

(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
2024 年 1 月 18 日 (18.01.2024)



(10) 国际公布号
WO 2024/012279 A1

(51) 国际专利分类号:
H04W 80/08 (2009.01)

(21) 国际申请号: PCT/CN2023/105066

(22) 国际申请日: 2023 年 6 月 30 日 (30.06.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210821506.3 2022年7月12日 (12.07.2022) CN

(71) 申请人: 维沃移动通信有限公司(VIVO MOBILE COMMUNICATION CO., LTD.) [CN/CN]; 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。

(72) 发明人: 潘翔(PAN, Xiang); 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。张宏平(ZHANG, Hongping); 中国广东省东莞市长安镇维沃路 1 号, Guangdong 523863 (CN)。

(74) 代理人: 北京路浩知识产权代理有限公司(CN-KNOWHOW INTELLECTUAL PROPERTY AGENT LIMITED); 中国北京市丰台区万丰路 68

号院和谐广场西侧写字楼 21 层 21009 号, Beijing 100161 (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)。

(54) Title: INFORMATION TRANSMISSION METHOD AND APPARATUS, AND DEVICE

(54) 发明名称: 信息传输方法、装置及设备

第一通信设备基于无线承载, 传输AI层信息;
其中, 所述AI层信息是第一通信设备的AI协议层生成的或所述第一通信设备的AI协议层待接收的;
所述AI层信息用于管理所述第一通信设备中的AI模型和/或第二通信设备中的AI模型, 所述第一通信设备的协议栈包括所述第一通信设备的AI协议层

500
图 5

500 A first communication device transmits AI layer information on the basis of a radio bearer; the AI layer information is generated by an AI protocol layer of the first communication device, or received by the AI protocol layer of the first communication device; the AI layer information is used to manage an AI model in the first communication device and/or an AI model in a second communication device, and a protocol stack of the first communication device comprises the AI protocol layer of the first communication device

(57) Abstract: The present application discloses an information transmission method and apparatus, and a device, in the technical field of communications. The information transmission method of an embodiment of the present application comprises: a first communication device transmitting AI layer information on the basis of a radio bearer, the AI layer information being generated by an AI protocol layer of the first communication device or received by the AI protocol layer of the first communication device. The AI layer information is used to manage an AI model in the first communication device and/or an AI model in a second communication device, and a protocol stack of the first communication device comprises the AI protocol layer of the first communication device.

(57) 摘要: 本申请公开了一种信息传输方法、装置及设备, 属于通信技术领域, 本申请实施例的信息传输方法包括: 第一通信设备基于无线承载, 传输AI层信息; 其中, 所述AI层信息是第一通信设备的AI协议层生成的或所述第一通信设备的AI协议层待接收的; 所述AI层信息用于管理所述第一通信设备中的AI模型和/或第二通信设备中的AI模型, 所述第一通信设备的协议栈包括所述第一通信设备的AI协议层。

[见续页]

WO 2024/012279 A1

本国际公布:

— 包括国际检索报告(条约第21条(3))。

信息传输方法、装置及设备

相关申请的交叉引用

5 本申请要求于 2022 年 07 月 12 日提交的申请号为 202210821506.3，发明名称为“信息传输方法、装置及设备”的中国专利申请的优先权，其通过引用方式全部并入本申请。

技术领域

本申请属于通信技术领域，具体涉及一种信息传输方法、装置及设备。

10

背景技术

人工智能（Artificial Intelligence，AI）目前在各个领域获得了广泛的应用，将 AI 融入通信网络，显著提升吞吐量、时延以及用户容量等技术指标是未来的通信网络的重要优化方向。

15

若模型训练功能部署在基站上，而模型推理功能部署在终端上，则 AI 模型需要在网络和终端之间进行交互。然而，部分 AI 模型较大，需要进行大量分段传输，导致其他控制信令传输受到影响，网络控制的实时性降低。

发明内容

20

本申请实施例提供一种信息传输方法、装置及设备，能够解决部分 AI 模型较大，需要进行大量分段传输，导致其他控制信令传输受到影响，网络控制的实时性降低的问题。

第一方面，提供了一种信息传输方法，该方法包括：

第一通信设备传输 AI 层信息；

25

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

第二方面，提供了一种信息传输方法，该方法包括：

30

第二通信设备传输 AI 层信息；

其中，所述 AI 层信息是所述第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

第三方面，提供了一种信息传输方法，该方法包括：

第三通信设备传输 AI 层信息；

- 5 其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

第四方面，提供了一种信息传输装置，该装置包括：

- 10 第一传输模块，用于传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

- 15 第五方面，提供了一种信息传输装置，该装置包括：

第二传输模块，用于传输 AI 层信息；

其中，所述 AI 层信息是第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

- 20 所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

第六方面，提供了一种信息传输装置，该装置包括：

第三传输模块，用于传输 AI 层信息；

其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

- 25 所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

第七方面，提供了一种第一通信设备，该第一通信设备包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如第一方面所述的方法的步骤。

- 30 第八方面，提供了一种第一通信设备，包括处理器及通信接口，其中，所述通信接口用于：

传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

- 35 所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的

AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

第九方面，提供了一种第二通信设备，该第二通信设备包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如第二方面所述的方法的步骤。

5 第十方面，提供了一种第二通信设备，包括处理器及通信接口，其中，所述通信接口用于：

传输 AI 层信息；

其中，所述 AI 层信息是所述第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

10 所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

第十一方面，提供了一种第三通信设备，该第三通信设备包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如第三方面所述的方法的步骤。

15 第十二方面，提供了一种第三通信设备，包括处理器及通信接口，其中，所述通信接口用于：

传输 AI 层信息；

其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

20 所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

第十三方面，提供了一种信息传输系统，包括：第一通信设备、第二通信设备和第三通信设备，所述第一通信设备可用于执行如第一方面所述的信息传输方法的步骤，所述第二通信设备可用于执行如第二方面所述的信息传输方法的步骤，所述
25 第三通信设备可用于执行如第三方面所述的信息传输方法的步骤。

第十四方面，提供了一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令被处理器执行时实现如第一方面所述的方法的步骤，或者实现如第二方面所述的方法的步骤，或者实现如第三方面所述的方法的步骤。

第十五方面，提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现如第一方面所述的方法的步骤，或者实现如第二方面所述的方法的步骤，或者实现如第三方面所述的方法的步骤。
30

第十六方面，提供了一种计算机程序/程序产品，所述计算机程序/程序产品被存储在存储介质中，所述计算机程序/程序产品被至少一个处理器执行以实现如第一方面所述的方法的步骤，或者实现如第二方面所述的方法的步骤，或者实现如第三方面
35

面所述的方法的步骤。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，

5 增强网络控制的实时性。

附图说明

图 1 示出本申请实施例可应用的一种无线通信系统的框图；

图 2 是相关技术提供的基于 AI 的通信网络优化的流程图；

10 图 3 是相关技术提供的用户面协议栈的示意图；

图 4 是相关技术提供的控制面协议栈的示意图；

图 5 是本申请实施例提供的信息传输方法的流程示意图之一；

图 6 是本申请实施例提供的 AI 协议层的示意图之一；

图 7 是本申请实施例提供的 AI 协议层的示意图之二；

15 图 8 是本申请实施例提供的 AI 协议层的示意图之三；

图 9 是本申请实施例提供的 AI 协议层的示意图之四；

图 10 是本申请实施例提供的信息传输方法的流程示意图之二；

图 11 是本申请实施例提供的 AI 协议层的示意图之五；

图 12 是本申请实施例提供的信息传输方法的流程示意图之三；

20 图 13 是本申请实施例提供的信息传输装置的结构示意图之一；

图 14 是本申请实施例提供的信息传输装置的结构示意图之二；

图 15 是本申请实施例提供的信息传输装置的结构示意图之三；

图 16 是本申请实施例提供的通信设备的结构示意图；

图 17 为实现本申请实施例的一种第一通信设备的硬件结构示意图；

25 图 18 为实现本申请实施例的一种第二通信设备的硬件结构示意图；

图 19 为实现本申请实施例的一种第三通信设备的硬件结构示意图。

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清楚描述，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所获得的所有其他实施例，都属于本申请保护的

30 范围。

本申请的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象，而不用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情

况下可以互换，以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施，且“第一”、“第二”所区别的对象通常为一类，并不限定对象的个数，例如第一对象可以是一个，也可以是多个。此外，说明书以及权利要求中“和/或”表示所连接对象的至少其中之一，字符“/”一般表示前后关联对象是一种“或”的关系。

值得指出的是，本申请实施例所描述的技术不限于长期演进型（Long Term Evolution, LTE）/LTE 的演进（LTE-Advanced, LTE-A）系统，还可用于其他无线通信系统，诸如码分多址（Code Division Multiple Access, CDMA）、时分多址（Time Division Multiple Access, TDMA）、频分多址（Frequency Division Multiple Access, FDMA）、正交频分多址（Orthogonal Frequency Division Multiple Access, OFDMA）、单载波频分多址（Single-carrier Frequency Division Multiple Access, SC-FDMA）和其他系统。本申请实施例中的术语“系统”和“网络”常被可互换地使用，所描述的技术既可用于以上提及的系统 and 无线电技术，也可用于其他系统和无线电技术。以下描述出于示例目的描述了新空口（New Radio, NR）系统，并且在以下大部分描述中使用 NR 术语，但是这些技术也可应用于 NR 系统应用以外的应用，如第 6 代（6th Generation, 6G）通信系统。

图 1 示出本申请实施例可应用的一种无线通信系统的框图。无线通信系统包括终端 11 和网络侧设备 12。其中，终端 11 可以是手机、平板电脑（Tablet Personal Computer）、膝上型电脑（Laptop Computer）或称为笔记本电脑、个人数字助理（Personal Digital Assistant, PDA）、掌上电脑、上网本、超级移动个人计算机（ultra-mobile personal computer, UMPC）、移动上网装置（Mobile Internet Device, MID）、增强现实(augmented reality, AR)/虚拟现实(virtual reality, VR)设备、机器人、可穿戴式设备（Wearable Device）、车载设备（VUE）、行人终端（PUE）、智能家居(具有无线通信功能的家居设备，如冰箱、电视、洗衣机或者家具等)、游戏机、个人计算机（personal computer, PC）、柜员机或者自助机等终端侧设备，可穿戴式设备包括：智能手表、智能手环、智能耳机、智能眼镜、智能首饰（智能手镯、智能手链、智能戒指、智能项链、智能脚镯、智能脚链等）、智能腕带、智能服装等。需要说明的是，在本申请实施例并不限定终端 11 的具体类型。网络侧设备 12 可以包括接入网设备或核心网设备，其中，接入网设备 12 也可以称为无线接入网设备、无线接入网（Radio Access Network, RAN）、无线接入网功能或无线接入网单元。接入网设备 12 可以包括基站、WLAN 接入点或 WiFi 节点等，基站可被称为节点 B、演进节点 B（eNB）、接入点、基收发机站（Base Transceiver Station, BTS）、无线电基站、无线电收发机、基本服务集（Basic Service Set, BSS）、扩展服务集（Extended Service Set, ESS）、家用 B 节点、家用演进型 B 节点、发送接收点（Transmitting Receiving Point, TRP）或所述领域中其他某个合适的

