

(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  
2023年12月14日 (14.12.2023)



(10) 国际公布号  
**WO 2023/236143 A1**

- (51) 国际专利分类号:  
*H04W 16/22* (2009.01)
- (21) 国际申请号: PCT/CN2022/097888
- (22) 国际申请日: 2022年6月9日 (09.06.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人:富士通株式会社(FUJITSU LIMITED) [JP/JP]; 日本神奈川县川崎市中原区上小田中4丁目1番1号, Kanagawa 〒211-8588 (JP)。
- (72) 发明人; 及
- (71) 申请人 (仅对US):孙刚(SUN, Gang) [CN/CN]; 中国北京市朝阳区建国门外大街8号IFC国际财源中心A座8层, Beijing 100022 (CN)。王昕(WANG, Xin) [CN/CN]; 中国北京市朝阳区建国门外大街8号IFC国际财源中心A座8层, Beijing 100022 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街35号国际企业大厦A座16层, Beijing 100033 (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, 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, ME, MG, MK, MN, 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, 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, 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: INFORMATION TRANSCIVING METHOD AND APPARATUS

(54) 发明名称: 信息收发方法与装置

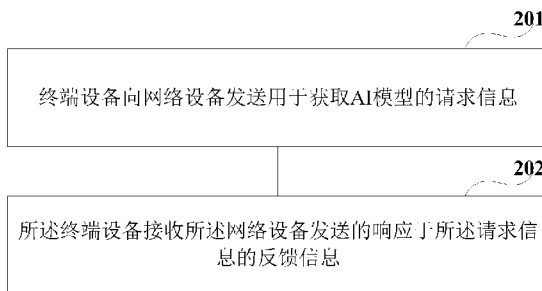


图 2

- 201 A terminal device sends to a network device request information which is used for acquiring an AI model
- 202 The terminal device receives feedback information which is sent by the network device in response to the request information

(57) Abstract: The embodiments of the present application provide an information transceiving method and apparatus. The method comprises: a terminal device sending to a network device request information which is used for acquiring an AI model, and receiving feedback information which is sent by the network device in response to the request information.

(57) 摘要: 本申请实施例提供一种信息收发方法以及装置。所述方法包括: 终端设备向网络设备发送用于获取AI模型请求信息, 并接收该网络设备发送的响应于该请求信息的反馈信息。



WO 2023/236143 A1

## 信息收发方法与装置

技术领域

本申请实施例涉及通信技术领域。

5

背景技术

为了能够支持不同的应用场景，提供不同的业务类型，无线网络希望能在设计，部署以及运营更加智能化。但伴随 5G 新无线（NR，New Radio）网络复杂度的增长，传统的网络设计，部署，运营等方法越来越难以满足智能化的需求。人工智能（AI，  
10 Artificial Intelligence）和机器学习（ML，Machine Learning）技术为 5G 新无线网络的优化提供了一种重要的手段。

伴随着 AI/ML 技术的发展，将 AI/ML 技术应用到无线通信物理层上，来优化现有系统中延时、负载、精度等难点成为当前一个技术方向。

应该注意，上面对技术背景的介绍只是为了方便对本申请的技术方案进行清楚、完整的说明，并方便本领域技术人员的理解而阐述的。不能仅仅因为这些方案在本申  
15 请的背景技术部分进行了阐述而认为上述技术方案为本领域技术人员所公知。

发明内容

但是，发明人发现：当网络中存储众多的功能，性能和/或复杂度等不同的 AI 模  
20 型时，终端设备如何获取合适的 AI 模型是一个需要解决的问题。

针对上述问题的至少之一，本申请实施例提供一种信息收发方法以及装置。

根据本申请实施例的一个方面，提供一种信息收发装置，包括：

第一发送单元，其向网络设备发送用于获取 AI 模型的请求信息；

第一接收单元，其接收所述网络设备发送的响应于所述请求信息的反馈信息。

25 根据本申请实施例的另一个方面，提供一种信息收发装置，包括：

第二接收单元，其接收终端设备发送的用于获取 AI 模型的请求信息；

第二发送单元，向所述终端设备发送响应于所述请求信息的反馈信息。

根据本申请实施例的另一个方面，提供一种通信系统，包括终端设备和/或网络设备，所述终端设备包括前述一个方面的信息收发装置，所述网络设备包括前述另一

个方面的信息收发装置。

本申请实施例的有益效果之一在于：终端设备向网络设备发送用于获取 AI 模型的请求，并接收网络设备发送的反馈信息，由此，终端设备能够从网络设备获取到合适的 AI 模型，可以使用获取到的 AI 模型对系统的负载和延迟进行优化。

5 参照后文的说明和附图，详细公开了本申请的特定实施方式，指明了本申请的原理可以被采用的方式。应该理解，本申请的实施方式在范围上并不因而受到限制。在所附权利要求的精神和条款的范围内，本申请的实施方式包括许多改变、修改和等同。

针对一种实施方式描述和/或示出的特征可以以相同或类似的方式在一个或多个其它实施方式中使用，与其它实施方式中的特征相组合，或替代其它实施方式中的特征。

10 特征。

应该强调，术语“包括/包含”在本文使用时指特征、整件、步骤或组件的存在，但并不排除一个或多个其它特征、整件、步骤或组件的存在或附加。

#### 附图说明

15 在本申请实施例的一个附图或一种实施方式中描述的元素和特征可以与一个或多个其它附图或实施方式中示出的元素和特征相结合。此外，在附图中，类似的标号表示几个附图中对应的部件，并可用于指示多于一种实施方式中使用的对应部件。

图 1 是本申请实施例的通信系统的示意图；

图 2 是本申请实施例的信息收发方法的示意图；

20 图 3 是本申请实施例的终端设备接收/发射链路的功能模块的示意图；

图 4 是本申请实施例的相关标识信息的示意图；

图 5 是本申请实施例的信息收发方法的示意图；

图 6 是本申请实施例的信息收发方法的示意图；

图 7 是本申请实施例的信息收发装置的示意图；

25 图 8 是本申请实施例的信息收发装置的示意图；

图 9 是本申请实施例的信息收发方法的示意图；

图 10 是本申请实施例的网络设备的示意图；

图 11 是本申请实施例的终端设备的示意图。

### 具体实施方式

参照附图，通过下面的说明书，本申请的前述以及其它特征将变得明显。在说明书和附图中，具体公开了本申请的特定实施方式，其表明了其中可以采用本申请的原则的部分实施方式，应了解的是，本申请不限于所描述的实施方式，相反，本申请包  
5 括落入所附权利要求的范围内的全部修改、变型以及等同物。

在本申请实施例中，术语“第一”、“第二”等用于对不同元素从称谓上进行区分，但并不表示这些元素的空间排列或时间顺序等，这些元素不应被这些术语所限制。术语“和/或”包括相关联列出的术语的一种或多个中的任何一个和所有组合。术语“包含”、“包括”、“具有”等是指所陈述的特征、元素、元件或组件的存在，但并不排除存在或  
10 添加一个或多个其他特征、元素、元件或组件。

在本申请实施例中，单数形式“一”、“该”等包括复数形式，应广义地理解为“一种”或“一类”而并不是限定为“一个”的含义；此外术语“所述”应理解为既包括单数形式也包括复数形式，除非上下文另外明确指出。此外术语“根据”应理解为“至少部分根据……”，术语“基于”应理解为“至少部分基于……”，除非上下文另外明确指出。

在本申请实施例中，术语“通信网络”或“无线通信网络”可以指符合如下任意通信标准的网络，例如长期演进（LTE, Long Term Evolution）、增强的长期演进（LTE-A, LTE- Advanced）、宽带码分多址接入（WCDMA, Wideband Code Division Multiple Access）、高速报文接入（HSPA, High-Speed Packet Access）等等。

并且，通信系统中设备之间的通信可以根据任意阶段的通信协议进行，例如可以  
20 包括但不限于如下通信协议：1G（generation）、2G、2.5G、2.75G、3G、4G、4.5G 以及 5G、新无线（NR, New Radio）、未来的 6G 等等，和/或其他目前已知或未来将被开发的通信协议。

在本申请实施例中，术语“网络设备”例如是指通信系统中将终端设备接入通信网络并为该终端设备提供服务的设备。网络设备可以包括但不限于如下设备：基站（BS, Base Station）、接入点（AP, Access Point）、发送接收点（TRP, Transmission Reception Point）、广播发射机、移动管理实体（MME, Mobile Management Entity）、网关、服务器、无线网络控制器（RNC, Radio Network Controller）、基站控制器（BSC, Base Station Controller）等等。

其中，基站可以包括但不限于：节点 B（NodeB 或 NB）、演进节点 B（eNodeB

或 eNB) 以及 5G 基站 (gNB), 等等, 此外还可包括远端无线头 (RRH, Remote Radio Head)、远端无线单元 (RRU, Remote Radio Unit)、中继 (relay) 或者低功率节点 (例如 femto、pico 等等)。并且术语“基站”可以包括它们的一些或所有功能, 每个基站可以对特定的地理区域提供通信覆盖。术语“小区”可以指的是基站和/或其覆盖区域, 这取决于使用该术语的上下文。

在本申请实施例中, 术语“用户设备” (UE, User Equipment) 或者“终端设备” (TE, Terminal Equipment 或 Terminal Device) 例如是指通过网络设备接入通信网络并接收网络服务的设备。终端设备可以是固定的或移动的, 并且也可以称为移动台 (MS, Mobile Station)、终端、用户台 (SS, Subscriber Station)、接入终端 (AT, Access Terminal)、站, 等等。

其中, 终端设备可以包括但不限于如下设备: 蜂窝电话 (Cellular Phone)、个人数字助理 (PDA, Personal Digital Assistant)、无线调制解调器、无线通信设备、手持设备、机器型通信设备、膝上型计算机、无绳电话、智能手机、智能手表、数字相机, 等等。

再例如, 在物联网 (IoT, Internet of Things) 等场景下, 终端设备还可以是进行监控或测量的机器或装置, 例如可以包括但不限于: 机器类通信 (MTC, Machine Type Communication) 终端、车载通信终端、设备到设备 (D2D, Device to Device) 终端、机器到机器 (M2M, Machine to Machine) 终端, 等等。

