息通过高层信令承载,则第一 AI 模型开始应用的时刻和/或模型应用时间可以根据携带该高层信令的 PUSCH 的 HARQ-ACK 信息所在的下行信道的子载波间隔确定。若第二信息通过 PRACH 承载,则第一 AI 模型开始应用的时刻和/或模型应用时间可以根据对应的 RAR 所在的下行信道的子载波间隔确定。

15 在又一种可能的实现方式中,第一 AI 模型开始应用的时刻和/或模型应用时间,可以根据应用第一 AI 模型的载波上所激活的 BWP 的子载波间隔。

20 在一些实施例,当应用第一 AI 模型的载波的数量为多个时,第一 AI 模型开始应用的时刻和/或模型应用时间可以根据第一子载波间隔确定。其中,第一子载波间隔为应用第一 AI 模型的多个载波上分别激活的 BWP 中任一个子载波间隔。第一子载波间隔可以是网络设备与终端设备约定好的。

示例性的,第一子载波间隔可以为应用第一 AI 模型的多个载波上分别激活的 BWP 中最小的子载波间隔。例如,应用第一 AI 模型的载波分别为 CC1, CC2 和 CC3,其中三个 CC 上激活的 BWP 分别为 BWP1, BWP2 和 BWP3,且 BWP1 配置的子载波间隔<BWP2 配置的子载波间隔<BWP3 配置的子载波间隔,此时,第一 AI 模型开始应用的时刻和/或模型应用时间可以根据 BWP1 配置的子载波间隔确定。

由此可见,通过终端设备与网络设备约定相同的子载波间隔用于确定第一 AI 模型开始应用的时刻和/或模型应用时间,可以确保终端设备和网络确定的模型应用时间,以及第一 AI 模型开始应用的时刻是相同的,从而达到同步更新模型的效果。反之,当承载第一信息/第二信息的信道的子载波间隔与应用第一 AI 模型的载波所用的子载波间隔不同时,若未约定使用哪个子载波间隔确定第一 AI 模型开始应用的时刻和/或模型应用时间,则终端设备和网络设备无法确定实际的时间长度,可能导致两侧确定的时间长度不同。

以下结合具体应用场景对本申请实施例提供的方法进行详细阐述。

实施例一:

- 35 在该实施例一中,网络设备可以通过广播信令为其覆盖范围内的终端设备配置 AI 模型(即第一信息通过广播信令承载)。该实施例可以包括以下步骤:

S1、终端设备通过 UE 能力信息上报所支持的模型应用时间。

需要说明的是,模型应用时间为终端设备接收到用于配置 AI 模型的广播信令到应用该 AI 模型之间需要的时间间隔。其中,时间间隔以时隙为单位。

40 在一些实施例,终端设备上报的模型应用时间可以包括:

终端设备支持的一个或多个 AI 模型分别对应的模型应用时间;

终端设备支持的一种或多种模型更新方式分别对应的模型应用时间;

终端设备支持的一个或多个模型类型分别对应的模型应用时间。

S2、网络设备根据终端设备上报的 UE 能力,为终端设备指示模型应用时间。

- 45 需要说明的是,网络设备需要考虑覆盖范围内多个终端设备分别上报的模型应用时间,来确定所指示的模型应用时间。示例性的,网络设备可以采用多个终端设备分别上报的模型应用时间中的最大值,作为覆盖范围内各个终端采用的模型应用时间,并指示给这些终端设备。

50 在一些实施例,模型应用时间可以通过广播信令来指示。示例性的,广播信令可以是 PBCH, SIB 或者 Group Common DCI。其中,指示模型应用时间的广播信令,与指示 AI 模型的广播信令可以以为同一信令,也可以为不同的信令,本申请实施例对此不做限制。

在一些实施例,模型应用时间也可以通过高层信令或者物理层信令来指示。示例性的,高层

信令可以是用于模型失败恢复响应 MAC 层信令指示, 或者用于模型更新的 MAC 层信令。

S3、终端设备接收网络设备发送的用于配置 AI 模型的广播信令。

在一些实施例中, 终端设备可以接收网络设备通过广播信令配置的 AI 模型的模型 ID, 功能 (functionality), 模型结构和模型参数中的至少一项。

可以理解的, 网络设备不一定要配置整个 AI 模型, 可以只配置或更新 AI 模型中的部分结构或者部分模型参数。

在一些实施例中, 如果网络设备通过广播信令配置了一个新的模型的模型 ID, 则所述 AI 模型开始应用的时刻为所更新的模型开始应用的时刻。在所述时刻之前, 终端设备可以使用之前采用的模型进行与网络设备之间的信息传输。

在一些实施例中, 如果网络设备通过广播信令配置了一个新的模型的功能 (functionality), 则所述 AI 模型开始应用的时刻为所更新的模型开始应用的时刻。在所述时刻之前, 终端设备可以使用之前的功能对应的模型进行与网络设备之间的信息传输。这里的功能也可以是功能 ID。

在一些实施例中, 如果网络设备通过广播信令更新了模型结构和模型参数, 则所述 AI 模型开始应用的时刻为所更新的模型结构和模型参数开始应用的时刻。在所述时刻之前, 终端设备可以使用之前采用的模型结构和模型参数进行与网络设备之间的信息传输。

在一些实施例中, 如果网络设备通过广播信令只更新了模型参数, 则所述 AI 模型开始应用的时刻为所更新的模型参数开始应用的时刻。在所述时刻之前, 终端设备可以使用之前采用的模型参数进行与网络设备之间的信息传输。

需要说明的是, 接收到用于配置 AI 模型的广播信令的终端设备都要应用该广播信令所配置的 AI 模型。例如, 网络设备可以通过 Group Common DCI 来给一组终端设备指示所用的 AI 模型, 所述 Group Common DCI 通过公共搜索空间 CSS 来传输, 采用公共的 RNTI 进行加扰。

在一些实施例中, 网络设备通过广播信令配置 AI 模型的同时, 也可以通过相同的信令指示相应的模型应用时间。

可以理解的, 网络设备可以给一组终端设备部署相同的 AI 模型, 并指示相同的模型应用时间, 这组终端设备就可以同时应用所述 AI 模型。这样在网络侧只需要部署一个 AI 模型作为对应的 encoder/decoder 模型, 不需要为每个终端设备分别部署一个对应的模型, 从而明显降低了网络设备的实现复杂度, 同时也减少了下行信令的开销。

S4、终端设备根据网络设备配置的模型应用时间, 确定 AI 模型开始应用的时刻。

其中, AI 模型开始应用的时刻为: 终端设备接收到用于配置 AI 模型的广播信令的时隙开始 k 个时隙之后的第一个时隙; k 为网络设备指示的模型应用时间 (以时隙为单位)。

示例性的, 参考图 8 所示, 假设终端设备在时隙 n 接收到用于配置 AI 模型的广播信令, 则该 AI 模型开始应用的时刻为时隙 $n+k$ 之后的第一个时隙, 即时隙 $n+k+1$ 开始应用。

需要说明的是, 本实施例中 AI 模型开始应用的时刻和/或模型应用时间是以时隙为单位的参数, 需要根据子载波间隔确定对应的时间长度/时刻, 其中, 不同子载波间隔对应的时间长度不同。

本申请实施例中, 网络设备配置的 AI 模型开始应用的时刻和/或模型应用时间可以根据承载该广播信令的信道的子载波间隔确定。

可以理解的, 通过终端设备与网络设备约定相同的子载波间隔用于确定 AI 模型开始应用的时刻和模型应用时间, 可以确保终端设备和网络设备确定的模型应用时间和模型开始应用的时刻是相同的, 从而达到同步更新模型的效果。反之, 当承载用于配置 AI 模型的广播信令的信道的子载波间隔与应用该 AI 模型的载波所用的子载波间隔不同时, 如果未确定基于哪个子载波间隔来确定 AI 模型开始应用的时刻和/或模型应用时间, 那么终端设备和网络设备无法确定 k 个时隙的长度。

S5、在 AI 模型开始应用的时刻之后, 终端设备采用广播信令配置的 AI 模型进行与网络设备之间的信息传输。

在一些实施例中, 在终端设备应用所述 AI 模型之前, 终端设备可以采用预定义的 AI 模型进行与网络设备之间的信息传输, 或者, 终端设备不基于 AI 模型进行与网络设备之间的信息传输。其中, 预定义的 AI 模型即终端设备和网络设备预先约定好的一个 AI 模型。终端设备不基于 AI 模型进行与网络设备之间的信息传输, 即终端设备采用传统的非 AI 的方式进行与网络设备之间的信息传输。

在一些实施例中, 在广播信令配置的 AI 模型开始应用的时刻之后, 终端设备可以将该 AI 模型用于如下任一过程:

下行 CSI 反馈; 或者,

下行数据的信道解码; 或者,

下行信号的解调；或者，
下行信道估计；或者，
下行射频信号处理；或者，
上行数据的信道编码；或者，
5 上行数据的调制；或者，
上行导频生成；或者，
上行射频信号处理。

S6、在 AI 模型开始应用的时刻之后，网络设备采用上述广播信令配置的 AI 模型对应的网络侧模型进行与终端设备之间的信息传输。

10 在一些实施例中，在广播信令配置的 AI 模型开始应用的时刻之后，网络设备可以将配置的 AI 模型对应的网络侧模型用于如下任一过程：

下行 CSI 反馈信息的解码；或者，
上行数据的信道解码；或者，
上行信号的解调；或者，
15 上行信道估计；或者，
上行射频信号处理；或者，
下行数据的信道编码；或者，
下行数据的调制；或者，
下行导频生成；或者，
20 下行射频信号处理。

需要说明的是，网络设备应用模型的模块与终端侧应用所述 AI 模型的模块是一一对应的，例如信道编码-信道解码，调制-解调，导频生成-信道估计，发端射频信号处理-收端射频信号处理等。

实施例二

25 在该实施例二中，网络设备可以通过高层信令为终端设备配置 AI 模型（即第一信息通过高层信令承载）。该实施例可以包括以下步骤：

S1、终端设备通过 UE 能力上报所支持的模型应用时间。

需要说明的是，模型应用时间为终端设备反馈用于配置 AI 模型的高层信令的 HARQ-ACK 信息之后到应用 AI 模型之间需要的时间间隔，时间间隔以时隙或 OFDM 符号为单位。

在一些实施例中，上述高层信令可以是 RRC 信令，也可以是 MAC 层信令（MAC CE）。

30 在一些实施例中，上述高层信令的 HARQ-ACK 信息即为承载第一信息的 PDSCH 的 HARQ-ACK 信息。

S2、终端设备通过高层信令接收网络设备配置的 AI 模型。

S3、终端设备根据 S1 中上报的模型应用时间，确定网络设备配置的 AI 模型开始应用的时刻。

35 在一种实施方式中，高层信令配置的 AI 模型开始应用的时刻可以为：携带该高层信令的 PDSCH 的 HARQ-ACK 信息所在时隙开始 k 个时隙之后的第一个时隙。

其中， k 为模型应用时间，可以由终端设备通过 UE 能力上报所支持的模型应用时间确定。

需要说明的是，该实施方式中典型的高层信令为 RRC 信令。

在一种实施方式中，高层信令配置的 AI 模型开始应用的时刻可以为：携带该高层信令的 PDSCH 的 HARQ-ACK 信息所在时隙开始 $(k+3N)$ 个时隙之后的第一个时隙。

40 其中， k 为所述模型应用时间， N 为 1ms 包含的时隙数量，即 $3N$ 个时隙为 3ms。

需要说明的是，典型的高层信令可以为 MAC 层信令。 $3N$ 个时隙（即 3ms）为 MAC 层信令生效至少需要的时间，通过间隔 $k+3N$ 个时隙从而保证 MAC 层信令有足够的生效时间。

在一种实施方式中，高层信令配置的 AI 模型开始应用的时刻可以为：携带该高层信令的 PDSCH 的 HARQ-ACK 信息所在时隙开始 p 个时隙之后的第一个时隙。

45 其中， p 为 k 和 $3N$ 中的较大值， k 为模型应用时间（时隙为单位）， N 为 1ms 包含的时隙数量，即 $3N$ 个时隙为 3ms。

需要说明的是，典型的高层信令可以为 MAC 层信令， $3N$ 个时隙（即 3ms）为 MAC 层信令生效至少需要的时间，通过取 k 和 $3N$ 中的较大值从而保证 MAC 层信令有足够的生效时间。

50 还需要说明的是，本实施例中 AI 模型开始应用的时刻和/或模型应用时间是以时隙为单位的参数，需要根据子载波间隔确定对应的时间长度/时刻，其中，不同子载波间隔对应的时间长度不同。

在一种实施方式中，网络设备配置的 AI 模型开始应用的时刻和/或模型应用时间可以根据携带

该高层信令的 PDSCH 的 HARQ-ACK 信息所在的上行信道的子载波间隔确定。

在另一种实施方式中,网络设备配置的 AI 模型开始应用的时刻和/或模型应用时间可以根据该 AI 模型的载波上所激活的 BWP 的子载波间隔确定。

如果该 AI 模型应用于多个载波,多个载波上都有激活的 BWP,则 AI 模型开始应用的时刻和/或模型应用时间根据所述多个载波上激活的 BWP 中最小的子载波间隔来确定。这样即使在多载波的场景中,网络设备和终端设备也能确定唯一的一个开始应用时刻。

可以理解的,通过终端设备与网络设备约定相同的子载波间隔用于确定 AI 模型开始应用的时刻和模型应用时间,可以确保终端设备和网络设备确定的模型应用时间和模型开始应用的时刻是相同的,从而达到同步更新模型的效果。反之,当承载用于配置 AI 模型的广播信令的信道的子载波间隔与应用该 AI 模型的载波所用的子载波间隔不同时,如果未确定基于哪个子载波间隔来确定 AI 模型开始应用的时刻和/或模型应用时间,那么终端设备和网络设备无法确定 k 个时隙的长度。

S4、在 AI 模型开始应用的时刻之后,终端设备采用上述高层信令配置的 AI 模型进行与网络设备之间的信息传输。

S5、在 AI 模型开始应用的时刻之后,网络设备采用上述高层信令配置的 AI 模型对应的网络侧模型进行与终端设备之间的信息传输。

实施例三

在该实施例一中,网络设备可以通过 DCI 为终端设备配置 AI 模型(即第一信息通过 DCI 承载)。该实施例可以包括以下步骤:

S1、终端设备通过 UE 能力上报所支持的模型应用时间。

需要说明的是,模型应用时间为终端设备反馈用于配置 AI 模型的 DCI 信令的 HARQ-ACK 信息之后到应用 AI 模型之间需要的时间间隔,时间间隔以时隙或 OFDM 符号为单位。

S2、终端设备通过 DCI 信令接收网络设备配置的 AI 模型。

需要说明的是,网络设备可以通过 UE 专属搜索空间传输的 DCI 信令,为终端设备配置 AI 模型。示例性的,该 DCI 信令可以为网络设备发送的用于模型失败恢复响应的 DCI,或者用于调度 PDSCH 或 PUSCH 的 DCI。其中,PDSCH 或 PUSCH 的传输可以基于 AI 模型(不一定是所配置的这个 AI 模型)。

S3、终端设备根据 S1 中上报的模型应用时间,确定网络设备配置的 AI 模型开始应用的时刻。

在一种实施方式中,DCI 信令配置的 AI 模型开始应用的时刻可以为:承载该 DCI 的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始至少 k 个 OFDM 符号之后的第一个时隙。其中, k 为模型应用时间(OFDM 符号为单位)。

示例性的,参考图 6A 所示,假设承载所述 DCI 信令的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号为符号 m ,则 AI 模型开始应用的时刻为符号 $m+k$ 所在时隙之后的第一个时隙。

在另一种实施方式中,DCI 信令配置的 AI 模型开始应用的时刻可以为:承载所述 DCI 的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始至少 k 个 OFDM 符号之后的第一个 OFDM 符号。其中, k 为模型应用时间(OFDM 符号为单位)。

示例性的,参考图 6N 所示,假设承载所述 DCI 的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号为符号 m ,则所述 AI 模型开始应用的时刻为符号 $m+k$ 之后的第一个符号,即符号 $m+k+1$ 。

还需要说明的是,本实施例中 AI 模型开始应用的时刻和/或模型应用时间是以时隙/OFDM 为单位的参数,需要根据子载波间隔确定对应的时间长度/时刻,其中,不同子载波间隔对应的时间长度不同。

其中,网络设备配置的 AI 模型开始应用的时刻和/或模型应用时间可以根据该 AI 模型的载波上所激活的 BWP 的子载波间隔确定。

如果该 AI 模型应用于多个载波,多个载波上都有激活的 BWP,则 AI 模型开始应用的时刻和/或模型应用时间根据所述多个载波上激活的 BWP 中最小的子载波间隔来确定。这样即使在多载波的场景中,网络设备和终端设备也能确定唯一的一个开始应用时刻。

S4、在 AI 模型开始应用的时刻之后,终端设备采用上述 DCI 信令配置的 AI 模型进行与网络设备之间的信息传输。

S5、在 AI 模型开始应用的时刻之后,网络设备采用上述 DCI 信令配置的 AI 模型对应的网络侧模型进行与终端设备之间的信息传输。

综上所述,在终端设备侧和网络设备侧采用相对应的双端 AI 模型时,如果终端设备侧模型发生更新或重配置,本申请实施例提供的无线通信方法可以确保两侧模型应用时刻是相同的,网络设

