



(12) 发明专利申请

(10) 申请公布号 CN 119166317 A

(43) 申请公布日 2024. 12. 20

(21) 申请号 202411673394.7

G06F 16/2457 (2019.01)

(22) 申请日 2024.11.21

(71) 申请人 闪互科技(广东)有限公司

地址 528200 广东省佛山市南海区桂城街
道简平路1号天安南海数码新城4栋
204室之一

(72) 发明人 孔繁斌

(74) 专利代理机构 广东海融科创知识产权代理
事务所(普通合伙) 44377

专利代理师 陈椅行

(51) Int. Cl.

G06F 9/48 (2006.01)

G06F 9/50 (2006.01)

G10L 17/26 (2013.01)

G06V 40/16 (2022.01)

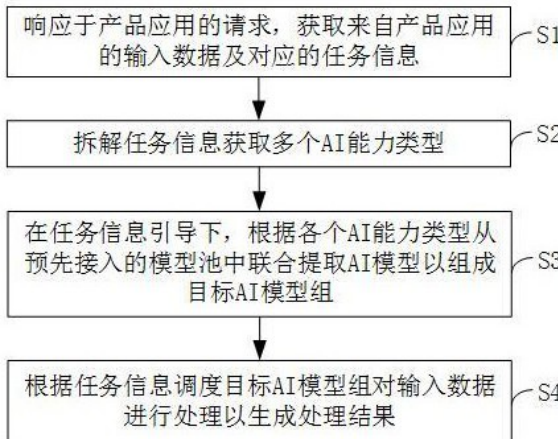
权利要求书2页 说明书12页 附图2页

(54) 发明名称

AI能力融合调度方法、装置、电子设备及存
储介质

(57) 摘要

本发明涉及人工智能技术领域,具体公开了一种AI能力融合调度方法、装置、电子设备及存储介质,其中,AI能力融合调度方法包括步骤:响应于产品应用的请求,获取来自产品应用的输入数据及对应的任务信息;拆解任务信息获取多个AI能力类型;在任务信息引导下,根据各个AI能力类型从预先接入的模型池中联合提取AI模型以组成目标AI模型组;根据任务信息调度目标AI模型组对输入数据进行处理以生成处理结果;该方法能无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策略来完成任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。



1. 一种AI能力融合调度方法,应用在中控平台中,其特征在于,所述AI能力融合调度方法包括以下步骤:

S1、响应于产品应用的请求,获取来自产品应用的输入数据及对应的任务信息;

S2、拆解所述任务信息获取多个AI能力类型;

S3、在所述任务信息引导下,根据各个所述AI能力类型从预先接入的模型池中联合提取AI模型以组成目标AI模型组,其中,所述模型池对应于每种AI能力类型具有多个AI模型;

S4、根据所述任务信息调度所述目标AI模型组对所述输入数据进行处理以生成处理结果。

2. 根据权利要求1所述的AI能力融合调度方法,其特征在于,步骤S3包括以下子步骤:

S31、根据任务信息确定其对应的各个AI能力类型的优先级别;

S32、根据优先级别和预先确定的冲突关系从所述模型池中提取匹配于所述任务信息的各个AI能力类型对应的AI模型以组成目标AI模型组。

3. 根据权利要求2所述的AI能力融合调度方法,其特征在于,步骤S31包括以下子步骤:

S311、根据任务信息和用户偏好需求信息确定所述任务信息对应的各个AI能力类型的优先级别。

4. 根据权利要求3所述的AI能力融合调度方法,其特征在于,在所述输入数据包括语音数据时,步骤S1和步骤S2之间还包括:

SA、根据所述输入数据获取用户的语音特征信息,并根据所述语音特征信息获取所述用户偏好需求信息。

5. 根据权利要求2所述的AI能力融合调度方法,其特征在于,步骤S32包括:

S321、基于用户需求或应用需求确定若干AI能力类型对应的指定性AI模型;

S322、根据优先级别和预先确定的冲突关系从所述模型池中提取匹配于所述任务信息对应的余下各个AI能力类型对应的AI模型,并与所述指定性AI模型组成所述目标AI模型组。

6. 根据权利要求1所述的AI能力融合调度方法,其特征在于,步骤S3和步骤S4之间还包括以下步骤:

SB、根据优先级别和预先确定的冲突关系从所述模型池中提取匹配于所述任务信息的各个AI能力类型且区别于目标AI模型组的多个AI模型以组成备用AI模型组;

S5、在执行步骤S4时,在后台根据所述任务信息调度所述备用AI模型组对所述输入数据进行处理以生成备用结果,并在用户对所述处理结果不满意或所述处理结果生成失败后,输出所述备用结果。

7. 根据权利要求1所述的AI能力融合调度方法,其特征在于,所述模型池基于标准化API接入的多个AI模型构成。

8. 一种AI能力融合调度装置,应用在中控平台中,其特征在于,所述AI能力融合调度装置包括:

获取模块,用于响应于产品应用的请求,获取来自产品应用的输入数据及对应的任务信息;

拆解模块,用于拆解所述任务信息获取多个AI能力类型;

提取模块,用于在所述任务信息引导下,根据各个所述AI能力类型从预先接入的模型

池中联合提取AI模型以组成目标AI模型组,其中,所述模型池对应于每种AI能力类型具有多个AI模型;

调度处理模块,用于根据所述任务信息调度所述目标AI模型组对所述输入数据进行处理以生成处理结果。

9.一种电子设备,其特征在于,包括处理器以及存储器,所述存储器存储有计算机可读指令,当所述计算机可读指令由所述处理器执行时,运行如权利要求1-7任一项所述方法中的步骤。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时运行如权利要求1-7任一项所述方法中的步骤。

AI能力融合调度方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及人工智能技术领域,具体而言,涉及一种AI能力融合调度方法、装置、电子设备及存储介质。

