一种数据标注方法及相关设备

本申请要求于 2024 年 03 月 27 日提交国家知识产权局、申请号为 202410361751.X、发明名称为“一种数据标注方法及相关设备”的中国专利申请的优先权，以及要求于 2024 年 07 月 12 日提交国家知识产权局、申请号为 202410939017.7、发明名称为“一种数据标注方法及相关设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及人工智能（artificial intelligence，AI）技术领域，尤其涉及一种数据标注方法、标注系统、计算设备集群、计算机可读存储介质以及计算机程序产品。

背景技术

随着 AI 技术的蓬勃发展，AI 代理（AI agent）成为一类独立的技术路线，AI agent 通过将大模型和特定的功能封装，从而支持更丰富的应用场景。例如，在游戏场景中，用户可以定义非玩家角色 (Non-player character) 的 AI 代理，也称作 AI NPC，通过 AI NPC 实现智能对话、聊天驱动故事线、设定 NPC 的背景、性格、记忆、情绪打分等功能。

AI NPC 等 AI 代理可以通过可扩展指令多世界智能体（Scalable Instructable Multiworld Agent，SIMA）实现。SIMA 无需定制化的应用程序编程接口（application programming interface，API），也能够遵循自然语言指令，在多款游戏中执行任务，并且展现其泛化的能力，将学习的技能和策略，应用在未曾见过的环境中。

其中，SIMA 的输入可以包括屏幕图像以及用户提供的简单的自然语言指令。SIMA 可以将上述图像、自然语言指令映射到键盘和鼠标操作（可以统称为键鼠操作）。然而，SIMA 的训练需要收集、整合人类专家的数据，如视频画面、自然语言指令、键鼠操作数据等，并进行标注。人工收集、标注上述数据的效率较低，并且成本比较高昂，难以满足业务需求。

发明内容

本申请提供了一种数据标注方法，该方法支持自动采集用户输入的自然语言指令、通过交互设备触发的交互操作数据以及对应的应用运行时的交互界面数据，基于各数据的时间戳对不同类型数据自动对齐，由此实现交互任务的数据自动标注，为训练或测试应用的智能代理提供高质量的标注数据。本申请还提供了与上述方法对应的数据标注系统、计算设备集群、计算机可读存储介质以及计算机程序产品。

第一方面，本申请提供一种数据标注方法。该方法可以由数据标注系统执行。数据标注系统用于进行数据标注，标注数据可以用于学习人类操作和机器反馈，从而训练或测试人工智能代理（artificial intelligence agent，AI agent）等智能体。数据标注系统可以是软件系统，该软件系统可以是独立的软件系统，或者是集成于其他软件系统的插件、功能模块。该软件系统可以部署在计算设备集群中，计算设备集群执行软件系统的程序代码，从而执行本申请的数据标注方法。其中，软件系统可以以软件包的形式提供给用户，用户自行部署软件包至本地的计算设备或者是私有云。或者，软件系统也可以以云服务的形式提供给用户，例如软件系统可以部署在公有云，云平台通过开放云服务的应用程序编程接口（application programming interface，API），以供用户使用。在一些可能的实现方式中，数据标注系统也可以是硬件系统，例如为具有数据标注能力的计算设备集群，计算设备集群运行时，执行本申请的数据标注方法。

具体地，数据标注系统可以获取用户输入的自然语言指令和用户通过交互设备触发的交互操作数据，以及获取应用运行时的交互界面数据，交互操作数据包括至少一个交互操作的操作描述，交互界面数据包括交互界面视频或多个交互界面图。然后数据标注系统根据自然语言指令提取至少一个任务的任务描述。接着数据标注系统根据自然语言指令的时间戳、交互操作数据的时间戳和交互界面数据的时间戳，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互

界面关联，获得标注数据。其中，目标交互界面从交互界面视频中提取或者是多个交互界面图中的至少一个，标注数据用于训练或测试人工智能 AI 代理。

该方法可以自动采集用户输入的自然语言指令、通过交互设备触发的交互操作数据以及对应的应用运行时的交互界面数据，基于各数据的时间戳对不同类型数据自动对齐，由此实现交互任务的数据自动标注，为训练或测试应用的智能代理提供高质量的标注数据，可直接用于训练或测试，降低成本。

在一些可能的实现方式中，数据标注系统还可以根据交互界面视频或多个交互界面图，提取至少一个交互操作导致的界面变化数据。数据标注系统对界面变化数据进行语义识别，获得至少一个交互操作关联的任务的效果描述。然后数据标注系统将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面、目标任务的效果描述关联。

该方法通过提取交互操作导致的界面变化数据，可以进一步提高数据标注的精准度，而且，该数据标注系统还通过语义识别，获得任务的效果描述，将任务的效果描述和任务描述、目标交互操作、目标交互界面关联，由此可以提高标注数据的信息量，为 AI 代理的训练或测试提供更丰富的信息。

在一些可能的实现方式中，用户包括多个用户，多个用户通过不同通道接入应用。数据标注系统可以根据自然语言指令的时间戳和通道标识、交互操作数据的时间戳和通道标识以及交互界面数据的时间戳和通道标识，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联。