此外, 术语“网络侧”或“网络设备侧”是指网络的一侧, 可以是某一基站, 也可以包括如上的一个或多个网络设备。术语“用户侧”或“终端侧”或“终端设备侧”是指用户或终端的一侧, 可以是某一 UE, 也可以包括如上的一个或多个终端设备。本文在没有特别指出的情况下, “设备”可以指网络设备, 也可以指终端设备。

以下通过示例对本申请实施例的场景进行说明, 但本申请不限于此。

图 1 是本申请实施例的通信系统的示意图, 示意性说明了以终端设备和网络设备为例的情况, 如图 1 所示, 通信系统 100 可以包括网络设备 101 和终端设备 102、103。为简单起见, 图 1 仅以两个终端设备和一个网络设备为例进行说明, 但本申请实施例不限于此。

在本申请实施例中, 网络设备 101 和终端设备 102、103 之间可以进行现有的业务或者未来可实施的业务发送。例如, 这些业务可以包括但不限于: 增强的移动宽带

(eMBB, enhanced Mobile Broadband)、大规模机器类型通信 (mMTC, massive Machine Type Communication) 和高可靠低时延通信 (URLLC, Ultra-Reliable and Low-Latency Communication), 等等。

值得注意的是, 图 1 示出了两个终端设备 102、103 均处于网络设备 101 的覆盖范围内, 但本申请不限于此。两个终端设备 102、103 可以均不在网络设备 101 的覆盖范围内, 或者一个终端设备 102 在网络设备 101 的覆盖范围之内而另一个终端设备 103 在网络设备 101 的覆盖范围之外。

在本申请实施例中, 高层信令例如可以是无线资源控制 (RRC) 信令; 例如称为 RRC 消息 (RRC message), 例如包括 MIB、系统信息 (system information)、专用 RRC 消息; 或者称为 RRC IE (RRC information element)。高层信令例如还可以是 MAC (Medium Access Control) 信令; 或者称为 MAC CE (MAC control element)。但本申请不限于此。

AI 模型包括但不限于: 输入层 (input)、多个卷积层、连接层 (concat)、全连接层 (FC) 以及量化器等。其中, 多个卷积层的处理结果在连接层进行合并, 关于 AI 模型的具体结构可以参考现有技术, 此处不再赘述。

发明人发现, 不同功能, 参数和/或复杂度的 AI 模型可以采用离线的方式进行训练, 在训练结束后, 由于终端设备存储容量的限制, 通常可以考虑将这些 AI 模型仅存储于网络设备中。但目前标准中还没有定义终端设备如何获取 AI 模型的方法。针对上述问题, 本申请实施例提供一种信息收发方法以及装置, 以下结合附图和实施例进行说明。

#### 第一方面的实施例

本申请实施例提供一种信息收发方法, 从终端设备侧进行说明。

图 2 是本申请实施例的信息收发方法的一示意图, 如图 2 所示, 该方法包括:

201, 终端设备向网络设备发送用于获取 AI 模型的请求信息;

202, 所述终端设备接收所述网络设备发送的响应于所述请求信息的反馈信息。

值得注意的是, 以上附图 2 仅对本申请实施例进行了示意性说明, 但本申请不限于此。例如可以适当地调整各个操作之间的执行顺序, 此外还可以增加其他的一些操作或者减少其中的某些操作。本领域的技术人员可以根据上述内容进行适当地变型,

而不仅限于上述附图 2 的记载。

在一些实施例中，网络设备中可以存储不同功能、参数和/或复杂度等的多个 AI 模型，也就是说网络设备中预先存储了多个 AI 模型。

5 在一些实施例中，AI 模型的功能是指终端设备的接收和/或发射链路中的部分功能，图 3 是本申请实施例中的终端设备接收和发射链路包括的各个功能模块示意图，如图 3 所示，在接收链路中，包含如下功能模块：接收模块、模数转换模块、傅里叶变换模块、资源解映射模块、信道状态信息（CSI）反馈模块、波束测量反馈模块、终端设备定位反馈模块，信道估计模块、多输入多输出（MIMO）检测模块、解码模块；在发射链路中，包括如下功能模块：发送模块、数模转换模块、傅里叶逆变换模  
10 块、资源映射模块、预编码模块、层映射模块、调制模块、编码模块。

在一些实施例中，所述 AI 模型的功能包括用于 CSI 压缩（或编码）的 AI 编码器模型或用于波束预测的 AI 模型或用于终端设备定位的 AI 模型等。

例如，如图 3 所示，在 CSI 反馈模块中，CSI 反馈如果采用现有标准定义的 typeII 或者 etypeII 码本的话，其反馈的负载是比较大的。可以使用 AI 模型来对信道系数矩  
15 阵或者信道系数矩阵的特征向量进行压缩的话，能够减少 CSI 反馈的负载，降低 CSI 反馈开销，因此在 CSI 反馈模块中，终端设备可以使用 AI 模型（也叫作 AI 编码器）进行 CSI 压缩（或编码）。

例如，如图 3 所示，在波束测量反馈模块中，波束测量和反馈采用现有标准定义的方法，在对所有同步信号块（SSB）进行波束扫描时，RS 的负载以及波束选择的  
20 延时都比较大，可以使用 AI 模型用少量波束的测量结果预测出空间上最优的波束，能够减少 RS 的负载以及波束选择的延时，因此在波束测量反馈模块中，终端设备可以使用 AI 模型预测最优波束。

例如，如图 3 所示，终端设备定位反馈模块中，终端设备的定位如果采用传统的方法，终端设备不能有效的识别视线传输（LOS）和非视线传输（NLOS）场景，会  
25 造成定位的精度较低。可以使用 AI 模型来对当前终端设备所处的场景是 LOS 或者 NLOS 进行有效的分类，能够提高定位的精度，因此在终端设备定位反馈模块中，终端设备可以使用 AI 模型进行终端设备定位。

以上仅示例说明，该 AI 模型还可以在终端设备的接收和/或发射链路中其他功能模块中使用，也就是说，该 AI 模型的功能还可以除上述示例外的其他终端设备接收

和/或发射链路中的部分功能，此处不再一一举例。

5 在一些实施例中，AI 模型的参数是指 AI 模型的输入参数和输出参数，该输入参数和输出参数包括输入或输出的维度、物理量。相同功能的 AI 模型的参数可以相同或不同（例如输入/输出参数中的维度可以相同或不同，输入/输出物理量可以相同或者不同）。例如针对用于 CSI 压缩的 AI 编码器模型，其输入参数的物理量可以是表示信道系数矩阵的特征向量，也可以为信道系数矩阵，其维度为  $X1 \times Y1 \times Z1 \times N$ ，输出参数的物理量可以是压缩后的信道特征向量，或者是信道系数矩阵，维度为  $X2$ 。例如，网络设备的发射天线端口的数目为 32，终端设备的接收天线端口的数目为 2，通信系统的带宽为 24 个资源块（RBs），信道状态信息-参考信号（CSI-RS）在频域上的密度为 0.5，即 2 个 RB 上有 1 个 CSI-RS 信号，那么频域上总共有 12 个 CSI-RS 信号。输入参数的物理量信道系数矩阵，其维度为  $12 \times 32 \times 2 \times 2$ （即，频域上 RS 数目  $\times$  网络设备发射天线端口的数目  $\times$  终端设备接收天线端口的数目  $\times$  I/Q 两路）。再例如针对用于波束预测的 AI 模型，其输入参数的物理量为部分波束对的 RSRP(Reference Signal Receiving Power, 参考信号接收功率)值，也可以为部分波束对的 SINR(Signal to Interference plus Noise Ratio, 信号与干扰加噪声比)值，输入维度为  $X1$ ，输出参数的物理量为所有波束对的 RSRP 或者 SINR，输出维度为  $X2$ ，例如下行发送波束有 12 个，接收波束有 8 个，总共有 96 波束对。通过配置，UE 只测量了其中 24 个波束对 RSRP。此时 AI 模型的输入参数的维度为 24，物理量为 RSRP，输出参数的维度为 96，物理量也为 RSRP。

20 在一些实施例中，AI 模型的复杂度是指部署所述 AI 模型实际所需的第二计算量和/或第二存储空间。该计算量可以使用每秒浮点运算数（FLOPs）表示，该部署所述 AI 模型实际所需的第二计算量与该 AI 模型的输入输出参数（维度、通道等）、卷积核大小等相关，具体确定方式可以参考现有技术，部署所述 AI 模型实际所需的第二存储空间与该 AI 模型的大小（即部署该 AI 模型所占的比特数/字节/兆等）以及特征消耗（中间或最后输出结果）相关，具体确定方式可以参考现有技术。

25 在一些实施例中，不同 AI 模型意味着 AI 模型的功能、参数和/或复杂度中至少一个不同，例如 AI 模型 A 的功能是用于 CSI 压缩，AI 模型 B 的功能是用于波束预测，则 AI 模型 A 和 AI 模型 B 是不同的 AI 模型，例如，AI 模型 A 和 B 的功能都是用于 CSI 压缩，但 AI 模型 A 和 B 的复杂度不同，和/或参数不同，则 AI 模型 A 和

AI 模型 B 是不同的 AI 模型。

5 在一些实施例中，终端设备在其接收和/或发射链路中执行部分功能时需要使用合适（对应）的 AI 模型，由于在网络设备中预先存储了多个 AI 模型，而终端设备没有存储该多个 AI 模型，因此，终端设备可以通过向网络设备发送用于获取 AI 模型的请求信息，来获取其所需要的模型，该请求信息包括所述 AI 模型的功能标识信息和/或 AI 模型的参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息。