术语，只要达到相同的技术效果，所述基站不限于特定技术词汇，需要说明的是，在本申请实施例中仅以 NR 系统中的基站为例进行介绍，并不限定基站的具体类型。核心网设备可以包含但不限于如下至少一项：核心网节点、核心网功能、移动管理实体（Mobility Management Entity, MME）、接入移动管理功能（Access and Mobility Management Function, AMF）、会话管理功能（Session Management Function, SMF）、用户平面功能（User Plane Function, UPF）、策略控制功能（Policy Control Function, PCF）、策略与计费规则功能单元（Policy and Charging Rules Function, PCRF）、边缘应用服务发现功能（Edge Application Server Discovery Function, EASDF）、统一数据管理（Unified Data Management, UDM）、统一数据仓储（Unified Data Repository, UDR）、归属用户服务器（Home Subscriber Server, HSS）、集中式网络配置（Centralized network configuration, CNC）、网络存储功能（Network Repository Function, NRF）、网络开放功能（Network Exposure Function, NEF）、本地 NEF（Local NEF, 或 L-NEF）、绑定支持功能（Binding Support Function, BSF）、应用功能（Application Function, AF）等。需要说明的是，在本申请实施例中仅以 NR 系统中的核心网设备为例进行介绍，并不限定核心网设备的具体类型。

首先对以下内容进行介绍：

（1）人工智能（Artificial Intelligence, AI）与通信网络的结合；

图 2 是相关技术提供的基于 AI 的通信网络优化的流程图，如图 2 所示，模型训练功能和模型推理功能是基于 AI 的通信网络优化整体流程中的关键功能；

其中，模型训练功能基于训练数据生成 AI 模型，完成模型有效性测试后，将模型部署至模型推理功能。在生成 AI 模型过程中，模型训练功能需要获取并分析大量数据，对硬件性能和算力的要求很高，主要部署在网络侧设备上，如操作维护管理节点（Operation Administration and Maintenance, OAM）、基站。

模型推理功能基于上述模型，以推理数据为输入，获得 AI 推理输出，如空口指标预测。模型推理功能对硬件性能和算力的要求相比模型训练功能更低。随着终端能力不断提升，甚至增加了 AI 专用芯片神经网络处理器（Neural Processing Unit, NPU），AI 驱动效率也不断提升，使得在终端侧进行模型推理，甚至部分模型训练成为可能。

（2）控制面和用户面的协议栈；

在移动通信系统中，需要传输的数据至少包括用户数据和控制信令，用户数据和控制信令的传输特征明显不同，比如，用户数据可能数据量很大，控制信令数据量相对更小，控制信令要求的时延和可靠性高，控制信令对网络可见并终结在网络，需要网络进行处理，而用户数据内容对网络透明。

基于上述不同的特征，引入了控制面（control plane）和用户面（user plane）的

概念。用户面功能主要用于支持用户数据传输，控制面功能主要用于支持控制信令传输，比如，RRC 连接建立，承载建立，无线资源配置等。在无线空口，用户面的数据通过数据无线承载（Data Radio Bearer, DRB）发送，控制面的数据通过信令无线承载（Signalling Radio Bearer, SRB）发送。

在通信系统中，用户面协议栈由服务数据适配协议(Service Data Adaptation Protocol, SDAP)、分组数据汇聚协议 (Packet Data Convergence Protocol, PDCP)、无线链路层控制协议 (Radio Link Control, RLC)、媒体接入控制协议 (Medium Access Control, MAC) 和物理层 (Physical Layer, PHY) 层组成，图 3 是相关技术提供的用户面协议栈的示意图，图 4 是相关技术提供的控制面协议栈的示意图，如图 3 和图 4 所示，控制面协议栈由非接入层 (Non-Access Stratum, NAS)、无线资源控制 (Radio Resource Control, RRC)、PDCP、RLC、MAC 和 PHY 层组成，终端和基站间的 RRC 消息封装在 PDCP 服务数据单元 (Service Data Unit, SDU) 中，单个 PDCP SDU 最大支持 9000 字节。为了满足更大 RRC 消息的传输，5G 支持 RRC 消息分段，下行最大支持 5 个分段，上行最大支持 16 个分段。

若模型训练功能部署在基站上，而模型推理功能部署在终端上，则 AI 模型需要在网络和终端之间进行交互。当前下行只支持 5 个分段，即可传输最大的下行 RRC 消息为 45000 字节。然而，部分 AI 模型较大，若仍采用现有 RRC 消息发送，将需要进行大量分段，如 9M 大小的 AI 模型，需要分接近 1000 段。大量的 RRC 分段可能会影响其他控制信令的传输，影响网络控制的实时性。

因此，本申请实施例提供一种信息传输方法、装置及设备。

下面结合附图，通过一些实施例及其应用场景对本申请实施例提供的信息传输方法、装置及设备进行详细地说明。

图 5 是本申请实施例提供的信息传输方法的流程示意图之一，如图 5 所示，该信息传输方法包括如下步骤：

步骤 500，第一通信设备基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

可选地，本申请实施例中，可以引入新的协议层，即引入 AI 协议层，负责模型能力管理、模型训练数据采集、模型训练、模型的分段与重组、模型维护和更新、以及模型推理。

可选地，可以引入新的无线承载类型，用于承载新的协议层的信息；

可选地，相比已有的无线承载，还可以新定义（新增）一个无线承载；

可选地，新增的无线承载可以是新增信令无线承载，比如可以称为 SRB5；

可选地，新增的无线承载可以是新增数据无线承载。

可选地，第一通信设备可以是终端；

可选地，第二通信设备可以是基站；

可选地，第一通信设备中可以保存有 AI 模型；

5 可选地，第二通信设备中可以保存有 AI 模型；

可选地，第一通信设备可以基于无线承载，传输 AI 层信息，以实现第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型的管理；

可选地，第一通信设备基于无线承载，传输 AI 层信息，可以是第一通信设备发送第一通信设备的 AI 协议层所生成的 AI 层信息；

10 可选地，第一通信设备基于无线承载，传输 AI 层信息，可以是第一通信设备接收第二通信设备和/或第三通信设备发送的 AI 层信息；

可选地，第三通信设备可以是核心网设备；

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，
15 避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

可选地，所述第一通信设备的 AI 协议层为：所述第一通信设备的第一 AI 协议层和/或所述第一通信设备的第二 AI 协议层，所述第一通信设备的第一 AI 协议层为所述第一通信设备和第二通信设备之间的协议层，所述第一通信设备的第二 AI 协议
20 层为所述第一通信设备和第三通信设备之间的协议层。

可选地，第一通信设备的 AI 协议层可以包括第一通信设备和第二通信设备之间的第一 AI 协议层；

可选地，第一通信设备的 AI 协议层可以包括第一通信设备和第三通信设备之间的第一 AI 协议层；

25 可选地，所述第一通信设备的第一 AI 协议层和所述第一通信设备的第二 AI 协议层为同一个协议层，或不同的协议层。

可选地，第一通信设备的第一 AI 协议层与第一通信设备的第二协议层可以是不同的协议层；

30 可选地，第一通信设备的第一 AI 协议层与第一通信设备的第二协议层可以是同一个协议层。

可选地，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上。

可选地，图 6 是本申请实施例提供的 AI 协议层的示意图之一，如图 6 所示，以第一通信设备为终端（比如，图 6 中的 UE），第二通信设备为基站（比如，图 6 中的 NG RAN），第三通信设备为核心网设备（比如，图 6 中的 AI 实体 AI Entity）为
35

例，第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上，即对于终端，第二 AI 层位于第一 AI 层的上方。

可选地，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上。

5 可选地，图 7 是本申请实施例提供的 AI 协议层的示意图之二，如图 7 所示，以第一通信设备为终端（比如，图 7 中的 UE），第二通信设备为基站（比如，图 7 中的 gNB）为例，第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上，即对于终端，第一 AI 协议层位于 RRC 层之上。

10 可选地，可以在 RRC 协议中引入新的消息，该消息中可以携带一个容器（字节流）信元，该容器信元可以与 AI 信息相对应。

可选地，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上。

15 可选地，图 8 是本申请实施例提供的 AI 协议层的示意图之三，如图 8 所示，以第一通信设备为终端（比如，图 8 中的 UE），第二通信设备为基站（比如，图 8 中的 gNB）为例，第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上，即对于终端，第一 AI 协议层位于 PDCP 层之上。

可选地，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的 NAS 层之上。

20 可选地，图 9 是本申请实施例提供的 AI 协议层的示意图之四，如图 9 所示，以第一通信设备为终端（比如，图 9 中的 UE），第二通信设备为基站（比如，图 9 中的 NG RAN），第三通信设备为核心网设备（比如，图 9 中的 AI 实体 AI Entity）为例，第一通信设备的第二 AI 协议层位于所述第一通信设备的 NAS 层之上，即对于终端，第二 AI 协议层位于 NAS 层之上。

可选地，所述方法还包括：

25 基于无线承载的配置信息，建立所述无线承载；

其中，所述无线承载的配置信息是基于网络侧的指示确定的或预配置的或协议预定义的。

可选地，可以首先基于无线承载的配置信息，建立无线承载，进而可以通过无线承载传输 AI 信息；

30 可选地，AI 层信息到达 PDCP 层后，可以由无线承载 SRB5 进行传输；

可选地，可以配置独立的配置信息，进行独立的传输优先级控制。比如，当 AI 相关传输对时延不敏感时，可以配置 SRB5 的传输优先级低于 SRB1、2、或 3，从而减少或避免对其他控制信令传输的影响。

35 比如，在第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下，包含有 AI 层信息的 RRC 消息可以通过无线承载 SRB5 进行传输；

比如，在第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下，AI 层信息可以直接递交给对应无线承载 SRB5 承载的 PDCP 实体进行传输。

5 可选地，AI 层消息直接递交给 PDCP 层，可以跳过 RRC 的处理（比如，ASN.1 编解码）和封装，从而显著降低终端和网络的处理开销，并降低了空口资源开销。

