

(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 7 月 4 日 (04.07.2024)



(10) 国际公布号
WO 2024/140455 A1

(51) 国际专利分类号:
G06F 18/21 (2023.01)

(21) 国际申请号: PCT/CN2023/140908

(22) 国际申请日: 2023 年 12 月 22 日 (22.12.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202211737215.2 2022年12月31日 (31.12.2022) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 严敏瑞 (YAN, Minrui); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。唐文 (TANG, Wen); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。秦钰昆 (QIN, Yukun); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。赵子俊 (ZHAO, Zijun); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。张宇 (ZHANG, Yu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。张瑞 (ZHANG, Rui); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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,

(54) Title: METHOD FOR GENERATING BILL-OF-MATERIAL FILE, AND RELATED DEVICE

(54) 发明名称: 一种生成物料清单文件的方法以及相关设备

生成设备获取目标信息, 目标信息包括训练依赖信息、模型组成信息以及模型元数据, 训练依赖信息为用于训练AI模型的训练资源的信息, 模型组成信息为训练AI模型的过程中的中间模型的信息, 模型元数据为AI模型的属性信息

201



生成设备生成AI模型的物料清单文件, 物料清单文件包括目标信息

202

图 2

- 201 A generation device acquires target information, which includes training dependency information, model composition information and model metadata, wherein the training dependency information is information of training resources used for training an AI model, the model composition information is information of an intermediate model during the process of training the AI model, and the model metadata is attribute information of the AI model
- 202 The generation device generates a bill-of-material file of the AI model, wherein the bill-of-material file comprises the target information

(57) Abstract: Disclosed in the embodiments of the present application are a method for generating a bill-of-material file, and a related device. The method comprises: a generation device acquiring target information, which includes training dependency information, model composition information and model metadata, wherein the training dependency information is information of training resources used for training an AI model, the model composition information is information of an intermediate model during the process of training the AI model, and the model metadata is attribute information of the AI model; and the generation device generating a bill-of-material

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))。
file of the AI model, wherein the bill-of-material file comprises the target information.

(57) 摘要: 本申请实施例公开了一种生成物料清单文件的方法以及相关设备，包括：生成设备获取目标信息，目标信息包括训练依赖信息、模型组成信息以及模型元数据，训练依赖信息为用于训练AI模型的训练资源的信息，模型组成信息为训练AI模型的过程中的中间模型的信息，模型元数据为AI模型的属性信息。生成设备生成AI模型的物料清单文件，物料清单文件包括目标信息。

一种生成物料清单文件的方法以及相关设备

技术领域

5 本申请实施例涉及人工智能领域，尤其涉及一种生成物料清单文件的方法以及相关设备。

背景技术

新一代基于深度神经网络的人工智能，是数据、算法以及算力的结合，需要汇聚大量数据，通过训练得到 AI 模型。在 AI 模型形成的过程中，所涉及的数据集、预训练模型、
10 训练参数设置以及训练算法等涉及到不同的获取途径以及知识产权问题，并且无法确定安全性。另一方面，AI 模型从训练到部署与运行的过程中涉及多方协作，极有可能因为人为操作导致 AI 模型偏离预期，影响业务运行。综上所述，AI 模型存在透明审计、风险管理以及在训练、测试、部署与运行过程中的一致性校验的需求。

传统的应用软件可以基于软件物料清单 (software bill of materials, SBOM) 满足
15 供应链管理以及风险管理需求，然而 AI 模型与应用软件存在较大差异，因此直接复用 SBOM 无法满足上述需求。

发明内容

本申请实施例提供了一种生成物料清单文件的方法以及相关设备，用于支持 AI 模型的透明审计、风险管理及在训练、测试、部署与运行过程中的一致性校验的需求。

20 本申请第一方面提供了一种生成物料清单文件的方法：

生成设备获取目标信息，目标信息包括训练依赖信息、模型组成信息以及模型元数据，训练依赖信息为用于训练 AI 模型的训练资源的信息，模型组成信息为训练 AI 模型的过程中的中间模型的信息，模型元数据为 AI 模型的属性信息。生成设备生成 AI 模型的物料清单文件，物料清单文件包括目标信息。

25 本申请中，由于模型元数据指示了 AI 模型的属性信息，因此基于 AI 模型的物料清单文件能够实现对 AI 模型在训练、测试、部署、运行的过程中的一致性校验，避免 AI 模型被替换。由于训练依赖信息指示了训练 AI 模型的训练资源的信息，模型组成信息指示了训练 AI 模型的过程中的中间模型的信息，因此基于 AI 模型的物料清单文件也能够实现 AI 模型的透明审计以及风险管理。

30 在一种可能的实现方式中，训练依赖信息包括用于训练 AI 模型的数据集的信息。

本申请中，训练依赖信息包括数据集的信息，因此当 AI 模型在实际的应用中出现问题时，若确定问题是由数据集所导致，则可以根据训练依赖信息进行追溯，从而控制风险影响范围。

35 在一种可能的实现方式中，训练依赖信息还包括用于训练 AI 模型的预训练模型的信息，和/或，用于训练 AI 模型的初始化参数的信息，和/或，用于训练 AI 模型的训练脚本的信息。

本申请中，训练依赖信息还包括预训练模型的信息和/或，初始化参数的信息，和/或，训练脚本的信息，因此当 AI 模型在实际的应用中出现问题时，若确定问题是由上述预训练

模型、初始化参数或训练脚本所导致，则可以根据训练依赖信息进行追溯，从而控制风险影响范围。

在一种可能的实现方式中，数据集的信息包括数据集的标识信息、数据集的获取途径、数据集的类型、数据集的用途、数据集的许可信息、数据集的大小、数据集的保存格式、数据集的保存位置、数据集的数据子集的信息、数据集的创建者的信息以及数据集的认证码中的一种或多种，预训练模型的信息包括预训练模型的标识信息、预训练模型的许可信息、预训练模型的创建者的信息以及预训练模型的认证码中的一种或多种，初始化参数的信息包括初始化参数的标识信息、初始化参数的大小、初始化参数的生成时间、初始化参数的保存链接以及初始化参数的认证码中的一种或多种，训练脚本的信息包括训练脚本的标识信息、训练脚本的文件类型、训练脚本的大小、训练脚本的保存路径、训练脚本的创建者的信息以及训练脚本的认证码中的一种或多种。

本申请中，限定了训练依赖信息所包括的具体数据，提高了训练依赖信息支持 AI 模型的透明审计以及风险管理的可行性。

在一种可能的实现方式中，数据子集的信息包括数据子集的名称、数据子集的类型、数据子集的样本的个数、数据子集的保存位置以及数据子集的标记的信息中的一种或多种。

在一种可能的实现方式中，模型组成信息包括中间模型的标识信息、中间模型的文件类型、中间模型的大小、中间模型的创建时间、中间模型的保存路径、中间模型的认证码、中间模型的性能信息以及中间模型的性能信息的认证码中的一种或多种。

本申请中，限定了模型组成信息所包括的具体数据，提高了模型组成信息支持 AI 模型的透明审计以及风险管理的可行性。

在一种可能的实现方式中，模型元数据包括 AI 模型的标识信息、AI 模型的版本信息、AI 模型的描述信息、AI 模型的物料清单文件的保存链接、AI 模型的许可信息、AI 模型的创建者的信息、AI 模型的算力平台的信息以及 AI 模型的认证码中的一种或多种。