背景技术

[0002] 在当今的科技浪潮中,中控平台凭借其卓越的功能性和便捷性,已经在众多领域获得了广泛的应用,而其中最引人注目的便是家居领域。随着智能家居的迅猛发展,中控平台不仅成为了连接各种智能设备的核心枢纽,更是提升居住体验、实现家居智能化的关键所在。

[0003] 然而,尽管中控平台在功能性和实用性上表现出色,但在其配置和应用上,仍然存在一些亟待解决的问题。特别是当涉及到AI模型与产品应用的配合时,现有的固定搭配形式显得尤为僵化。这种搭配方式意味着,每当有新的用户需求或产品应用出现时,都需要重新调整或开发与之相匹配的AI模型,这无疑增加了开发成本和时间投入,这种固定搭配的形式也限制了AI模型在处理用户需求时的灵活性和处理效果,使得AI模型在处理复杂或多样化的用户需求时,不仅影响了用户体验,也限制了中控平台在智能家居领域中的进一步发展和应用。

[0004] 针对上述问题,目前尚未有有效的技术解决方案。

发明内容

[0005] 本申请的目的在于提供一种AI能力融合调度方法、装置、电子设备及存储介质,以便于中控平台进行AI能力低成本拓展,并提高AI模型的灵活性和处理效果,以优化用户体验。

[0006] 第一方面,本申请提供了一种AI能力融合调度方法,应用在中控平台中,所述AI能力融合调度方法包括以下步骤:

S1、响应于产品应用的请求,获取来自所述产品应用的输入数据及对应的任务信息;

S2、拆解所述任务信息获取多个AI能力类型;

S3、在所述任务信息引导下,根据各个所述AI能力类型从预先接入的模型池中联合提取AI模型以组成目标AI模型组,其中,所述模型池对应于每种AI能力类型具有多个AI模型;

S4、根据所述任务信息调度所述目标AI模型组对所述输入数据进行处理以生成处理结果。

[0007] 本申请的AI能力融合调度方法应用在中控平台中,能根据产品应用发布的任务信息解析出处理输入数据所需的多个AI能力类型,并在任务信息的引导下,根据这些AI能力类型从模型池调度出合适的AI模型以组成目标AI模型组来处理输入数据,无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策略来完

成任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。

[0008] 所述的AI能力融合调度方法,其中,步骤S3包括以下子步骤:

S31、根据任务信息确定其对应的各个AI能力类型的优先级别;

S32、根据优先级别和预先确定的冲突关系从所述模型池中提取匹配于所述任务信息的各个AI能力类型对应的AI模型以组成目标AI模型组。

[0009] 本申请的AI能力融合调度方法对不同的AI能力类型设定优先级别,以在提取AI模型来组成目标AI模型组的过程中,以基于优先级别高的AI能力类型挑选合适的AI模型,并以该AI模型作为主导来挑选优先级别低的AI能力类型对应的AI模型作为辅助模型,以构成高核心能力的目标AI模型组,以提高处理结果的处理效果。

[0010] 所述的AI能力融合调度方法,其中,步骤S31包括以下子步骤:

S311、根据任务信息和用户偏好需求信息确定所述任务信息对应的各个AI能力类型的优先级别。

[0011] 在该示例中,步骤S311可以是根据任务信息确定任务信息对应的各个AI能力类型的初始的优先级别,再基于用户偏好需求信息补偿调整这些AI能力类型的优先级别,还可以是根据用户偏好需求信息将对应的AI能力类型设置为最高优先级别,再根据任务信息设置其余的AI能力类型的优先级别,以使基于优先级别确定的目标AI模型组能更满足用户的使用需求。

[0012] 所述的AI能力融合调度方法,其中,在所述输入数据包括语音数据时,步骤S1和步骤S2之间还包括:

SA、根据所述输入数据获取用户的语音特征信息,并根据所述语音特征信息获取所述用户偏好需求信息。

[0013] 所述的AI能力融合调度方法,其中,步骤S32包括:

S321、基于用户需求或应用需求确定若干AI能力类型对应的指定性AI模型;

S322、根据优先级别和预先确定的冲突关系从所述模型池中提取匹配于所述任务信息对应的余下各个AI能力类型对应的AI模型,并与所述指定性AI模型组成所述目标AI模型组。

[0014] 所述的AI能力融合调度方法,其中,步骤S3和步骤S4之间还包括以下步骤:

SB、根据优先级别和预先确定的冲突关系从所述模型池中提取匹配于所述任务信息的各个AI能力类型且区别于目标AI模型组的多个AI模型以组成备用AI模型组;

S5、在执行步骤S4时,在后台根据所述任务信息调度所述备用AI模型组对所述输入数据进行处理以生成备用结果,并在用户对所述处理结果不满意或所述处理结果生成失败后,输出所述备用结果。

[0015] 所述的AI能力融合调度方法,其中,步所述模型池基于标准化API接入的多个AI模型构成。

[0016] 第二方面,本申请还提供了一种AI能力融合调度装置,应用在中控平台中,所述AI能力融合调度装置包括:

获取模块,用于响应于产品应用的请求,获取来自所述产品应用的输入数据及对应的任务信息;

拆解模块,用于拆解所述任务信息获取多个AI能力类型;

提取模块,用于在所述任务信息引导下,根据各个所述AI能力类型从预先接入的模型池中联合提取AI模型以组成目标AI模型组,其中,所述模型池对应于每种AI能力类型具有多个AI模型;

调度处理模块,用于根据所述任务信息调度所述目标AI模型组对所述输入数据进行处理以生成处理结果。

[0017] 本申请的AI能力融合调度装置应用在中控平台中,能根据产品应用发布的任务信息解析出处理输入数据所需的多个AI能力类型,并在任务信息的引导下,根据这些AI能力类型从模型池调度出合适的AI模型以组成目标AI模型组来处理输入数据,无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策略来完成任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。