可选地，可以采用专用控制信道(Dedicated Control Channel, DCCH)逻辑信道由无线承载 SRB5 进行 AI 层信息的传输。

可选地，配置信息可以是在接入层安全激活后由网络侧配置给第一通信设备的。

10 可选地，配置信息可以是在接入层安全激活后由第二通信设备配置给第一通信设备的；

可选地，第二通信设备在与第一通信设备交互 AI 信息之前，可以触发建立与第一通信设备的 SRB5 承载。

15 可选地，第一通信设备基于无线承载，传输人工智能 AI 层信息，包括：在接入层安全激活后，所述第一通信设备通过无线承载传输 AI 层信息。

可选地，无线承载 SRB5 可以是默认存在的，不需要建立，但是在接入层安全激活后才能传输 AI 层信息。

20 可选地，在所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上的情况下，所述第一通信设备基于无线承载，传输人工智能 AI 层信息，包括：

所述第一通信设备通过无线承载传输 AI 层信息，其中，所述 AI 层信息中包括所述第二 AI 协议层生成的 AI 层信息和第一 AI 协议层生成的 AI 层信息。

25 可选地，如图 6 所示，在所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上的情况下，第一通信设备通过无线承载传输 AI 层信息可以包括第二 AI 协议层生成的 AI 层信息和第一 AI 协议层生成的 AI 层信息；

可选地，所述第二 AI 协议层生成的 AI 层信息包含在所述第一 AI 协议层生成的 AI 层信息中。

30 可选地，以第一通信设备为终端，第二通信设备为基站，第三通信设备为核心网设备为例，对于终端，第二 AI 层位于第一 AI 层的上方，即在空口，第二 AI 协议层生成的 AI 层信息承载在第一 AI 协议层生成的 AI 层信息里发送给基站。

可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下，所述第一通信设备基于无线承载，传输人工智能 AI 层信息，包括：

35 所述第一通信设备将所述第一 AI 协议层生成的 AI 层信息放入上行 RRC 消息中的容器信元中，并发送至底层，通过无线承载发送至第二通信设备的无线资源控制

(无线资源控制, RRC) 层。

可选地, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下, 第一 AI 协议层与 RRC 层基于新引入的 RRC 消息类型 (上行 RRC 消息和下行 RRC 消息) 来关联;

5 可选地, 以第一通信设备为终端, 第二通信设备为基站, 第三通信设备为核心网设备为例, 对于终端, RRC 层可以向基站的 RRC 层发送上行 RRC 消息, 可以将终端的 AI 层收到的 AI 层信息放入上行 RRC 消息中的容器信元中, 并发送到底层, 通过无线承载发送。

10 可选地, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下, 所述第一通信设备基于无线承载, 传输人工智能 AI 层信息, 包括:

所述第一通信设备中的 RRC 层接收第二通信设备的 RRC 层发送的下行 RRC 消息;

15 所述第一通信设备将所述下行 RRC 消息中的容器信元传递至所述第一通信设备的第一 AI 协议层。

可选地, 以第一通信设备为终端, 第二通信设备为基站, 第三通信设备为核心网设备为例, 对于终端, RRC 层可以收到基站的 RRC 层发送的下行 RRC 消息, 可以将所述下行 RRC 消息中的容器信元发送到底层的 AI 层。

20 可选地, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下, 所述第一通信设备基于无线承载, 传输人工智能 AI 层信息, 包括:

所述第一通信设备将所述第一 AI 协议层生成的 AI 层信息放入所述无线承载的 PDCP SDU 中, 并发送到底层, 向第二通信设备的 PDCP 层发送。

25 可选地, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下, 第一 AI 协议层与 PDCP 层基于无线承载 (SRB5) 来关联;

可选地, 以第一通信设备为终端, 第二通信设备为基站, 第三通信设备为核心网设备为例, 对于终端, 通过 PDCP 层向基站的 PDCP 层发送上行的 AI 层信息, 可以将终端的 AI 层收到的 AI 层消息放入 SRB5 承载的 PDCP SDU 中, 并发送到底层, 向第二通信设备的 PDCP 层发送。

30 可选地, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下, 所述第一通信设备通过无线承载传输 AI 层信息, 包括:

所述第一通信设备中的 PDCP 层接收第二通信设备的 PDCP 层发送的下行的 AI 层信息;

35 所述第一通信设备将所述下行的 AI 层信息传递至所述第一通信设备的第一 AI 协议层。

可选地，以第一通信设备为终端，第二通信设备为基站，第三通信设备为核心网设备为例，对于终端，PDCP层可以收到基站PDCP层发送的下行的AI层信息，可以将下行PDCP SDU发送至终端的第一AI协议层。

5 可选地，承载在无线承载上的AI层信息使用针对RRC的完整性保护密钥进行完整性保护，使用针对RRC的加密密钥进行加密和解密。

可选的，承载在无线承载上的AI层信息使用针对RRC的完整性保护密钥KRRCin进行完整性保护，使用KRRCenc进行加密和解密。

可选地，所述AI层信息包括以下至少一项：

分段序号；

10 分段类型指示，用于指示所述AI层信息是否为最后一个分段；

容器信元。

可选地，容器信元可以包括一个模型分段。

在本申请实施例中，通过引入新的协议层即AI协议层，并通过AI协议层对应的无线承载传输AI协议层的AI层信息，以实现管理所述各通信设备中的AI模型，
15 避免AI层信息的传输受限于底层的字节限制，避免在传递AI模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

图10是本申请实施例提供的信息传输方法的流程示意图之二，如图10所示，该信息传输方法包括如下步骤：

步骤1000，第二通信设备基于无线承载，传输AI层信息；

20 其中，所述AI层信息是所述第二通信设备的AI协议层生成的或所述第二通信设备的AI协议层待接收的；

所述AI层信息用于管理所述第二通信设备中的AI模型和/或第一通信设备中的AI模型，所述第一通信设备的协议栈包括所述第二通信设备的AI协议层。

25 可选地，本申请实施例中，可以引入新的协议层，即引入AI协议层，负责模型能力管理、模型训练数据采集、模型训练、模型的分段与重组、模型维护和更新、以及模型推理。

可选地，可以引入新的无线承载类型，用于承载新的协议层的信息；

可选地，相比已有的无线承载，还可以新定义（新增）一个无线承载；

可选地，新增的无线承载可以是新增信令无线承载，比如可以称为SRB5；

30 可选地，新增的无线承载可以是新增数据无线承载。

可选地，第一通信设备可以是终端；

可选地，第二通信设备可以是基站；

可选地，第一通信设备中可以保存有AI模型；

可选地，第二通信设备中可以保存有AI模型；

35 可选地，第二通信设备可以基于无线承载，传输AI层信息，以实现第一通信

设备中的 AI 模型和/或第二通信设备中的 AI 模型的管理;

可选地, 第二通信设备基于无线承载, 传输 AI 层信息, 可以是第二通信设备发送第二通信设备的 AI 协议层所生成的 AI 层信息;

5 可选地, 第二通信设备基于无线承载, 传输 AI 层信息, 可以是第二通信设备接收第一通信设备和/或第三通信设备发送的 AI 层信息;

可选地, 第三通信设备可以是核心网设备;

10 在本申请实施例中, 通过引入新的协议层即 AI 协议层, 并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息, 以实现管理所述各通信设备中的 AI 模型, 避免 AI 层信息的传输受限于底层的字节限制, 避免在传递 AI 模型时底层需要分段和重组的缺陷, 避免影响其他控制信令的传输, 增强网络控制的实时性。

可选地, 所述第二通信设备的 AI 协议层包括: 所述第二通信设备的第一 AI 协议层和/或所述第二通信设备的第三 AI 协议层, 所述第二通信设备的第一 AI 协议层为第一通信设备和所述第二通信设备之间的协议层, 所述第二通信设备的第三 AI 协议层为所述第二通信设备和第三通信设备之间的协议层。

15 可选地, 第二通信设备的第一 AI 协议层可以包括第一通信设备和第二通信设备之间的第一 AI 协议层;

可选地, 第二通信设备的第三 AI 协议层可以包括第二通信设备和第三通信设备之间的第三 AI 协议层。

20 可选地, 第二通信设备的第一 AI 协议层和第二通信设备的第三 AI 协议层可以是同一个协议层;

可选地, 第二通信设备的第一 AI 协议层和第二通信设备的第三 AI 协议层可以是不同的协议层。

可选地, 所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 RRC 层之上。

25 可选地, 以第一通信设备为终端, 第二通信设备为基站为例, 如图 7 所示, 第二通信设备的第一 AI 协议层位于所述第二通信设备的 RRC 层之上, 即对于基站, 第一 AI 协议层位于 RRC 层之上。

可选地, 可以在 RRC 协议中引入新的消息, 该消息中可以携带一个容器 (字节流) 信元, 该容器信元可以与 AI 信息相对应。

30 可选地, 所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 PDCP 层之上。

可选地, 以第一通信设备为终端, 第二通信设备为基站为例, 如图 8 所示, 第二通信设备的第一 AI 协议层位于所述第二通信设备的 PDCP 层之上, 即对于基站, 第一 AI 协议层位于 PDCP 层之上。

35 可选地, 所述第二通信设备的第三 AI 协议层位于所述第二通信设备的 NGAP 层

之上。

可选地，图 11 是本申请实施例提供的 AI 协议层的示意图之五，如图 11 所示，以第一通信设备为终端，第二通信设备为基站（比如，图 11 中的 NG RAN），第三通信设备为核心网设备（比如，图 11 中的 AI 实体 AI entity）为例，第二通信设备的第三 AI 协议层位于所述第二通信设备的 NG 接口应用协议（NG Application Protocol，NGAP）层之上，即对于基站，第三 AI 协议层位于 NGAP 层之上。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

图 12 是本申请实施例提供的信息传输方法的流程示意图之三，如图 12 所示，该信息传输方法包括如下步骤：

步骤 1200，第三通信设备基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

可选地，本申请实施例中，可以引入新的协议层，即引入 AI 协议层，负责模型能力管理、模型训练数据采集、模型训练、模型的分段与重组、模型维护和更新、以及模型推理。

可选地，可以引入新的无线承载类型，用于承载新的协议层的信息；

可选地，相比已有的无线承载，还可以新定义（新增）一个无线承载；

可选地，新增的无线承载可以是新增信令无线承载，比如可以称为 SRB5；

可选地，新增的无线承载可以是新增数据无线承载。

可选地，第一通信设备可以是终端；

可选地，第二通信设备可以是基站；

可选地，第三通信设备可以是核心网设备；

可选地，第一通信设备中可以保存有 AI 模型；

可选地，第二通信设备中可以保存有 AI 模型；

可选地，第三通信设备可以基于无线承载，传输 AI 层信息，以实现第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型的管理；