10 在一些实施例中，AI 模型的功能标识信息用于标识 AI 模型的功能，例如，该功能标识信息为 3 个比特，不同的比特值表示不同的 AI 模型的功能，终端设备和网络设备中可以预定义好比特位的值和 AI 模型的功能的对应关系，根据该功能标识信息指示终端设备所需的（所请求的）AI 模型的功能，例如该功能标识信息为 001 时，所需的（所请求的）AI 模型的功能为用于 CSI 压缩，该功能标识信息为 010 时，所需的（所请求的）AI 模型的功能为用于波束预测，该功能标识信息为 011 时，所需的（所请求的）AI 模型的功能为用于终端设备定位，或者，例如，该功能标识信息可以是比特位图，每个比特位对应指示一个 AI 模型的功能，在比特位的值为 1（或者 0）时，指示所需的（所请求的）AI 模型的功能为与所述比特位对应的 AI 模型的功能，例如，该功能标识信息为 3 比特的比特位图，终端设备和网络设备中可以预定义好比特位和 AI 模型的功能的对应关系，根据该功能标识信息指示终端设备所需的（所请求的）AI 模型的功能，例如该功能标识信息为 001 时，所需的（所请求的）AI 模型的功能为用于 CSI 压缩，该功能标识信息为 010 时，所需的（所请求的）AI 模型的功能为用于波束预测，该功能标识信息为 100 时，所需的（所请求的）AI 模型的功能为用于终端设备定位，以上仅为示例说明，但本申请实施例并不以此作为限制。

25 在一些实施例中，如前所述，AI 模型的参数信息包括所述 AI 模型输入参数和输出参数信息，该输入参数和输出参数信息包括指示输入和输出的维度的第一信息，和/或指示输入和输出的物理量的第二信息。其中，该第一信息指示维度的数量和/或各个维度的具体数值，该维度的数量可以显式或隐式的指示，例如使用第一预定数量比特（例如 2 个）指示维度的数量，其中，01 指示输入维度为 X2，11 指示输入维度为 X1×Y1×Z1×N，使用第二预定数量比特分别指示各个维度的具体数值，即使用二进制编码来表示各个维度的具体数值，例如，使用 3 比特指示 X1 的值，再使用 3 比特指

示 Y1 的值,再使用 3 比特指示 Z1 的值,再使用 3 比特指示 N 的值,以上仅为示例,本申请并不以此作为限制。例如,第一预定数量比特也可以缺省,使用前述第二预定数量比特隐式的指示维度的数量,此处不再一一举例。该第二信息使用第三预定数量比特表示,不同的比特值表示不同的物理量,终端设备和网络设备中可以预定义好比特位的值和物理量的对应关系,例如该第二信息为 001 时,指示物理量为信道系数矩阵,该第二信息为 010 时,指示物理量为 RSRP,此处不再一一举例。。在一些实施例中,该能力信息包括所述终端设备针对所述 AI 模型部署所能支持的最大第一计算量和/或第一存储空间。与前述类似的,该第一计算量可以使用每秒浮点运算数(FLOPs)表示(例如二进制编码),第一存储空间可以使用比特数/字节/兆等表示(例如二进制编码),该终端设备针对所述 AI 模型部署所能支持的最大第一计算量和/或最大第一存储空间的具体值由该终端设备自身的能力,例如硬件(例如处理器和存储器等)性能和当前正在运行/执行的程序或业务或功能确定,此处不再一一举例。

例如,终端设备接收到网络设备发送的 CSI-RS 后,需要进行 CSI 估计并上报,为了能够减少 CSI 反馈的负载,降低 CSI 反馈开销,终端设备需要获取用于 CSI 压缩的 AI 模型,可以使用 AI 模型得到压缩后的 CSI,因此,终端设备可以向网络设备发送用于获取用于 CSI 压缩的 AI 模型请求信息,或者说,该终端设备可以向网络设备请求获取用于 CSI 压缩的 AI 模型,该请求信息中包括用于 CSI 压缩的 AI 模型的功能标识信息和/或 AI 模型的参数信息和/或所述终端设备所能支持的所述用于 CSI 压缩的 AI 模型的能力信息。

例如,终端设备接收网络设备发送的用于波束测量的参考信号,为了减少 RS 的负载以及波束选择的延时,终端设备需要获取用于波束预测的 AI 模型,使用该 AI 模型预测出最优的波束,并向网络设备发送最优波束的信息,因此,终端设备可以向网络设备发送用于获取用于波束预测的 AI 模型请求信息,或者说,该终端设备可以向网络设备请求获取用于波束预测的 AI 模型,该请求信息中包括用于波束预测的 AI 模型的功能标识信息和/或 AI 模型的参数信息和/或所述终端设备所能支持的所述用于波束预测的 AI 模型的能力信息。

例如,为了提高定位精度,终端设备需要获取用于终端设备定位的 AI 模型,使用 AI 模型来对当前终端设备所处的场景是 LOS 或者 NLOS 进行有效的分类,因此,终端设备可以向网络设备发送用于获取用于终端设备定位的 AI 模型请求信息,或

者说，该终端设备可以向网络设备请求获取用于终端设备定位的 AI 模型，该请求信息中包括用于终端设备定位的 AI 模型的功能标识信息和/或 AI 模型参数信息和/或所述终端设备所能支持的所述用于终端设备定位的 AI 模型的能力信息。

5 在一些实施例中，该请求信息由 RRC 或 MAC CE 或 UCI 承载，例如，该请求信息可以为 UCI 或现有 RRC 信令中新增的信息元（域），或者，该请求信息也可以由新增的 RRC 信令承载，此处不再一一举例。以上请求信息中各类信息的比特数仅为示例说明，本申请实施例并不以此作为限制。

在一些实施例中，网络设备在接收到该请求信息后，可以向终端设备发送反馈信息，以响应该请求信息。

10 在一些实施例中，网络设备在接收到该请求信息后，向终端设备发送反馈信息，网络设备根据该请求信息，在其预先存储的多个 AI 模型中进行匹配，如果没有匹配到满足终端设备请求的 AI 模型（匹配失败）时，通过该反馈信息告知终端设备不支持所述终端设备请求，如果匹配到满足终端设备请求的 AI 模型（匹配成功）时，通过该反馈信息告知终端设备支持所述终端设备请求，以及告知终端设备该匹配到的  
15 AI 模型的相关信息。

以下先说明网络设备匹配 AI 模型的方法。

例如，请求消息中包括所述 AI 模型的功能标识信息和/或 AI 模型参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息，网络设备在接收到请求信息后，根据其中的功能标识信息，在其预先存储的多个 AI 模型中进行匹配，尝试匹配与该  
20 功能标识信息指示的 AI 模型的功能相同的 AI 模型，如果没有与该功能标识信息指示的 AI 模型的功能相同的 AI 模型，即匹配失败，则表示网络设备不支持终端设备的请求。如果有与该功能标识信息指示的 AI 模型的功能相同的多个 AI 模型，则再根据请求消息中 AI 模型参数信息，在其预先存储该功能相同的多个 AI 模型中进行匹配，尝试匹配与该 AI 模型参数信息指示的 AI 模型参数相同的 AI 模型，如果没有与  
25 该参数信息指示的 AI 模型参数相同的 AI 模型，即匹配失败，则表示网络设备不支持终端设备的请求。如果在功能相同的多个 AI 模型中有参数也匹配（参数相同）的多个 AI 模型，则再次比较该功能相同，参数匹配的多个 AI 模型的第二计算量和/或第二存储空间与该能力信息中的第一计算量和/或第一存储空间，如果该功能相同，参数能够匹配的多个 AI 模型的第二计算量和/或第二存储空间都大于第一计算量和/

或第一存储空间，则表示终端设备的能力不能支持部署该 AI 模型（匹配失败），如果有至少一个 AI 模型的第二计算量和/或第二存储空间小于第一计算量和/或第一存储空间，则表示终端设备的能力可以支持部署该 AI 模型，即表示匹配到满足终端设备请求的 AI 模型（匹配成功），从该至少一个（M 个）AI 模型中选择一个 AI 模型作为匹配到的 AI 模型（以下也叫作合适的 AI 模型）。如果 M 等于 1，则该一个 AI 模型即作为匹配到的 AI 模型，如果 M 大于 1，则可以从该 M 个 AI 模型中任意选择一个 AI 模型作为匹配到的 AI 模型，也可以根据预定规则从 M 个 AI 模型中选择一个 AI 模型作为匹配到的 AI 模型，例如，该预定规则可以是 M 个 AI 模型中第二计算量和/或第二存储空间最小的 AI 模型或最大的 AI 模型，以上仅为示例说明，本申请实施例并不以此作为限制。

以上匹配过程仅为示例说明，本申请实施例并不以此作为限制，例如在请求信息中包括功能标识信息和能力信息中的一种时，仅根据功能标识信息或者能力信息进行上述部分匹配过程即可，此处不再重复赘述。

在一些实施例中，该反馈信息包括是否支持所述终端设备请求的指示信息和/或所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度。其中，该指示信息包括 1 比特，在该 1 比特的值为 1 时，指示网络设备支持该终端设备的请求（即匹配成功），在该 1 比特的值为 0 时，指示网络设备不支持终端设备的请求（即匹配失败），反之亦可，本申请并不以此作为限制。

在一些实施例中，在网络设备支持该终端设备的请求（即匹配成功）时，该反馈信息还可以包括所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度，所述 AI 模型即为匹配到的 AI 模型（合适的 AI 模型）。

在一些实施例中，所述 AI 模型的相关标识信息用于标识所述 AI 模型的功能和/或相同功能的 AI 模型的参数信息和/或相同功能相同参数的多个 AI 模型中所述 AI 模型的序号。例如，所述 AI 模型的相关标识信息包括第一标识信息和/或第二标识信息和/或第三标识信息，所述第一标识信息是所述 AI 模型的功能标识，所述第二标识信息是相同功能的 AI 模型的参数信息，所述第三标识信息是所述相同功能，相同参数的多个 AI 模型中所述 AI 模型的序号。该第一标识信息可以参考前述功能标识信息的实施方式，该第二标识信息可以包括预定数量个比特，该预定数量个比特的值标识相同功能的 AI 模型的参数信息，该第二标识信息与前述请求信息中的参数信息的不

同之处在于，相同功能的 AI 模型参数信息不同时，第二标识信息也不同，但不同功能的 AI 模型参数信息相同或不同时，可以使用相同的第二标识信息，例如，针对用于 CSI 压缩的 AI 模型，在第二标识信息为 1000 时，指示输入维度为 8，在第二标识信息为 1010 时，指示输入维度为 10，针对用于波束预测的 AI 模型，在第二标识信息为 1000 时，指示输入维度为 8，在第二标识信息为 1010 时，指示输入维度为 12，也就是说第二标识信息仅用于唯一标识同一功能的 AI 模型的不同参数信息。该第三标识信息可以包括预定数量个比特，该预定数量个比特的值标识所述相同功能相同参数的各个 AI 模型的序号，例如，相同功能相同参数的 AI 模型有 4 个，该 4 个 AI 模型的第三标识信息分别为 00,01,10,11。