如此可以实现基于用户的通道标识对相关数据进行区分，实现细粒度的数据标注，如此可以提高标注数据的质量，进而提高通过该标注数据训练的 AI 代理的推理精度。

在一些可能的实现方式中，多个用户通过不同终端接入应用，通道标识包括终端标识。如此可以实现多用户通过各自终端交互的应用场景下的数据标注，满足业务需求。

在一些可能的实现方式中，多个用户通过相同终端接入应用。多个用户的交互设备连接至相同终端，通道标识包括交互设备的设备标识。例如，多人在线游戏场景中，多个用户可以通过各自的手柄连接至同一游戏主机进行游戏。如此可以实现多用户通过各自交互设备进行交互的应用场景下的数据标注，满足业务需求。

在一些可能的实现方式中，数据标注系统还可以获取应用的驱动逻辑。数据标注系统可以根据交互操作数据的时间戳和交互界面数据的时间戳，将至少一个交互操作中目标交互操作与交互界面数据中目标交互界面关联，然后根据应用的驱动逻辑将至少一个任务中目标任务的任务描述与目标交互操作或目标交互界面关联。

该方法通过结合应用的驱动逻辑和相关数据的时间戳对数据进行关联，进一步提高了标注数据的精准度。

在一些可能的实现方式中，交互操作包括键盘操作、鼠标操作、屏幕触摸操作、体感操作或控制器操作中的至少一种。如此可以实现对不同交互方式的交互数据标注，具有较高可用性、灵活性。

在一些可能的实现方式中，应用包括游戏应用、设计应用或绘图应用中的一种或多种。该方法可以适用于通用的人机交互任务的数据标注，而不局限于游戏应用，具有高可用性。

第二方面，本申请提供一种数据标注系统。所述数据标注系统包括：

采集子系统，用于获取用户输入的自然语言指令和所述用户通过交互设备触发的交互操作数据，以及获取所述应用运行时的交互界面数据，所述交互操作数据包括至少一个交互操作的操作描述，所述交互界面数据包括交互界面视频或多个交互界面图；

标注子系统，用于根据所述自然语言指令提取至少一个任务的任务描述；

所述标注子系统，还用于根据所述自然语言指令的时间戳、所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，获得标注数据，所述目标交互界面从所述交互界面视频中提取或者是多个交互界面图中的至少一个，所述标注数据用于训练或测试人工智能 AI 代理。

在一些可能的实现方式中，所述标注子系统还用于：

根据所述交互界面视频或所述多个交互界面图，提取所述至少一个交互操作导致的界面变化数据；

对所述界面变化数据进行语义识别，获得所述至少一个交互操作关联的任务的效果描述；

所述标注子系统具体用于：

将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面、所述目标任务的效果描述关联。

在一些可能的实现方式中，所述用户包括多个用户，所述多个用户通过不同通道接入所述应用；

所述标注子系统具体用于：

根据所述自然语言指令的时间戳和通道标识、所述交互操作数据的时间戳和通道标识以及所述交互界面数据的时间戳和通道标识，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联。

在一些可能的实现方式中，所述多个用户通过不同终端接入所述应用，所述通道标识包括终端标识。

在一些可能的实现方式中，所述多个用户通过相同终端接入所述应用，所述多个用户的交互设备连接至相同终端，所述通道标识包括所述交互设备的设备标识。

在一些可能的实现方式中，所述标注子系统还用于：

获取所述应用的驱动逻辑；

所述标注子系统具体用于：

根据所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个交互操作中目标交互操作与所述交互界面数据中目标交互界面关联；

根据所述应用的驱动逻辑将所述至少一个任务中目标任务的任务描述与所述目标交互操作或所述目标交互界面关联。

在一些可能的实现方式中，所述交互操作包括键盘操作、鼠标操作、屏幕触摸操作、体感操作或控制器操作中的至少一种。

在一些可能的实现方式中，所述应用包括游戏应用、设计应用或绘图应用中的一种或多种。

第三方面，本申请提供一种计算设备集群。该计算设备集群包括至少一台计算设备，至少一台计算设备包括至少一个处理器和至少一个存储器。至少一个处理器、至少一个存储器进行相互的通信。至少一个处理器用于执行至少一个存储器中存储的指令，以使得计算设备或计算设备集群执行如第一方面或第一方面的任一种实现方式所述的数据标注方法。

第四方面，本申请提供一种计算机可读存储介质，计算机可读存储介质中存储有指令，该指令指示计算设备或计算设备集群执行上述第一方面或第一方面的任一种实现方式所述的数据标注方法。

第五方面，本申请提供了一种包含指令的计算机程序产品，当其在计算设备或计算设备集群上运行时，使得计算设备或计算设备集群执行上述第一方面或第一方面的任一种实现方式所述的数据标注方法。

本申请在上述各方面提供的实现方式的基础上，还可以进行进一步组合以提供更多实现方式。

附图说明

为了更清楚地说明本申请的技术方法，下面将对所需使用的附图进行简单介绍。

图 1 为本申请提供的一种数据标注系统的架构示意图；

图 2 为本申请提供的一种数据标注方法的流程图；

图 3 为本申请提供的一种采集的数据的示意图；

图 4 为本申请提供的一种标注数据的示意图；

图 5 为本申请提供的一种多人在线游戏场景中数据标注方法的流程示意图；

图 6 为本申请提供的一种设计应用中生成场景的流程示意图；

图 7 为本申请提供的一种计算设备的结构示意图；

图 8 为本申请提供的一种计算设备集群的结构示意图；

图 9 为本申请提供的另一种计算设备集群的结构示意图；

图 10 为本申请提供的又一种计算设备集群的结构示意图。

具体实施方式

本申请实施例中的术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或

者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。

首先对本申请实施例中所涉及到的一些技术术语进行介绍。

人工智能代理（artificial intelligence agent，AI agent），也可以称作智能代理或智能主体、智能体。在AI领域，AI agent指感知环境并采取行动（如执行操作）以实现特定目标的自主实体。AI agent通常可以是一个软件程序，例如是独立的软件，或者是集成于软件的功能模块，AI agent也可以是物理实体，例如是部署有上述软件的计算设备。