可选地，第三通信设备基于无线承载，传输 AI 层信息，可以是第三通信设备发送第三通信设备的 AI 协议层所生成的 AI 层信息；

可选地，第三通信设备基于无线承载，传输 AI 层信息，可以是第三通信设备接收第一通信设备和/或第二通信设备发送的 AI 层信息；

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

5 可选地，所述第三通信设备的 AI 协议层为：所述第三通信设备的第二 AI 协议层和/或所述第三通信设备的第三 AI 协议层，所述第三通信设备的第二 AI 协议层为所述第一通信设备和所述第三通信设备之间的协议层，所述第三通信设备的第三 AI 协议层为所述第二通信设备和所述第三通信设备之间的协议层。

10 可选地，第三通信设备的 AI 协议层可以包括第一通信设备和所述第三通信设备之间的第二 AI 协议层；

可选地，第三通信设备的 AI 协议层可以包括第二通信设备和所述第三通信设备之间的第三 AI 协议层；

可选地，所述第三通信设备的第二 AI 协议层位于超文本传输协议（Hyper Text Transfer Protocol，HTTP）/2 层之上。

15 可选地，以第一通信设备为终端，第二通信设备为基站，第三通信设备为核心网设备为例，如图 9 和图 10 所示，第三通信设备的第二 AI 协议层位于所述第三通信设备的 HTTP/2 层之上，即对于核心网设备，第二 AI 协议层位于 HTTP/2 层之上。

可选地，所述第三通信设备的第三 AI 协议层位于 HTTP/2 层之上。

20 可选地，以第一通信设备为终端，第二通信设备为基站，第三通信设备为核心网设备为例，如图 11 所示，第三通信设备的第三 AI 协议层位于所述第三通信设备的 HTTP/2 层之上，即对于核心网设备，第三 AI 协议层位于 HTTP/2 层之上。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

25 本申请实施例提供的信息传输方法，执行主体可以为信息传输装置。本申请实施例中以信息传输装置执行信息传输方法为例，说明本申请实施例提供的信息传输装置。

30 图 13 是本申请实施例提供的信息传输装置的结构示意图之一，如图 13 所示，该装置 1300 包括：第一传输模块 1310；其中：

第一传输模块 1310 用于基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

35 所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的

AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

本申请实施例提供的信息传输装置能够实现上述各方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

可选地，所述第一通信设备的 AI 协议层为：所述第一通信设备的第一 AI 协议层和/或所述第一通信设备的第二 AI 协议层，所述第一通信设备的第一 AI 协议层为所述第一通信设备和第二通信设备之间的协议层，所述第一通信设备的第二 AI 协议层为所述第一通信设备和第三通信设备之间的协议层。

可选地，所述第一通信设备的第一 AI 协议层和所述第一通信设备的第二 AI 协议层为同一个协议层，或不同的协议层。

可选地，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上。

可选地，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上。

可选地，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上。

可选地，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的 NAS 层之上。

可选地，所述装置还包括：

第一建立模块，用于基于无线承载的配置信息，建立所述无线承载；

其中，所述无线承载的配置信息是基于网络侧的指示确定的或预配置的或协议预定义的。

可选地，所述第一传输模块 1310 用于：

在接入层安全激活后，通过无线承载传输 AI 层信息。

可选地，在所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上的情况下，所述第一传输模块 1310 用于：

通过无线承载传输 AI 层信息，其中，所述 AI 层信息中包括所述第二 AI 协议层生成的 AI 层信息和第一 AI 协议层生成的 AI 层信息。

可选地，所述第二 AI 协议层生成的 AI 层信息包含在所述第一 AI 协议层生成的 AI 层信息中。

可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下，所述第一传输模块 1310 用于：

将所述第一 AI 协议层生成的 AI 层信息放入上行 RRC 消息中的容器信元中，并发送到底层，通过无线承载发送至第二通信设备的 RRC 层。

在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下，所述第一传输模块 1310 用于：

- 5 所述第一通信设备中的 RRC 层接收第二通信设备的 RRC 层发送的下行 RRC 消息；

将所述下行 RRC 消息中的容器信元传递至所述第一通信设备的第一 AI 协议层。

- 10 可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下，所述第一传输模块 1310 用于：

将所述第一 AI 协议层生成的 AI 层信息放入所述无线承载的 PDCP SDU 中，并发送到底层，向第二通信设备的 PDCP 层发送。

可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下，所述第一传输模块 1310 用于：

- 15 所述第一通信设备中的 PDCP 层接收第二通信设备的 PDCP 层发送的下行的 AI 层信息；

将所述下行的 AI 层信息传递至所述第一通信设备的第一 AI 协议层。

可选地，承载在无线承载上的 AI 层信息使用针对 RRC 的完整性保护密钥进行完整性保护，使用针对 RRC 的加密密钥进行加密和解密。

- 20 可选地，所述 AI 层信息包括以下至少一项：

分段序号；

分段类型指示，用于指示所述 AI 层信息是否为最后一个分段；

容器信元。

- 25 在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

- 30 本申请实施例中的信息传输装置可以是电子设备，例如具有操作系统的电子设备，也可以是电子设备中的部件，例如集成电路或芯片。该电子设备可以是终端，也可以为除终端之外的其他设备。示例性的，终端可以包括但不限于上述所列举的终端 11 的类型，其他设备可以为服务器、网络附属存储器（Network Attached Storage，NAS）等，本申请实施例不作具体限定。

本申请实施例提供的信息传输装置能够实现图 5 至图 9 的方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

- 35 图 14 是本申请实施例提供的信息传输装置的结构示意图之二，如图 14 所示，

该装置 1400 包括：第二传输模块 1410；其中：

第二传输模块 1410 用于基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

5 所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

本申请实施例提供的信息传输装置能够实现上述各方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

10 在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

15 可选地，所述第二通信设备的 AI 协议层包括：所述第二通信设备的第一 AI 协议层和/或所述第二通信设备的第三 AI 协议层，所述第二通信设备的第一 AI 协议层为第一通信设备和所述第二通信设备之间的协议层，所述第二通信设备的第三 AI 协议层为所述第二通信设备和第三通信设备之间的协议层。

可选地，所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 RRC 层之上。

20 可选地，所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 PDCP 层之上。

可选地，所述第二通信设备的第三 AI 协议层位于所述第二通信设备的 NGAP 层之上。

25 在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

30 本申请实施例中的信息传输装置可以是电子设备，例如具有操作系统的电子设备，也可以是电子设备中的部件，例如集成电路或芯片。该电子设备可以是终端，也可以为除终端之外的其他设备。示例性的，终端可以包括但不限于上述所列举的终端 11 的类型，其他设备可以为服务器、网络附属存储器（Network Attached Storage, NAS）等，本申请实施例不作具体限定。

本申请实施例提供的信息传输装置能够实现图 10 至图 11 的方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

35 图 15 是本申请实施例提供的信息传输装置的结构示意图之三，如图 15 所示，该装置 1500 包括第三传输模块 1510；其中：

第三传输模块 1510 用于基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

本申请实施例提供的信息传输装置能够实现上述各方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

可选地，所述第三通信设备的 AI 协议层为：所述第三通信设备的第二 AI 协议层和/或所述第三通信设备的第三 AI 协议层，所述第三通信设备的第二 AI 协议层为所述第一通信设备和所述第三通信设备之间的协议层，所述第三通信设备的第三 AI 协议层为所述第二通信设备和所述第三通信设备之间的协议层。

可选地，所述第三通信设备的第二 AI 协议层位于 HTTP/2 层之上。

可选地，所述第三通信设备的第三 AI 协议层位于 HTTP/2 层之上。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

本申请实施例中的信息传输装置可以是电子设备，例如具有操作系统的电子设备，也可以是电子设备中的部件，例如集成电路或芯片。该电子设备可以是终端，也可以为除终端之外的其他设备。示例性的，终端可以包括但不限于上述所列举的终端 11 的类型，其他设备可以为服务器、网络附属存储器（Network Attached Storage，NAS）等，本申请实施例不作具体限定。

本申请实施例提供的信息传输装置能够实现图 12 的方法实施例实现的各个过程，并达到相同的技术效果，为避免重复，这里不再赘述。

可选的，图 16 是本申请实施例提供的通信设备的结构示意图，如图 16 所示，本申请实施例还提供一种通信设备 1600，包括处理器 1601 和存储器 1602，存储器 1602 上存储有可在所述处理器 1601 上运行的程序或指令，例如，该通信设备 1600 为第一通信设备时，该程序或指令被处理器 1601 执行时实现上述第一通信设备对应的信息传输方法实施例的各个步骤，且能达到相同的技术效果。该通信设备 1600 为第二通信设备时，该程序或指令被处理器 1601 执行时实现上述第二通信设备对应的信息传输方法实施例的各个步骤，且能达到相同的技术效果。该通信设备 1600 为第

三通信设备时，该程序或指令被处理器 1601 执行时实现上述第三通信设备对应的信息传输方法实施例的各个步骤，且能达到相同的技术效果。为避免重复，这里不再赘述。

本申请实施例还提供一种第一通信设备，包括处理器及通信接口，其中，所述通信接口用于：

基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

该第一通信设备实施例与上述第一通信设备侧方法实施例对应，上述方法实施例的各个实施过程和实现方式均可适用于该第一通信设备实施例中，且能达到相同的技术效果。具体地，图 17 为实现本申请实施例的一种第一通信设备的硬件结构示意图。

该终端 1700 包括但不限于：射频单元 1701、网络模块 1702、音频输出单元 1703、输入单元 1704、传感器 1705、显示单元 1706、用户输入单元 1707、接口单元 1708、存储器 1709 以及处理器 1710 等中的至少部分部件。

本领域技术人员可以理解，终端 1700 还可以包括给各个部件供电的电源（比如电池），电源可以通过电源管理系统与处理器 1710 逻辑相连，从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。图 17 中示出的终端结构并不构成对终端的限定，终端可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置，在此不再赘述。

应理解的是，本申请实施例中，输入单元 1704 可以包括图形处理单元（Graphics Processing Unit，GPU）17041 和麦克风 17042，图形处理器 17041 对在视频捕获模式或图像捕获模式中由图像捕获装置（如摄像头）获得的静态图片或视频的图像数据进行处理。显示单元 1706 可包括显示面板 17061，可以采用液晶显示器、有机发光二极管等形式来配置显示面板 17061。用户输入单元 1707 包括触控面板 17071 以及其他输入设备 17072 中的至少一种。触控面板 17071，也称为触摸屏。触控面板 17071 可包括触摸检测装置和触摸控制器两个部分。其他输入设备 17072 可以包括但不限于物理键盘、功能键（比如音量控制按键、开关按键等）、轨迹球、鼠标、操作杆，在此不再赘述。