[0018] 第三方面,本申请还提供了一种电子设备,包括处理器以及存储器,所述存储器存储有计算机可读取指令,当所述计算机可读取指令由所述处理器执行时,运行如上述第一方面提供的所述方法中的步骤。

[0019] 第四方面,本申请还提供了一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时运行如上述第一方面提供的所述方法中的步骤。

[0020] 由上可知,本申请提供了一种AI能力融合调度方法、装置、电子设备及存储介质,其中,AI能力融合调度方法应用在中控平台中,能根据产品应用发布的任务信息解析出处理输入数据所需的多个AI能力类型,并在任务信息的引导下,根据这些AI能力类型从模型池调度出合适的AI模型以组成目标AI模型组来处理输入数据,无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策略来完成任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。

附图说明

[0021] 图1为本申请实施例提供的AI能力融合调度方法的流程图。

[0022] 图2为本申请实施例提供的AI能力融合调度装置的结构示意图。

[0023] 图3为本申请实施例提供的电子设备的结构示意图。

[0024] 附图标记:201、获取模块;202、拆解模块;203、提取模块;204、调度处理模块;301、处理器;302、存储器;303、通信总线。

具体实施方式

[0025] 下面将结合本申请实施例中附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。因此,以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围,而是仅仅表示本申请的选定实施例。基于本申请的实施例,本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0026] 应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步定义和解释。同时，在本申请的描述中，术语“第一”、“第二”等仅用于区分描述，而不能理解为指示或暗示相对重要性。

[0027] 第一方面，请参照图1，本申请一些实施例提供了一种AI能力融合调度方法，应用在中控平台中，AI能力融合调度方法包括以下步骤：

S1、响应于产品应用的请求，获取来自产品应用的输入数据及对应的任务信息；

S2、拆解任务信息获取多个AI能力类型；

S3、在任务信息引导下，根据各个AI能力类型从预先接入的模型池中联合提取AI模型以组成目标AI模型组，其中，模型池对应于每种AI能力类型具有多个AI模型；

S4、根据任务信息调度目标AI模型组对输入数据进行处理以生成处理结果。

[0028] 具体地，中控平台即中央控制平台，广泛应用在汽车设计、家居、工业、医疗和交通监控等领域的系统中，本申请实施例的AI能力融合调度方法旨在为中控平台提供AI功能决策，以使中控平台能根据使用需求自动选择合适的AI模型来执行相关功能。

[0029] 更具体地，本申请实施例的AI能力融合调度方法尤其适合于应用在家居智能终端的中控平台中，以为用户提供更人性化的交互体验以及轻松便捷的数字家庭生活。

[0030] 更具体地，产品应用为中控平台连接相应前台产品的应用程序或应用端口，前台产品为与用户交互的设备端，如手机、遥控、麦克风、摄像头等，同一前台产品可以通过一类或多类产品应用与中控平台进行连接，本申请实施例的AI能力融合调度方法实际上在产品应用基于相应的任务信息访问中控平台的AI能力时配置合适的目标AI模型组以执行相应任务来生成处理结果。

[0031] 更具体地，任务信息为需要对输入数据执行的一系列处理行为，在本申请实施例中，主要指的是涉及AI模型处理的一系列处理行为，一般情况下，一类产品应用对应于一类或多类任务信息，任务信息实际上是由产品应用的类型来决定，使得步骤S1可以根据产品应用的类型来获取任务信息，也可以基于产品应用的标准化接口的接口文档的内容来获取任务信息，还可以基于产品应用直接输入任务信息。

[0032] 更具体地，产品应用的请求为相应前台产品需要通过中控平台执行相关操作时对中控平台发出的请求命令，可以是在前台产品被唤醒（可以通过按键或语音进行唤醒）时发出，还可以是在产品应用接收到用户的输入数据时发出，本申请实施例的AI能力融合调度方法在获取到产品应用的请求后再接收来自产品应用的输入数据或通过产品应用接收前台产品实时输入的输入数据，并同时确定对应产品应用的任务信息。

[0033] 更具体地，由于任务信息为需要对输入数据执行的、涉及AI模型处理的一系列处理行为，中控平台需要配置多个AI模型来对输入数据完成相应的任务处理，如针对语音数据生成相应图像的任务信息而言，中控平台需要配置语音识别、关键词提取/语义理解、文生图等多个AI模型来完成相应的任务处理，不同类型的AI模型具有不同的AI能力，该AI能力指的是AI模型的功能，如语音合成、语音识别、人脸识别等，任务信息可以分为多个不同的具有依存关系的子任务，这些子任务需要利用具有不同AI能力的AI模型来执行，步骤S2相当于是拆解任务信息获取对应的多个子任务，并确定不同子任务所需要使用的AI能力类型。

[0034] 更具体地，属于不同AI能力类型的AI模型能处理不同的子任务，属于相同AI能力

类型的AI模型在对相同的数据执行子任务时产生的结果也存在差异,尤其是不同厂商提供的或微调的AI模型的处理效果差异甚大,且对于整体的任务信息而言,相互依存的子任务对应采用不同搭配方式的AI模型来完成任务最终会产生不同的处理结果,本申请实施例的AI能力融合调度方法可以通过搭配不同的AI模型组成目标AI模型组来处理输入数据以产生不同处理效果的处理结果。