10 图 4 是本申请实施例的 AI 模型的相关标识信息示意图，如图 4 所示，该相关标识信息为 8 比特，其中，前 3 个比特为第一标识信息，指示该 AI 模型的功能，中间 3 个比特为第二标识信息，指示该模型的参数信息，后 2 个比特为第三标识信息，指示该 AI 模型在同功能和参数的多个 AI 模型中的序号。

15 以上比特数仅为示例说明，本申请并不以此作为限制，另外，该相关标识信息还可以使用其他形式表示，例如该相关标识信息可以包括第四标识信息，该第四标识信息用于唯一标识多个 AI 模型中的各个 AI 模型，例如该相关标识信息为 8 比特，其中，该 8 个比特为各 AI 模型在多个 AI 模型中的序号，此处不再一一示例。

20 在一些实施例中，所述 AI 模型的复杂度包括部署所述 AI 模型实际所需的第二计算量和/或第二存储空间，关于该第二计算量和/或该第二存储空间的含义和确定方式请参考前述说明，该第二计算量和/或第二存储空间可以使用二进制编码后包含在该反馈信息中，此处不再一一举例。

在一些实施例中，该反馈信息由 RRC 或 MAC CE 或 DCI 承载，例如，该反馈信息可以为 DCI 或现有 RRC 信令中新增的信息元（域），或者，该反馈信息也可以由新增的 RRC 信令承载，此处不再一一举例。

25 图 5 是本申请实施例中信息收发方法示意图，如图 5 所示，该方法包括：

501，终端设备向网络设备发送用于获取 AI 模型的请求信息；

502，所述终端设备接收所述网络设备发送的响应于所述请求信息的反馈信息。

503，该终端设备接收所述网络设备发送的所述 AI 模型传输资源分配信息，所述资源分配信息用于指示传输所述 AI 模型所需的时频域资源；

504, 该终端设备接收所述网络设备在所述时频域资源上发送的所述 AI 模型。

在一些实施例中, 501-502 的实施方式可以参考 201-202, 重复之处不再赘述。

在一些实施例中, 在 503 中, 在所述网络设备支持所述终端设备请求(匹配成功)时, 网络设备向终端设备发送该资源分配信息, 该资源分配信息用于指示传输所述 AI 模型所需的时频域资源(空口上的资源), 该时频域资源的大小可以根据该匹配到的 AI 模型的大小(尺寸, 或者说第二存储空间)确定, 该 AI 模型可以在 PDSCH 上传输, 该资源分配信息可以由调度 PDSCH 的 DCI 承载, 例如该资源分配信息可以是 DCI 中的时域和/或频率资源分配字段, 具体可以参考现有技术, 此处不再赘述。在网络设备不支持该终端设备请求(匹配失败)时, 不需要执行 503-504。

10 在一些实施例中, 在 504 中, 网络设备在为 AI 模型传输分配资源后, 即在该分配的时频域资源上向终端设备发送该 AI 模型, 其中, 发送该 AI 模型是指发送该 AI 模型的网络结构, 节点数量、AI 模型各节点系数等, 例如, 可以在 pytorch 或 tensorflow 等开发环境中以预定的存储格式将训练好的多个 AI 模型分别保存为对应的的多个预定格式的文件, 该文件中包含有对应 AI 模型的网络结构, 节点数量、AI 模型各节点系数等, 将与该多个 AI 模型对应的多个预定格式的文件预先存储在网络设备中, 在 15 502 中匹配到合适的 AI 模型后, 在分配的时频域资源上将与该合适的 AI 模型对应的文件发送给终端设备。

在一些实施例中, 该方法还可以包括(未图示): 该终端设备使用该 AI 模型进行相应的处理, 例如 CSI 压缩, 预测最优波束, 定位终端设备(对当前终端设备所处的场景是 LOS 或者 NLOS 进行有效的分类)等, 具体可以参考现有技术, 此处不再一一赘述。

以上各个实施例仅对本申请实施例进行了示例性说明, 但本申请不限于此, 还可以在以上各个实施例的基础上进行适当的变型。例如, 可以单独使用上述各个实施例, 也可以将以上各个实施例中的一种或多种结合起来。

25 由上述实施例可知, 终端设备向网络设备发送用于获取 AI 模型的请求, 并接收网络设备发送的反馈信息, 由此, 终端设备能够从网络设备获取到合适的 AI 模型, 可以使用获取到的 AI 模型对系统的负载和延迟进行优化。

## 第二方面的实施例

本申请实施例提供一种信息收发方法，从网络设备侧进行说明，与第一方面的实施例相同的内容不再赘述。

图 6 是本申请实施例的信息收发方法的一示意图，如图 6 所示，该方法包括：

601，网络设备接收终端设备发送的用于获取 AI 模型请求信息；

5 602，所述网络设备向所述终端设备发送响应于所述请求信息的反馈信息。

在一些实施例中，所述请求信息包括所述 AI 模型的功能标识信息和/或 AI 模型参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息。

在一些实施例中，所述反馈信息包括是否支持所述终端设备请求的指示信息和/或所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度。

10 关于 601-602 的实施方式可以参考第一方面实施例的 201-202，关于该请求信息和反馈信息可以参考第一方面的实施例，此处不再赘述。

在一些实施例中，该方法还包括：

该网络设备根据所述终端设备请求的所述 AI 模型的功能和/或 AI 模型参数信息和/或所述能力信息与所述网络设备存储的各个 AI 模型进行匹配，以确定所述网络设备存储的 AI 模型中是否有满足所述终端设备请求的 AI 模型；

15 并且，该网络设备基于所述处理单元的匹配结果发送所述反馈信息。

在一些实施例中，上述匹配的实施方式可以参考第一方面实施例的匹配过程，在匹配成功时确定网络设备存储的 AI 模型中有满足所述终端设备请求的 AI 模型；在匹配失败时，确定网络设备存储的 AI 模型中没有满足所述终端设备请求的 AI 模型。网络设备基于所述处理单元的匹配结果发送所述反馈信息。具体可以参考第一方面的实施例，此处不再赘述。

在一些实施例中，该方法还包括：

25 该网络设备向所述终端设备发送所述 AI 模型传输资源分配信息，所述资源分配信息用于指示传输所述 AI 模型所需的时频域资源，以及在所述时频域资源上向所述终端设备发送所述 AI 模型。

值得注意的是，以上附图 6 仅对本申请实施例进行了示意性说明，但本申请不限于此。例如可以适当地调整各个操作之间的执行顺序，此外还可以增加其他的一些操作或者减少其中的某些操作。本领域的技术人员可以根据上述内容进行适当地变形，而不仅限于上述附图 6 的记载。

以上各个实施例仅对本申请实施例进行了示例性说明，但本申请不限于此，还可以在以上各个实施例的基础上进行适当的变型。例如，可以单独使用上述各个实施例，也可以将以上各个实施例中的一种或多种结合起来。

由上述实施例可知，网络设备接收终端设备发送的用于获取 AI 模型的请求，向终端设备发送反馈信息，由此，终端设备能够从网络设备获取到合适的 AI 模型，可以使用获取到的 AI 模型对系统的负载和延迟进行优化。

### 第三方面的实施例

本申请实施例提供一种信息收发装置。该装置例如可以是终端设备，也可以是配置于终端设备的某个或某些部件或者组件，与第一方面的实施例相同的内容不再赘述。

图 7 是本申请实施例的信息收发装置的一示意图。如图 7 所示，信息收发装置 700 包括：

第一发送单元 701，其向网络设备发送用于获取 AI 模型的请求信息；

第一接收单元 702，其接收该网络设备发送的响应于该请求信息的反馈信息。

在一些实施例中，该请求信息包括该 AI 模型的功能标识信息和/或 AI 模型的性能信息、或该终端设备所能支持的该 AI 模型的能力信息。

在一些实施例中，该能力信息包括该终端设备针对该 AI 模型部署所能支持的最大第一计算量和/或第一存储空间。

在一些实施例中，该反馈信息包括是否支持该终端设备请求的指示信息和/或该 AI 模型的相关标识信息和/或该 AI 模型的复杂度。

在一些实施例中，该 AI 模型的相关标识信息用于标识该 AI 模型的功能和/或相同功能的 AI 模型的性能信息和/或相同功能相同参数的多个 AI 模型中该 AI 模型的序号。

在一些实施例中，该 AI 模型的相关标识信息包括第一标识信息和/或第二标识信息和/或第三标识信息，该第一标识信息是该 AI 模型的功能标识，该第二标识信息是相同功能的 AI 模型的性能信息，该第三标识信息是相同功能相同参数的多个 AI 模型中该 AI 模型的序号。

在一些实施例中，该 AI 模型的复杂度包括部署该 AI 模型实际所需的第二计算量和/或第二存储空间。

在一些实施例中，该请求信息由 RRC 或 MAC CE 或 UCI 承载。

在一些实施例中，该反馈信息由 RRC 或 MAC CE 或 DCI 承载。

在一些实施例中，该 AI 模型的功能是指该终端设备的接收和/或发射链路中的部分功能。

5 在一些实施例中，该 AI 模型的功能包括用于 CSI 压缩的 AI 编码器模型或用于波束预测的 AI 模型或用于终端设备定位的 AI 模型。

在一些实施例中，该第一接收单元还用于接收该网络设备发送的该 AI 模型传输资源分配信息，该资源分配信息用于指示传输该 AI 模型所需的时频域资源。

10 在一些实施例中，该第一接收单元还用于接收该网络设备在该时频域资源上发送的该 AI 模型。

在一些实施例中，在该网络设备支持该终端设备请求时，该第一接收单元接收该资源分配信息。

15 以上各个实施例仅对本申请实施例进行了示例性说明，但本申请不限于此，还可以在以上各个实施例的基础上进行适当的变型。例如，可以单独使用上述各个实施例，也可以将以上各个实施例中的一种或多种结合起来。