AI agent可以广泛应用于各个领域，包括但不限于机器人、游戏、虚拟助手、自动驾驶汽车或者下一代应用开发平台等。以游戏领域为例，在设置有非玩家角色（non-player character，NPC）的游戏场景中，用户可以通过AI agent技术配置AI NPC，实现智能对话、驱动故事线、设定NPC的背景、性格、记忆、情绪等功能。

AI agent通常可以通过大量标注数据训练得到。仍以游戏领域示例说明，AI agent可以为可扩展指令多世界智能体（Scalable Instructable Multiworld Agent，SIMA）。SIMA的输入可以包括屏幕图像以及用户提供的简单的自然语言指令，SIMA可以将上述图像、自然语言指令映射到键鼠操作。基于此，SIMA可以通过标注键鼠操作的自然语言指令、屏幕图像训练得到。

然而，相关技术主要是人工收集、整合人类专家的数据，例如是收集视频画面、自然语言指令、键鼠操作数据等，并进行标注。人工收集、标注上述数据的效率较低，并且成本比较高昂，难以满足业务需求。

有鉴于此，本申请提供一种数据标注方法。该方法可以由数据标注系统执行。数据标注系统用于进行数据标注，标注数据可以用于学习人类操作和机器反馈，从而训练AI agent。其中，学习人类操作以及机器反馈的技术不仅可以用于游戏，而是一类通用的人机交互任务的代理开发方式，键鼠操作等交互操作数据可以映射到屏幕的交互界面（例如是交互界面变化数据）形成标注数据，标注数据可以被学习，以替代人类的工作任务。

数据标注系统可以是软件系统，该软件系统可以是独立的软件系统，或者是集成于其他软件系统的插件、功能模块。该软件系统可以部署在计算设备集群中，计算设备集群执行软件系统的程序代码，从而执行本申请的数据标注方法。其中，软件系统可以以软件包的形式提供给用户，用户自行部署软件包至本地的计算设备或者是私有云。或者，软件系统也可以以云服务的形式提供给用户，例如软件系统可以部署在公有云，云平台通过开放云服务的应用程序编程接口（application programming interface，API），以供用户使用。在一些可能的实现方式中，数据标注系统也可以是硬件系统，例如为具有数据标注能力的计算设备集群，计算设备集群运行时，执行本申请的数据标注方法。

具体地，数据标注系统可以获取用户输入的自然语言指令和用户通过交互设备触发的交互操作数据，以及获取应用运行时的交互界面数据。其中，交互操作数据包括至少一个交互操作的操作描述，交互界面数据包括交互界面视频或多个交互界面图。数据标注系统可以根据自然语言指令提取至少一个任务的任务描述，根据自然语言指令的时间戳、交互操作数据的时间戳和交互界面数据的时间戳，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联，获得用于训练或测试AI代理的标注数据。其中，目标交互界面为从交互界面视频中提取，或者是多个交互界面图中的至少一个。

该方法可以自动采集用户输入的自然语言指令、通过交互设备触发的交互操作数据以及对应的应用运行时的交互界面数据，基于各数据的时间戳对不同类型数据自动对齐，由此实现交互任务的数据自动标注，为训练或测试应用的智能代理提供高质量的标注数据，可直接用于训练或测试，降低成本。

为了使得本申请的技术方案更加清楚、易于理解，下面结合附图对本申请的数据标注系统的系统架构进行介绍。

参见图1所示的一种数据标注系统的系统架构图，数据标注系统10包括采集子系统100和标注子系统200。其中，采集子系统100、标注子系统200可以集中或分布式地部署在云环境中，云环境可以是多台计算设备形成的云计算集群，云环境可以包括公有云、私有云或者混合云。在一些示例中，采集

子系统 100、标注子系统 200 也可以部署本地的计算设备，例如是本地的服务器。为了便于描述，图 1 以采集子系统 100 和标注子系统 200 部署在云环境示例说明。其中，数据标注系统 10 可以以提供数据标注能力的云服务（如数据标注服务）提供给用户使用，例如数据标注系统 10 可以开放数据标注服务的 API，以供调用。

具体地，采集子系统 100 用于获取用户输入的自然语言指令和用户通过交互设备触发的交互操作数据，以及获取应用运行时的交互界面数据。其中，自然语言指令可以是基于自然语言的指令，自然语言指令可以包括文本指令或语音指令。交互操作数据包括至少一个交互操作的操作描述，该操作描述可以包括操作标识（identifier，ID），例如为操作名。操作描述还可以包括操作参数，例如是操作持续时间。其中，操作描述还可以包括触发交互操作的时间戳。交互界面数据包括交互界面视频或多个交互界面图。交互界面视频可以是对应用运行时的交互界面进行屏幕录制所得的视频，交互界面图可以是在运行时对交互界面进行截图获得的图像。如图 1 所示，采集子系统 100 可以使用 hook 或者直接在应用内截取获得交互界面图，例如采集子系统 100 可以基于 unreal 引擎直接在渲染模块写插件或者红绿蓝（red green blue，RGB）图像，以获得交互界面图。如此可以实现自动获取交互操作数据、自然语言指令（如语音形式的自然语言指令，可简称为语音指令）、交互界面数据（如自动截取驱动的画面图像），无需专门的人工标注和记录。