备与不同终端之间能够同步进行模型更新，从而保证对端的模型输出是可用的。

以上结合附图详细描述了本申请的优选实施方式，但是，本申请并不限于上述实施方式中的具体细节，在本申请的技术构思范围内，可以对本申请的技术方案进行多种简单变型，这些简单变型均属于本申请的保护范围。例如，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合，为了避免不必要的重复，本申请对各种可能的组合方式不再另行说明。又例如，本申请的各种不同的实施方式之间也可以进行任意组合，只要其不违背本申请的思想，其同样应当视为本申请所公开的内容。又例如，在不冲突的前提下，本申请描述的各个实施例和/或各个实施例中的技术特征可以和现有技术任意的相互组合，组合之后得到的技术方案也应落入本申请的保护范围。

还应理解，在本申请的各种方法实施例中，上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请实施例的实施过程构成任何限定。此外，在本申请实施例中，术语“下行”、“上行”和“侧行”用于表示信号或数据的传输方向，其中，“下行”用于表示信号或数据的传输方向为从站点发送至小区的用户设备的第一方向，“上行”用于表示信号或数据的传输方向为从小区的用户设备发送至站点的第二方向，“侧行”用于表示信号或数据的传输方向为从用户设备 1 发送至用户设备 2 的第三方向。例如，“下行信号”表示该信号的传输方向为第一方向。另外，本申请实施例中，术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系。具体地，A 和/或 B 可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

图 10 是本申请实施例提供的无线通信装置 1000 的结构组成示意图，应用于终端设备，如图 10 所示，所述无线通信装置 1000 包括：

第一确定单元 1001，配置为根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

第一通信单元 1002，配置为在所述时刻之后，采用所述第一 AI 模型与所述网络设备进行通信。

在一些实施例中，所述模型应用时间为以下中的任意一项：

终端设备接收到第一信息之后到应用所述第一 AI 模型的时刻之间的时间间隔；

终端设备发送第一信息的 HARQ-ACK 信息之后到应用第一 AI 模型的时刻之间的时间间隔；

终端设备接收到第一信息之后到终端设备发送基于第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

终端设备发送第一信息的 HARQ-ACK 信息之后到终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

其中，所述第一信息为用于配置所述第一 AI 模型的下行信息，所述时间间隔以时隙或 OFDM 符号为单位。

在一些实施例中，所述第一 AI 模型开始应用的时刻为以下中的任意一项：

承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号；

承载第一信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号；

所述终端设备接收到第一信息的时隙开始 k 个时隙之后的第一个时隙；

承载第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 k 个时隙，或 $(k+3N)$ 个时隙，或 $\max(k, 3N)$ 个时隙之后的第一个时隙；

其中，所述第一信息为用于配置所述第一 AI 模型的下行信息，k 为所述模型应用时间，N 为 1ms 包含的时隙数量。

在一些实施例中，所述无线通信装置 1000 还包括接收单元，配置为接收第一信息，所述第一信息是用于配置所述第一 AI 模型的下行信息，其中，所述第一信息包括以下中的一项或多项：

所述第一 AI 模型的标识信息；

所述第一 AI 模型的功能；

所述第一 AI 模型的模型结构；以及，

所述第一 AI 模型的模型参数。

在一些实施例中，所述模型应用时间为以下中的任意一项：

所述终端设备发送第二信息之后到应用所述第一 AI 模型的时刻之间的时间间隔；

所述终端设备发送第二信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的

时刻之间的时间间隔;

所述终端设备接收第二信息的响应信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

所述终端设备接收第二信息的响应信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

5 其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, 所述时间间隔以时隙或 OFDM 符号为单位。

在一些实施例中, 所述第一 AI 模型开始应用的时刻为以下中的任意一项:

承载第二信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

10 承载第二信息的上行信道的响应信息的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

所述终端设备发送第二信息的时隙开始 k 个时隙之后的第一个时隙;

承载第二信息的响应信息的下行信道所在时隙开始 k 个时隙, 或 $(k+3N)$ 个时隙, 或 $\max(k, 3N)$ 个时隙之后的第一个时隙;

15 其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, k 为所述模型应用时间, N 为 1ms 包含的时隙数量。

在一些实施例中, 所述无线通信装置 1000 还包括发送单元, 配置为发送第二信息, 所述第二信息是用于指示所述第一 AI 模型的上行信息;

其中, 所述第二信息包括以下中的一项或多项:

20 所述第一 AI 模型的标识信息;

所述第一 AI 模型的模型功能;

所述第一 AI 模型的模型结构; 以及,

所述第一 AI 模型的模型参数。

在一些实施例中, 所述上报给网络设备的模型应用时间包括以下中的一项或多项:

25 所述终端设备支持的一个或多个 AI 模型分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个 AI 模型包括所述第一 AI 模型;

所述终端设备支持的一个或多个模型功能分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个模型功能包括所述第一 AI 模型的功能;

30 所述终端设备支持的一种或多种模型更新方式分别对应的模型应用时间; 其中, 所述终端设备支持的一种或多种模型更新方式包括所述第一 AI 模型的更新方式;

所述终端设备支持的一个或多个模型类型分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个模型类型包括所述第一 AI 模型的类型。

在一些实施例中, 所述模型更新方式包括: 更新模型参数和/或更新模型结构。

在一些实施例中, 所述模型类型包括:

35 高精度模型和低精度模型; 或者, 已知模型和未知模型。

在一些实施例中, 所述第一 AI 模型开始应用的时刻和/或所述模型应用时间, 根据以下任意一项确定:

承载第一信息的下行信道的子载波间隔;

承载第二信息的上行信道的子载波间隔;

40 承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔;

承载第二信息的响应信息的下行信道的子载波间隔;

应用所述第一 AI 模型的载波上所激活的 BWP 的子载波间隔;

其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, 所述第二信息为用于指示所述第一 AI 模型的上行信息。

45 在一些实施例中, 应用所述第一 AI 模型的载波的数量为多个,

所述第一 AI 模型开始应用的时刻和/或所述模型应用时间根据第一子载波间隔确定, 所述第一子载波间隔为应用所述第一 AI 模型的多个载波上分别激活的 BWP 中最小的子载波间隔。

在一些实施例中, 在所述终端设备应用所述第一 AI 模型之前, 所述终端设备基于第三 AI 模型进行与所述网络设备之间的信息传输, 或者, 所述终端设备不基于 AI 模型进行与所述网络设备之间的信息传输;

50 其中, 所述第三 AI 模型为预定义的 AI 模型, 或者历史 AI 模型。

图 11 是本申请实施例提供的无线通信装置 1100 的结构组成示意图, 应用于网络设备, 如图 11 所示, 所述无线通信装置 1100 包括:

第二确定单元 1101, 配置为根据终端设备上报的模型应用时间, 或者, 根据为所述终端设备配置的模型应用时间, 确定第一 AI 模型开始应用的时刻;

5 第二通信单元 1102, 配置为在所述时刻之后, 采用所述第一 AI 模型对应的第二 AI 模型与所述终端设备进行通信。

在一些实施例中, 所述模型应用时间为以下中的任意一项:

所述终端设备接收到第一信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

10 所述终端设备发送第一信息的 HARQ-ACK 信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

所述终端设备接收到第一信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

所述终端设备发送第一信息的 HARQ-ACK 信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

15 其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, 所述时间间隔以时隙或 OFDM 符号为单位。

在一些实施例中, 所述第一 AI 模型开始应用的时刻为以下中的任意一项:

承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

20 承载第一信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

所述终端设备接收到第一信息的时隙开始 k 个时隙之后的第一个时隙;

承载第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 k 个时隙, 或 $(k+3N)$ 个时隙, 或 $\max(k, 3N)$ 个时隙之后的第一个时隙;

25 其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, k 为所述模型应用时间, N 为 1ms 包含的时隙数量。

在一些实施例中, 所述无线通信装置 1100 还包括发送单元, 配置为向所述终端设备发送第一信息, 所述第一信息为用于配置所述第一 AI 模型的下行, 其中, 所述第一信息包括以下中的一项或多项:

30 所述第一 AI 模型的标识信息;

所述第一 AI 模型的功能;

所述第一 AI 模型的模型结构; 以及,

所述第一 AI 模型的模型参数。

在一些实施例中, 所述模型应用时间为以下中的任意一项:

35 所述终端设备发送第二信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

所述终端设备发送第二信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

所述终端设备接收第二信息的响应信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

40 所述终端设备接收第二信息的响应信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, 所述时间间隔以时隙或 OFDM 符号为单位。

在一些实施例中, 所述第一 AI 模型开始应用的时刻为以下中的任意一项:

45 承载第二信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

承载第二信息的上行信道的响应信息的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

所述终端设备发送第二信息的时隙开始 k 个时隙之后的第一个时隙;

50 承载第二信息的响应信息的下行信道所在时隙开始 k 个时隙, 或 $(k+3N)$ 个时隙, 或 $\max(k, 3N)$ 个时隙之后的第一个时隙;

其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, k 为所述模型应用时间, N 为 1ms

包含的时隙数量。

在一些实施例中,所述无线通信装置 1100 还包括接收单元,配置为接收所述终端设备发送的第二信息,所述第二信息是用于指示所述第一 AI 模型的上行信息;

其中,所述第二信息包括以下中的一项或多项:

所述第一 AI 模型的标识信息;
所述第一 AI 模型的模型功能;
所述第一 AI 模型的模型结构;以及,
所述第一 AI 模型的模型参数。

在一些实施例中,所述终端设备上报的模型应用时间包括以下中的一项或多项:

所述终端设备支持的一个或多个 AI 模型分别对应的模型应用时间;其中,所述终端设备支持的一个或多个 AI 模型包括所述第一 AI 模型;

所述终端设备支持的一个或多个模型功能分别对应的模型应用时间;其中,所述终端设备支持的一个或多个模型功能包括所述第一 AI 模型的功能;

所述终端设备支持的一种或多种模型更新方式分别对应的模型应用时间;其中,所述终端设备支持的一种或多种模型更新方式包括所述第一 AI 模型的更新方式;

所述终端设备支持的一个或多个模型类型分别对应的模型应用时间;其中,所述终端设备支持的一个或多个模型类型包括所述第一 AI 模型的类型。

在一些实施例中,所述模型更新方式包括:更新模型参数和/或更新模型结构。

在一些实施例中,所述模型类型包括:

高精度模型和低精度模型;或者,已知模型和未知模型。

在一些实施例中,所述网络设备为所述终端设备配置的模型应用时间,基于所述终端设备上报的模型应用时间确定。

在一些实施例中,所述网络设备为所述终端设备配置的模型应用时间的时长,大于或等于所述终端设备上报的模型应用时间的时长。

在一些实施例中,所述第一 AI 模型开始应用的时刻和/或所述模型应用时间,根据以下任意一项确定:

承载第一信息的下行信道的子载波间隔;
承载第二信息的上行信道的子载波间隔;
承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔;
承载第二信息的响应信息的下行信道的子载波间隔;
应用所述第一 AI 模型的载波上所激活的 BWP 的子载波间隔;

其中,所述第一信息为用于配置所述第一 AI 模型的下行信息,所述第二信息为用于指示所述第一 AI 模型的上行信息。

在一些实施例中,应用所述第一 AI 模型的载波的数量为多个;

所述第一 AI 模型开始应用的时刻和/或所述模型应用时间,根据第一子载波间隔确定,所述第一子载波间隔为应用所述第一 AI 模型的多个载波上分别激活的 BWP 中最小的子载波间隔。

在一些实施例中,在所述网络设备应用所述第一 AI 模型之前,所述网络设备基于第三 AI 模型对应的第四 AI 模型进行与所述终端设备之间的信息传输,或者,所述网络设备不基于 AI 模型进行与所述终端设备之间的信息传输;

其中,所述第三 AI 模型为预定义的模型,或者历史 AI 模型。

本领域技术人员应当理解,本申请实施例的上述无线通信装置的相关描述可以参照本申请实施例的无线通信方法的相关描述进行理解。

图 12 是本申请实施例提供的一种通信设备 1200 示意性结构图。该通信设备可以是终端设备,也可以是网络设备。图 12 所示的通信设备 1200 包括处理器 1210,处理器 1210 可以从存储器中调用并运行计算机程序,以实现本申请实施例中的方法。

可选地,如图 12 所示,通信设备 1200 还可以包括存储器 1220。其中,处理器 1210 可以从存储器 1220 中调用并运行计算机程序,以实现本申请实施例中的方法。

其中,存储器 1220 可以是独立于处理器 1210 的一个单独的器件,也可以集成在处理器 1210 中。

可选地,如图 12 所示,通信设备 1200 还可以包括收发器 1230,处理器 1210 可以控制该收发器 1230 与其他设备进行通信,具体地,可以向其他设备发送信息或数据,或接收其他设备发送的信息或数据。

其中,收发器 1230 可以包括发射机和接收机。收发器 1230 还可以进一步包括天线,天线的数量可以为一个或多个。

可选地,该通信设备 1200 具体可为本申请实施例的网络设备,并且该通信设备 1800 可以实现本申请实施例的各个方法中由网络设备实现的相应流程,为了简洁,在此不再赘述。

5 可选地,该通信设备 1200 具体可为本申请实施例的移动终端/终端设备,并且该通信设备 1200 可以实现本申请实施例的各个方法中由移动终端/终端设备实现的相应流程,为了简洁,在此不再赘述。

图 13 是本申请实施例的芯片的示意性结构图。图 13 所示的芯片 1300 包括处理器 1910,处理器 1310 可以从存储器中调用并运行计算机程序,以实现本申请实施例中的方法。

10 可选地,如图 13 所示,芯片 1300 还可以包括存储器 1320。其中,处理器 1310 可以从存储器 1320 中调用并运行计算机程序,以实现本申请实施例中的方法。

其中,存储器 1320 可以是独立于处理器 1310 的一个单独的器件,也可以集成在处理器 1310 中。

可选地,该芯片 1300 还可以包括输入接口 1330。其中,处理器 1310 可以控制该输入接口 1330 与其他设备或芯片进行通信,具体地,可以获取其他设备或芯片发送的信息或数据。

15 可选地,该芯片 1300 还可以包括输出接口 1340。其中,处理器 1310 可以控制该输出接口 1340 与其他设备或芯片进行通信,具体地,可以向其他设备或芯片输出信息或数据。

可选地,该芯片可应用于本申请实施例中的网络设备,并且该芯片可以实现本申请实施例的各个方法中由网络设备实现的相应流程,为了简洁,在此不再赘述。

20 可选地,该芯片可应用于本申请实施例中的移动终端/终端设备,并且该芯片可以实现本申请实施例的各个方法中由移动终端/终端设备实现的相应流程,为了简洁,在此不再赘述。

应理解,本申请实施例提到的芯片还可以称为系统级芯片,系统芯片,芯片系统或片上系统芯片等。

本申请实施例还提供了一种计算机存储介质,所述计算机存储介质存储有一个或者多个程序,所述一个或者多个程序可被一个或者多个处理器执行,以实现本申请实施例中的方法。

25 图 14 是本申请实施例提供的一种通信系统 1400 的示意性框图。如图 14 所示,该通信系统 1400 包括终端设备 1410 和网络设备 1420。

其中,该终端设备 1410 可以用于实现上述方法中由终端设备实现的相应的功能,以及该网络设备 1420 可以用于实现上述方法中由网络设备实现的相应的功能为了简洁,在此不再赘述。

30 应理解,本申请实施例的处理器可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法实施例的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可以是通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现成可编程门阵列(Field Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法的步骤。

可以理解,本申请实施例中的存储器可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(Read-Only Memory, ROM)、可编程只读存储器(Programmable ROM, PROM)、可擦除可编程只读存储器(Erasable PROM, EPROM)、电可擦除可编程只读存储器(Electrically EPROM, EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory, RAM),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的 RAM 可用,例如静态随机存取存储器(Static RAM, SRAM)、动态随机存取存储器(Dynamic RAM, DRAM)、同步动态随机存取存储器(Synchronous DRAM, SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM, DDR SDRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM, ESDRAM)、同步连接动态随机存取存储器(Synchlink DRAM, SLDRAM)和直接内存总线随机存取存储器(Direct Rambus RAM, DR RAM)。应注意,本文描述的系统和方法的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

50 应理解,上述存储器为示例性但不是限制性说明,例如,本申请实施例中的存储器还可以是静态随机存取存储器(static RAM, SRAM)、动态随机存取存储器(dynamic RAM, DRAM)、同步动

态随机存取存储器 (synchronous DRAM, SDRAM)、双倍数据速率同步动态随机存取存储器 (double data rate SDRAM, DDR SDRAM)、增强型同步动态随机存取存储器 (enhanced SDRAM, ESDRAM)、同步连接动态随机存取存储器 (synch link DRAM, SLDRAM) 以及直接内存总线随机存取存储器 (Direct Rambus RAM, DR RAM) 等等。也就是说, 本申请实施例中的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

本申请实施例还提供了一种计算机可读存储介质, 用于存储计算机程序。

可选的, 该计算机可读存储介质可应用于本申请实施例中的网络设备, 并且该计算机程序使得计算机执行本申请实施例的各个方法中由网络设备实现的相应流程, 为了简洁, 在此不再赘述。