本申请实施例中，射频单元 1701 接收来自网络侧设备的下行数据后，可以传输给处理器 1710 进行处理；另外，射频单元 1701 可以向网络侧设备发送上行数据。通常，射频单元 1701 包括但不限于天线、放大器、收发信机、耦合器、低噪声放大器、双工器等。

存储器 1709 可用于存储软件程序或指令以及各种数据。存储器 1709 可主要包括存储程序或指令的第一存储区和存储数据的第二存储区，其中，第一存储区可存储操作系统、至少一个功能所需的应用程序或指令（比如声音播放功能、图像播放功能等）等。此外，存储器 1709 可以包括易失性存储器或非易失性存储器，或者，
5 存储器 1709 可以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)、静态随机存取存储器(Static RAM, SRAM)、动
10 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, DDRSDRAM)、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)和直接内存总线随机存取存储器(Direct Rambus RAM, DRRAM)。本申请实施例中的存储器
15 1709 包括但不限于这些和任意其它适合类型的存储器。

处理器 1710 可包括一个或多个处理单元；可选的，处理器 1710 集成应用处理器和调制解调处理器，其中，应用处理器主要处理涉及操作系统、用户界面和应用程序等的操作，调制解调处理器主要处理无线通信信号，如基带处理器。可以理解的是，上述调制解调处理器也可以不集成到处理器 1710 中。

20 其中，处理器 1710 用于：

基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

25 所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

30 可选地，所述第一通信设备的 AI 协议层为：所述第一通信设备的第一 AI 协议层和/或所述第一通信设备的第二 AI 协议层，所述第一通信设备的第一 AI 协议层为所述第一通信设备和第二通信设备之间的协议层，所述第一通信设备的第二 AI 协议层为所述第一通信设备和第三通信设备之间的协议层。

35 可选地，所述第一通信设备的第一 AI 协议层和所述第一通信设备的第二 AI 协议层为同一个协议层，或不同的协议层。

可选地，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上。

可选地，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上。

5 可选地，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上。

可选地，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的 NAS 层之上。

可选地，处理器 1710 用于：

10 基于无线承载的配置信息，建立所述无线承载；

其中，所述无线承载的配置信息是基于网络侧的指示确定的或预配置的或协议预定义的。

可选地，处理器 1710 用于：

在接入层安全激活后，通过无线承载传输 AI 层信息。

15 可选地，在所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上的情况下，处理器 1710 用于：

通过无线承载传输 AI 层信息，其中，所述 AI 层信息中包括所述第二 AI 协议层生成的 AI 层信息和第一 AI 协议层生成的 AI 层信息。

20 可选地，所述第二 AI 协议层生成的 AI 层信息包含在所述第一 AI 协议层生成的 AI 层信息中。

可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下，处理器 1710 用于：

将所述第一 AI 协议层生成的 AI 层信息放入上行 RRC 消息中的容器信元中，并发送到底层，通过无线承载发送至第二通信设备的 RRC 层。

25 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下，处理器 1710 用于：

所述第一通信设备中的 RRC 层接收第二通信设备的 RRC 层发送的下行 RRC 消息；

30 将所述下行 RRC 消息中的容器信元传递至所述第一通信设备的第一 AI 协议层。

可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下，处理器 1710 用于：

将所述第一 AI 协议层生成的 AI 层信息放入所述无线承载的 PDCP SDU 中，并发送到底层，向第二通信设备的 PDCP 层发送。

35 可选地，在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP

层之上的情况下，处理器 1710 用于：

所述第一通信设备中的 PDCP 层接收第二通信设备的 PDCP 层发送的下行的 AI 层信息；

将所述下行的 AI 层信息传递至所述第一通信设备的第一 AI 协议层。

5 可选地，承载在无线承载上的 AI 层信息使用针对 RRC 的完整性保护密钥进行完整性保护，使用针对 RRC 的加密密钥进行加密和解密。

可选地，所述 AI 层信息包括以下至少一项：

分段序号；

分段类型指示，用于指示所述 AI 层信息是否为最后一个分段；

10 容器信元。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

15 本申请实施例还提供一种第二通信设备，包括处理器及通信接口，其中，所述通信接口用于：

基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是所述第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

20 所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

该第二通信设备实施例与上述第二通信设备方法实施例对应，上述方法实施例的各个实施过程和实现方式均可适用于该第二通信设备实施例中，且能达到相同的技术效果。

25 具体地，本申请实施例还提供了一种第二通信设备。图 18 为实现本申请实施例的一种第二通信设备的硬件结构示意图，如图 18 所示，该网络侧设备 1800 包括：天线 1801、射频装置 1802、基带装置 1803、处理器 1804 和存储器 1805。天线 1801 与射频装置 1802 连接。在上行方向上，射频装置 1802 通过天线 1801 接收信息，将接收的信息发送给基带装置 1803 进行处理。在下行方向上，基带装置 1803 对要发送的信息进行处理，并发送给射频装置 1802，射频装置 1802 对收到的信息进行处理后经过天线 1801 发送出去。

以上实施例中网络侧设备执行的方法可以在基带装置 1803 中实现，该基带装置 1803 包括基带处理器。

35 基带装置 1803 例如可以包括至少一个基带板，该基带板上设置有多个芯片，如图 18 所示，其中一个芯片例如为基带处理器，通过总线接口与存储器 1805 连接，

以调用存储器 1805 中的程序，执行以上方法实施例中所示的网络设备操作。

该网络侧设备还可以包括网络接口 1806，该接口例如为通用公共无线接口（common public radio interface，CPRI）。

具体地，本发明实施例的网络侧设备 1800 还包括：存储在存储器 1805 上并可
5 在处理器 1804 上运行的指令或程序，处理器 1804 调用存储器 1805 中的指令或程序
执行图 14 所示各模块执行的方法，并达到相同的技术效果，为避免重复，故不在此
赘述。

其中，处理器 1804 用于：

基于无线承载，传输 AI 层信息；

10 其中，所述 AI 层信息是第二通信设备的 AI 协议层生成的或所述第二通信设备
的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的
AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应
15 的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，
避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段
和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

可选地，所述第二通信设备的 AI 协议层包括：所述第二通信设备的第一 AI 协
议层和/或所述第二通信设备的第三 AI 协议层，所述第二通信设备的第一 AI 协议层
20 为第一通信设备和所述第二通信设备之间的协议层，所述第二通信设备的第三 AI 协
议层为所述第二通信设备和第三通信设备之间的协议层。

可选地，所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 RRC 层
之上。

可选地，所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 PDCP 层
25 之上。

可选地，所述第二通信设备的第三 AI 协议层位于所述第二通信设备的 NGAP 层
之上。

在本申请实施例中，通过引入新的协议层即 AI 协议层，并通过 AI 协议层对应
的无线承载传输 AI 协议层的 AI 层信息，以实现管理所述各通信设备中的 AI 模型，
30 避免 AI 层信息的传输受限于底层的字节限制，避免在传递 AI 模型时底层需要分段
和重组的缺陷，避免影响其他控制信令的传输，增强网络控制的实时性。

具体地，本申请实施例还提供了一种第三通信设备，包括处理器及通信接口，
其中，所述通信接口用于：

基于无线承载，传输 AI 层信息；

35 其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备

的 AI 协议层待接收的;

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型, 所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

图 19 为实现本申请实施例的一种第三通信设备的硬件结构示意图, 如图 19 所示, 该网络侧设备 1900 包括: 处理器 1901、网络接口 1902 和存储器 1903。其中, 网络接口 1902 例如为通用公共无线接口 (common public radio interface, CPRI)。

具体地, 本发明实施例的网络侧设备 1900 还包括: 存储在存储器 1903 上并可在处理器 1901 上运行的指令或程序, 处理器 1901 调用存储器 1903 中的指令或程序执行图 15 所示各模块执行的方法, 并达到相同的技术效果, 为避免重复, 故不在此赘述。

其中, 处理器 1901 用于:

基于无线承载, 传输 AI 层信息;

其中, 所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的;

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型, 所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

在本申请实施例中, 通过引入新的协议层即 AI 协议层, 并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息, 以实现管理所述各通信设备中的 AI 模型, 避免 AI 层信息的传输受限于底层的字节限制, 避免在传递 AI 模型时底层需要分段和重组的缺陷, 避免影响其他控制信令的传输, 增强网络控制的实时性。

可选地, 所述第三通信设备的 AI 协议层为: 所述第三通信设备的第二 AI 协议层和/或所述第三通信设备的第三 AI 协议层, 所述第三通信设备的第二 AI 协议层为所述第一通信设备和所述第三通信设备之间的协议层, 所述第三通信设备的第三 AI 协议层为所述第二通信设备和所述第三通信设备之间的协议层。

可选地, 所述第三通信设备的第二 AI 协议层位于 HTTP/2 层之上。

可选地, 所述第三通信设备的第三 AI 协议层位于 HTTP/2 层之上。

在本申请实施例中, 通过引入新的协议层即 AI 协议层, 并通过 AI 协议层对应的无线承载传输 AI 协议层的 AI 层信息, 以实现管理所述各通信设备中的 AI 模型, 避免 AI 层信息的传输受限于底层的字节限制, 避免在传递 AI 模型时底层需要分段和重组的缺陷, 避免影响其他控制信令的传输, 增强网络控制的实时性。

本申请实施例还提供一种可读存储介质, 所述可读存储介质上存储有程序或指令, 该程序或指令被处理器执行时实现上述信息传输方法实施例的各个过程, 且能达到相同的技术效果, 为避免重复, 这里不再赘述。

其中, 所述处理器为上述实施例中所述的终端中的处理器。所述可读存储介质, 包括计算机可读存储介质, 如计算机只读存储器 ROM、随机存取存储器

RAM、磁碟或者光盘等。

本申请实施例另提供了一种芯片，所述芯片包括处理器和通信接口，所述通信接口和所述处理器耦合，所述处理器用于运行程序或指令，实现上述信息传输方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再赘述。

5 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。

本申请实施例另提供了一种计算机程序/程序产品，所述计算机程序/程序产品被存储在存储介质中，所述计算机程序/程序产品被至少一个处理器执行以实现上述信息传输方法实施例的各个过程，且能达到相同的技术效果，为避免重复，这里不再
10 赘述。

本申请实施例还提供了一种信息传输系统，包括：第一通信设备、第二通信设备、和第三通信设备，所述第一通信设备可用于执行第一通信设备对应的信息传输方法的步骤，所述第二通信设备可用于执行第二通信设备对应的信息传输方法的步骤，所述第三通信设备可用于执行第三通信设备对应的信息传输方法的步骤。