标注子系统 200 用于根据自然语言指令提取至少一个任务的任务描述，然后根据自然语言指令的时间戳、交互操作数据的时间戳和交互界面数据的时间戳，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联，获得标注数据。其中，目标交互界面从交互界面视频中提取或者是多个交互界面图中的至少一个。标注数据用于训练或测试 AI 代理。通过抽取任务描述（如语音指令中抽取的语音任务的任务描述）以及时间序列上关联的交互操作和画面帧列表，实现任务、操作、画面的自动对齐。

进一步地，上述采集子系统 100、标注子系统 200 可以包括多个功能模块。下面对采集子系统 100、标注子系统 200 的功能模块进行详细介绍。

在一些可能的实现方式中，采集子系统 100 可以包括指令采集模块 102、操作采集模块 104 和界面采集模块 106。指令采集模块 102 用于采集自然语言指令，操作采集模块 104 用于采集交互操作数据，例如是至少一个交互操作的操作描述，界面采集模块 106 用于采集交互界面数据，例如是交互界面视频。

其中，终端可以部署通信软件开发套件（software development kit，SDK）和驱动层，驱动层可以包括交互设备的驱动，例如是键盘、鼠标的驱动，驱动层也包括麦克风的驱动，云环境中可以部署应用的代码，终端可以通过 SDK 和驱动层接入应用，用户可以通过终端部署的通信 SDK 触发交互操作，如键鼠操作，或者是输入语音形式的自然语言指令。指令采集模块 102 可以基于驱动层中麦克风的驱动，采集语音形式的自然语言指令，操作采集模块 104 可以基于驱动层中键鼠的驱动，采集键鼠操作的操作数据。其中，操作数据可以保存在操作记录（例如是操作记录文件），语音形式的自然语言指令可以保存在音频记录（例如是音频文件）。界面采集模块 106 可以与应用对接，采集交互界面数据。例如，界面采集模块 106 可以使用钩子 hook 或者直接在应用内截取以获得交互界面视频。

在一些可能的实现方式中，标注子系统 200 可以包括任务提取模块 202、关联模块 204。任务提取模块 202 用于根据自然语言指令提取至少一个任务的任务描述，关联模块 204 用于根据自然语言指令的时间戳、交互操作数据的时间戳和交互界面数据的时间戳，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联，获得标注数据。

进一步地，标注子系统 200 还可以包括映射模块 206，映射模块 206 用于将操作映射至交互界面数据。例如，映射模块 206 可以从交互界面数据中识别交互操作导致画面变化的至少两帧交互界面，例如是变化前的交互界面、变化后的交互界面，将交互操作与上述至少两帧交互界面进行映射。进一步地，映射模块 206 还可以生成画面变化内容的理解，将操作与画面变化内容的理解进行映射。

需要说明的是，本申请的数据标注系统 10 可以针对计算机操作类的应用实现自动化数据采集与标注，获得标注数据。该标注数据可以输入代理训练系统 20 用于根据标注数据训练 AI agent。其中，代理训练系统 20 可以部署在云环境，从而实现在云上训练 AI agent。AI agent 可以基于转换器 transformer 构建。在训练 AI agent 时，可以获取多模态的向量 embedding，多模态的 embedding 与标注数据对齐，代

理训练系统 20 可以基于多模态的 embedding 训练 AI agent。进一步地，代理训练系统 20 还可以回放语音或随机任务，以验证交互操作（action）以及效果对应性。

还需要说明的是，图 1 仅仅是数据标注系统 10 的一种示意性划分方式，在本申请实施例其他可能的实现方式，数据标注系统 10 还可以按照功能划分为不同的子系统或模块。

基于前述的数据标注系统 10，本申请还提供一种数据标注方法。下面结合附图，对本申请的数据标注方法进行详细描述。

参见图 2 所示的一种数据标注方法的流程图，该方法包括如下步骤：

S202、数据标注系统 10 获取用户输入的自然语言指令和用户通过交互设备触发的交互操作数据。

自然语言指令是指通过自然语言发出的指令。以游戏应用为例，自然语言指令可以包括点塔，其中，点塔表示在游戏中攻击敌方塔的行为。自然语言指令可以是语音形式的指令，或者是文本形式的指令。例如，用户可以发出“点塔”的语音，或者在交互界面的文本输入框输入“点塔”。

交互操作数据包括至少一个交互操作的操作描述。其中，操作描述可以包括操作类型。在一些示例中，操作类型可以基于交互设备划分。具体地，交互设备可以包括键盘、鼠标、体感设备、控制器或者触摸屏。其中，控制器可以为手柄控制器或操控杆控制器。相应地，交互操作可以包括键盘操作、鼠标操作、屏幕触摸操作、体感操作或控制器操作中的至少一种。基于此，交互操作数据包括用户通过交互设备触发的交互操作的操作类型。例如，交互操作数据可以为键盘 A/虚拟操控杆/鼠标。