值得注意的是，以上仅对与本申请相关的各部件或模块进行了说明，但本申请不限于此。信息收发装置 700 还可以包括其他部件或者模块，关于这些部件或者模块的具体内容，可以参考相关技术。

20 此外，为了简单起见，图 7 中仅示例性示出了各个部件或模块之间的连接关系或信号走向，但是本领域技术人员应该清楚的是，可以采用总线连接等各种相关技术。上述各个部件或模块可以通过例如处理器、存储器、发射机、接收机等硬件设施来实现；本申请实施并不对此进行限制。

25 由上述实施例可知，终端设备向网络设备发送用于获取 AI 模型的请求，并接收网络设备发送的反馈信息，由此，终端设备能够从网络设备获取到合适的 AI 模型，可以使用获取到的 AI 模型对系统的负载和延迟进行优化。

#### 第四方面的实施例

本申请实施例提供一种信息收发装置。该装置例如可以是网络设备，也可以是配置于网络设备的某个或某些部件或者组件，与第二方面的实施例相同的内容不再赘述。

图 8 是本申请实施例的信息收发装置的一示意图。如图 8 所示，信息收发装置 800 包括：

第二接收单元 801，其接收终端设备发送的用于获取 AI 模型请求信息；

第二发送单元 802，向该终端设备发送响应于该请求信息的反馈信息。

5 在一些实施例中，该请求信息包括该 AI 模型的功能标识信息和/或 AI 模型的参数信息和/或该终端设备所能支持的该 AI 模型的能力信息。

在一些实施例中，该反馈信息包括是否支持该终端设备请求的指示信息和/或该 AI 模型的相关标识信息和/或该 AI 模型的复杂度。

在一些实施例中，该装置还包括：（未图示，可选）

10 处理单元，其用于根据该终端设备请求的该 AI 模型的功能和/或 AI 模型的参数信息和/或该能力信息与该网络设备存储的各个 AI 模型进行匹配，以确定该网络设备存储的 AI 模型中是否有满足该终端设备请求的 AI 模型；

并且，该第二发送单元基于该处理单元的匹配结果发送该反馈信息。

15 在一些实施例中，该第二发送单元还用于向该终端设备发送该 AI 模型传输资源分配信息，该资源分配信息用于指示传输该 AI 模型所需的时频域资源，以及在该时频域资源上向该终端设备发送该 AI 模型。

以上各个实施例仅对本申请实施例进行了示例性说明，但本申请不限于此，还可以在以上各个实施例的基础上进行适当的变型。例如，可以单独使用上述各个实施例，也可以将以上各个实施例中的一种或多种结合起来。

20 值得注意的是，以上仅对与本申请相关的各部件或模块进行了说明，但本申请不限于此。信息收发装置 800 还可以包括其他部件或者模块，关于这些部件或者模块的具体内容，可以参考相关技术。

25 此外，为了简单起见，图 8 中仅示例性示出了各个部件或模块之间的连接关系或信号走向，但是本领域技术人员应该清楚的是，可以采用总线连接等各种相关技术。上述各个部件或模块可以通过例如处理器、存储器、发射机、接收机等硬件设施来实现；本申请实施并不对此进行限制。

由上述实施例可知，网络设备接收终端设备发送的用于获取 AI 模型请求，向终端设备发送反馈信息，由此，终端设备能够从网络设备获取到合适的 AI 模型，可以使用获取到的 AI 模型对系统的负载和延迟进行优化。

### 第五方面的实施例

本申请实施例还提供一种通信系统，可以参考图 1，与第一至四方面的实施例相同的内容不再赘述。

5 在一些实施例中，通信系统 100 至少可以包括：

终端设备 102，其向网络设备发送用于获取 AI 模型的请求信息，接收所述网络设备发送的响应于所述请求信息的反馈信息；以及

网络设备 101，其接收终端设备发送的用于获取 AI 模型的请求信息；向所述终端设备发送响应于所述请求信息的反馈信息。

10 图 9 是本申请实施例的信息收发方法示意图，如图 9 所示，该方法包括：

901，终端设备向网络设备发送用于获取 AI 模型的请求信息；

902，网络设备根据该请求信息进行 AI 模型的匹配（进行 AI 模型的选择）；

903，网络设备根据匹配结果向终端设备发送反馈信息；

904，网络设备向终端设备发送所述 AI 模型传输资源分配信息；

15 905，网络设备在所述时频域资源上向终端设备发送所述 AI 模型；

906，该终端设备使用该 AI 模型执行相应的（接收和/或发射链路中的部分）功能。

关于 901-906 的实施方式可以参考前述第一和第二方面的实施例，此处不再赘述。

20 本申请实施例还提供一种网络设备，例如可以是基站，但本申请不限于此，还可以是其他的网络设备。

图 10 是本申请实施例的网络设备的构成示意图。如图 10 所示，网络设备 1000 可以包括：处理器 1010（例如中央处理器 CPU）和存储器 1020；存储器 1020 耦合到处理器 1010。其中该存储器 1020 可存储各种数据；此外还存储信息处理的程序 1030，并且在处理器 1010 的控制下执行该程序 1030。

25 例如，处理器 1010 可以被配置为执行程序而实现如第二方面的实施例所述的信息收发方法。例如处理器 1010 可以被配置为进行如下的控制：接收终端设备发送的用于获取 AI 模型的请求信息；向所述终端设备发送响应于所述请求信息的反馈信息

此外，如图 10 所示，网络设备 1000 还可以包括：收发机 1040 和天线 1050 等；其中，上述部件的功能与现有技术类似，此处不再赘述。值得注意的是，网络设备

1000 也并不是必须要包括图 10 中所示的所有部件；此外，网络设备 1000 还可以包括图 10 中没有示出的部件，可以参考现有技术。

本申请实施例还提供一种终端设备，但本申请不限于此，还可以是其他的设备。

5 图 11 是本申请实施例的终端设备的示意图。如图 11 所示，该终端设备 1100 可以包括处理器 1110 和存储器 1120；存储器 1120 存储有数据和程序，并耦合到处理器 1110。值得注意的是，该图是示例性的；还可以使用其他类型的结构，来补充或代替该结构，以实现电信功能或其他功能。

例如，处理器 1110 可以被配置为执行程序而实现如第一方面的实施例所述的信息收发方法。例如处理器 1110 可以被配置为进行如下的控制：向网络设备发送用于  
10 获取 AI 模型的请求信息；接收所述网络设备发送的响应于所述请求信息的反馈信息。

如图 11 所示，该终端设备 1100 还可以包括：通信模块 1130、输入单元 1140、显示器 1150、电源 1160。其中，上述部件的功能与现有技术类似，此处不再赘述。值得注意的是，终端设备 1100 也并不是必须要包括图 11 中所示的所有部件，上述部件并不是必需的；此外，终端设备 1100 还可以包括图 11 中没有示出的部件，可以参  
15 考现有技术。

本申请实施例还提供一种计算机程序，其中当在终端设备中执行所述程序时，所述程序使得所述终端设备执行第一方面的实施例所述的信息收发方法。

本申请实施例还提供一种存储有计算机程序的存储介质，其中所述计算机程序使得终端设备执行第一方面的实施例所述的信息收发方法。

20 本申请实施例还提供一种计算机程序，其中当在网络设备中执行所述程序时，所述程序使得所述网络设备执行第二方面的实施例所述的信息收发方法。

本申请实施例还提供一种存储有计算机程序的存储介质，其中所述计算机程序使得网络设备执行第二方面的实施例所述的信息收发方法。

25 本申请以上的装置和方法可以由硬件实现，也可以由硬件结合软件实现。本申请涉及这样的计算机可读程序，当该程序被逻辑部件所执行时，能够使该逻辑部件实现上文所述的装置或构成部件，或使该逻辑部件实现上文所述的各种方法或步骤。本申请还涉及用于存储以上程序的存储介质，如硬盘、磁盘、光盘、DVD、flash 存储器等。

结合本申请实施例描述的方法/装置可直接体现为硬件、由处理器执行的软件模块或二者组合。例如，图中所示的功能框图中的一个或多个和/或功能框图的一个或

多个组合，既可以对应于计算机程序流程的各个软件模块，亦可以对应于各个硬件模块。这些软件模块，可以分别对应于图中所示的各个步骤。这些硬件模块例如可利用现场可编程门阵列（FPGA）将这些软件模块固化而实现。

软件模块可以位于 RAM 存储器、闪存、ROM 存储器、EPROM 存储器、EEPROM 存储器、寄存器、硬盘、移动磁盘、CD-ROM 或者本领域已知的任何其它形式的存储介质。可以将一种存储介质耦接至处理器，从而使处理器能够从该存储介质读取信息，且可向该存储介质写入信息；或者该存储介质可以是处理器的组成部分。处理器和存储介质可以位于 ASIC 中。该软件模块可以存储在移动终端的存储器中，也可以存储在可插入移动终端的存储卡中。例如，若设备（如移动终端）采用的是较大容量的 MEGA-SIM 卡或者大容量的闪存装置，则该软件模块可存储在该 MEGA-SIM 卡或者大容量的闪存装置中。

针对附图中描述的功能方框中的一个或多个和/或功能方框的一个或多个组合，可以实现为用于执行本申请所描述功能的通用处理器、数字信号处理器（DSP）、专用集成电路（ASIC）、现场可编程门阵列（FPGA）或者其它可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件或者其任意适当组合。针对附图描述的功能方框中的一个或多个和/或功能方框的一个或多个组合，还可以实现为计算设备的组合，例如，DSP 和微处理器的组合、多个微处理器、与 DSP 通信结合的一个或多个微处理器或者任何其它这种配置。

以上结合具体的实施方式对本申请进行了描述，但本领域技术人员应该清楚，这些描述都是示例性的，并不是对本申请保护范围的限制。本领域技术人员可以根据本申请的精神和原理对本申请做出各种变型和修改，这些变型和修改也在本申请的范围

关于包括以上实施例的实施方式，还公开下述的附记：

1. 一种信息收发方法，其特征在于，所述方法包括：  
终端设备向网络设备发送用于获取 AI 模型请求信息；  
所述终端设备接收所述网络设备发送的响应于所述请求信息的反馈信息。
2. 根据附记 1 所述的方法，其中，所述请求信息包括所述 AI 模型的功能标识信息和/或 AI 模型的参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息。
3. 根据附记 2 所述的方法，其中，所述能力信息包括所述终端设备针对所述 AI