15 需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者装置中还存
20 在另外的相同要素。此外，需要指出的是，本申请实施方式中的方法和装置的范围不限按示出或讨论的顺序来执行功能，还可包括根据所涉及的功能按基本同时的方式或按相反的顺序来执行功能，例如，可以按不同于所描述的次序来执行所描述的方法，并且还可以添加、省去、或组合各种步骤。另外，参照某些示例所描述的特征可在其他示例中被组合。

25 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。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可以以计算机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（如 ROM/RAM、磁碟、光盘）中，包括若干指令用以
30 使得一台终端（可以是手机，计算机，服务器，空调器，或者网络设备等）执行本申请各个实施例所述的方法。

上面结合附图对本申请的实施例进行了描述，但是本申请并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本申请的启示下，在不脱离本申请宗旨和权利要求所保护的范围情况
35 下，还可做出很多形式，均属于本申请的保护之内。

权 利 要 求 书

1. 一种信息传输方法，包括：

第一通信设备基于无线承载，传输人工智能 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

2. 根据权利要求 1 所述的信息传输方法，其中，所述第一通信设备的 AI 协议层为：所述第一通信设备的第一 AI 协议层和/或所述第一通信设备的第二 AI 协议层，所述第一通信设备的第一 AI 协议层为所述第一通信设备和第二通信设备之间的协议层，所述第一通信设备的第二 AI 协议层为所述第一通信设备和第三通信设备之间的协议层。

3. 根据权利要求 2 所述的信息传输方法，其中，所述第一通信设备的第一 AI 协议层和所述第一通信设备的第二 AI 协议层为同一个协议层，或不同的协议层。

4. 根据权利要求 2 或 3 所述的信息传输方法，其中，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上。

5. 根据权利要求 2-4 任一项所述的信息传输方法，其中，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的无线资源控制 RRC 层之上。

6. 根据权利要求 2-4 任一项所述的信息传输方法，其中，所述第一通信设备的第一 AI 协议层位于所述第一通信设备的分组数据汇聚协议 PDCP 层之上。

7. 根据权利要求 2 或 3 或 5 或 6 所述的信息传输方法，其中，所述第一通信设备的第二 AI 协议层位于所述第一通信设备的非接入层 NAS 层之上。

8. 根据权利要求 1-7 任一项所述的信息传输方法，其中，所述方法还包括：

所述第一通信设备基于无线承载的配置信息，建立所述无线承载；

其中，所述无线承载的配置信息是基于网络侧的指示确定的或预配置的或协议预定义的。

9. 根据权利要求 1-7 任一项所述的信息传输方法，其中，所述第一通信设备基于无线承载，传输人工智能 AI 层信息，包括：

在接入层安全激活后，所述第一通信设备通过所述无线承载传输 AI 层信息。

10. 根据权利要求 8 或 9 所述的信息传输方法，其中，在所述第一通信设备的第二 AI 协议层位于所述第一通信设备的第一 AI 协议层之上的情况下，所述第一通信设备基于无线承载，传输人工智能 AI 层信息，包括：

所述第一通信设备通过无线承载传输 AI 层信息，其中，所述 AI 层信息中包括所述第二 AI 协议层生成的 AI 层信息和第一 AI 协议层生成的 AI 层信息。

11. 根据权利要求 10 所述的信息传输方法, 其中, 所述第二 AI 协议层生成的 AI 层信息包含在所述第一 AI 协议层生成的 AI 层信息中。

12. 根据权利要求 8-11 任一项所述的信息传输方法, 其中, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下, 所述第一通信设备基于无线承载, 传输人工智能 AI 层信息, 包括:

所述第一通信设备将所述第一 AI 协议层生成的 AI 层信息放入上行 RRC 消息中的容器信元中, 并发送到底层, 通过无线承载发送至第二通信设备的 RRC 层。

13. 根据权利要求 8-11 任一项所述的信息传输方法, 其中, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 RRC 层之上的情况下, 所述第一通信设备基于无线承载, 传输人工智能 AI 层信息, 包括:

所述第一通信设备中的 RRC 层接收第二通信设备的 RRC 层发送的下行 RRC 消息;

所述第一通信设备将所述下行 RRC 消息中的容器信元传递至所述第一通信设备的第一 AI 协议层。

14. 根据权利要求 8-11 任一项所述的信息传输方法, 其中, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下, 所述第一通信设备基于无线承载, 传输人工智能 AI 层信息, 包括:

所述第一通信设备将所述第一 AI 协议层生成的 AI 层信息放入所述无线承载的 PDCP SDU 中, 并发送到底层, 向第二通信设备的 PDCP 层发送。

15. 根据权利要求 8-11 任一项所述的信息传输方法, 其中, 在所述第一通信设备的第一 AI 协议层位于所述第一通信设备的 PDCP 层之上的情况下, 所述第一通信设备基于无线承载, 传输人工智能 AI 层信息, 包括:

所述第一通信设备中的 PDCP 层接收第二通信设备的 PDCP 层发送的下行的 AI 层信息;

所述第一通信设备将所述下行的 AI 层信息传递至所述第一通信设备的第一 AI 协议层。

16. 根据权利要求 8-15 任一项所述的信息传输方法, 其中, 承载在无线承载上的 AI 层信息使用针对 RRC 的完整性保护密钥进行完整性保护, 使用针对 RRC 的加密密钥进行加密和解密。

17. 根据权利要求 1-16 任一项所述的信息传输方法, 其中, 所述 AI 层信息包括以下至少一项:

分段序号;

分段类型指示, 用于指示所述 AI 层信息是否为最后一个分段;

容器信元。

18. 一种信息传输方法, 包括:

第二通信设备基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是所述第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

19. 根据权利要求 18 所述的信息传输方法，其中，所述第二通信设备的 AI 协议层包括：所述第二通信设备的第一 AI 协议层和/或所述第二通信设备的第三 AI 协议层，所述第二通信设备的第一 AI 协议层为第一通信设备和所述第二通信设备之间的协议层，所述第二通信设备的第三 AI 协议层为所述第二通信设备和第三通信设备之间的协议层。

20. 根据权利要求 19 所述的信息传输方法，其中，所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 RRC 层之上。

21. 根据权利要求 19 所述的信息传输方法，其中，所述第二通信设备的第一 AI 协议层位于所述第二通信设备的 PDCP 层之上。

22. 根据权利要求 19 所述的信息传输方法，其中，所述第二通信设备的第三 AI 协议层位于所述第二通信设备的 NG 接口应用协议 NGAP 层之上。

23. 一种信息传输方法，包括：

第三通信设备基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

24. 根据权利要求 23 所述的信息传输方法，其中，所述第三通信设备的 AI 协议层为：所述第三通信设备的第二 AI 协议层和/或所述第三通信设备的第三 AI 协议层，所述第三通信设备的第二 AI 协议层为所述第一通信设备和所述第三通信设备之间的协议层，所述第三通信设备的第三 AI 协议层为所述第二通信设备和所述第三通信设备之间的协议层。

25. 根据权利要求 24 所述的信息传输方法，其中，所述第三通信设备的第二 AI 协议层位于超文本传输协议 HTTP/2 层之上。

26. 根据权利要求 24 所述的信息传输方法，其中，所述第三通信设备的第三 AI 协议层位于 HTTP/2 层之上。

27. 一种信息传输装置，包括：

第一传输模块，用于基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第一通信设备的 AI 协议层生成的或所述第一通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第一通信设备中的 AI 模型和/或第二通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第一通信设备的 AI 协议层。

28. 一种信息传输装置，包括：

第二传输模块，用于基于无线承载，传输 AI 层信息；

5 其中，所述 AI 层信息是第二通信设备的 AI 协议层生成的或所述第二通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理所述第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第一通信设备的协议栈包括所述第二通信设备的 AI 协议层。

29. 一种信息传输装置，包括：

10 第三传输模块，用于基于无线承载，传输 AI 层信息；

其中，所述 AI 层信息是第三通信设备的 AI 协议层生成的或所述第三通信设备的 AI 协议层待接收的；

所述 AI 层信息用于管理第二通信设备中的 AI 模型和/或第一通信设备中的 AI 模型，所述第三通信设备的协议栈包括所述第三通信设备的 AI 协议层。

15 30. 一种第一通信设备，包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如权利要求 1-17 任一项所述的信息传输方法。

31. 一种第二通信设备，包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如权利要求 18-22 任
20 一项所述的信息传输方法。

32. 一种第三通信设备，包括处理器和存储器，所述存储器存储可在所述处理器上运行的程序或指令，所述程序或指令被所述处理器执行时实现如权利要求 23-26 任
一项所述的信息传输方法。