可选地, 该计算机可读存储介质可应用于本申请实施例中的移动终端/终端设备, 并且该计算机程序使得计算机执行本申请实施例的各个方法中由移动终端/终端设备实现的相应流程, 为了简洁, 在此不再赘述。

本申请实施例还提供了一种计算机程序产品, 包括计算机程序指令。

可选的, 该计算机程序产品可应用于本申请实施例中的网络设备, 并且该计算机程序指令使得计算机执行本申请实施例的各个方法中由网络设备实现的相应流程, 为了简洁, 在此不再赘述。

可选地, 该计算机程序产品可应用于本申请实施例中的移动终端/终端设备, 并且该计算机程序指令使得计算机执行本申请实施例的各个方法中由移动终端/终端设备实现的相应流程, 为了简洁, 在此不再赘述。

本申请实施例还提供了一种计算机程序。

可选的, 该计算机程序可应用于本申请实施例中的网络设备, 当该计算机程序在计算机上运行时, 使得计算机执行本申请实施例的各个方法中由网络设备实现的相应流程, 为了简洁, 在此不再赘述。

可选地, 该计算机程序可应用于本申请实施例中的移动终端/终端设备, 当该计算机程序在计算机上运行时, 使得计算机执行本申请实施例的各个方法中由移动终端/终端设备实现的相应流程, 为了简洁, 在此不再赘述。

本领域普通技术人员可以意识到, 结合本文中所公开的实施例描述的各示例的单元及算法步骤, 能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行, 取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能, 但是这种实现不应认为超出本申请的范围。

所属领域的技术人员可以清楚地了解到, 为描述的方便和简洁, 上述描述的系统、装置和单元的具体工作过程, 可以参考前述方法实施例中的对应过程, 在此不再赘述。

在本申请所提供的几个实施例中, 应该理解到, 所揭露的系统、装置和方法, 可以通过其它的方式实现。例如, 以上所描述的装置实施例仅仅是示意性的, 例如, 所述单元的划分, 仅仅为一种逻辑功能划分, 实际实现时可以有另外的划分方式, 例如多个单元或组件可以结合或者可以集成到另一个系统, 或一些特征可以忽略, 或不执行。另一点, 所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口, 装置或单元的间接耦合或通信连接, 可以是电性, 机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的, 作为单元显示的部件可以是或者也可以不是物理单元, 即可以位于一个地方, 或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外, 在本申请各个实施例中的各功能单元可以集成在一个处理单元中, 也可以是各个单元单独物理存在, 也可以两个或两个以上单元集成在一个单元中。

所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时, 可以存储在一个计算机可读存储介质中。基于这样的理解, 本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来, 该计算机软件产品存储在一个存储介质中, 包括若干指令用以使得一台计算机设备 (可以是个人计算机, 服务器, 或者网络设备等) 执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括: U 盘、移动硬盘、只读存储器 (Read-Only Memory, ROM)、随机存取存储器 (Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述, 仅为本申请的具体实施方式, 但本申请的保护范围并不局限于此, 任何熟悉本技术领域的技术人员在本申请揭露的技术范围内, 可轻易想到变化或替换, 都应涵盖在本申请的保护范围之内。因此, 本申请的保护范围应所述以权利要求的保护范围为准。

权利要求书

1、一种无线通信方法，所述方法包括：

终端设备根据上报给网络设备的模型应用时间，或者，根据网络设备配置的模型应用时间，确定第一 AI 模型开始应用的时刻；

所述终端设备在所述时刻之后，采用所述第一 AI 模型与所述网络设备进行通信。

2、根据权利要求 1 所述的方法，其中，所述模型应用时间为以下中的任意一项：

所述终端设备接收到第一信息之后到应用所述第一 AI 模型的时刻之间的时间间隔；

所述终端设备发送第一信息的 HARQ-ACK 信息之后到应用所述第一 AI 模型的时刻之间的时间间隔；

所述终端设备接收到第一信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

所述终端设备发送第一信息的 HARQ-ACK 信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

其中，所述第一信息为用于配置所述第一 AI 模型的下行信息，所述时间间隔以时隙或 OFDM 符号为单位。

3、根据权利要求 1 或 2 所述的方法，其中，所述第一 AI 模型开始应用的时刻为以下中的任意一项：

承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号；

承载第一信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号；

所述终端设备接收到第一信息的时隙开始 k 个时隙之后的第一个时隙；

承载第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 k 个时隙，或 $(k+3N)$ 个时隙，或 $\max(k, 3N)$ 个时隙之后的第一个时隙；

其中，所述第一信息为用于配置所述第一 AI 模型的下行信息，k 为所述模型应用时间，N 为 1ms 包含的时隙数量。

4、根据权利要求 2 或 3 所述的方法，其中，所述方法还包括：

所述终端设备接收第一信息，所述第一信息是用于配置所述第一 AI 模型的下行信息，

其中，所述第一信息包括以下中的一项或多项：

所述第一 AI 模型的标识信息；

所述第一 AI 模型的模型功能；

所述第一 AI 模型的模型结构；以及，

所述第一 AI 模型的模型参数。

5、根据权利要求 1 所述的方法，其中，所述模型应用时间为以下中的任意一项：

所述终端设备发送第二信息之后到应用所述第一 AI 模型的时刻之间的时间间隔；

所述终端设备发送第二信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

所述终端设备接收第二信息的响应信息之后到应用所述第一 AI 模型的时刻之间的时间间隔；

所述终端设备接收第二信息的响应信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔；

其中，所述第二信息为用于指示所述第一 AI 模型的上行信息，所述时间间隔以时隙或 OFDM 符号为单位。

6、根据权利要求 1 或 5 所述的方法，其中，所述第一 AI 模型开始应用的时刻为以下中的任意一项：

承载第二信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号；

承载第二信息的上行信道的响应信息的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号；

所述终端设备发送第二信息的时隙开始 k 个时隙之后的第一个时隙;

承载第二信息的响应信息的下行信道所在时隙开始 k 个时隙, 或 $(k+3N)$ 个时隙, 或 $\max(k, 3N)$ 个时隙之后的第一个时隙;

其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, k 为所述模型应用时间, N 为 1ms 包含的时隙数量。

7、根据权利要求 5 或 6 所述的方法, 其中, 所述方法还包括:

所述终端设备发送第二信息, 所述第二信息是用于指示所述第一 AI 模型的上行信息;

其中, 所述第二信息包括以下中的一项或多项:

所述第一 AI 模型的标识信息;

所述第一 AI 模型的模型功能;

所述第一 AI 模型的模型结构; 以及,

所述第一 AI 模型的模型参数。

8、根据权利要求 1-7 任一项所述的方法, 其中, 所述上报给网络设备的模型应用时间包括以下中的一项或多项:

所述终端设备支持的一个或多个 AI 模型分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个 AI 模型包括所述第一 AI 模型;

所述终端设备支持的一个或多个模型功能分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个模型功能包括所述第一 AI 模型的功能;

所述终端设备支持的一种或多种模型更新方式分别对应的模型应用时间; 其中, 所述终端设备支持的一种或多种模型更新方式包括所述第一 AI 模型的更新方式;

所述终端设备支持的一个或多个模型类型分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个模型类型包括所述第一 AI 模型的类型。

9、根据权利要求 8 所述的方法, 其中, 所述模型更新方式包括: 更新模型参数和/或更新模型结构。

10、根据权利要求 8 或 9 所述的方法, 其中, 所述模型类型包括:

高精度模型和低精度模型; 或者, 已知模型和未知模型。

11、根据权利要求 1-10 任一项所述的方法, 其中, 所述第一 AI 模型开始应用的时刻和/或所述模型应用时间, 根据以下任意一项确定:

承载第一信息的下行信道的子载波间隔;

承载第二信息的上行信道的子载波间隔;

承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔;

承载第二信息的响应信息的下行信道的子载波间隔;

应用所述第一 AI 模型的载波上所激活的 BWP 的子载波间隔;

其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, 所述第二信息为用于指示所述第一 AI 模型的上行信息。

12、根据权利要求 11 所述的方法, 其中, 应用所述第一 AI 模型的载波的数量为多个,

所述第一 AI 模型开始应用的时刻和/或所述模型应用时间根据第一子载波间隔确定, 所述第一子载波间隔为应用所述第一 AI 模型的多个载波上分别激活的 BWP 中最小的子载波间隔。

13、根据权利要求 1-12 任一项所述的方法, 其中, 在所述终端设备应用所述第一 AI 模型之前, 所述终端设备基于第三 AI 模型进行与所述网络设备之间的信息传输, 或者, 所述终端设备不基于 AI 模型进行与所述网络设备之间的信息传输;