[0035] 更具体地,与传统的一对多的固定配置方式相比,本申请实施例的AI能力融合调度方法根据产品应用的请求统一调度所有AI模型,并根据产品应用的任务信息的特点调取多个合适的AI模型组成目标AI模型组来完成对应任务以产生合适的处理结果;步骤S3的提取的目标AI模型组所包含的AI模型为模型池中匹配于对应AI能力类型的、适合于对应任务信息的AI模型,模型池为预先整合的包含所有可用的AI模型的集合,其中,步骤S3提取AI模型的过程可以是使用学习模型或深度学习模型来根据AI能力类型自动从模型池中提取合适的AI模型,也可以是基于产品应用和AI模型的供应商关系结合AI能力类型自动从模型池中提取合适的AI模型,还可以是基于用户使用习惯和/或使用需求和/或不同AI能力类型的不同AI模型执行对应的任务信息的能力强弱(或得分)自动从模型池中提取合适的AI模型,以组成目标AI模型组来完成对应产品应用发布的任务;在这些提取方式下,本申请实施例的AI能力融合调度方法均能从模型池中获取到完成产品应用的任务信息对应的目标AI模型组,且在新增产品应用后或在模型池中新增AI模型后,本申请实施例的AI能力融合调度方法均无需人工重新根据产品应用的使用特性固定配置合适的多个AI模型,便于中控平台进行AI能力拓展,也提高了AI模型的灵活性和处理效果,以优化用户体验。

[0036] 更具体地,在步骤S3中,任务信息用于引导AI模型提取的过程,可以是引导AI模型的提取顺序,如依照任务信息中AI能力类型的使用顺序进行AI模型提取,也可以是引导产品应用对AI能力类型的选择倾向,如某产品应用更着重于语义理解,则任务信息优选引导选择优质的关于语义理解的AI模型,再在此基础上选择关于其他的AI能力类型且与该AI模型能搭配使用(使用不冲突)的AI模型,基于任务信息的引导,本申请实施例的AI能力融合调度方法调度获取的目标AI模型组高度匹配于产品应用使用,能满足其使用需求。

[0037] 更具体地,输入数据为基于前台产品输入至产品应用的数据,可以为各类模态的数据或任意组合,取决于前台产品的使用方式,如包括语音数据、静态图像数据、视频数据、文字数据中的一种或多种,处理结果为经由目标AI模型组执行产品应用提供的任务信息而生成的最终结果,其可以通过中控平台反馈给对应的交互产品(可以是前台产品本身,也可以是其他产品)进行输出。

[0038] 需要说明的是,本申请实施例的AI能力融合调度方法主要应用在需要调度多个AI模型来完成产品应用发布的任务信息的场合,尤其适用于不同AI模型组合下产生的处理结果差异性较为明显的场合。

[0039] 需要说明的是,本申请实施例的AI能力融合调度方法适用于语音方向、图像方向、视频方向、NLP(自然语言处理)方向的数据处理,优选为针对语音方向的数据处理。

[0040] 本申请实施例的AI能力融合调度方法应用在中控平台中,能根据产品应用发布的任务信息解析出处理输入数据所需的多个AI能力类型,并在任务信息的引导下,根据这些AI能力类型从模型池调度出合适的AI模型以组成目标AI模型组来处理输入数据,无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策

略来完成任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。

[0041] 在一些优选的实施方式中,步骤S3包括以下子步骤:

S31、根据任务信息确定其对应的各个AI能力类型的优先级别;

S32、根据优先级别和预先确定的冲突关系从模型池中提取匹配于任务信息的各个AI能力类型对应的AI模型以组成目标AI模型组。

[0042] 具体地,步骤S31中基于任务信息确定的各个AI能力类型的优先级别用于引导AI模型提取,不同的任务信息具有不同的任务需求,其对于不同的AI能力具有不同的倾向性,在本申请实施例中,倾向性更高的AI能力类型的优先级别更高,优先级别高的AI能力类型对应于任务信息中的核心的子任务,决定了整个任务处理效果的核心能力;考虑到不同AI模型配合使用具有不同的效果,本申请实施例的AI能力融合调度方法对不同的AI能力类型设定优先级别,以在提取AI模型来组成目标AI模型组的过程中,以基于优先级别高的AI能力类型挑选合适的AI模型,并以该AI模型作为主导来挑选优先级别低的AI能力类型对应的AI模型作为辅助模型,以构成高核心能力的目标AI模型组,以提高处理结果的处理效果;另外,一些AI模型结合使用时可能存在冲突(主要表现为bug)而导致无法完成对应任务,在步骤S32中,预先确定的冲突关系为不同AI能力类型的AI模型的冲突标记,如A能力的a模型与B能力的b模型存在冲突标记,则在提取a模型后则将b模型视为不可提取的对象,以避免步骤S4执行失败;由此,步骤S32结合优先级别和预先确定的冲突关系获取多个AI模型,以保证基于这些AI模型构成的目标AI模型组能顺利、高质量地对输入数据进行处理以获取优质的处理结果。

[0043] 更具体地,在该实施方式中,任务信息包括能力偏重信息,能力偏重信息表征该产品应用发布的任务中所需的各个AI能力类型的重要程度,其由前台产品的类型决定,如一个前台产品的产品应用发布任务为基于语音数据生成图像,其需要的AI能力类型包括语音识别处理、语义理解处理、关键词提取处理和文生图处理,若该前台产品侧重于提供高质量或高分辨率图像,则文生图处理为该任务的核心能力,能力偏重信息更倾向于文生图处理,使得步骤S31能赋予文生图处理更高的优先级别,从而使得步骤S32在更倾向于提供优质的文生图处理的AI模型的情况下搭配其他AI能力类型的AI模型以组成目标AI模型组。

[0044] 更具体地,模型池中的AI模型具有关于对应AI能力类型的评分,步骤S32可以根据优先级别和预先确定的冲突关系结合AI能力类型的评分来提取AI模型,如以优先级别作为提取次序,提取各个AI能力类型对应的评分最高、且相互不冲突的AI模型来组成目标AI模型组,又如基于优先级别对不同AI能力类型赋予不同的权重,再基于AI模型的评分和权重的乘积确定对应各个AI模型的偏置得分,再提取偏置得分总和最高且相互不冲突的AI模型来组成目标AI模型组(相当于评价不同AI模型组合下得分总和,再选择得分总和最高的AI模型组合作为目标AI模型组),前一种处理方式能快速确定高核心能力的目标AI模型组,后一种处理方式能确定综合性能最强的目标AI模型组,两者均能顺利、高质量地对输入数据进行处理以获取优质的处理结果。