具体实现时，数据标注系统 10 启动云上应用（application，APP）环境部署，其中，应用可以是游戏应用、设计应用或绘图应用。在一些示例中，数据标注系统 10 为云上自动标注服务时，数据标注系统 10 还可以启动关联的采集子系统 100、标注子系统 200 的部署。当玩家通过客户端（如浏览器）接入云上部署的 APP 操作时，数据标注系统 10 可以获取用户输入的自然语言指令和用户通过交互设备触发的交互操作数据。其中，数据标注系统 10 可以通过麦克风的驱动，获取用户输入的语音形式的自然语言指令，或者通过键盘、鼠标的驱动，获取用户输入的文本形式的自然语言指令。

需要说明的是，数据标注系统 10 可以向用户呈现授权许可界面，授权许可界面可以包授权许可控件，用户（玩家）可以在阅读授权许可协议后，触发授权许可控件，从而触发授权许可操作。相应地，数据标注系统 10 可以在获取用户授权许可的情况下，获取该用户输入的自然语言指令、该用户通过交互设备触发的交互操作数据。其中，应用开发商可以将应用的可执行程序提供至云上的数据标注系统 10（如云上自动操作采集及数据理解服务），以便在云环境部署应用。应用开发商还可以选择特定用户或用户群（例如是游戏的玩家、玩家群），以针对上述特定用户或用户群启动自动数据采集及标注（包括数据理解）。

S204、数据标注系统 10 获取应用运行时的交互界面数据。

交互界面数据包括交互界面视频或多个交互界面图。数据标注系统 10 可以通过对应用运行时的交互界面进行屏幕录制，获得交互界面视频，或者是对应用运行时的交互界面进行截图，获得交互界面图。当应用部署在云环境时，数据标注系统 10 可以基于云的托管环境，获得应用运行时的交互界面视频或多个交互界面图。在一些示例中，交互界面图可以是视频的视频帧。其中，视频帧的获取可以基于操作系统截取，也可以在 APP 或者引擎等底层软件加载插件来获取。

在一些可能的实现方式中，自然语言指令、交互操作数据、交互界面数据还可以包括时间戳（timestamp）。自然语言指令的时间戳、交互操作数据的时间戳、交互界面数据的时间戳可以用于数据对齐、标注。

数据标注系统 10 可以按时间戳存储相应时间戳的自然语言指令、交互操作数据、交互界面数据。为了便于理解，本申请提供一示例进行说明。如图 3 所示，数据标注系统 10 可以获取时间戳为 10:00:00 的自然语言指令、交互操作数据、交互界面数据，其中，自然语言指令包括内容为“去点塔、点塔”的音频，交互操作数据可以为“键盘 A/虚拟操控杆/鼠标”，交互界面数据可以应用运行时的视频画面，具体为帧 n-1、帧 n。

进一步地，针对多用户协作应用，例如是多人在线游戏，数据标注系统 10 还可以获取数据的通道标识（channel ID），每个客户端连接可以视为一个 channel，通过该客户端连接触发或传输的数据可以关

联至该客户端的用户。数据标注系统 10 还可以基于 channel ID 实现用户粒度的数据对齐。

S206、数据标注系统 10 根据自然语言指令提取至少一个任务的任务描述。

具体地，当自然语言指令为语音形式时，数据标注系统 10 可以通过语音转文本（speech to text），将上述语音形式的自然语言指令转换为文本形式的自然语言指令，接着数据标注系统 10 可以通过文本识别，从文本形式的自然语言指令中提取任务至少一个任务的任务描述。其中，任务描述可以包括任务类型，例如自然语言指令为点塔时，数据标注系统 10 基于该自然语言指令可以提取任务的任务描述，具体为“攻击防御塔和水晶”。在一些可能的实现方式中，数据标注系统 10 也可以直接从语音形式的自然语言指令中提取至少一个任务的任务描述。例如，数据标注系统 10 可以利用标注有任务描述的语音，训练任务描述提取模型，数据标注系统 10 可以将语音形式的自然语言指令输入训练好的任务描述提取模型，从而提取至少一个任务的任务描述。

当自然语言指令为文本形式时，数据标注系统 10 可以通过文本识别，从文本形式的自然语言指令中提取至少一个任务的任务描述。其中，在提取至少一个任务的任务描述时，还可以结合先验知识进行提取。先验知识可以来自应用的业务逻辑，或者是人工自定义。例如，在游戏应用中，先验知识可以包括“塔”表示防御塔。

S208、数据标注系统 10 根据自然语言指令的时间戳、交互操作数据的时间戳和交互界面数据的时间戳，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联，获得标注数据。

任务描述是从自然语言指令中提取的，数据标注系统 10 可以根据自然语言指令的时间戳、交互操作数据的时间戳和交互界面数据的时间戳，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联。其中，目标交互界面从所述交互界面视频中提取或者是多个交互界面图中的至少一个。目标任务的任务描述可以是从目标时间的自然语言指令提取的，目标时间与目标交互操作的时间戳、目标交互界面的时间戳匹配。本申请中，时间戳匹配可以是时间戳相等或接近，时间戳的差值的绝对值小于设定值，表示时间戳接近。关联后的数据为标注数据，标注数据用于训练或测试 AI 代理。