其中, 所述第三 AI 模型为预定义的 AI 模型, 或者历史 AI 模型。

14、一种无线通信方法, 所述方法包括:

网络设备根据终端设备上报的模型应用时间, 或者, 根据为所述终端设备配置的模型应用时间, 确定第一 AI 模型开始应用的时刻;

所述网络设备在所述时刻之后, 采用所述第一 AI 模型对应的第二 AI 模型与所述终端设备进行通信。

15、根据权利要求 14 所述的方法, 其中, 所述模型应用时间为以下中的任意一项:

所述终端设备接收到第一信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

所述终端设备发送第一信息的 HARQ-ACK 信息之后到应用所述第一 AI 模型的时刻之间的时间

间隔;

所述终端设备接收到第一信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

所述终端设备发送第一信息的 HARQ-ACK 信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, 所述时间间隔以时隙或 OFDM 符号为单位。

16、根据权利要求 14 或 15 所述的方法, 其中, 所述第一 AI 模型开始应用的时刻为以下中的任一项:

承载第一信息的 HARQ-ACK 信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

承载第一信息的下行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

所述终端设备接收到第一信息的时隙开始 k 个时隙之后的第一个时隙;

承载第一信息的 HARQ-ACK 信息的上行信道所在时隙开始 k 个时隙, 或 $(k+3N)$ 个时隙, 或 $\max(k, 3N)$ 个时隙之后的第一个时隙;

其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, k 为所述模型应用时间, N 为 1ms 包含的时隙数量。

17、根据权利要求 15 或 16 所述的方法, 其中, 所述方法还包括:

所述网络设备向所述终端设备发送第一信息, 所述第一信息为用于配置所述第一 AI 模型的下行, 其中, 所述第一信息包括以下中的一项或多项:

所述第一 AI 模型的标识信息;

所述第一 AI 模型的功能;

所述第一 AI 模型的模型结构; 以及,

所述第一 AI 模型的模型参数。

18、根据权利要求 14 所述的方法, 其中, 所述模型应用时间为以下中的任意一项:

所述终端设备发送第二信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

所述终端设备发送第二信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

所述终端设备接收第二信息的响应信息之后到应用所述第一 AI 模型的时刻之间的时间间隔;

所述终端设备接收第二信息的响应信息之后到所述终端设备发送基于所述第一 AI 模型得到的反馈信息的时刻之间的时间间隔;

其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, 所述时间间隔以时隙或 OFDM 符号为单位。

19、根据权利要求 14 或 18 所述的方法, 其中, 所述第一 AI 模型开始应用的时刻为以下中的任一项:

承载第二信息的上行信道的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

承载第二信息的上行信道的响应信息的最后一个 OFDM 符号开始 k 个 OFDM 符号之后的第一个时隙或者第一个 OFDM 符号;

所述终端设备发送第二信息的时隙开始 k 个时隙之后的第一个时隙;

承载第二信息的响应信息的下行信道所在时隙开始 k 个时隙, 或 $(k+3N)$ 个时隙, 或 $\max(k, 3N)$ 个时隙之后的第一个时隙;

其中, 所述第二信息为用于指示所述第一 AI 模型的上行信息, k 为所述模型应用时间, N 为 1ms 包含的时隙数量。

20、根据权利要求 18 或 19 所述的方法, 其中, 所述方法还包括:

所述网络设备接收所述终端设备发送的第二信息, 所述第二信息是用于指示所述第一 AI 模型的上行信息;

其中, 所述第二信息包括以下中的一项或多项:

所述第一 AI 模型的标识信息;

所述第一 AI 模型的模型功能;

所述第一 AI 模型的模型结构; 以及,

所述第一 AI 模型的模型参数。

21、根据权利要求 14-20 任一项所述的方法, 其中, 所述终端设备上报的模型应用时间包括以下中的一项或多项:

5 所述终端设备支持的一个或多个 AI 模型分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个 AI 模型包括所述第一 AI 模型;

所述终端设备支持的一个或多个模型功能分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个模型功能包括所述第一 AI 模型的功能;

10 所述终端设备支持的一种或多种模型更新方式分别对应的模型应用时间; 其中, 所述终端设备支持的一种或多种模型更新方式包括所述第一 AI 模型的更新方式;

所述终端设备支持的一个或多个模型类型分别对应的模型应用时间; 其中, 所述终端设备支持的一个或多个模型类型包括所述第一 AI 模型的类型。

22、根据权利要求 21 所述的方法, 其中, 所述模型更新方式包括: 更新模型参数和/或更新模型结构。

15 23、根据权利要求 21 或 22 所述的方法, 其中, 所述模型类型包括: 高精度模型和低精度模型; 或者, 已知模型和未知模型。

24、根据权利要求 14-23 任一项所述的方法, 其中,

所述网络设备为所述终端设备配置的模型应用时间, 基于所述终端设备上报的模型应用时间确定。

20 25、根据权利要求 14-24 任一项所述的方法, 其中,

所述网络设备为所述终端设备配置的模型应用时间的时长, 大于或等于所述终端设备上报的模型应用时间的时长。

26、根据权利要求 14-25 任一项所述的方法, 其中, 所述第一 AI 模型开始应用的时刻和/或所述模型应用时间, 根据以下任意一项确定:

25 承载第一信息的下行信道的子载波间隔;

承载第二信息的上行信道的子载波间隔;

承载第一信息的 HARQ-ACK 信息的上行信道的子载波间隔;

承载第二信息的响应信息的下行信道的子载波间隔;

应用所述第一 AI 模型的载波上所激活的 BWP 的子载波间隔;

30 其中, 所述第一信息为用于配置所述第一 AI 模型的下行信息, 所述第二信息为用于指示所述第一 AI 模型的上行信息。

27、根据权利要求 26 所述的方法, 其中, 应用所述第一 AI 模型的载波的数量为多个;

所述第一 AI 模型开始应用的时刻和/或所述模型应用时间, 根据第一子载波间隔确定, 所述第一子载波间隔为应用所述第一 AI 模型的多个载波上分别激活的 BWP 中最小的子载波间隔。

35 28、根据权利要求 14-27 任一项所述的方法, 其中, 在所述网络设备应用所述第一 AI 模型之前, 所述网络设备基于第三 AI 模型对应的第四 AI 模型进行与所述终端设备之间的信息传输, 或者, 所述网络设备不基于 AI 模型进行与所述终端设备之间的信息传输;

其中, 所述第三 AI 模型为预定义的模型, 或者历史 AI 模型。

29、一种无线通信装置, 应用于终端设备, 所述装置包括:

40 第一确定单元, 配置为根据上报给网络设备的模型应用时间, 或者, 根据网络设备配置的模型应用时间, 确定第一 AI 模型开始应用的时刻;

第一通信单元, 配置为在所述时刻之后, 采用所述第一 AI 模型与所述网络设备进行通信。

30、一种无线通信装置, 应用于网络设备, 所述装置包括:

45 第二确定单元, 配置为根据终端设备上报的模型应用时间, 或者, 根据为所述终端设备配置的模型应用时间, 确定第一 AI 模型开始应用的时刻;

第二通信单元, 配置为在所述时刻之后, 采用所述第一 AI 模型对应的第二 AI 模型与所述终端设备进行通信。

31、一种终端设备, 包括:

存储器, 用于存储计算机可执行指令;

50 处理器, 与所述存储器连接, 用于通过执行所述计算机可执行指令, 实现权利要求 1 至 13 中任一项所述的方法。

32、一种网络设备，包括：

存储器，用于存储计算机可执行指令；

处理器，与所述存储器连接，用于通过执行所述计算机可执行指令，实现权利要求 14 至 28 中任一项所述的方法。

5 33、一种芯片，所述芯片包括：

处理器，用于从存储器中调用并运行计算机程序，使得安装有所述芯片的设备执行如权利要求 1 至 13 中任一项所述的方法，或者，执行如权利要求 14 至 28 中任一项所述的方法。

10 34、一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被至少一个处理器执行时实现如权利要求 1 至 13 中任一项所述的方法，或者，实现如权利要求 14 至 28 中任一项所述的方法。

35、一种计算机程序产品，所述计算机程序产品包括计算机存储介质，所述计算机存储介质存储计算机程序，所述计算机程序包括能够由至少一个处理器执行的指令，当所述指令由所述至少一个处理器执行时实现权利要求 1 至 13 中任一项所述的方法，或者，实现如权利要求 14 至 28 中任一项所述的方法。