模型部署所能支持的最大第一计算量和/或第一存储空间。

4. 根据附记 1 至 3 任一项所述的方法，其中，所述反馈信息包括是否支持所述终端设备请求的指示信息和/或所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度。

5 5. 根据附记 4 所述的方法，其中，所述 AI 模型的相关标识信息用于标识所述 AI 模型的功能和/或相同功能的 AI 模型的参数信息和/或相同功能相同参数的多个 AI 模型中所述 AI 模型的序号。

6. 根据附记 5 所述的方法，其中，所述 AI 模型的相关标识信息包括第一标识信息和/或第二标识信息和/或第三标识信息，所述第一标识信息是所述 AI 模型的功能标识，所述第二标识信息是相同功能的 AI 模型的参数信息，所述第三标识信息是相同功能相同参数的多个 AI 模型中所述 AI 模型的序号。

7. 根据附记 4 至 6 任一项所述的方法，其中，所述 AI 模型的复杂度包括部署所述 AI 模型实际所需的第二计算量和/或第二存储空间。

8. 根据附记 1 至 7 任一项所述的方法，其中，所述请求信息由 RRC 或 MAC CE 或 UCI 承载。

9. 根据附记 1 至 8 任一项所述的方法，其中，所述反馈信息由 RRC 或 MAC CE 或 DCI 承载。

10. 根据附记 1 至任一项所述的方法，其中，所述 AI 模型的功能是指所述终端设备的接收和/或发射链路中的部分功能。

11. 根据附记 10 所述的方法，其中，所述 AI 模型的功能包括用于 CSI 压缩的 AI 编码器模型或用于波束预测的 AI 模型或用于终端设备定位的 AI 模型。

12. 根据附记 1 至 11 任一项所述的方法，其中，所述方法还包括：

所述终端设备接收所述网络设备发送的所述 AI 模型传输资源分配信息，所述资源分配信息用于指示传输所述 AI 模型所需的时频域资源。

13. 根据附记 12 所述的方法，其中，所述方法还包括：

所述终端设备还接收所述网络设备在所述时频域资源上发送的所述 AI 模型。

14. 根据附记 12 所述的方法，其中，在所述网络设备支持所述终端设备请求时，所述终端设备接收所述资源分配信息。

15. 一种信息收发方法，其特征在于，所述方法包括：

网络设备接收终端设备发送的用于获取 AI 模型请求信息；

所述网络设备向所述终端设备发送响应于所述请求信息的反馈信息。

16. 根据附记 15 所述的方法，其中，所述请求信息包括所述 AI 模型的功能标识信息和/或 AI 模型参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息。

5 17. 根据附记 15 或 16 所述的方法，其中，所述反馈信息包括是否支持所述终端设备请求的指示信息和/或所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度。

18. 根据附记 16 或 17 所述的方法，其中，所述方法还包括：

所述网络设备根据所述终端设备请求的所述 AI 模型的功能和/或 AI 模型参数信息和/或所述能力信息与所述网络设备存储的各个 AI 模型进行匹配，以确定所述网络设备存储的 AI 模型中是否有满足所述终端设备请求的 AI 模型；

10

并且，所述网络设备基于匹配结果发送所述反馈信息。

19. 根据附记 15 至 18 任一项所述的方法，其中，所述方法还包括：

所述网络设备向所述终端设备发送所述 AI 模型传输资源分配信息，所述资源分配信息用于指示传输所述 AI 模型所需的时频域资源，以及在所述时频域资源上向所述终端设备发送所述 AI 模型。

15

20. 一种网络设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器被配置为执行所述计算机程序而实现如附记 15 至 19 任一项所述的方法。

21. 一种终端设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器被配置为执行所述计算机程序而实现如附记 1 至 14 任一项所述的方法。

## 权利要求书

1. 一种信息收发装置，应用于终端设备，其特征在于，所述装置包括：  
第一发送单元，其向网络设备发送用于获取 AI 模型的请求信息；  
5 第一接收单元，其接收所述网络设备发送的响应于所述请求信息的反馈信息。
2. 根据权利要求 1 所述的装置，其中，所述请求信息包括所述 AI 模型的功能标识信息和/或 AI 模型参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息。
3. 根据权利要求 2 所述的装置，其中，所述能力信息包括所述终端设备针对所述 AI 模型部署所能支持的最大第一计算量和/或第一存储空间。  
10
4. 根据权利要求 1 所述的装置，其中，所述反馈信息包括是否支持所述终端设备请求的指示信息和/或所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度。
5. 根据权利要求 4 所述的装置，其中，所述 AI 模型的相关标识信息用于标识所述 AI 模型的功能和/或相同功能的 AI 模型参数信息和/或相同功能相同参数的多个  
15 AI 模型中所述 AI 模型的序号。
6. 根据权利要求 5 所述的装置，其中，所述 AI 模型的相关标识信息包括第一标识信息和/或第二标识信息和/或第三标识信息，所述第一标识信息是所述 AI 模型的功能标识，所述第二标识信息是相同功能的 AI 模型参数信息，所述第三标识信息是相同功能相同参数的多个 AI 模型中所述 AI 模型的序号。  
20
7. 根据权利要求 4 所述的装置，其中，所述 AI 模型的复杂度包括部署所述 AI 模型实际所需的第二计算量和/或第二存储空间。
8. 根据权利要求 1 所述的装置，其中，所述请求信息由 RRC 或 MAC CE 或 UCI 承载。
9. 根据权利要求 1 所述的装置，其中，所述反馈信息由 RRC 或 MAC CE 或 DCI  
25 承载。
10. 根据权利要求 1 所述的装置，其中，所述 AI 模型的功能是指所述终端设备的接收和/或发射链路中的部分功能。
11. 根据权利要求 10 所述的装置，其中，所述 AI 模型的功能包括用于 CSI 压缩的 AI 编码器模型或用于波束预测的 AI 模型或用于终端设备定位的 AI 模型。

12. 根据权利要求 1 所述的装置，其中，

所述第一接收单元还用于接收所述网络设备发送的所述 AI 模型传输资源分配信息，所述资源分配信息用于指示传输所述 AI 模型所需的时频域资源。

13. 根据权利要求 12 所述的装置，其中，

5 所述第一接收单元还用于接收所述网络设备在所述时频域资源上发送的所述 AI 模型。

14. 根据权利要求 12 所述的装置，其中，在所述网络设备支持所述终端设备请求时，所述第一接收单元接收所述资源分配信息。

15. 一种信息收发装置，应用于网络设备，其特征在于，所述装置包括：

10 第二接收单元，其接收终端设备发送的用于获取 AI 模型的请求信息；

第二发送单元，向所述终端设备发送响应于所述请求信息的反馈信息。

16. 根据权利要求 15 所述的装置，其中，所述请求信息包括所述 AI 模型的功能标识信息和/或 AI 模型参数信息和/或所述终端设备所能支持的所述 AI 模型的能力信息。

15 17. 根据权利要求 15 所述的装置，其中，所述反馈信息包括是否支持所述终端设备请求的指示信息和/或所述 AI 模型的相关标识信息和/或所述 AI 模型的复杂度。

18. 根据权利要求 16 所述的装置，其中，所述装置还包括：

20 处理单元，其用于根据所述终端设备请求的所述 AI 模型的功能和/或 AI 模型参数信息和/或所述能力信息与所述网络设备存储的各个 AI 模型进行匹配，以确定所述网络设备存储的 AI 模型中是否有满足所述终端设备请求的 AI 模型；

并且，所述第二发送单元基于所述处理单元的匹配结果发送所述反馈信息。

19. 根据权利要求 15 所述的装置，其中，

25 所述第二发送单元还用于向所述终端设备发送所述 AI 模型传输资源分配信息，所述资源分配信息用于指示传输所述 AI 模型所需的时频域资源，以及在所述时频域资源上向所述终端设备发送所述 AI 模型。