[0045] 另外,由于不同前台产品输入相同的输入数据时可能会由于产品特性或环境特性或噪音干扰等因素导致输入数据产生偏差,也会使得同样的目标AI模型组输出的处理结果产生差异,因此,在一些实施方式中,步骤S1获取的任务信息还包括了产品应用的特性信息

和/或类型信息,以使步骤S31能进一步基于特性信息或类型信息设定合适的优先级别,即步骤S31综合能力偏重信息以及产品应用的特性信息和/或类型信息来确定对应的各个AI能力类型的优先级别,具体地处理方式可以是:基于能力偏重信息初步确定对应的各个AI能力类型的优先级别,再根据产品应用的特性信息和/或类型信息来补偿调节部分AI能力类型的优先级别。

[0046] 更具体地, AI模型的评分可以是用户评分,如综合各个用户评定的分值确定的均值评分,又如根据用户对处理结果的满意程度进行调整的分值,该用户可以是使用当前中控平台的用户,也可以是包含其他中控平台的用户,该实施方式获取的目标AI模型组更能满足用户使用需求; AI模型的评分还可以是模型的性能评分,如结合AI模型的精度得分、时效得分和负荷得分的综合得分,其中,精度得分为AI模型处理对应子任务的处理精度的得分,时效得分为AI模型处理对应子任务的处理速度的得分,负荷得分为AI模型处理对应子任务占用或消耗的计算资源的得分,基于这些得分确定的综合得分能准确表征不同AI模型的性能差异,该实施方式获取的目标AI模型组更能提供优质性能的处理能力。

[0047] 更具体地,在AI模型的评分为模型的性能评分的实施方式中,综合得分还包含配套得分,配套得分包括AI模型之间的关联性搭配得分以及AI模型与前台产品的搭配得分;其中,部分特定的多个AI模型搭配使用下会产生更优异的处理效果,普遍表现为同一厂商针对其提供的多个AI模型进行适配微调,使得这些AI模型组合使用能产生更优异的处理效果,在该情况下,对同时提取的这些AI模型增加对应的配套得分(AI模型之间的关联性搭配得分),以更准确地区分不同AI模型组合的优劣;另外,也有部分AI模型搭配前台产品会产生更优异的处理效果,普遍表现为同一厂商针对其提供的前台产品的获取的输入数据的特性对AI模型进行微调,使得其提供的AI模型针对该前台应用提供的输入处理能产生更优异的处理效果,在该情况下,对提取的这些AI模型增加配套得分(AI模型与前台产品的搭配得分),以更准确地区分不同AI模型组合的优劣。

[0048] 在一些优选的实施方式中,步骤S31包括以下子步骤:

S311、根据任务信息和用户偏好需求信息确定任务信息对应的各个AI能力类型的优先级别。

[0049] 具体地,在该实施方式中,步骤S311可以是根据任务信息确定任务信息对应的各个AI能力类型的初始的优先级别,再基于用户偏好需求信息补偿调整这些AI能力类型的优先级别,还可以是根据用户偏好需求信息将对应的AI能力类型设置为最高优先级别,再根据任务信息设置其余的AI能力类型的优先级别,以使基于优先级别确定的目标AI模型组更能满足用户的使用需求。

[0050] 更具体地,用户偏好需求信息可以是用户预先设定的用户标签,还可以是基于用户使用习惯或大数据信息确定的用户标签,如一个前台产品的产品应用发布任务为基于语音数据生成图像,若用户标签倾向于提供高画质的图像,则步骤S311则基于该用户偏好需求信息调高文生图处理的AI能力类型的优先级别,若用户标签倾向于提供画面内容更准确的图像,则步骤S311则基于该用户偏好需求信息调高语义理解处理的AI能力类型的优先级别。

[0051] 在一些实施方式中,任务信息还包括用于拆解获取多个AI能力类型的子任务信息,或者,任务信息包括任务类型,子任务信息基于任务类型分析获取。

[0052] 在一些优选的实施方式中,在输入数据包括语音数据时,步骤S1和步骤S2之间还包括:

SA、根据输入数据获取用户的声音特征信息,并根据声音特征信息获取用户偏好需求信息。

[0053] 具体地,步骤SA中的声音特征信息为基于语音数据提取的、表征用户声音特征的声音指纹数据,其用于确定用户身份,以使步骤SA能从预先存储关联于用户身份对应的用户偏好需求信息的数据库中提取对应的用户偏好需求信息;本申请实施例的AI能力融合调度方法能基于声音特征信息确定用户身份,并调用对应的用户偏好需求信息作为AI能力类型的优先级别的设置规则,以提供满足用户偏好需求的目标AI模型组来处理输入数据,以满足用户个性化使用需求。

[0054] 在一些别的实施方式中,在输入数据包括语音数据时,步骤S1和步骤S2之间还包括:

SA'、根据输入数据获取用户的声音特征信息,根据声音特征信息推断用户的年龄信息和性别信息,并根据年龄信息和性别信息获取用户偏好需求信息。

[0055] 具体地,对于用户为陌生用户或不具备预先存储关联于用户身份对应的用户偏好需求信息的数据库的场合,本申请实施例的AI能力融合调度方法能根据语音数据提取声音特征信息,并以此为基准推断用户的年龄信息和性别信息,再根据大数据方法获取对应年龄信息和性别信息的用户偏好需求信息或根据年龄信息和性别信息从预先存储关联于不同年龄、性别的用户偏好需求信息的数据库中提取对应的用户偏好需求信息,以提供满足用户偏好需求的目标AI模型组来处理输入数据,以基本满足陌生用户个性化使用需求。