15 36、一种计算机程序，所述计算机程序使得计算机执行如权利要求 1 至 13 中任一项所述的方法，或者，实现如权利要求 14 至 28 中任一项所述的方法。

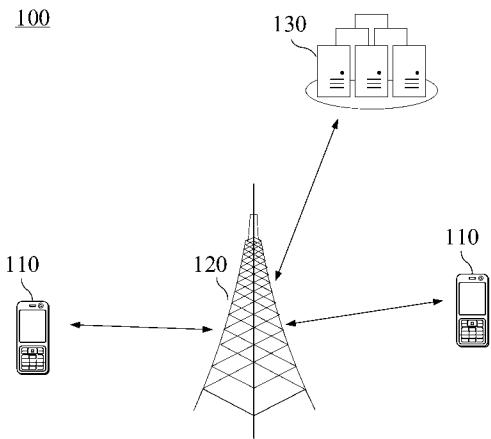


图 1

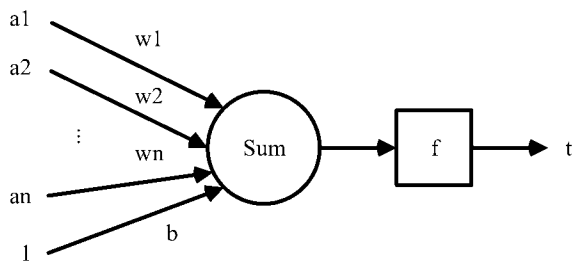


图 2

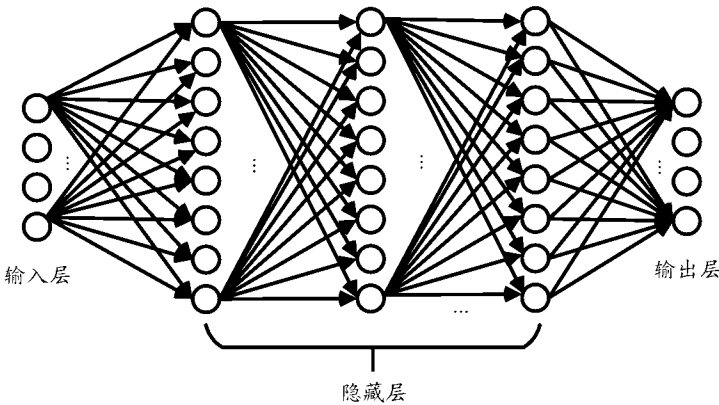


图 3

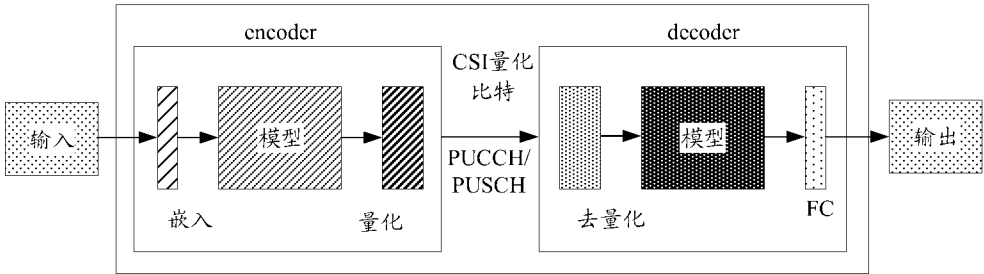


图 4

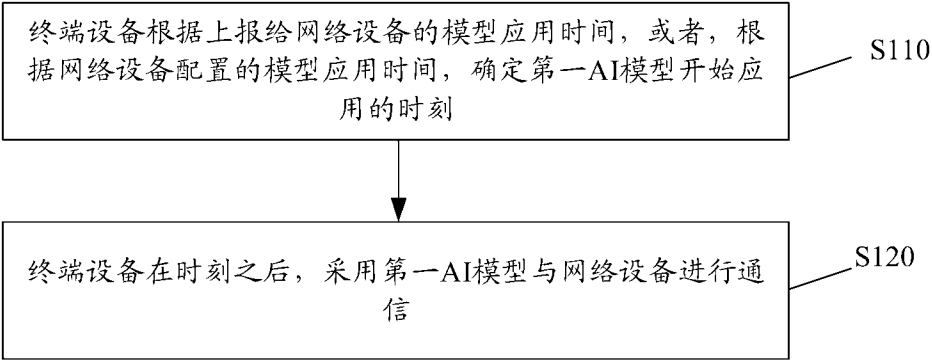


图 5

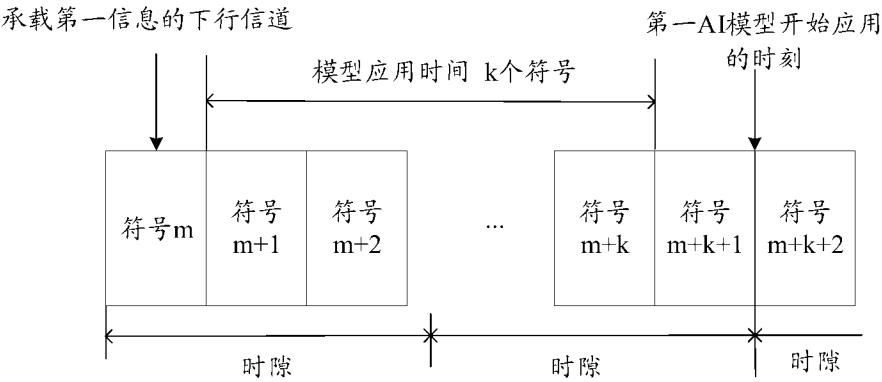


图 6A

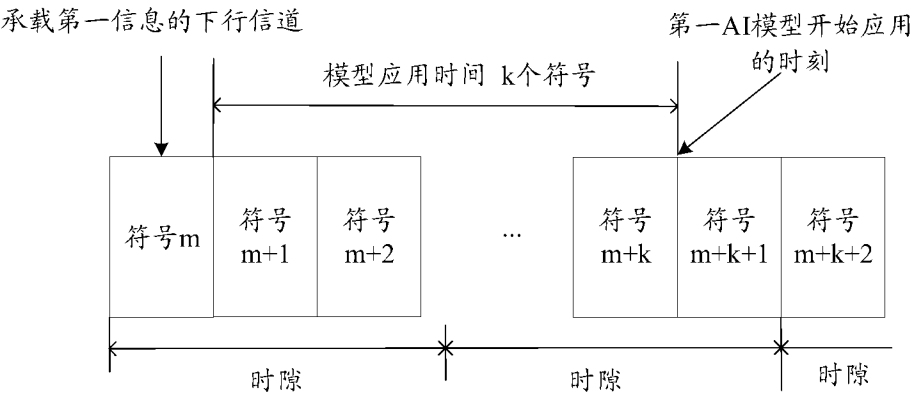


图 6B

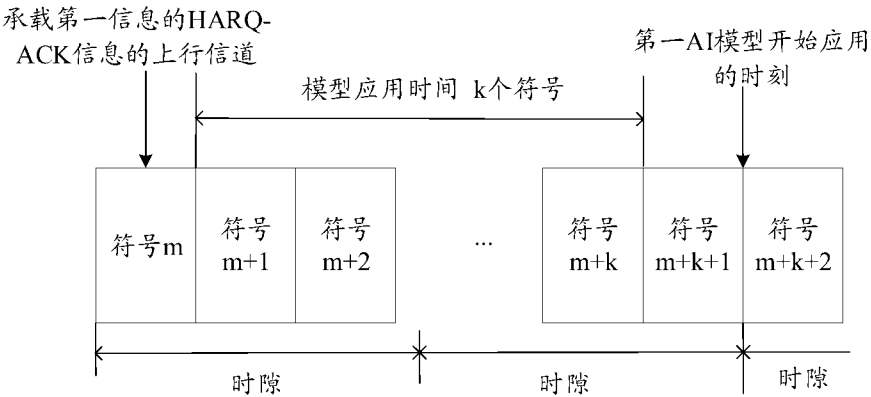


图 7A

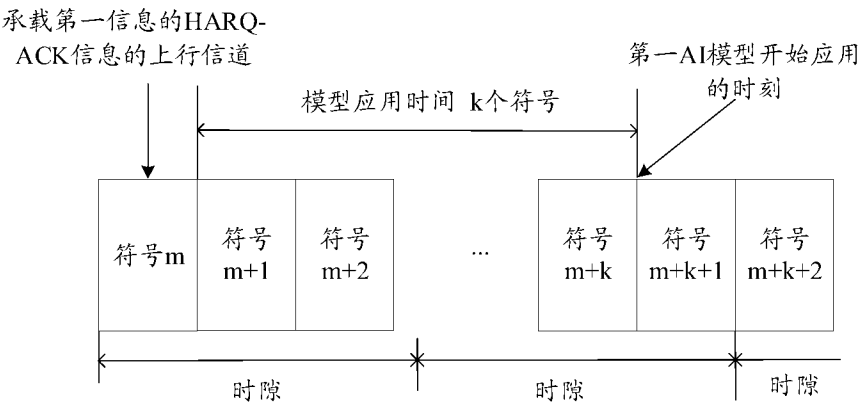


图 7B

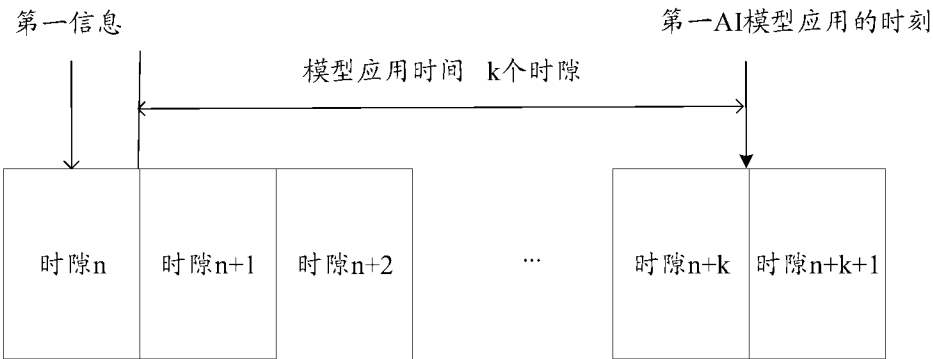


图 8



图 9A

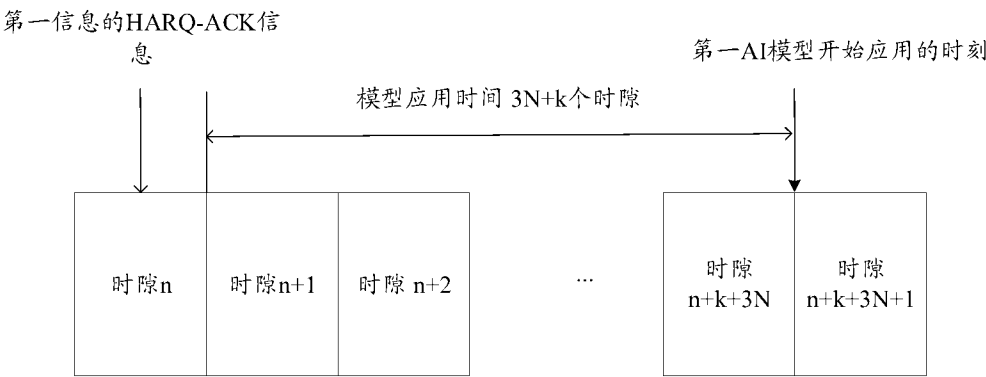


图 9B



图 9C

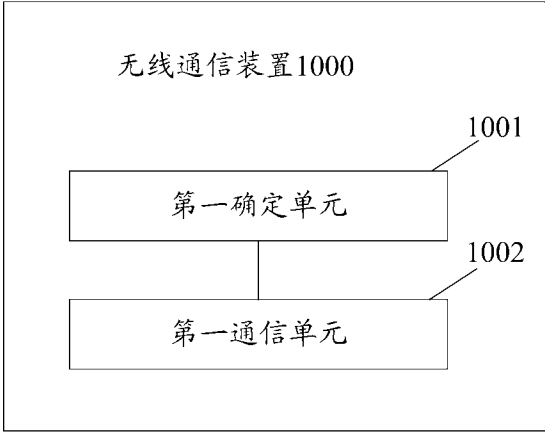


图 10

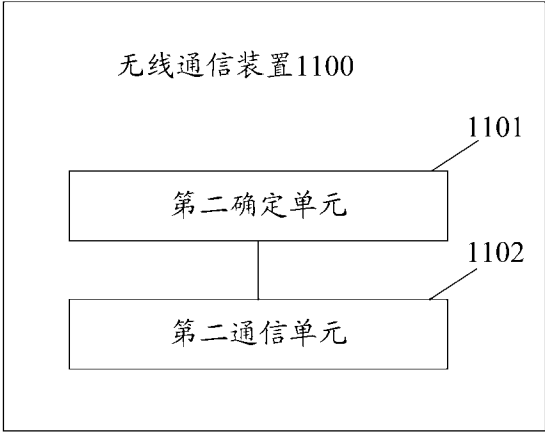


图 11

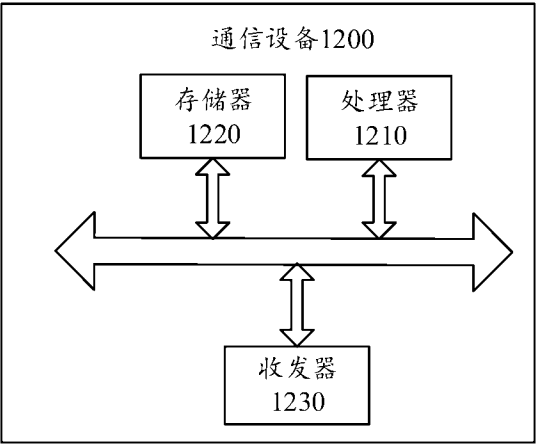


图 12

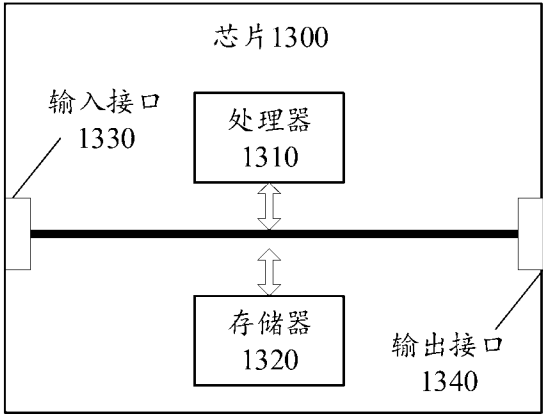


图 13

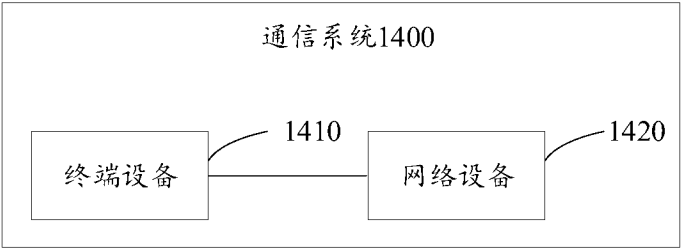


图 14

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/112522

A. CLASSIFICATION OF SUBJECT MATTER H04W 72/0446(2023.01)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) IPC:H04W 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) CNABS; CNKI; CNTXT; VEN; USTXT; WOTXT; EPTXT; 3GPP; IEEE; IETF: 人工智能, 神经网络, 应用, 启用, 时间, 时刻, 更新, 通信, Artificial Intelligence, AI, Neural network, DNN, enable, time, update, communication		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 114175051 A (GOOGLE INC.) 11 March 2022 (2022-03-11) description, paragraphs [0110]-[0136]	1, 13, 14, 28-36
A	CN 110636020 A (PEKING UNIVERSITY) 31 December 2019 (2019-12-31) entire document	1-36
A	CN 112997435 A (GOOGLE INC.) 18 June 2021 (2021-06-18) entire document	1-36