20. 一种通信系统，包括终端设备和/或网络设备，所述终端设备包括权利要求 1 中所述的信息收发装置，所述网络设备包括权利要求 15 所述的信息收发装置。

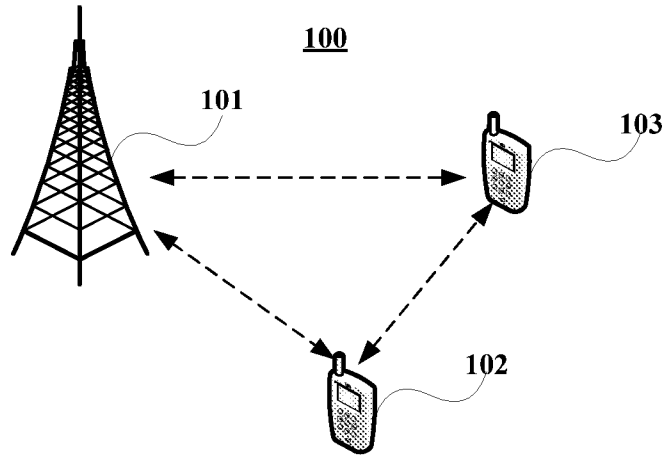


图 1

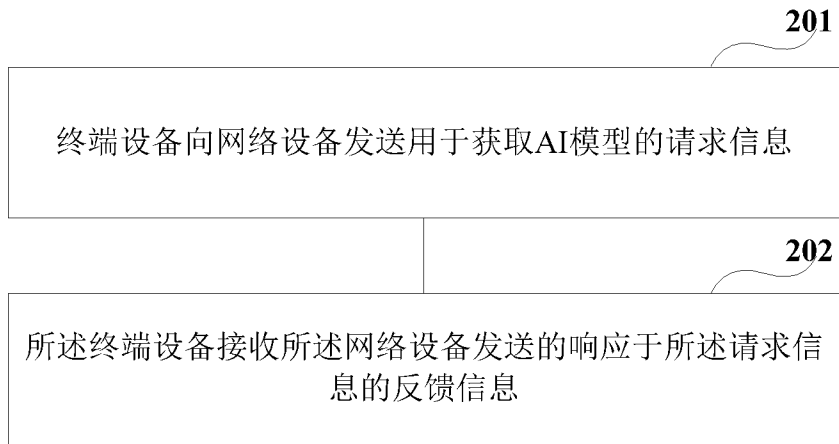


图 2

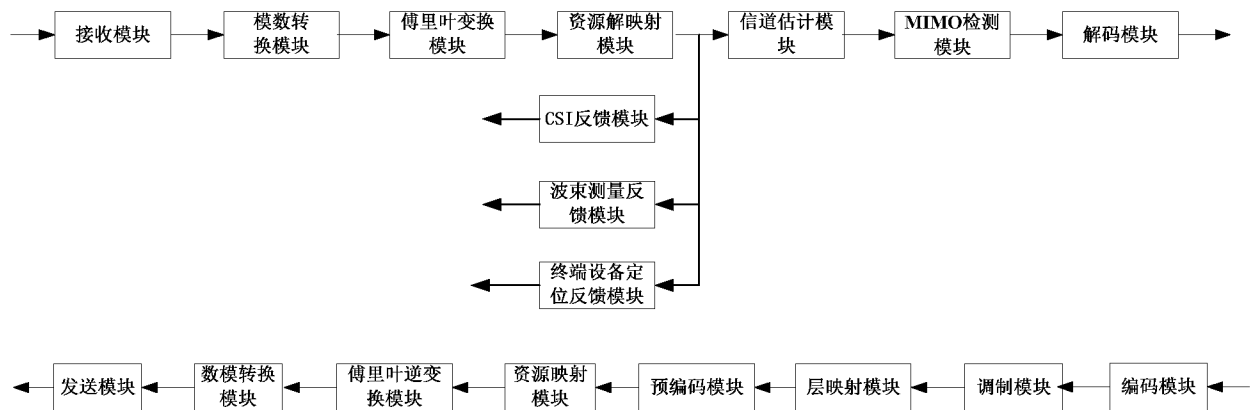


图 3

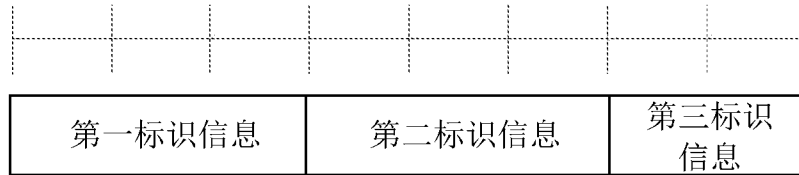


图 4

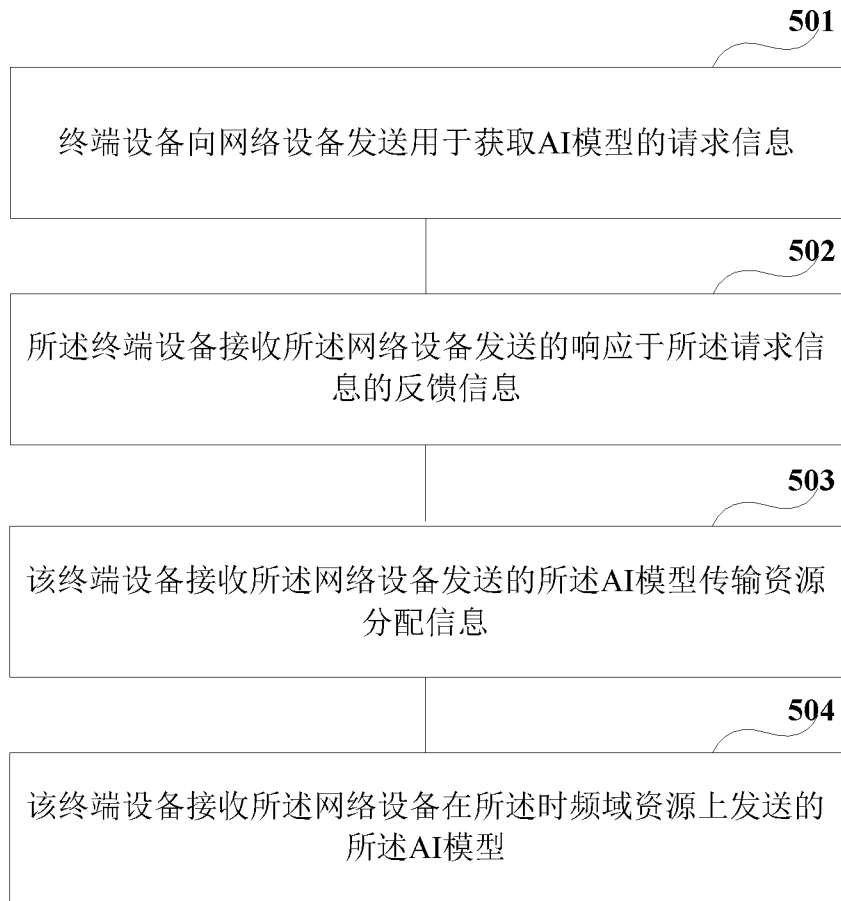


图 5

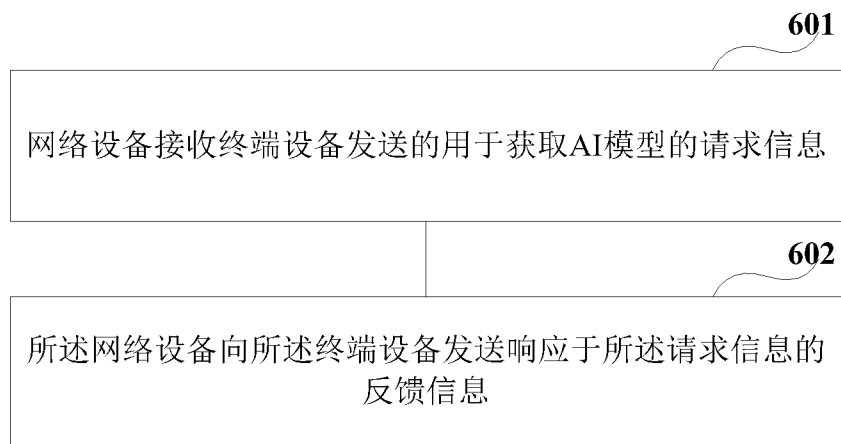


图 6

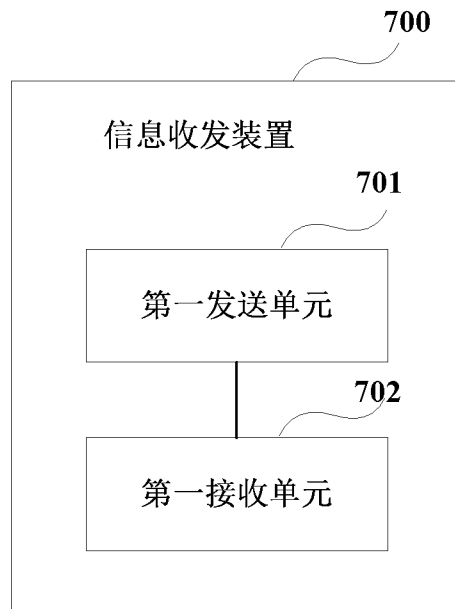


图 7

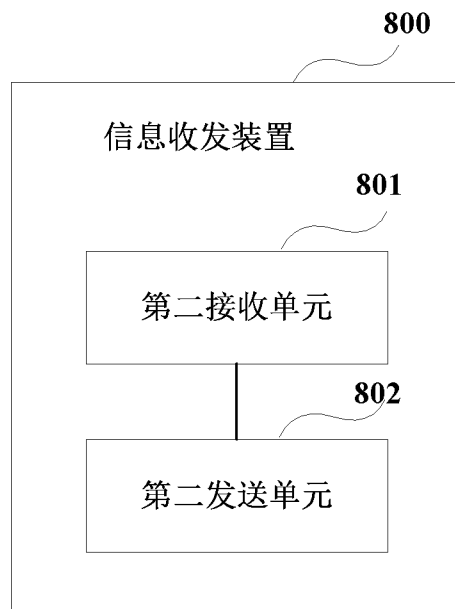


图 8

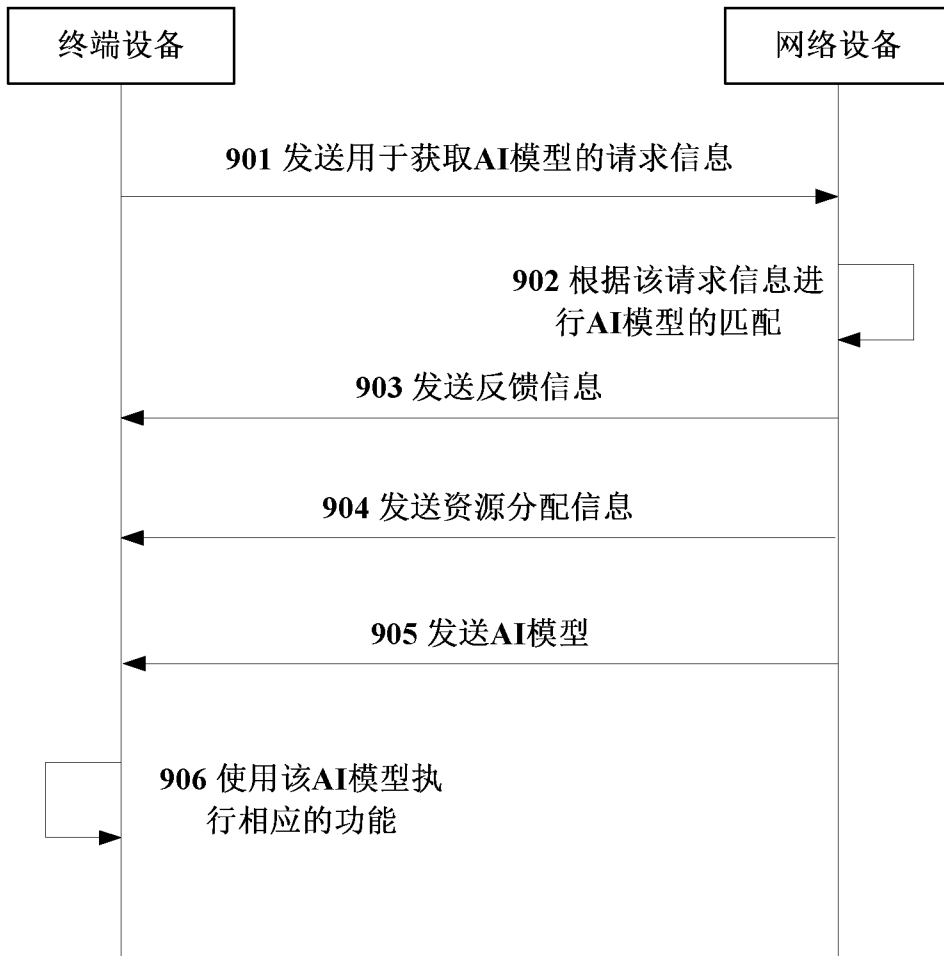


图 9

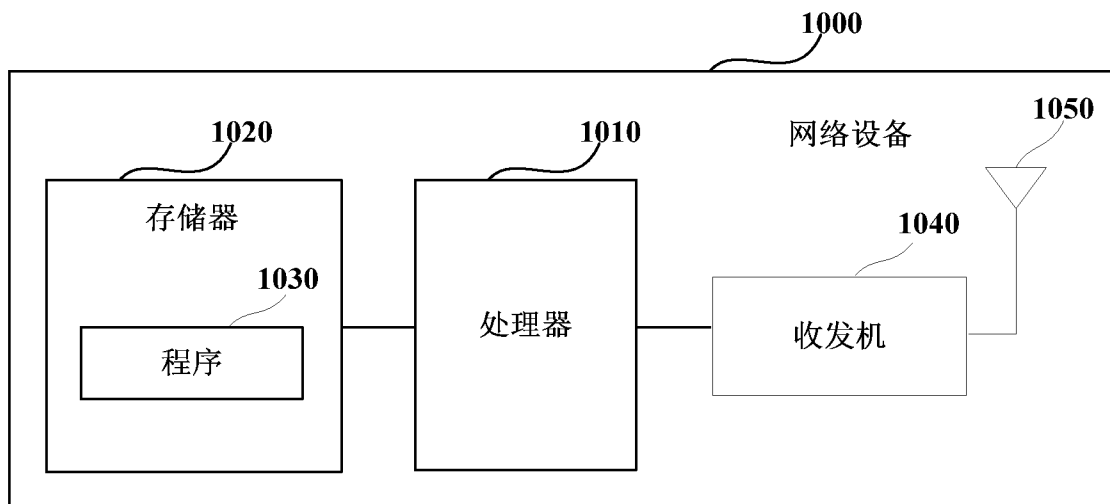


图 10

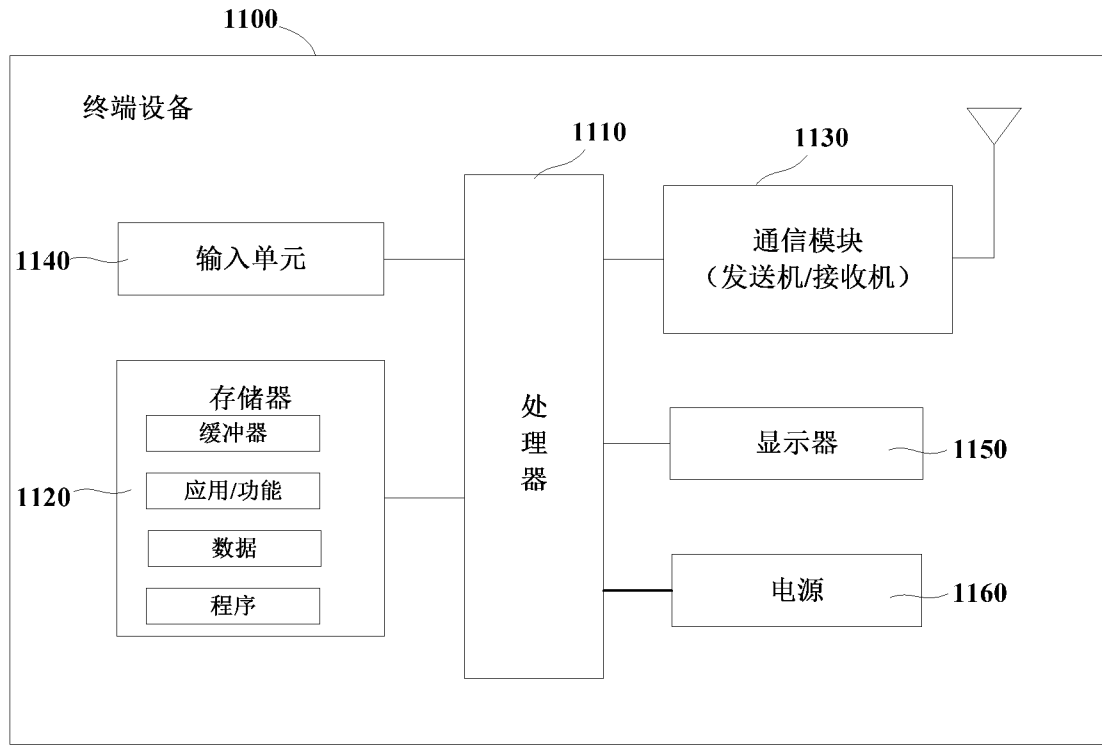


图 11

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/097888

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
H04W 16/22(2009.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols)		
H04W; H04L		
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; CNKI; WPABS; ENTXT; 3GPP: 模型, 人工智能, 获取, 请求, 标识, 参数, 功能, 能力, 网络, 终端, model, artificial intelligence, AI, obtain+, request+, requir+, identif+, ID, parameter, function, capability, network, terminal		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2021163895 A1 (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP. LTD.) 26 August 2021 (2021-08-26) description, page 3, line 11 to page 10, line 29	1-20
X	WO 2022022334 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 03 February 2022 (2022-02-03) description, page 12, line 33 to page 34, line 3	1-20
A	WO 2021248371 A1 (BEIJING XIAOMI MOBILE SOFTWARE CO., LTD.) 16 December 2021 (2021-12-16) entire document	1-20
A	WO 2022028450 A1 (SPREADTRUM SEMICONDUCTOR (NANJING) CO., LTD.) 10 February 2022 (2022-02-10) entire document	1-20
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>		
Date of the actual completion of the international search		Date of mailing of the international search report
21 November 2022		18 January 2023
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2022/097888**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2021163895	A1	26 August 2021	CN	114787793	A	22 July 2022
WO	2022022334	A1	03 February 2022	CN	114071484	A	18 February 2022
WO	2021248371	A1	16 December 2021	CN	111837425	A	27 October 2020
WO	2022028450	A1	10 February 2022	CN	114070676	A	18 February 2022

<p><b>A. 主题的分类</b></p> <p>H04W 16/22 (2009.01) i</p> <p>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</p>																	
<p><b>B. 检索领域</b></p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>H04W; H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNXTX;CNKI;WPABS;ENTXT;3GPP: 模型, 人工智能, 获取, 请求, 标识, 参数, 功能, 能力, 网络, 终端, model, artificial intelligence, AI, obtain+, request+, requir+, identif+, ID, parameter, function, capability, network, terminal</p>																	