[0056] 更具体地,根据声音特征信息推断用户的年龄信息和性别信息的处理过程实际上为分类过程,可以利用基于语音输入的分类模型来实现,在此不做赘述。

[0057] 在一些别的实施方式中,步骤SA和步骤SA'可以结合使用,具体为:在获取到根据输入数据获取的声音特征信息后,根据声音特征信息判断当前用户是否为私有用户,是则根据声音特征信息从预先存储关联于用户身份对应的用户偏好需求信息的数据库中提取对应的用户偏好需求信息,否则根据声音特征信息推断用户的年龄信息和性别信息,并根据年龄信息和性别信息获取用户偏好需求信息。

[0058] 具体地,私有用户指中控平台记录有对应用户身份信息用户,该实施方式相当于先判断当前用户是否曾经基于语音输入的方式使用过该中控平台,是则直接调用对应的用户偏好需求信息,否则推断该新的陌生用户的声特点推断其年龄和性别以提供满足对应用户偏好需求的目标AI模型组来处理输入数据。

[0059] 在一些别的实施方式中,在输入数据包括人像数据时,步骤S1和步骤S2之间还包括:

SA''、根据输入数据获取用户的人脸信息,并根据人脸信息获取用户偏好需求信息。

[0060] 具体地,步骤SA''中的人脸信息为基于人像数据截取的关于人物的脸部图像,其用于确定用户身份,以使步骤SA''能从预先存储关联于用户身份对应的用户偏好需求信息的数据库中提取对应的用户偏好需求信息;本申请实施例的AI能力融合调度方法能基于人脸信息确定用户身份,并调用对应的用户偏好需求信息作为AI能力类型的优先级别的设置

规则,以提供满足用户偏好需求的目标AI模型组来处理输入数据,以满足用户个性化使用需求。

[0061] 在一些别的实施方式中,在输入数据包括人像数据时,步骤S1和步骤S2之间还包括:

SA''、根据输入数据获取用户的人脸信息,根据人脸信息推断用户的年龄信息和性别信息,并根据年龄信息和性别信息获取用户偏好需求信息。

[0062] 具体地,对于用户为陌生用户或不具备预先存储关联于用户身份对应的用户偏好需求信息的数据库的场合,本申请实施例的AI能力融合调度方法能输入数据提取人脸信息,并以此为基准推断用户的年龄信息和性别信息,再根据大数据方法获取对应年龄信息和性别信息的用户偏好需求信息或根据年龄信息和性别信息从预先存储关联于不同年龄、性别的用户偏好需求信息的数据库中提取对应的用户偏好需求信息,以提供满足用户偏好需求的目标AI模型组来处理输入数据,以基本满足陌生用户个性化使用需求。

[0063] 更具体地,根据人脸信息推断用户的年龄信息和性别信息的处理过程实际上为分类过程,可以利用基于语音输入的分类模型来实现,在此不做赘述。

[0064] 在一些别的实施方式中,步骤SA'和步骤SA''可以结合使用,具体为:在获取到根据输入数据获取的人脸信息后,根据人脸信息判断当前用户是否为私有用户,是则根据人脸信息从预先存储关联于用户身份对应的用户偏好需求信息的数据库中提取对应的用户偏好需求信息,否则根据人脸信息推断用户的年龄信息和性别信息,并根据年龄信息和性别信息获取用户偏好需求信息。

[0065] 在一些优选的实施方式中,步骤S32包括:

S321、基于用户需求或应用需求确定若干AI能力类型对应的指定性AI模型;

S322、根据优先级别和预先确定的冲突关系从模型池中提取匹配于任务信息对应的余下各个AI能力类型对应的AI模型,并与指定性AI模型组成目标AI模型组。

[0066] 具体地,基于用户需求或应用需求,部分产品应用必须配合特定的一类或多类的AI能力类型对应的指定性AI模型,本申请实施例的AI能力融合调度方法能在满足该需求场景下匹配余下AI能力类型对应的AI模型以与这些基于用户需求或应用需求确定的指定性AI模型组成目标AI模型组来处理输入数据,使得本申请实施例的AI能力融合调度方法具备灵活使用的优势,能满足各类场景使用需求。

[0067] 更具体地,基于用户需求确定的指定性AI模型,可以是用户预先设定的关于使用该对应AI能力类型必须使用的AI模型,还可以是用户预先设定的关于使用该产品应用下的该对应AI能力类型必须使用的AI模型,使得用户可以针对AI功能或者产品应用设定指定性AI模型,使得本申请实施例的AI能力融合调度方法具备灵活使用的优势。

[0068] 更具体地,基于应用需求确定的指定性AI模型,可以是厂商预先设定的关于使用该产品应用下的该对应AI能力类型必须使用的AI模型。

[0069] 在一些优选的实施方式中,步骤S3和步骤S4之间还包括以下步骤:

SB、根据优先级别和预先确定的冲突关系从模型池中提取匹配于任务信息的各个AI能力类型且区别于目标AI模型组的多个AI模型以组成备用AI模型组;

S5、在执行步骤S4时,在后台根据任务信息调度备用AI模型组对输入数据进行处理以生成备用结果,并在用户对处理结果不满意或处理结果生成失败后,输出备用结果。

[0070] 具体地,步骤S3获取的目标AI模型组相当于是当前状态下处理相应产品应用发布的任务信息的最优选的AI模型搭配组合,步骤S4获取的备用AI模型组相当于是当前状态下处理相应产品应用发布的任务信息的完全区别于目标AI模型组的、次优选的AI模型搭配组合;本申请实施例的AI能力融合方法相当于提供了两种处理效果存在一定差异的、用于处理产品应用发布的任务信息的决策,并在作为主要决策的目标AI模型组输出的处理结果不符合用户需求或生成失败时,输出作为次要决策的备用AI模型组的备用结果,使得用户在对处理结果不满意或处理结果生成失败后能及时获取区别于处理结果的备用结果,一来避免了用户需要耗费更多时间等待新的处理结果的生成,二来避免了重新生成的处理结果与前面用户不满意的处理结果重复度过高。