A	CN 115485995 A (HUAWEI TECHNOLOGIES CO., LTD.) 16 December 2022 (2022-12-16) entire document	1-36
☐ 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 17 April 2024		Date of mailing of the international search report 29 April 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/ CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/112522

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	114175051	A	11 March 2022	None	
CN	110636020	A	31 December 2019	None	
CN	112997435	A	18 June 2021	None	
CN	115485995	A	16 December 2022	None	

A. 主题的分类

H04W 72/0446(2023.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:H04W

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNKI;CNTXT;VEN;USTXT;WOTXT;EPTXT;3GPP;IEEE;IETF: 人工智能, 神经网络, 应用, 启用, 时间, 时刻, 更新, 通信, Artificial Intelligence, AI, Neural network, DNN, enable, time, update, communication

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 114175051 A (谷歌有限责任公司) 2022年3月11日 (2022 - 03 - 11) 说明书第[0110]–[0136]段	1、13、14、28–36
A	CN 110636020 A (北京大学) 2019年12月31日 (2019 - 12 - 31) 全文	1–36
A	CN 112997435 A (谷歌有限责任公司) 2021年6月18日 (2021 - 06 - 18) 全文	1–36
A	CN 115485995 A (华为技术有限公司) 2022年12月16日 (2022 - 12 - 16) 全文	1–36

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

★ 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2024年4月17日

国际检索报告邮寄日期

2024年4月29日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

陈晓霞

电话号码 (+86) 0512-88996090

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/112522

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	114175051	A	2022年3月11日	无	
CN	110636020	A	2019年12月31日	无	
CN	112997435	A	2021年6月18日	无	
CN	115485995	A	2022年12月16日	无	