在一些可能的实现方式中，数据标注系统 10 还可以根据交互操作数据的时间戳和交互界面数据的时间戳，将至少一个交互操作中目标交互操作与交互界面数据中目标交互界面关联，并根据应用的驱动逻辑将至少一个任务中目标任务的任务描述与目标交互操作或目标交互界面关联，如此可以实现将将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联，获得标注数据。其中，驱动逻辑定义了业务流，基于驱动逻辑可以实现任务到交互界面的关联。为了保障标注的准确性，数据标注系统 10 可以根据时间戳，结合应用的驱动逻辑将至少一个任务中目标任务的任务描述与目标交互操作或目标交互界面关联。

进一步地，数据标注系统 10 还可以根据交互界面视频或多个交互界面图，提取至少一个交互操作导致的界面变化数据。例如，数据标注系统 10 可以根据交互操作的时间戳，获取与该时间戳对应的交互界面视频，例如是相同时间戳的交互界面视频或邻近界面的交互界面视频，然后数据标注系统 10 可以对该交互界面视频进行语义分析或语义识别，从而提取上述交互操作导致画面变化的交互界面图，数据标注系统 10 可以对上述交互界面图进行处理，例如是作差处理，从而获得界面变化数据。数据标注系统 10 可以对界面变化数据进行语义识别，获得至少一个交互操作关联的任务的效果描述。相应地，数据标注系统 10 可以将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面、目标任务的效果描述关联。

需要说明的是，数据标注系统 10 可以对某一时间戳的数据进行关联，从而实现多维度数据自动对齐。在一些示例中，数据标注系统 10 也可以对一段时间的数据进行关联，从而实现多维度数据自动对齐。

如图 4 所示，对时间戳为 10:00:00 的数据进行关联可以得到如下标注数据：

数据理解后数据：10:00:00

Task 任务：攻击防御塔和水晶

Action 关联操作：键盘 A/虚拟操控杆/鼠标

Result 关联结果画面：帧 n-1→帧 n

任务效果理解：键盘 A 一次点击，防御塔掉血 20%

对时间戳 10:00:00 至时间戳 10:02:00 的数据进行关联可以得到如下标注数据：

长视野任务分析（Long-Horizon）：10:00:00-10:02:00

Task 任务：去上路推塔

Action 关联操作：键盘 A/虚拟操控杆/鼠标

Result 关联结果画面：帧 n-1→帧 n

任务效果理解：键盘 A 一次点击，防御塔掉血 20%

进一步地，数据标注系统 10 还可以将标注数据输入 AI 代理进行训练，例如是通过微调（fine-tuning）方式对 AI 代理进行微调。其中，应用开发商可以随时查询数据内容，并启动 AI 代理的训练或微调。当应用开发商已有针对该应用的 AI 代理，数据标注系统 10 也可以将标注数据用于对 AI 代理进行有效性验证。

该方法支持自动采集用户输入的自然语言指令、通过交互设备触发的交互操作数据以及对应的应用运行时的交互界面数据，基于各数据的时间戳对不同类型数据自动对齐，由此实现交互任务的数据自动标注，为训练或测试应用的智能代理提供高质量的标注数据，可直接用于训练或测试，降低成本。而且，该方法可以增强应用适应性，例如一些领域的特定术语如游戏领域的“去上路”可以和操作、界面直接对应，具有高可用性。

其中，自然语言指令、交互操作数据、应用运行时的交互界面数据等数据采集可以是在应用（如游戏服务）和最终用户授权的情况下自动采集，能够满足安全、隐私需求。

图 2 所示的数据标注方法可以应用于单人交互的应用，也可以应用于多人协作的应用。在多人协作应用中，多个用户可以通过不同通道接入应用。数据标注系统 10 也可以根据自然语言指令的时间戳和通道标识、交互操作数据的时间戳和通道标识以及交互界面数据的时间戳和通道标识，将至少一个任务中目标任务的任务描述与至少一个交互操作中目标交互操作、交互界面数据中目标交互界面关联。

需要说明的是，多个用户可以通过不同终端接入应用。例如，在多人在线手游场景中，多个用户可以通过各自的智能手机接入应用。在该情况下，通道标识可以为终端标识。在一些示例中，多个用户也可以通过相同终端接入应用。例如，在多人在线主机游戏场景中，多个用户可以通过相同终端接入应用。该终端例如可以是游戏主机。多个用户的交互设备可以连接至相同终端。其中，多个用户的交互设备可以是多个用户的手柄控制器。在该情况下，通道标识可以包括交互设备的设备标识。

为了便于理解，下面以多人在线游戏场景对本申请的数据标注方法进行示例说明。

参见图 5 所示的一种多人在线游戏场景中数据标注方法的流程示意图，在该场景中，当有多个玩家在同一个的游戏 APP 时，需要能够区分不同玩家的数据，其中，每个玩家的接入连接（如客户端连接）是一个独立的 channel，可以将该玩家的键鼠操作、声音与 channel ID 关联，从而区分不同用户的数据。