本申请中，限定了模型元数据所包括的具体数据，提高了模型元数据支持 AI 模型在训练、测试、部署、运行的过程中的一致性校验的可行性。

在一种可能的实现方式中，目标信息还包括训练过程信息，训练过程信息为 AI 模型的训练过程中的处理步骤的信息。

本申请中，由于训练过程信息指示了 AI 模型的训练过程中的处理步骤的信息，因此 AI 模型的物料清单文件能够进一步支持 AI 模型的透明审计以及风险管理。

在一种可能的实现方式中，训练过程信息包括处理步骤的标识信息、处理步骤的类型、处理步骤的时间戳、处理步骤的参数以及经过处理步骤后得到的中间模型的性能信息的认证码中的一种或多种。

本申请中，限定了训练过程信息所包括的具体数据，提高了训练过程信息支持 AI 模型的透明审计以及风险管理的可行性。

在一种可能的实现方式中，目标信息还包括训练环境信息，训练环境信息为训练 AI 模型所使用的软件与硬件的信息。

本申请中，由于训练环境信息指示了训练 AI 模型所使用的软件与硬件的信息，因此 AI 模型的物料清单文件能够进一步支持 AI 模型的透明审计以及风险管理。

在一种可能的实现方式中，训练环境信息包括软件的标识信息、软件的版本信息、软件的类型、软件的用途以及软件的许可信息中的一种或多种与硬件的标识信息、硬件的型号以及硬件的类型中的一种或多种。

本申请中，限定了训练环境信息所包括的具体数据，提高了训练环境信息支持 AI 模型的透明审计以及风险管理的可行性。

本申请第二方面提供了一种生成设备：

包括获取单元，用于获取目标信息，目标信息包括训练依赖信息、模型组成信息以及模型元数据，训练依赖信息为用于训练 AI 模型的训练资源的信息，模型组成信息为训练 AI 模型的过程中的中间模型的信息，模型元数据为 AI 模型的属性信息。处理单元，用于生成 AI 模型的物料清单文件，物料清单文件包括目标信息。

在一种可能的实现方式中，训练依赖信息包括用于训练 AI 模型的数据集的信息。

在一种可能的实现方式中，训练依赖信息还包括用于训练 AI 模型的预训练模型的信息，和/或，用于训练 AI 模型的初始化参数的信息，和/或，用于训练 AI 模型的训练脚本的信息。

在一种可能的实现方式中，数据集的信息包括数据集的标识信息、数据集的获取途径、数据集的类型、数据集的用途、数据集的许可信息、数据集的大小、数据集的保存格式、数据集的保存位置、数据集的数据子集的信息、数据集的创建者的信息以及数据集的认证码中的一种或多种，预训练模型的信息包括预训练模型的标识信息、预训练模型的许可信息、预训练模型的创建者的信息以及预训练模型的认证码中的一种或多种，初始化参数的信息包括初始化参数的标识信息、初始化参数的大小、初始化参数的生成时间、初始化参数的保存链接以及初始化参数的认证码中的一种或多种，训练脚本的信息包括训练脚本的标识信息、训练脚本的文件类型、训练脚本的大小、训练脚本的保存路径、训练脚本的创建者的信息以及训练脚本的认证码中的一种或多种。

在一种可能的实现方式中，数据子集的信息包括数据子集的名称、数据子集的类型、数据子集的样本的个数、数据子集的保存位置以及数据子集的标记的信息中的一种或多种。

在一种可能的实现方式中，模型组成信息包括中间模型的标识信息、中间模型的文件类型、中间模型的大小、中间模型的创建时间、中间模型的保存路径、中间模型的认证码、中间模型的性能信息以及中间模型的性能信息的认证码中的一种或多种。

在一种可能的实现方式中，模型元数据包括 AI 模型的标识信息、AI 模型的版本信息、AI 模型的描述信息、AI 模型的物料清单文件的保存链接、AI 模型的许可信息、AI 模型的创建者的信息、AI 模型的算力平台的信息以及 AI 模型的认证码中的一种或多种。

在一种可能的实现方式中，目标信息还包括训练过程信息，训练过程信息为 AI 模型的训练过程中的处理步骤的信息。

在一种可能的实现方式中，训练过程信息包括处理步骤的标识信息、处理步骤的类型、处理步骤的时间戳、处理步骤的参数以及经过处理步骤后得到的中间模型的性能信息的认证码中的一种或多种。

在一种可能的实现方式中，目标信息还包括训练环境信息，训练环境信息为训练 AI 模型所使用的软件与硬件的信息。

在一种可能的实现方式中，训练环境信息包括软件的标识信息、软件的版本信息、软件的类型、软件的用途以及软件的许可信息中的一种或多种与硬件的标识信息、硬件的型号以及硬件的类型中的一种或多种。

本申请第三方面提供了一种生成设备：

5 包括存储器、收发器、处理器，其中，存储器用于存储程序。处理器用于执行存储器中的程序，处理器用于根据程序的代码中的指令，使得生成设备执行前述第一方面中的方法。

本申请第四方面提供了一种计算机可读存储介质，其上存储有计算机指令或程序，当计算机指令或程序被执行时，使得计算机执行如前述第一方面中的方法。

10 本申请第五方面提供了一种计算机程序产品，包括计算机指令或程序，当计算机指令或程序被执行时，使得计算机执行如前述第一方面中的方法。

附图说明

图 1 为本申请中的生成物料清单文件的方法所应用的系统架构的示意图；

图 2 为本申请中的生成物料清单文件的方法的一个流程示意图；

15 图 3a 为本申请中的生成物料清单文件的方法的一个示意图；

图 3b 为本申请中的生成物料清单文件的方法的另一示意图；

图 4 为本申请中的生成物料清单文件的方法的应用场景的示意图；

图 5 为本申请中的生成设备的一个结构示意图；

图 6 为本申请中的生成设备的另一结构示意图。

20 具体实施方式

下面结合附图，对本申请的实施例进行描述，显然，所描述的实施例仅仅是本申请一部分的实施例，而不是全部的实施例。本领域普通技术人员可知，随着技术发展和新场景的出现，本申请提供的技术方案对于类似的技术问题，同样适用。

25 本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的实施例能够以除了在这里图示或描述的内容以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它

30 步骤或单元。

为了便于对本申请的理解，下面先对本申请涉及的概念进行介绍：

AI 模型：用于描述机器学习推理过程的算法结构文件，通常包含算子、连接关系与权重值等信息。机器学习通过输入推理请求数据，执行模型描述的复杂计算过程，最终获得运算结果，实现推理功能。

35 预训练模型 (Pre-trained Model)：已经用数据集训练好的模型，但其性能等还未达到直接使用的程度。

模型微调 (Model Fine-tuning)：基于给定预训练模型进行微调 (Fine Tune)，将预训练模型变为可直接投入业务应用的成熟模型。相对于从头开始训练，模型微调将节省大量

计算资源和计算时间，提高了计算效率，甚至提高准确率。

机器学习（Machine Learning）：是人工智能学科领域中的重要分支，研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。深度学习则是机器学习中重要的研究方向。AI 模型是机器学习中最重要也是最有价值的资产之一。

物料清单（Bill of Material）：也称为产品结构，是构建、制造或维修产品或服务所需的所有物料的列表。物料清单包含创建产品或服务所涉及的原材料、子装配体、中间装配体、零部件和零件的广泛列表，以及每种零件的数量、成本和如何装配它们的说明。