[0071] 更具体地,目标AI模型组生成的处理结果直接通过中控平台反馈给对应的交互产品进行输出,备用AI模型组生成的备用结果保留在中控平台的后台中,并在用户对处理结果不满意或处理结果生成失败后,通过中控平台反馈给对应的交互产品进行输出。

[0072] 更具体地,目标AI模型组中的AI模型和备用AI模型组中的AI模型完全不同;完全不同的AI模型组成的备用AI模型组能产生与目标AI模型组获取的处理结果差异性较为明显且依然能满足用户需求的备用结果,在处理过程中也无需复制出相同的AI模型来执行子任务以节省处理时间、空间,并确保了目标AI模型组的处理过程和备用AI模型组的处理过程不会相互干扰。

[0073] 更具体地,备用AI模型组可以是在目标AI模型组处理输入数据时同时开始处理输入数据,还可以是在目标AI模型组处理输入数据延时后开始处理输入数据,在本申请实施例中,优选为后者,以错开目标AI模型组和备用AI模型组的启用节点,以降低中控平台的运行负荷。

[0074] 在一些实施方式中,当用户不满意处理结果或处理结果生成失败时,表明当前目标AI模型组不符合用户偏好或不适用于当前产品应用;因此,本申请实施例的AI能力融合调度方法还包括步骤:

S6、在用户不满意处理结果或处理结果生成失败时,降低目标AI模型组对应的AI模型的综合得分。

[0075] 具体地,在该实施方中,综合得分可以基于固定分值或百分比分值进行降低,优选为后者。

[0076] 更具体地,本申请实施例的AI能力融合调度方法降低目标AI模型组对应的AI模型的综合得分能降低后续执行产品应用发布的任务信息时挑选这些AI模型来组成AI模型组的概率,以使得本申请实施例的AI能力融合调度方法调度出的目标AI模型组能更满足用户需求。

[0077] 在一些实施方式中,本申请实施例的AI能力融合调度方法还包括步骤:

S7、在用户满意备用结果时,提高备用AI模型组对应的AI模型的综合得分。

[0078] 具体地,在该实施方中,综合得分可以基于固定分值或百分比分值进行提高,优选为后者。

[0079] 更具体地,本申请实施例的AI能力融合调度方法提高备用AI模型组对应的AI模型的综合得分能提高后续执行产品应用发布的任务信息时挑选这些AI模型来组成AI模型组的概率,以使得本申请实施例的AI能力融合调度方法后续调度出的目标AI模型组能更满足

用户需求。

[0080] 在一些实施方式中,模型池中的AI模型对应于不同产品应用或任务信息的综合得分独立计算。

[0081] 具体地,在该实施方式中,每个AI模型对应于不同产品应用或任务信息具有对应的、独立计算的综合得分,这些综合得分可以是相同的,也可以是不同的;基于前述内容可知,不同前台产品输入相同的输入数据时可能会由于产品特性或环境特性或噪音干扰等因素导致输入数据产生偏差,也会使得同样的目标AI模型组输出的处理结果产生差异,且对于不同任务信息,同样的AI模型会产生不同的处理效果,因此,本申请实施例的AI能力融合调度方法将模型池中的AI模型对应于不同产品应用或任务信息的综合得分设计为独立计算,使得本申请实施例的AI能力融合调度方法能针对不同的产品应用或任务信息为用户提供更优质的目标AI模型组来完成相应任务。

[0082] 更具体地,在该实施方式中,步骤S6和步骤S7改变的综合得分为AI模型在对应产品应用或任务信息下的综合得分。

[0083] 在一些优选的实施方式中,模型池基于标准化API接入的多个AI模型构成。

[0084] 具体地,多个AI模型基于标准化API接入中控平台中以构成模型池,实现了多源AI的能力接入,能对外提供标准统一的服务。

[0085] 更具体地,在该实施方式中,标准化API接入AI模型的过程为:加入AI模型,根据AI模型的AI能力类型分辨生成标准API,并生成风格统一的接口文档,使得本申请实施例的AI能力融合调度方法在使用时能根据接口文档的内容调用对应的AI模型。

[0086] 更具体地,本申请实施例的AI能力融合调度方法基于上述接入方式可以便捷地接入或删除AI模型,以实现AI能力的优化、扩展或缩减,减少中控平台的设备更新维护成本。

[0087] 第二方面,请参照图2,本申请一些实施例还提供了一种AI能力融合调度装置,应用在中控平台中,AI能力融合调度装置包括:

获取模块201,用于响应于产品应用的请求,获取来自产品应用的输入数据及对应的任务信息;

拆解模块202,用于拆解任务信息获取多个AI能力类型;

提取模块203,用于在任务信息引导下,根据各个AI能力类型从预先接入的模型池中联合提取AI模型以组成目标AI模型组,其中,模型池对应于每种AI能力类型具有多个AI模型;

调度处理模块204,用于根据任务信息调度目标AI模型组对输入数据进行处理以生成处理结果。

[0088] 本申请实施例的AI能力融合调度装置应用在中控平台中,能根据产品应用发布的任务信息解析出处理输入数据所需的多个AI能力类型,并在任务信息的引导下,根据这些AI能力类型从模型池调度出合适的AI模型以组成目标AI模型组来处理输入数据,无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策略来完成的任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。