33. 一种可读存储介质，所述可读存储介质上存储程序或指令，所述程序或指令
25 被处理器执行时实现如权利要求 1 至 26 任一项所述的信息传输方法。

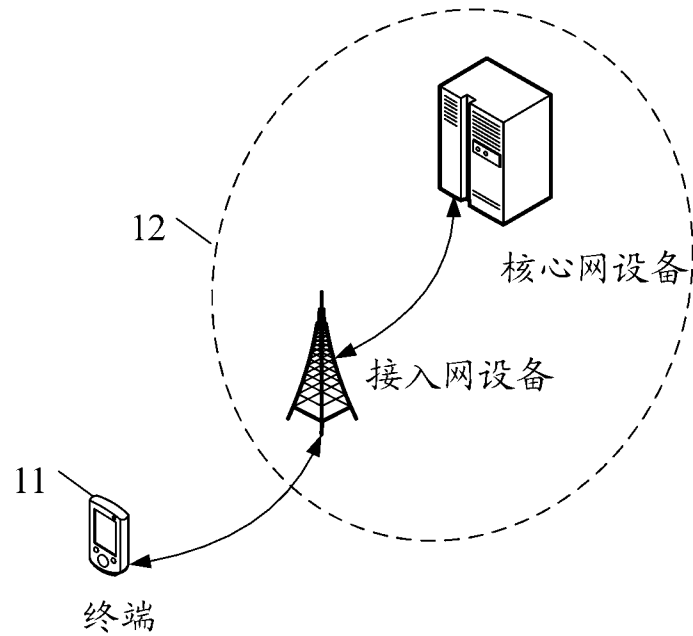


图 1

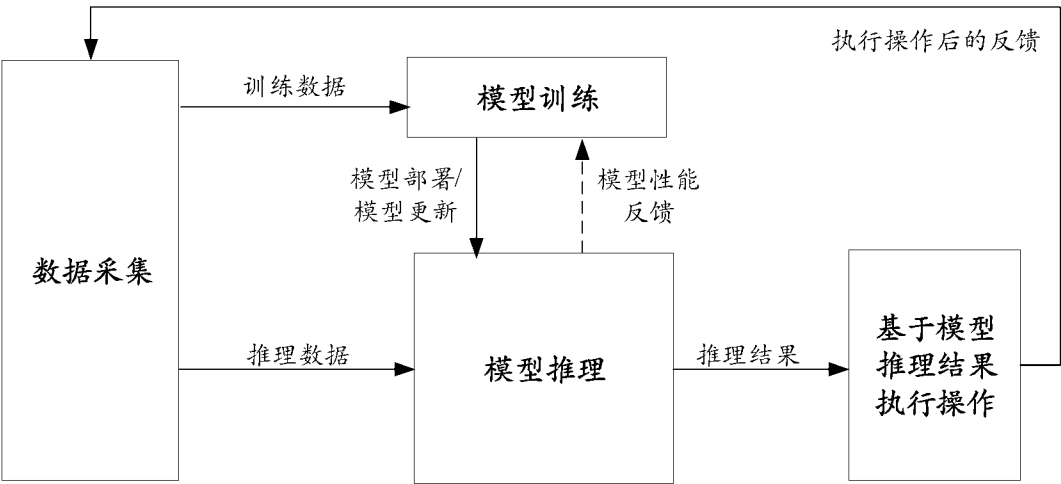


图 2

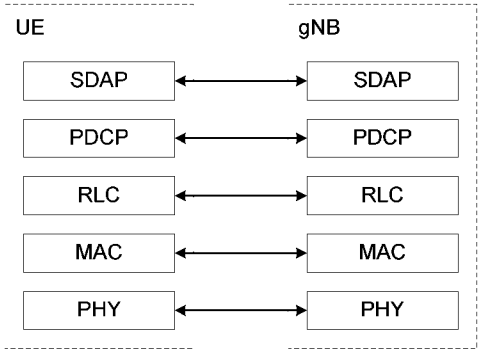


图 3

2/6

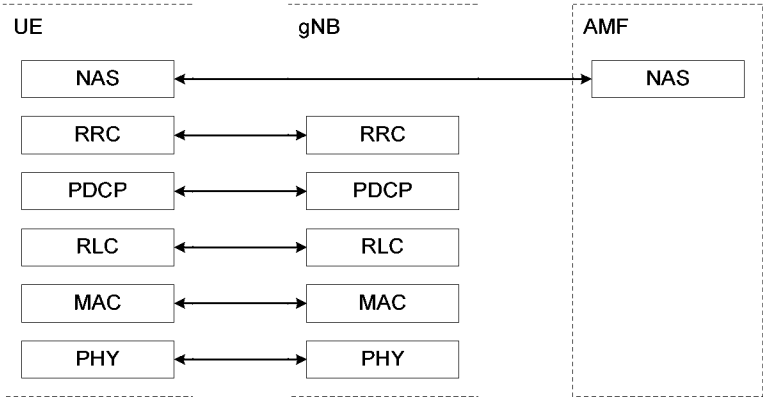


图 4

第一通信设备基于无线承载，传输AI层信息；
其中，所述AI层信息是第一通信设备的AI协议层生成的或所述第一通信设备的AI协议层待接收的；
所述AI层信息用于管理所述第一通信设备中的AI模型和/或第二通信设备中的AI模型，所述第一通信设备的协议栈包括所述第一通信设备的AI协议层

500

图 5

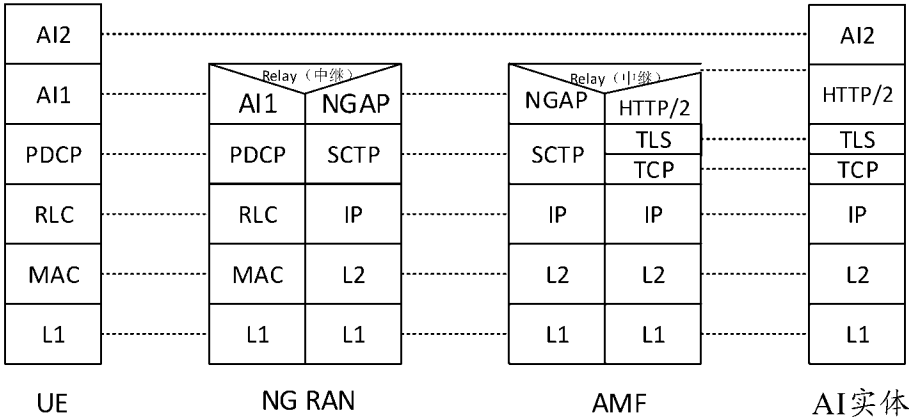


图 6

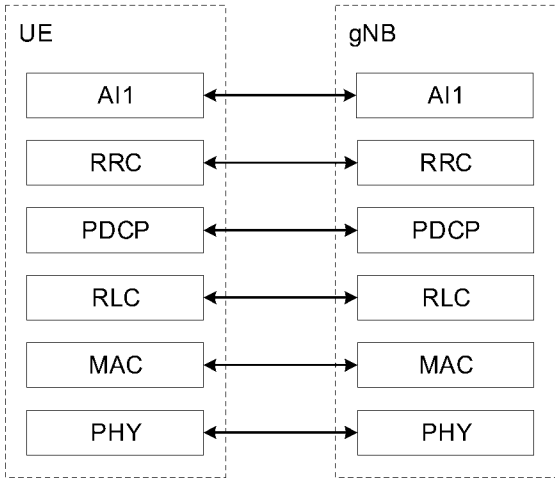


图 7

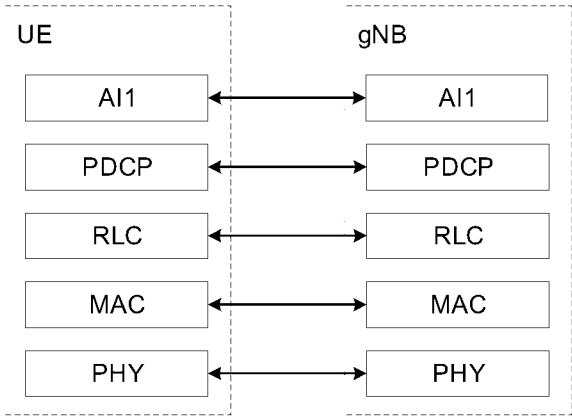


图 8

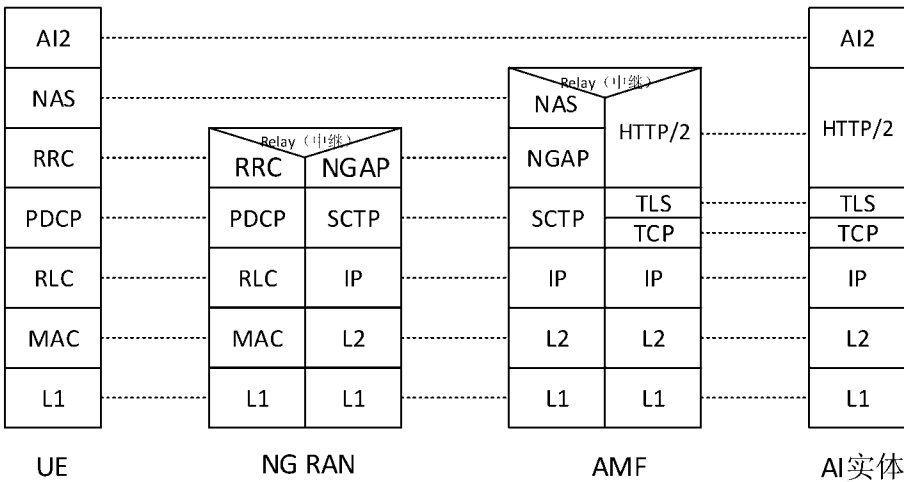


图 9

第二通信设备基于无线承载，传输AI层信息；
其中，所述AI层信息是所述第二通信设备的AI协议层生成的或所述第二通信设备的AI协议层待接收的；
所述AI层信息用于管理所述第二通信设备中的AI模型和/或第一通信设备中的AI模型，所述第一通信设备的协议栈包括所述第二通信设备的AI协议层

1000

图 10

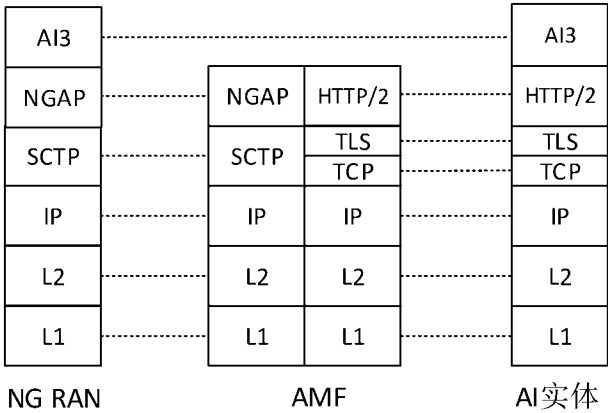


图 11

第三通信设备基于无线承载，传输AI层信息；
其中，所述AI层信息是第三通信设备的AI协议层生成的或所述第三通信设备的AI协议层待接收的；
所述AI层信息用于管理第二通信设备中的AI模型和/或第一通信设备中的AI模型，所述第三通信设备的协议栈包括所述第三通信设备的AI协议层