AI 模型形成的过程中，涉及一个或多个数据集、预训练模型、训练参数设置、训练算法等，其中所依赖的数据集、预训练模型及训练算法可能是私有、开源或外购的，涉及不同的知识产权问题，且数据集、预训练模型以及训练算法是否可能潜藏恶意样本、后门及漏洞等，都可能直接影响 AI 模型的安全性。另一方面，AI 模型从训练到部署与运行的过程中，可能涉及数据提供方、数据加工方、预训练模型创建者、算力平台提供方、AI 应用开发运行方等多个相关参与方以及 AI 模型训练方内部的开发、测试及应用等多个团队协作，并且 AI 模型训练过程涉及多轮迭代优化，会形成诸多的 AI 模型版本，有极大概率因为参与方的误操作或恶意注入后门等恶意样本导致最终部署应用的 AI 模型偏离预期，影响业务运行。

同时，产业也出现了非常强烈的 AI 模型透明可审计、可追溯、一致性校验相关需求。例如要求在算法设计、实现、应用等环节，提升透明性、可解释性、可理解性、可靠性、可控性，增强人工智能系统的韧性、自适应性和抗干扰能力，逐步实现可验证、可审核、可监督、可追溯、可预测、可信赖。或者要求训练数据应有完整签名或校验码、应记录部署的软硬件环境配置信息、对 AI 模型进行管理，提供对 AI 模型的可回溯方面功能、对 AI 模型的版本进行管理以及对版本可回溯等。在实际的 AI 模型从训练到部署使用过程中，涉及数据集构建团队、模型训练团队、模型测试团队、模型部署使用团队等多个开发团队协作，且最终 AI 模型还可能被部署到边缘环境使用，面临的风险较大，且与管理员之间弱交互，存在 AI 模型的训练、测试或部署使用过程中，AI 模型被篡改、替换或人为误操作等部署了非预期的 AI 模型版本等风险，因此，存在强烈的 AI 模型一致性追溯需求。另一方面，在 AI 模型形成过程中涉及多个团队协作，可能存在内部恶意人员在数据集中插入了恶意样本等未被及时发现的情况，并且含有恶意样本的数据集可能在多个 AI 模型的训练过程中复用，那么如在后续的 AI 模型使用过程中发现 AI 模型出现问题，应能够追溯出导致问题的数据集，同时确定出该数据集还影响了哪些 AI 模型，及时更新所有受影响的 AI 模型，把对业务的影响降低到最低。除此之外，国际上也已经有多项法规及标准提出类似要求，例如，欧盟 AI regulation Article 12 要求确保人工智能系统全生命周期运作可追溯，用途符合预期，监控高风险人工智能系统的运行。欧洲人工智能高级别专家组也要求应确保 AI 系统相关元素的可追溯性，包括数据、系统和商业模型。ETSI SAI GR010 Traceability of AI Models 要求 AI 模型的可追溯性包括数据的可追溯性、the processing pipeline 可追溯性、模型输出、模型元数据和生命周期管理。

综上所述，AI 模型存在透明审计、风险管理以及在训练、测试、部署与运行过程中的

一致性校验的需求。传统的应用软件能够基于 SBOM 满足供应链管理以及风险管理需求，诸多标准组织也定义了 SBOM 的组成规范，例如 Linux Foundation 制定了 SPDX 规范，旨在创建一种数据交换格式，以便以共同格式收集和共享有关软件包和相关内容的信息，以节省时间和提高数据准确性，使公司和组织能够共享人类可读和机器可处理的软件包元数据，以促进软件供应链管理。SPDX 包括如下信息：

SPDX 文档创建信息：用于将分析结果与 SPDX 文件的特定版本和许可证关联起来的元数据，并提供有关 SPDX 文件创建方式、时间和由谁创建的信息；

包信息：整个软件包相关的共性属性信息；

文件信息：软件包中可能包含的文件相关信息；

代码段信息：特定文件的部分信息；

其他 license 信息：捕获不在 SPDX 许可证列表中的相关 license 信息；

SPDX 元素之间的关系：关于文档、包和文件如何相互关联的信息；

注释：有关 SPDX 文件审查的时间和由谁审查的信息。

然而 AI 模型与应用软件存在较大区别，直接复用 SBOM 无法满足上述需求。

本申请提供了一种生成物料清单文件的方法以及相关设备，用于支持 AI 模型的透明审计、风险管理及在训练、测试、部署与运行过程中的一致性校验的需求。

本申请可以应用于如图 1 所示的场景中，AI 模型的产生主要依赖于模型厂商以及算力平台，算力平台负责提供并管理 AI 模型训练设备，上面安装了基本的 AI 训练或 AI 推理的软件栈，包括 AI 处理器加速库、训练框架、AI 应用使能开发套件等，模型厂商会向 AI 模型训练设备上传训练 AI 模型所需要的训练资源，例如初始化参数、训练脚本、数据集、神经网络结构、预训练模型以及超参数等，从而进行模型的训练迭代，并经过一个或多个中间模型的迭代后最终得到所需的 AI 模型。

本申请中的生成设备为上述 AI 模型训练设备，请参阅图 2，下面对本申请中的生成物料清单文件的方法的流程进行介绍：

201、生成设备获取目标信息，目标信息包括训练依赖信息、模型组成信息以及模型元数据，训练依赖信息为用于训练 AI 模型的训练资源的信息，模型组成信息为训练 AI 模型的过程中的中间模型的信息，模型元数据为 AI 模型的属性信息；

在训练开始之前，模型厂商会向 AI 模型训练设备上传用于训练 AI 模型的训练资源，请参阅图 3a，AI 模型训练设备根据训练资源自动抽取训练依赖信息，该训练依赖信息也即为训练资源的信息。示例性的，训练依赖信息包括用于训练 AI 模型的数据集的信息，可选的，训练依赖信息也可以进一步包括训练 AI 模型的预训练模型的信息，和/或，初始化参数的信息，和/或，训练脚本的信息。

示例性的，数据集的信息包括数据集的标识信息、数据集的获取途径、数据集的类型、数据集的用途、数据集的许可信息、数据集的大小、数据集的保存格式、数据集的保存位置、数据集中的数据子集的信息、数据集的创建者的信息以及数据集的认证码中的一种或多种。上述数据集的标识信息例如包括数据集的标识 ID 以及数据集的名称，数据集的获取途径例如为 imagenet 或者 COCO，数据集的类型例如为图片、视频、语音或者文本，数据集的用途例如为用于图片分类或者用于目标检测，数据集的创建者的信息例如包括数据集

的创建时间以及创建者的公钥信息，数据集的认证码例如可以是数据集对应的散列码，或者也可以是用私钥对该散列码进行加密后的签名。

数据子集的信息例如包括数据子集的名称、数据子集的类型、数据子集中的样本的个数、数据子集的保存位置以及数据子集的标记的信息中的一种或多种，上述数据子集的类型例如包括用于训练、验证或测试的数据子集，标记的信息例如包括标记的文件格式、标记的保存位置中的一种或多种。上述标记的文件格式，可以是.json、.xml 等多种类型，例如，用于图像分类的数据子集中的标记通常采用.json 的文件格式。