在该示例中，用户 1 的数据与 channel 1 的 channel ID 关联，用户 2 的数据与 channel 2 的 channel ID 关联。部署在云环境中的数据标注系统 10 中的采集子系统 100 可以分别采集不同用户的自然语言指令，存储在音频记录中，其中，采集子系统 100 还记录音频采集时间。采集子系统 100 还分别采集不同用户的交互操作数据以及时间戳，存储在操作记录中。此外，采集子系统 100 还可以截取应用运行时的界面，存储至视频帧记录。其中，视频帧记录中还记录有交互界面数据的时间戳。数据标注系统 10 中的标注子系统 200 可以对不同用户的自然语言指令进行任务提取获得不同用户的任务的任务描述，对不同用户的视频帧记录进行视频信息提取，并将视频帧与操作进行映射，标注子系统 200 可以将任务与操作、视频帧进行关联，从而获得标注数据。基于该标注数据训练或测试 AI 代理的具体实现可以参考图 1 实施例相关内容描述，在此不再赘述。

其中，交互界面数据（游戏画面）与玩家的操作的关联关系可以通过多种方式实现。一种方式可以为，数据标注系统 10 从游戏应用开发商以服务接口或人工方式获得通道与玩家标识（player ID）的关联关系，当画面变化时，数据标注系统 10 可以通过图像分割与识别技术，获得该玩家的画面变化情况。另一种方式可以为，游戏应用的交互界面中可以显示玩家的名称，由此可以通过视觉大模型直接学习玩

家的操作与其操作的对象的映射关系。其中，视觉大模型可以通过文本识别、图像识别学习玩家的操作与其操作的对象的映射关系。

本申请的数据标注方法不仅可以应用于游戏应用，还可以泛化至除游戏外的所有人机交互式应用软件智能代理场景。例如该数据标注方法可以应用于常见的桌面应用。下面以桌面应用为设计应用示例说明，如图 6 所示，在生成家居场景时，如果需要调整光照，例如用户输入改变场景的光照为暗光，通常可以通过集成 blender 的 API，通过 Python 脚本实现光照的控制。其中，blender 是一款开源的三维计算机图形软件，可以用于创建和渲染三维模型、动画、游戏、视觉效果等。基于此，在本申请的方案中，数据标注系统 10 可以采集用户的自然语言指令、鼠标操作和交互界面数据，进行自动标注，获得标注数据，基于该数据训练 AI 代理。该 AI 代理可以模拟鼠标操作，例如是模拟用户通过鼠标调用 blender API 的操作，实现场景编辑（包括但不限于生成家居场景）。

基于上述内容描述，可知本申请的数据标注方法可以针对计算机操作类的应用实现自动化数据采集与标注，降低数据采集难度并提升精度，基于标注数据能够实现针对场景的有效学习，可泛化除游戏外的所有人机交互式应用软件智能代理场景。

基于图2至图6所示实施例的数据标注方法，本申请还提供一种数据标注系统10。如图1所示，该数据标注系统10包括：

采集子系统 100，用于获取用户输入的自然语言指令和所述用户通过交互设备触发的交互操作数据，以及获取所述应用运行时的交互界面数据，所述交互操作数据包括至少一个交互操作的操作描述，所述交互界面数据包括交互界面视频或多个交互界面图；

标注子系统 200，用于根据所述自然语言指令提取至少一个任务的任务描述；

所述标注子系统 200，还用于根据所述自然语言指令的时间戳、所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，获得标注数据，所述目标交互界面从所述交互界面视频中提取或者是多个交互界面图中的至少一个，所述标注数据用于训练或测试人工智能 AI 代理。

示例性地，上述采集子系统100、标注子系统200可以通过硬件实现，或者可以通过软件实现。

当通过软件实现时，采集子系统 100、标注子系统 200 可以是运行在计算机设备上的应用程序，如计算引擎等。上述应用程序还可以进行虚拟化，以虚拟化服务形式提供给用户使用。虚拟化服务可以包括虚拟机（virtual machine，VM）服务、裸金属服务器（bare metal server，BMS）服务或容器（container）服务。其中，VM 服务可以是通过虚拟化技术在多个物理主机上虚拟出虚拟机（virtual machine，VM）资源池以为用户按需提供 VM 进行使用的服务。BMS 服务是在多个物理主机上虚拟出 BMS 资源池以为用户按需提供 BMS 进行使用的服务。容器服务是在多个物理主机上虚拟出容器资源池以为用户按需提供容器进行使用的服务。VM 是模拟出来的一台虚拟的计算机，也即逻辑上的一台计算机。BMS 是一种可弹性伸缩的高性能计算服务，计算性能与传统物理机无差别，具有安全物理隔离的特点。容器是一种内核虚拟化技术，可以提供轻量级的虚拟化，以达到隔离用户空间、进程和资源的目的。应理解，上述虚拟化服务中的 VM 服务、BMS 服务以及容器服务仅仅是作为具体的事例，在实际应用中，虚拟化服务还可以其他轻量级或者重量级的虚拟化服务，此处不作具体限定。

当通过硬件实现时，采集子系统 100、标注子系统 200 中可以包括至少一个计算设备，如服务器等。或者，采集子系统 100、标注子系统 200 也可以是利用专用集成电路（application-specific integrated circuit，ASIC）实现、或可编程逻辑器件（programmable logic device，PLD）实现的设备等。其中，上述 PLD 可以是复杂程序逻辑器件（complex programmable logical device，CPLD）、现场可编程门阵列（field-programmable gate array，FPGA）、通用阵列逻辑（generic array logic，GAL）或其任意组合实现。