[0089] 在一些优选的实施方式中,本申请实施例的AI能力融合调度装置用于执行上述第一方面提供的AI能力融合调度方法。

[0090] 第三方面,请参照图3,本申请一些实施例还提供了一种电子设备的结构示意图,本申请提供一种电子设备,包括:处理器301和存储器302,处理器301和存储器302通过通信总线303和/或其他形式的连接机构(未标出)互连并相互通讯,存储器302存储有处理器301可执行的计算机可读取指令,当电子设备运行时,处理器301执行该计算机可读取指令,以执行时执行上述实施例的任一可选的实现方式中的方法。

[0091] 第四方面,本申请实施例提供一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时,执行上述实施例的任一可选的实现方式中的方法。其中,计算机可读存储介质可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(Static Random Access Memory, 简称SRAM),电可擦除可编程只读存储器(Electrically Erasable Programmable Read-Only Memory, 简称EEPROM),可擦除可编程只读存储器(Erasable Programmable Read Only Memory, 简称EPROM),可编程只读存储器(Programmable Red-Only Memory, 简称PROM),只读存储器(Read-Only Memory, 简称ROM),磁存储器,快闪存储器,磁盘或光盘。

[0092] 综上,本申请实施例提供了一种AI能力融合调度方法、装置、电子设备及存储介质,其中,AI能力融合调度方法应用在中控平台中,能根据产品应用发布的任务信息解析出处理输入数据所需的多个AI能力类型,并在任务信息的引导下,根据这些AI能力类型从模型池调度出合适的AI模型以组成目标AI模型组来处理输入数据,无需根据产品应用设计定制化AI模型的搭配,能根据产品应用的使用需求自动提供AI模型的搭配策略来完成任务,便于中控平台进行AI能力拓展且降低成本投入,并提高了AI模型的灵活性和处理效果,以优化用户体验。

[0093] 在本申请所提供的实施例中,应该理解到,所揭露装置和方法,可以通过其它的方式实现。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,又例如,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通信接口,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。

[0094] 另外,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。

[0095] 再者,在本申请各个实施例中的各功能模块可以集成在一起形成一个独立的部分,也可以是各个模块单独存在,也可以两个或两个以上模块集成形成一个独立的部分。

[0096] 在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。

[0097] 以上所述仅为本申请的实施例而已,并不用于限制本申请的保护范围,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

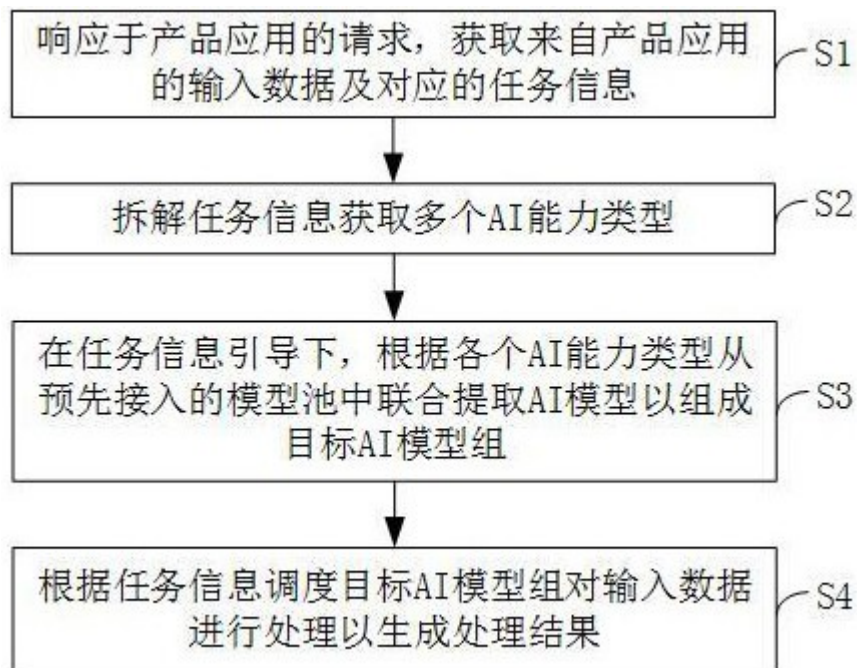


图 1

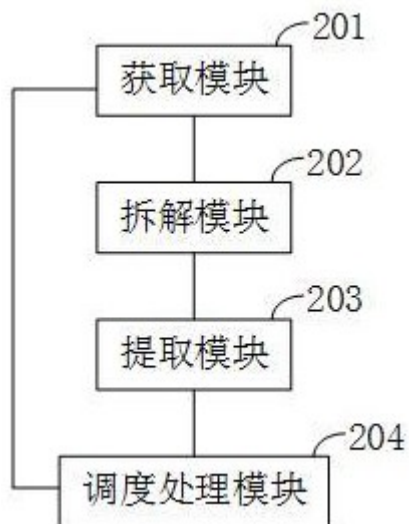


图 2

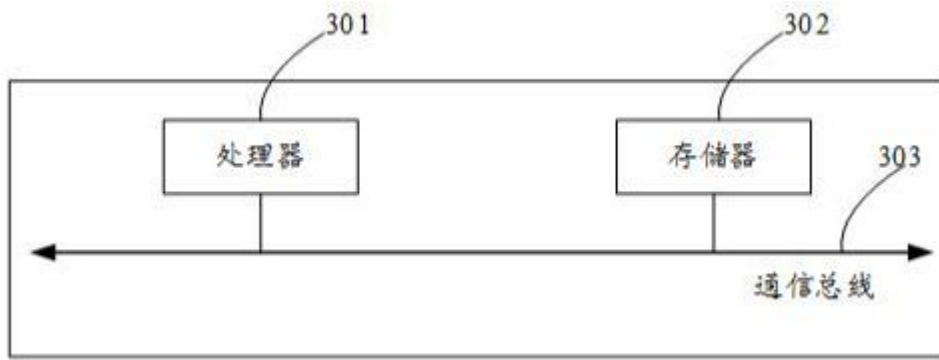


图 3