预训练模型的信息包括预训练模型的标识信息、预训练模型的许可信息、预训练模型的创建者的信息以及预训练模型的认证码中的一种或多种，预训练模型的创建者的信息例如包括预训练模型的创建时间以及创建者的公钥信息。初始化参数的信息包括初始化参数的标识信息、初始化参数的大小、初始化参数的生成时间、初始化参数的保存链接以及初始化参数的认证码中的一种或多种。训练脚本的信息包括训练脚本的标识信息、训练脚本的文件类型、训练脚本的大小、训练脚本的保存路径、训练脚本的创建者的信息以及训练脚本的认证码中的一种或多种。

可选的，AI 模型训练设备还会自动抽取训练环境信息，训练环境信息为训练 AI 模型所使用的软件与硬件的信息。示例性的，训练环境信息包括软件的标识信息、软件的版本信息、软件的类型、软件的用途以及软件的许可信息中的一种或多种与硬件的标识信息、硬件的型号以及硬件的类型中的一种或多种。

在训练迭代的过程中，AI 模型训练设备还会自动抽取模型组成信息，模型组成信息为训练 AI 模型的过程中的一个或多个中间模型的信息。示例性的，模型组成信息包括中间模型的标识信息、中间模型的文件类型、中间模型的大小、中间模型的创建时间、中间模型的保存路径、中间模型的认证码、中间模型的性能信息以及中间模型的性能信息的认证码中的一种或多种。上述性能信息可以是借助主流训练框架提供的性能记录文件，例如在Tensorflow 训练框架内置了一个可视化工具开发的 AI 模型内部参数跟踪和可视化的调试工具 Tensorboard，TensorBoard 能够保存 AI 模型训练的性能数据，性能数据的文件格式为 tfevents 文件，PyTorch 框架自 1.2.0 也开始支持 TensorBoard。使用 TensorBoard 可以监控 AI 模型的各种指标的变化，如 acc、loss 的动态变化，可以将 AI 模型结构可视化，可以可视化词嵌入空间，可以分析模型性能，可以分析数据集的公平性等等。上述中间模型的文件类型，基于不同训练框架导出的模型文件类型存在差异，例如 TensorFlow 框架导出的模型文件包含模型参数.data 文件，模型网络结构.meta 文件以及映射信息.index 文件，.index 文件保存的是.data 文件中数据和 .meta 文件中结构图之间的对应关系。而在 pytorch 框架中最常见的模型保存使用 .pt 或者是 .pth 作为模型文件扩展名。

此外，对于上述训练依赖信息、训练环境信息以及模型组成信息，AI 模型训练设备均会获取其对应的认证码，该认证码可以是散列码，或者也可以是用算力平台的私钥对散列码进行加密所得到的签名，例如 AI 模型训练设备会获取训练依赖信息对应的散列码，或者进一步用算力平台的私钥对上述散列码进行加密从而得到签名。

模型厂商可以通过 AI 模型训练设备所提供的接口查看上述训练依赖信息、训练环境信息、模型组成信息以及上述每种信息对应的认证码，并进一步获通过该接口向 AI 模型训练

设备输入模型元数据，该模型元数据为 AI 模型的属性信息，示例性的，模型元数据包括 AI 模型的标识信息、AI 模型的版本信息、AI 模型的描述信息、物料清单文件的保存链接、AI 模型的许可信息、AI 模型的创建者的信息、训练 AI 模型的算力平台的信息以及 AI 模型的认证码中的一种或多种，上述训练 AI 模型的算力平台的信息包括该算力平台的公钥信息，由于 AI 模型的创建者为模型厂商，因此 AI 模型的创建者的信息包括 AI 模型的创建时间以及模型厂商的公钥信息。

可选的，模型厂商还会通过接口向 AI 模型训练设备输入训练过程信息，训练过程信息为 AI 模型的训练过程中的处理步骤的信息。示例性的，训练过程信息包括处理步骤的标识信息、经过一个处理步骤后得到的中间模型的性能信息的认证码、处理步骤的类型、处理步骤的时间戳以及处理步骤的参数中的一种或多种，上述处理步骤的类型例如包括数据采用或数据增强，处理步骤的参数例如包括 epoch、batch、损失函数定义以及最大模型大小设置。

202、生成设备生成 AI 模型的物料清单文件，物料清单文件包括目标信息。

AI 模型训练设备将获取到的训练依赖信息、训练环境信息、模型组成信息、模型元数据以及训练过程信息构建成为物料清单文件，该物料清单文件可以为 XML 格式或者 JSON 格式，此外 AI 模型训练设备还会使用模型厂商的私钥对该物料清单文件进行签名。

在一种情况中，前述预训练模型的信息中还可以包括预训练模型的物料清单文件的保存链接，预训练模型的物料清单文件与 AI 模型的物料清单文件所包含的信息类似，具体不再赘述。

又或者，请参阅图 3b，在另外一种实现方式中，模型厂商向 AI 模型训练设备上传训练资源，AI 模型训练设备根据训练资源进行模型的训练迭代，然而 AI 模型训练设备不会获取训练依赖信息、训练环境信息以及模型组成信息，而是在 AI 模型训练完成之后，由模型厂商自行获取训练依赖信息、训练环境信息、模型组成信息、模型元数据以及训练过程信息，并根据模型厂商的私钥对上述每种信息进行签名，将上述每种信息输入到 AI 模型训练设备中，使得 AI 模型训练设备生成 AI 模型的物料清单文件，并使用模型厂商的私钥对 AI 模型的物料清单文件进行签名。当然，上述训练环境信息以及训练过程信息同样为可选的。

当然，在实现的实现当中，AI 模型训练设备获取到训练依赖信息、训练环境信息、模型组成信息、模型元数据以及训练过程信息也可以通过其他方式实现，具体不再展开说明。

请参阅图 4，在得到 AI 模型的物料清单文件后，能够应用于实现 AI 模型在训练、测试、部署、运行的一致性校验以及在 AI 模型的生命周期中实现透明审计以及风险管理的两大应用场景，下面分别进行介绍：

对于实现 AI 模型在训练、测试、部署、运行的过程中的一致性校验，例如首先根据模型元数据中的模型厂商的公钥信息以及 AI 模型的物料清单文件的认证码对 AI 模型的物料清单文件的真实性以及完整性进行校验，在 AI 模型发布、集成、部署、运行与维护过程中，获取模型元数据中的 AI 模型的认证码，从而对 AI 模型的完整性进行验证。也可以根据模型元数据确定 AI 模型是否为业务所预期部署的模型，例如对部署在端侧/边缘侧的 AI 模型，在加载运行之前，根据 AI 模型的标识信息以及版本信息校验 AI 模型是否正确，避免模型

被替换，必要时可进一步利用模型组成信息以及模型依赖信息对 AI 模型进行校验。

对于在 AI 模型的生命周期中实现透明审计以及风险管理，首先可以根据模型元数据中的模型厂商的公钥信息以及 AI 模型的物料清单文件的认证码对 AI 模型的物料清单文件的真实性以及完整性进行校验。其次在 AI 模型的集成阶段，从模型组成信息中获得中间模型的信息，从模型依赖信息中获得用于训练 AI 模型的训练资源的信息与其对应的许可信息，并利用模型元数据中的算力平台的公钥信息校验上述信息的真实性与完整性，从而支持供应链的合规管理及透明审计。在 AI 模型的部署、运行与维护阶段，基于模型依赖信息、模型组成信息、训练环境信息以及训练过程信息支持漏洞影响分析与风险管理，例如若 AI 模型在实际的应用过程中出现问题后，若是训练资源导致的问题，则可批量追溯出与上述训练资源存在关联的其他 AI 模型，从而控制影响范围，基于模型组成信息以及训练过程信息能够向模型使用方呈现模型的训练过程以及中间结果，提升模型训练过程透明性，支撑模型使用者在发现模型偏离预期时，能够基于上述信息回溯问题，发现一些人为或误操作的问题，例如，发现模型推理结果表现为过拟合，可回溯模型训练迭代次数，是否设置过大等。训练环境信息可支撑模型使用方了解和维护模型训练的软硬件环境信息，从而提升模型训练过程的可重复性，支撑其审计和发现模型训练过程及结果的准确性、分析模型偏离预期的原因等