1200

图 12

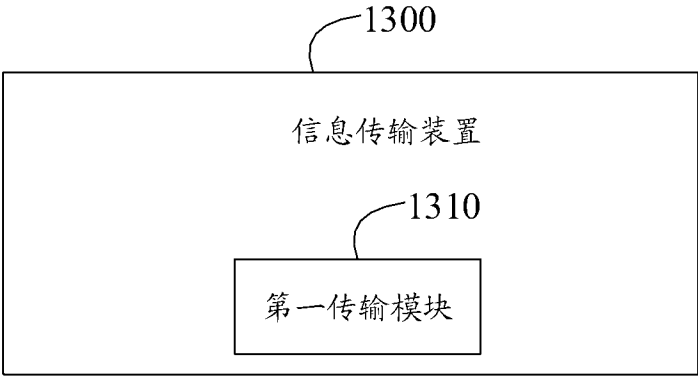


图 13

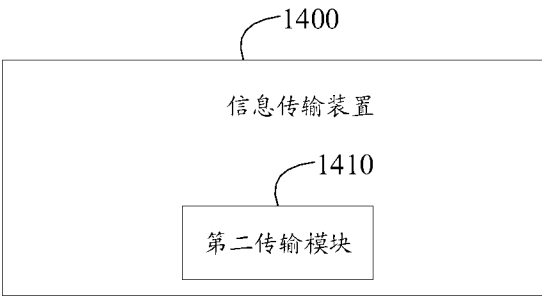


图 14

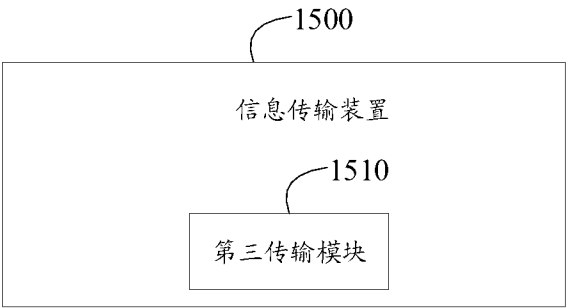


图 15

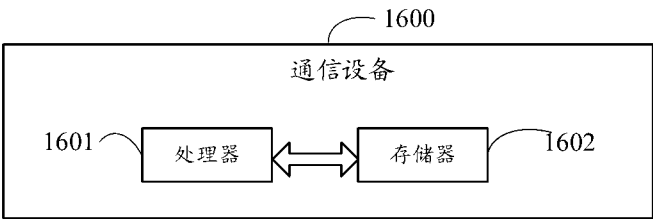


图 16

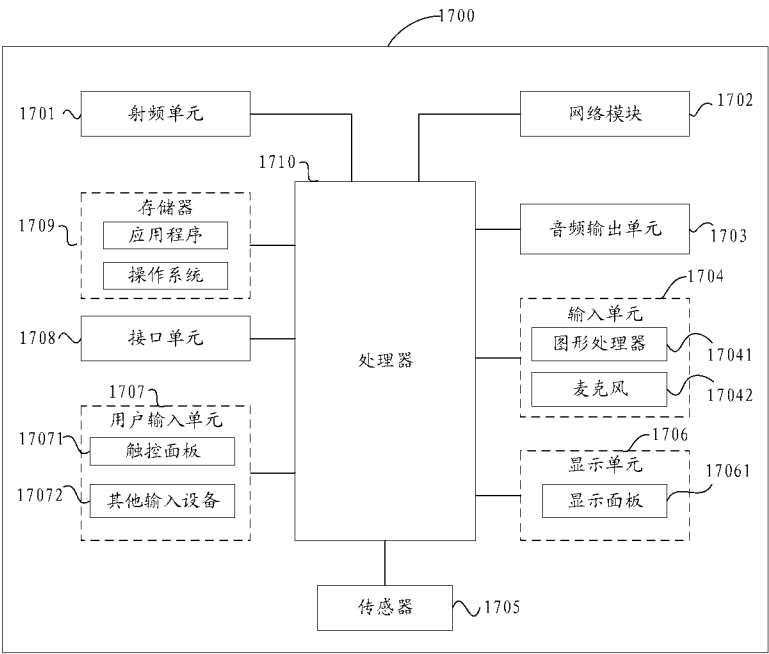


图 17

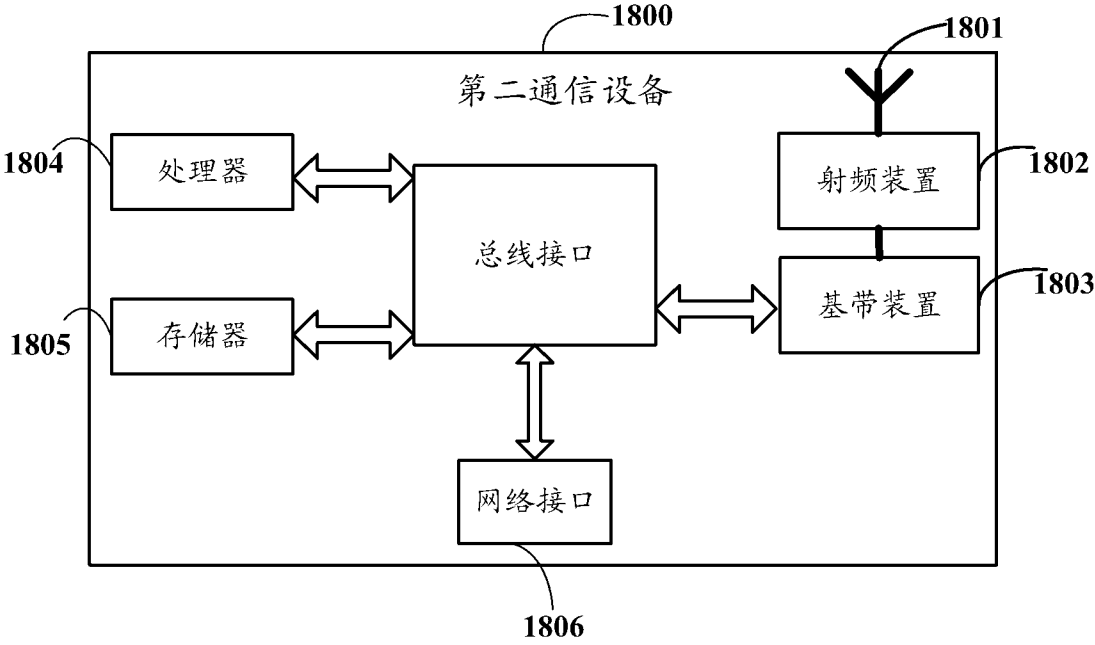


图 18

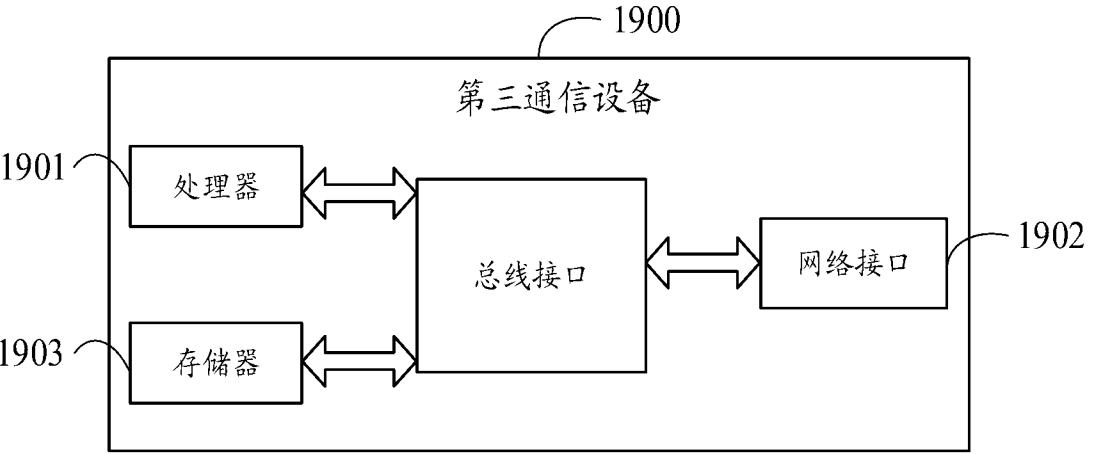


图 19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/105066

A. CLASSIFICATION OF SUBJECT MATTER

H04W80/08(2009.01)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)

CNTXT, ENTXTC, DWPI, CNKI, IEEE, 3GPP: 人工智能, AI, AIML, AI/ML, 层, 协议层, 无线承载, layer, bearer, 终端, UE, 接入网, RAN, gNB, 核心网

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 113365287 A (HUAWEI TECHNOLOGIES CO., LTD.) 07 September 2021 (2021-09-07) description, paragraphs [0125]-[0301], and figures 3-17	1-33
A	WO 2021134597 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 08 July 2021 (2021-07-08) entire document	1-33
A	WO 2022050432 A1 (LG ELECTRONICS INC.) 10 March 2022 (2022-03-10) entire document	1-33
A	WO 2022078276 A1 (VIVO COMMUNICATION TECHNOLOGY CO., LTD.) 21 April 2022 (2022-04-21) entire document	1-33
A	FUTUREWEI. "ML/AI for Wireless" 3GPP TSG RAN Rel-18 Workshop RWS-210038, 02 July 2021 (2021-07-02), entire document	1-33



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

12 September 2023

Date of mailing of the international search report

25 September 2023

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/CN2023/105066

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113365287	A	07 September 2021	WO	2021175103	A1	10 September 2021
				EP	4090064	A1	16 November 2022
				US	2022408282	A1	22 December 2022
WO	2021134597	A1	08 July 2021	US	2022330072	A1	13 October 2022
				EP	4075843	A1	19 October 2022
				CN	114868412	A	05 August 2022
WO	2022050432	A1	10 March 2022	KR	20230066020	A	12 May 2023
WO	2022078276	A1	21 April 2022	US	2023254711	A1	10 August 2023
				EP	4231693	A1	23 August 2023
				CN	114363921	A	15 April 2022

国际检索报告

国际申请号

PCT/CN2023/105066

A. 主题的分类

H04W80/08 (2009.01) 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))

CNTXT, ENTXTC, DWPI, CNKI, IEEE, 3GPP: 人工智能, AI, AIML, AI/ML, 层, 协议层, 无线承载, layer, bearer, 终端, UE, 接入网, RAN, gNB, 核心网

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 113365287 A (华为技术有限公司) 2021年9月7日 (2021 - 09 - 07)
说明书第[0125]-[0301]段, 图3-17

1-33

A

WO 2021134597 A1 (华为技术有限公司) 2021年7月8日 (2021 - 07 - 08)
全文

1-33

A

WO 2022050432 A1 (LG ELECTRONICS INC.) 2022年3月10日 (2022 - 03 - 10)
全文

1-33

A

WO 2022078276 A1 (维沃移动通信有限公司) 2022年4月21日 (2022 - 04 - 21)
全文

1-33

A

FUTUREWEI. "ML/AI for Wireless"
3GPP TSG RAN Rel-18 Workshop RWS-210038, 2021年7月2日 (2021 - 07 - 02),
全文

1-33

☐ 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

2023年9月12日

国际检索报告邮寄日期

2023年9月25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

李文娟

电话号码 (+86) 010-53961585

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/105066

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	113365287	A	2021年9月7日	WO	2021175103	A1	2021年9月10日
				EP	4090064	A1	2022年11月16日
				US	2022408282	A1	2022年12月22日
WO	2021134597	A1	2021年7月8日	US	2022330072	A1	2022年10月13日
				EP	4075843	A1	2022年10月19日
				CN	114868412	A	2022年8月5日
WO	2022050432	A1	2022年3月10日	KR	20230066020	A	2023年5月12日
WO	2022078276	A1	2022年4月21日	US	2023254711	A1	2023年8月10日
				EP	4231693	A1	2023年8月23日
				CN	114363921	A	2022年4月15日