<p><b>C. 相关文件</b></p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>WO 2021163895 A1 (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP. LTD.) 2021年8月26日 (2021 - 08 - 26) 说明书第3页第11行-第10页第29行</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>WO 2022022334 A1 (HUAWEI TECH. CO., LTD.) 2022年2月3日 (2022 - 02 - 03) 说明书第12页第33行-第34页第3行</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>WO 2021248371 A1 (BEIJING XIAOMI MOBILE SOFTWARE CO., LTD.) 2021年12月16日 (2021 - 12 - 16) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>WO 2022028450 A1 (SPREADTRUM SEMICONDUCTOR NANJING CO., LTD.) 2022年2月10日 (2022 - 02 - 10) 全文</td> <td>1-20</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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          “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          “&amp;” 同族专利的文件</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	WO 2021163895 A1 (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP. LTD.) 2021年8月26日 (2021 - 08 - 26) 说明书第3页第11行-第10页第29行	1-20	X	WO 2022022334 A1 (HUAWEI TECH. CO., LTD.) 2022年2月3日 (2022 - 02 - 03) 说明书第12页第33行-第34页第3行	1-20	A	WO 2021248371 A1 (BEIJING XIAOMI MOBILE SOFTWARE CO., LTD.) 2021年12月16日 (2021 - 12 - 16) 全文	1-20	A	WO 2022028450 A1 (SPREADTRUM SEMICONDUCTOR NANJING CO., LTD.) 2022年2月10日 (2022 - 02 - 10) 全文	1-20
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	WO 2021163895 A1 (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP. LTD.) 2021年8月26日 (2021 - 08 - 26) 说明书第3页第11行-第10页第29行	1-20															
X	WO 2022022334 A1 (HUAWEI TECH. CO., LTD.) 2022年2月3日 (2022 - 02 - 03) 说明书第12页第33行-第34页第3行	1-20															
A	WO 2021248371 A1 (BEIJING XIAOMI MOBILE SOFTWARE CO., LTD.) 2021年12月16日 (2021 - 12 - 16) 全文	1-20															
A	WO 2022028450 A1 (SPREADTRUM SEMICONDUCTOR NANJING CO., LTD.) 2022年2月10日 (2022 - 02 - 10) 全文	1-20															
国际检索实际完成的日期	国际检索报告邮寄日期																
2022年11月21日	2023年1月18日																
ISA/CN的名称和邮寄地址	授权官员																
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	申砾 电话号码 86-(010)-62411421																

国际检索报告  
关于同族专利的信息

国际申请号  
PCT/CN2022/097888

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)
WO	2021163895	A1	2021年8月26日	CN	114787793	A	2022年7月22日
WO	2022022334	A1	2022年2月3日	CN	114071484	A	2022年2月18日
WO	2021248371	A1	2021年12月16日	CN	111837425	A	2020年10月27日
WO	2022028450	A1	2022年2月10日	CN	114070676	A	2022年2月18日