本申请中，由于模型元数据指示了 AI 模型的属性信息，因此基于 AI 模型的物料清单文件能够实现对 AI 模型在训练、测试、部署、运行的过程中的一致性校验，避免 AI 模型被替换。由于训练依赖信息指示了训练 AI 模型的训练资源的信息，模型组成信息指示了训练 AI 模型的过程中的中间模型的信息，因此基于 AI 模型的物料清单文件也能够实现 AI 模型的透明审计以及风险管理。

上面对本申请中的生成物料清单文件的方法进行了介绍，下面对本申请中的生成设备进行介绍：

请参阅图 5，本申请中的生成设备 500 包括获取单元 501 以及处理单元 502。

获取单元 501，用于获取目标信息，目标信息包括训练依赖信息、模型组成信息以及模型元数据，训练依赖信息为用于训练 AI 模型的训练资源的信息，模型组成信息为训练 AI 模型的过程中的中间模型的信息，模型元数据为 AI 模型的属性信息。

处理单元 502，用于生成 AI 模型的物料清单文件，物料清单文件包括目标信息。

在一种可能的实现方式中，训练依赖信息包括用于训练 AI 模型的数据集的信息。

在一种可能的实现方式中，训练依赖信息还包括用于训练 AI 模型的预训练模型的信息，和/或，用于训练 AI 模型的初始化参数的信息，和/或，用于训练 AI 模型的训练脚本的信息。

在一种可能的实现方式中，数据集的信息包括数据集的标识信息、数据集的获取途径、数据集的类型、数据集的用途、数据集的许可信息、数据集的大小、数据集的保存格式、数据集的保存位置、数据集的数据子集的信息、数据集的创建者的信息以及数据集的认证码中的一种或多种，预训练模型的信息包括预训练模型的标识信息、预训练模型的许可信息、预训练模型的创建者的信息以及预训练模型的认证码中的一种或多种，初始化参数的信息包括初始化参数的标识信息、初始化参数的大小、初始化参数的生成时间、初始化参

数的保存链接以及初始化参数的认证码中的一种或多种，训练脚本的信息包括训练脚本的标识信息、训练脚本的文件类型、训练脚本的大小、训练脚本的保存路径、训练脚本的创建者的信息以及训练脚本的认证码中的一种或多种。

在一种可能的实现方式中，数据子集的信息包括数据子集的名称、数据子集的类型、数据子集的样本的个数、数据子集的保存位置、数据子集的标记的信息中的一种或多种。

在一种可能的实现方式中，模型组成信息包括中间模型的标识信息、中间模型的文件类型、中间模型的大小、中间模型的创建时间、中间模型的保存路径、中间模型的认证码、中间模型的性能信息以及中间模型的性能信息的认证码中的一种或多种。

在一种可能的实现方式中，模型元数据包括 AI 模型的标识信息、AI 模型的版本信息、AI 模型的描述信息、AI 模型的物料清单文件的保存链接、AI 模型的许可信息、AI 模型的创建者的信息、AI 模型的算力平台的信息以及 AI 模型的认证码中的一种或多种。

在一种可能的实现方式中，目标信息还包括训练过程信息，训练过程信息为 AI 模型的训练过程中的处理步骤的信息。

在一种可能的实现方式中，训练过程信息包括处理步骤的标识信息、处理步骤的类型、处理步骤的时间戳、处理步骤的参数以及经过处理步骤后得到的中间模型的性能信息的认证码中的一种或多种。

在一种可能的实现方式中，目标信息还包括训练环境信息，训练环境信息为训练 AI 模型所使用的软件与硬件的信息。

在一种可能的实现方式中，训练环境信息包括软件的标识信息、软件的版本信息、软件的类型、软件的用途以及软件的许可信息中的一种或多种与硬件的标识信息、硬件的型号以及硬件的类型中的一种或多种。

图 6 是本申请实施例提供的生成设备的结构示意图，该生成设备 600 可以包括一个或一个以上中央处理器（central processing units, CPU）601 和存储器 605，该存储器 605 中存储有一个或一个以上的应用程序或数据。

其中，存储器 605 可以是易失性存储或持久存储。存储在存储器 605 的程序可以包括一个或一个以上模块，每个模块可以包括一系列指令操作。更进一步地，中央处理器 601 可以设置为与存储器 605 通信，在生成设备 600 上执行存储器 605 中的一系列指令操作。

生成设备 600 还可以包括一个或一个以上电源 602，一个或一个以上有线或无线网络接口 603，一个或一个以上输入输出接口 604，和/或，一个或一个以上操作系统。中央处理器 601 可以执行前述图 2 所示实施例中生成设备所执行的操作，具体此处不再赘述。

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。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元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，可以存储在一个计算机可读取存储介质中。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全部或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（ROM，read-only memory）、随机存取存储器（RAM，random access memory）、磁碟或者光盘等各种可以存储程序代码的介质。

权 利 要 求 书

1. 一种生成物料清单文件的方法，其特征在于，包括：

生成设备获取目标信息，所述目标信息包括训练依赖信息、模型组成信息以及模型元数据，所述训练依赖信息为用于训练所述 AI 模型的训练资源的信息，所述模型组成信息为训练所述 AI 模型的过程中的中间模型的信息，所述模型元数据为所述 AI 模型的属性信息；
所述生成设备生成所述 AI 模型的物料清单文件，所述物料清单文件包括所述目标信息。

2. 根据权利要求 1 所述的方法，其特征在于，所述训练依赖信息包括用于训练所述 AI 模型的数据集的信息。

3. 根据权利要求 2 所述的方法，其特征在于，所述训练依赖信息还包括用于训练所述 AI 模型的预训练模型的信息，和/或，用于训练所述 AI 模型的初始化参数的信息，和/或，用于训练所述 AI 模型的训练脚本的信息。

4. 根据权利要求 3 所述的方法，其特征在于，所述数据集的信息包括所述数据集的标识信息、获取途径、类型、用途、许可信息、大小、保存格式、保存位置、数据子集的信息、创建者的信息以及认证码中的一种或多种，所述预训练模型的信息包括所述预训练模型的标识信息、许可信息、创建者的信息以及认证码中的一种或多种，所述初始化参数的信息包括所述初始化参数的标识信息、大小、生成时间、保存链接以及认证码中的一种或多种，所述训练脚本的信息包括所述训练脚本的标识信息、文件类型、大小、保存路径、创建者的信息以及认证码中的一种或多种。

5. 根据权利要求 4 所述的方法，其特征在于，所述数据子集的信息包括所述数据子集的名称、类型、样本的个数、保存位置以及标记的信息中的一种或多种。

6. 根据权利要求 1 至 5 中任一项所述的方法，其特征在于，所述模型组成信息包括所述中间模型的标识信息、文件类型、大小、创建时间、保存路径、认证码、性能信息以及所述性能信息的认证码中的一种或多种。

7. 根据权利要求 6 所述的方法，其特征在于，所述模型元数据包括所述 AI 模型的标识信息、版本信息、描述信息、所述物料清单文件的保存链接、许可信息、创建者的信息、算力平台的信息以及认证码中的一种或多种。

8. 根据权利要求 7 所述的方法，其特征在于，所述目标信息还包括训练过程信息，所述训练过程信息为所述 AI 模型的训练过程中的处理步骤的信息。

9. 根据权利要求 8 所述的方法，其特征在于，所述训练过程信息包括所述处理步骤的标识信息、类型、时间戳、参数以及经过所述处理步骤后得到的所述中间模型的所述性能信息的认证码中的一种或多种。

10. 根据权利要求 9 所述的方法，其特征在于，所述目标信息还包括训练环境信息，所述训练环境信息为训练所述 AI 模型所使用的软件与硬件的信息。

11. 根据权利要求 10 所述的方法，其特征在于，所述训练环境信息包括所述软件的标识信息、版本信息、类型、用途以及许可信息中的一种或多种与所述硬件的标识信息、型号以及类型中的一种或多种。

12. 一种生成设备，其特征在于，包括：

获取单元，用于获取目标信息，所述目标信息包括训练依赖信息、模型组成信息以及模型元数据，所述训练依赖信息为用于训练所述 AI 模型的训练资源的信息，所述模型组成信息为训练所述 AI 模型的过程中的中间模型的信息，所述模型元数据为所述 AI 模型的属性信息；

5 处理单元，用于生成所述 AI 模型的物料清单文件，所述物料清单文件包括所述目标信息。

13. 根据权利要求 12 所述的生成设备，其特征在于，所述训练依赖信息包括用于训练所述 AI 模型的数据集的信息。

10 14. 根据权利要求 13 所述的生成设备，其特征在于，所述训练依赖信息还包括用于训练所述 AI 模型的预训练模型的信息，和/或，用于训练所述 AI 模型的初始化参数的信息，和/或，用于训练所述 AI 模型的训练脚本的信息。

15 15. 根据权利要求 14 所述的生成设备，其特征在于，所述数据集的信息包括所述数据集的标识信息、获取途径、类型、用途、许可信息、大小、保存格式、保存位置、数据子集的信息、创建者的信息以及认证码中的一种或多种，所述预训练模型的信息包括所述预训练模型的标识信息、许可信息、创建者的信息以及认证码中的一种或多种，所述初始化参数的信息包括所述初始化参数的标识信息、大小、生成时间、保存链接以及认证码中的一种或多种，所述训练脚本的信息包括所述训练脚本的标识信息、文件类型、大小、保存路径、创建者的信息以及认证码中的一种或多种。

20 16. 根据权利要求 15 所述的生成设备，其特征在于，所述数据子集的信息包括所述数据子集的名称、类型、样本的个数、保存位置以及标记的信息中的一种或多种。

17. 根据权利要求 12 至 16 中任一项所述的生成设备，其特征在于，所述模型组成信息包括所述中间模型的标识信息、文件类型、大小、创建时间、保存路径、认证码、性能信息以及所述性能信息的认证码中的一种或多种。

25 18. 根据权利要求 17 所述的生成设备，其特征在于，所述模型元数据包括所述 AI 模型的标识信息、版本信息、描述信息、所述物料清单文件的保存链接、许可信息、创建者的信息、算力平台的信息以及认证码中的一种或多种。

19. 根据权利要求 18 所述的生成设备，其特征在于，所述目标信息还包括训练过程信息，所述训练过程信息为所述 AI 模型的训练过程中的处理步骤的信息。

30 20. 根据权利要求 19 所述的生成设备，其特征在于，所述训练过程信息包括所述处理步骤的标识信息、类型、时间戳、参数以及经过所述处理步骤后得到的所述中间模型的所述性能信息的认证码中的一种或多种。

21. 根据权利要求 20 所述的生成设备，其特征在于，所述目标信息还包括训练环境信息，所述训练环境信息为训练所述 AI 模型所使用的软件与硬件的信息。

35 22. 根据权利要求 21 所述的生成设备，其特征在于，所述训练环境信息包括所述软件的标识信息、版本信息、类型、用途以及许可信息中的一种或多种与所述硬件的标识信息、型号以及类型中的一种或多种。

23. 一种生成设备，其特征在于，包括存储器、收发器、处理器；
其中，所述存储器用于存储程序；

所述处理器用于执行所述存储器中的程序，所述处理器用于根据所述程序的代码中的指令，使得所述生成设备执行权利要求 1 至 11 中任一项所述的方法。

24. 一种计算机可读存储介质，其上存储有计算机指令或程序，其特征在于，所述计算机指令或程序被执行时，使得计算机执行如权利要求 1 至 11 中任一项所述的方法。

5 25. 一种计算机程序产品，包括计算机指令或程序，其特征在于，所述计算机指令或程序被执行时，使得计算机执行如权利要求 1 至 11 中任一项所述的方法。

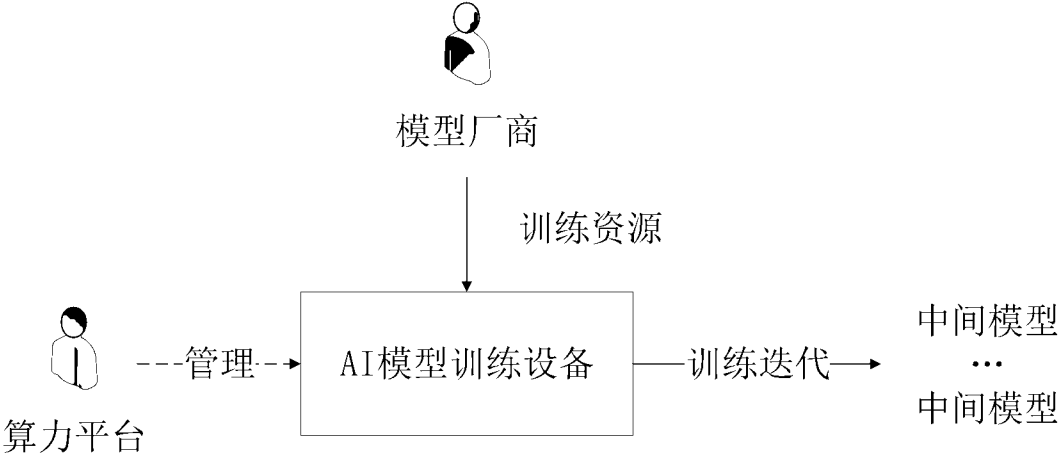


图 1

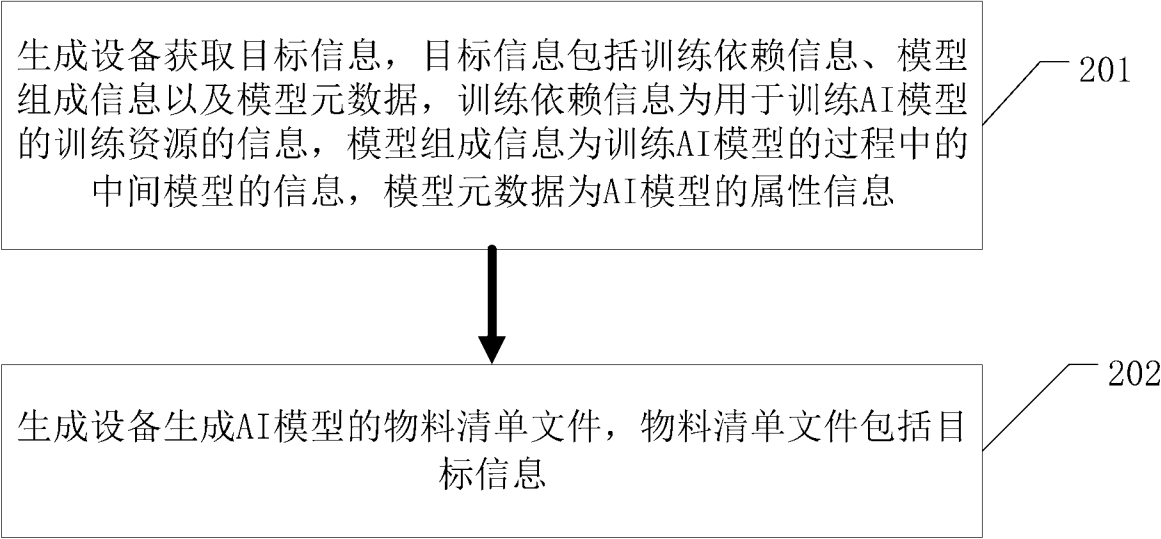


图 2

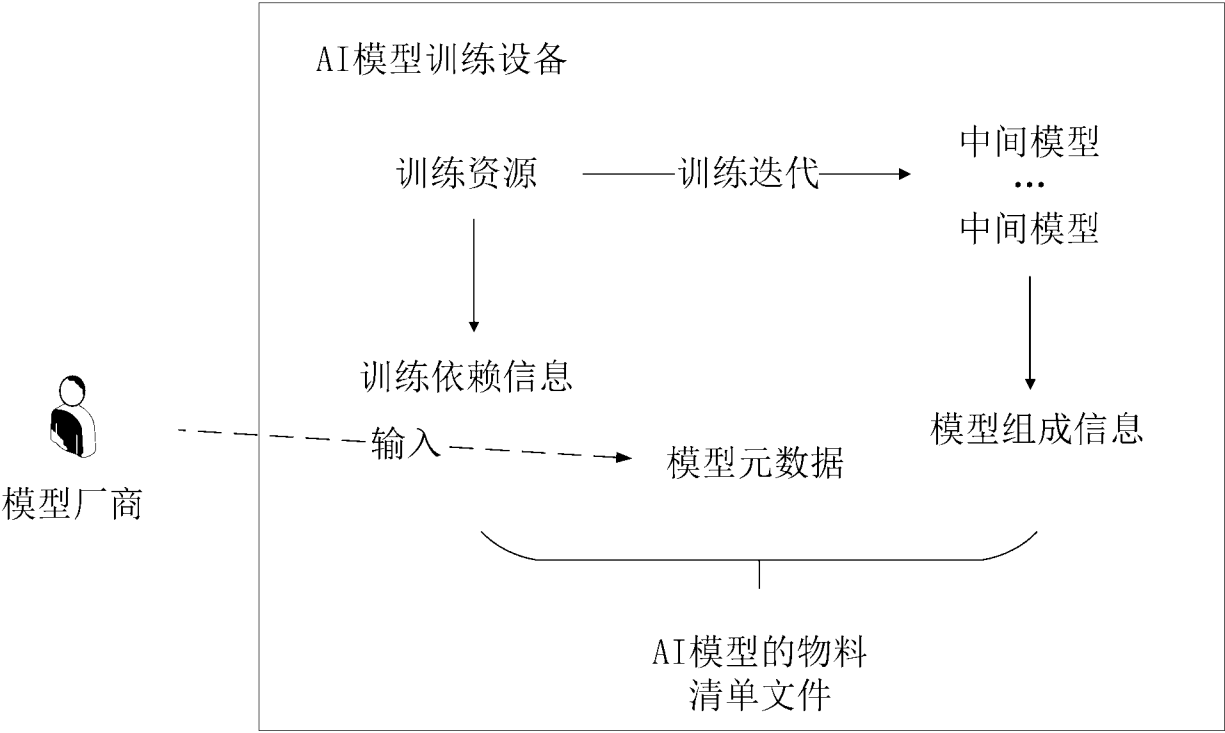


图 3a

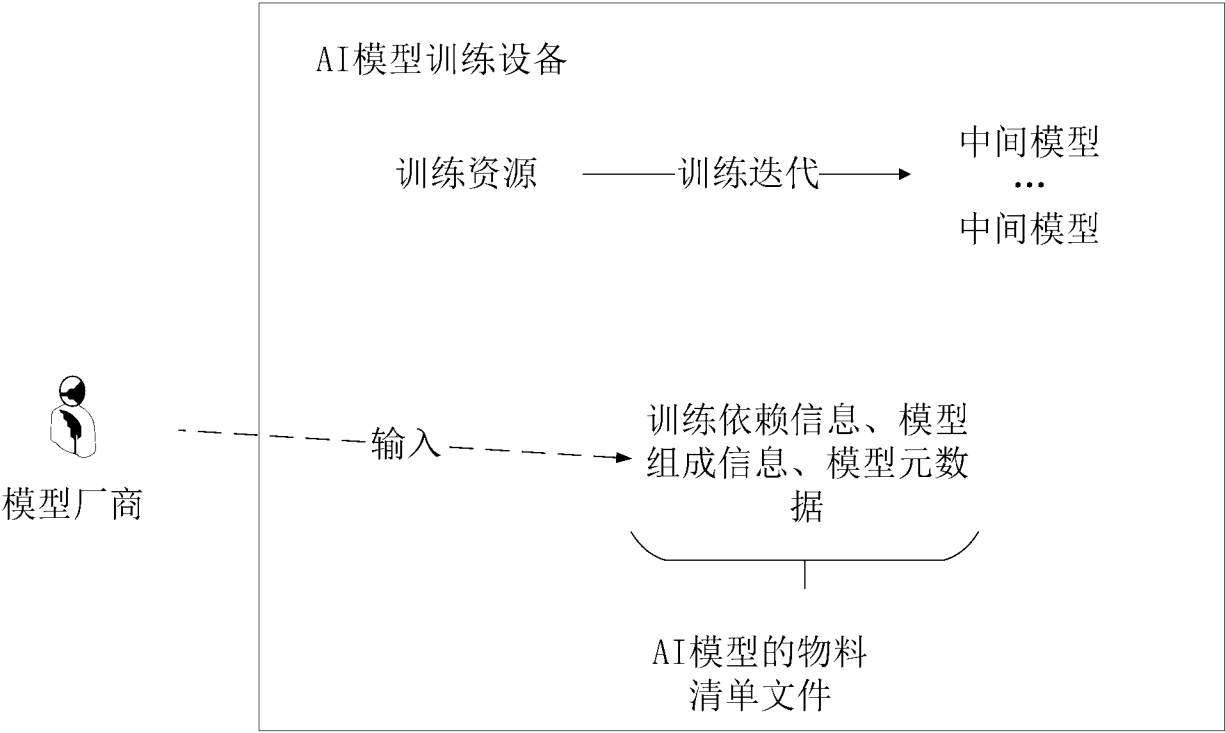


图 3b

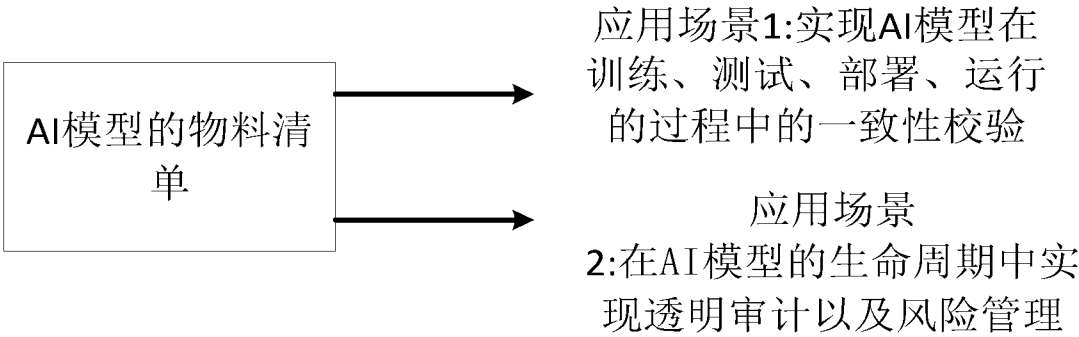


图 4

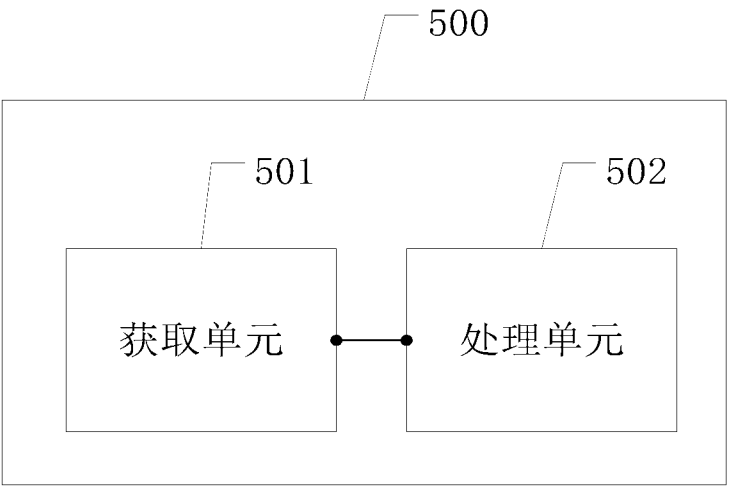


图 5

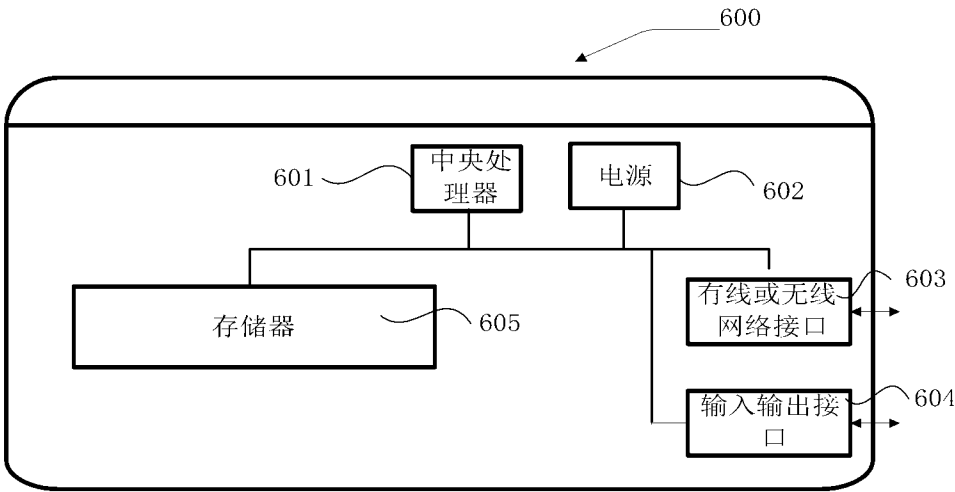


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/140908

A. CLASSIFICATION OF SUBJECT MATTER

G06F 18/21(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: G06F

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

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

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 模型, 人工智能, 软件物料清单, 训练, 元数据, BOM, SBOM, software, artificial intelligence, software bill+ of material+, train+, metadata

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015262124 A1 (AIRBUS OPERATIONS (SAS)) 17 September 2015 (2015-09-17) description, paragraphs 25-50	1-25
A	US 11436330 B1 (SOOS LLC) 06 September 2022 (2022-09-06) entire document	1-25
A	CN 115543406 A (DING (BEIJING) INTELLIGENT TECHNOLOGY CO., LTD.) 30 December 2022 (2022-12-30) entire document	1-25
A	CN 115454496 A (SHENZHEN SECZONE INTERNET SECURITY TECHNOLOGY CO., LTD.) 09 December 2022 (2022-12-09) entire document	1-25
A	US 2022318675 A1 (AIXPLAIN, INC.) 06 October 2022 (2022-10-06) entire document	1-25
A	US 2022253871 A1 (ASSENT INC. et al.) 11 August 2022 (2022-08-11) entire document	1-25



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

28 February 2024

Date of mailing of the international search report

05 March 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/CN2023/140908

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2015262124	A1	17 September 2015	US	9489654	B2	08 November 2016
US	11436330	B1	06 September 2022	US	2023019837	A1	19 January 2023
				US	11698967	B2	11 July 2023
CN	115543406	A	30 December 2022	None			
CN	115454496	A	09 December 2022	None			
US	2022318675	A1	06 October 2022	None			
US	2022253871	A1	11 August 2022	WO	2022087497	A1	28 April 2022
				US	11568423	B2	31 January 2023
				EP	4232910	A1	30 August 2023

A. 主题的分类

G06F 18/21(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: G06F

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))

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 模型, 人工智能, 软件物料清单, 训练, 元数据, BOM, SBOM, software, artificial intelligence, software bill+ of material+, train+, metadata

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
A	US 2015262124 A1 (AIRBUS OPERATIONS(SAS)) 2015年9月17日 (2015 - 09 - 17) 说明书第25-50段	1-25
A	US 11436330 B1 (SOOS LLC) 2022年9月6日 (2022 - 09 - 06) 全文	1-25
A	CN 115543406 A (鼎道智联(北京)科技有限公司) 2022年12月30日 (2022 - 12 - 30) 全文	1-25
A	CN 115454496 A (深圳开源互联网安全技术有限公司) 2022年12月9日 (2022 - 12 - 09) 全文	1-25
A	US 2022318675 A1 (AIXPLAIN, INC.) 2022年10月6日 (2022 - 10 - 06) 全文	1-25
A	US 2022253871 A1 (ASSENT INC等) 2022年8月11日 (2022 - 08 - 11) 全文	1-25

☐ 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2月28日

国际检索报告邮寄日期

2024年3月5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

张楚慧

电话号码 (+86) 010-53961496

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/CN2023/140908

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)
US	2015262124	A1	2015年9月17日	US	9489654	B2	2016年11月8日
US	11436330	B1	2022年9月6日	US	2023019837	A1	2023年1月19日
				US	11698967	B2	2023年7月11日
CN	115543406	A	2022年12月30日		无		
CN	115454496	A	2022年12月9日		无		
US	2022318675	A1	2022年10月6日		无		
US	2022253871	A1	2022年8月11日	WO	2022087497	A1	2022年4月28日
				US	11568423	B2	2023年1月31日
				EP	4232910	A1	2023年8月30日