在一些可能的实现方式中，所述标注子系统 200 还用于：

根据所述交互界面视频或所述多个交互界面图，提取所述至少一个交互操作导致的界面变化数据；

对所述界面变化数据进行语义识别，获得所述至少一个交互操作关联的任务的效果描述；

所述标注子系统 200 具体用于：

将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界

面数据中目标交互界面、所述目标任务的效果描述关联。

在一些可能的实现方式中，所述用户包括多个用户，所述多个用户通过不同通道接入所述应用；

所述标注子系统200具体用于：

根据所述自然语言指令的时间戳和通道标识、所述交互操作数据的时间戳和通道标识以及所述交互界面数据的时间戳和通道标识，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联。

在一些可能的实现方式中，所述多个用户通过不同终端接入所述应用，所述通道标识包括终端标识。

在一些可能的实现方式中，所述多个用户通过相同终端接入所述应用，所述多个用户的交互设备连接至相同终端，所述通道标识包括所述交互设备的设备标识。

在一些可能的实现方式中，所述标注子系统200还用于：

获取所述应用的驱动逻辑；

所述标注子系统200具体用于：

根据所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个交互操作中目标交互操作与所述交互界面数据中目标交互界面关联；

根据所述应用的驱动逻辑将所述至少一个任务中目标任务的任务描述与所述目标交互操作或所述目标交互界面关联。

在一些可能的实现方式中，所述交互操作包括键盘操作、鼠标操作、屏幕触摸操作、体感操作或控制器操作中的至少一种。

在一些可能的实现方式中，所述应用包括游戏应用、设计应用或绘图应用中的一种或多种。

本申请还提供一种计算设备700。如图7所示，计算设备700包括：总线702、处理器704、存储器706和通信接口708。处理器704、存储器706和通信接口708之间通过总线702通信。计算设备700可以是服务器或终端设备。应理解，本申请不限定计算设备700中的处理器、存储器的个数。

总线702可以是外设部件互连标准（peripheral component interconnect，PCI）总线或扩展工业标准结构（extended industry standard architecture，EISA）总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图7中仅用一条线表示，但并不表示仅有一根总线或一种类型的总线。总线702可包括在计算设备700各个部件（例如，存储器706、处理器704、通信接口708）之间传送信息的通路。

处理器704可以包括中央处理器(central processing unit，CPU)、图形处理器（graphics processing unit，GPU）、微处理器（micro processor，MP）或者数字信号处理器（digital signal processor，DSP）等处理器中的任意一种或多种。

存储器706可以包括易失性存储器(volatile memory)，例如随机存取存储器(random access memory，RAM)。存储器706还可以包括非易失性存储器(non-volatile memory)，例如只读存储器(read-only memory，ROM)，快闪存储器，机械硬盘（hard disk drive，HDD）或固态硬盘（solid state drive，SSD）。存储器706中存储有可执行的程序代码，处理器704执行该可执行的程序代码以实现前述数据标注方法。具体的，存储器706上存有数据标注系统10用于执行数据标注方法的指令。

通信接口708使用例如但不限于网络接口卡、收发器一类的收发模块，来实现计算设备700与其他设备或通信网络之间的通信。

本申请实施例还提供了一种计算设备集群。该计算设备集群包括至少一台计算设备。该计算设备可以是服务器，例如是中心服务器、边缘服务器，或者是本地数据中心中的本地服务器。在一些实施例中，计算设备也可以是台式机、笔记本电脑或者智能手机等终端设备。

如图8所示，所述计算设备集群包括至少一个计算设备700。计算设备集群中的一个或多个计算设备700中的存储器706中可以存有相同的数据标注系统10用于执行数据标注方法的指令。

在一些可能的实现方式中，该计算设备集群中的一个或多个计算设备700也可以用于执行数据标注系统10用于执行数据标注方法的部分指令。换言之，一个或多个计算设备700的组合可以共同

执行数据标注系统10用于执行数据标注方法的指令。

需要说明的是，计算设备集群中的不同的计算设备700中的存储器706可以存储不同的指令，用于执行数据标注系统10的部分功能。

图9示出了一种可能的实现方式。如图9所示，两个计算设备700A和700B通过通信接口708实现连接。计算设备700A中的存储器上存有用于执行功能分析子系统202的功能的指令。计算设备700B中的存储器上存有用于执行采集子系统100、标注子系统200的功能的指令。换言之，计算设备700A和700B的存储器706共同存储了数据标注系统10用于执行数据标注方法的指令。

图9所示的计算设备集群之间的连接方式可以是考虑到本申请提供的数据标注方法需要较多算力资源对采集的数据进行分析、关联，从而实现数据标注。因此，考虑将标注子系统200实现的功能交由独立的计算设备执行。例如，采集子系统100实现的功能可以由计算设备700A执行，标注子系统200实现的功能交由计算设备700B执行。

应理解，图9中示出的计算设备700A的功能也可以由多个计算设备700完成。同样，计算设备700B的功能也可以由多个计算设备700完成。

在一些可能的实现方式中，计算设备集群中的一个或多个计算设备可以通过网络连接。其中，所述网络可以是广域网或局域网等等。图10示出了一种可能的实现方式。如图10所示，两个计算设备700C和700D之间通过网络进行连接。具体地，通过各个计算设备中的通信接口与所述网络进行连接。在这一类可能的实现方式中，计算设备700C中的存储器706中存有执行采集子系统100的功能的指令。同时，计算设备700D中的存储器706中存有执行标注子系统200的功能的指令。

图10所示的计算设备集群之间的连接方式可以是考虑到本申请提供的数据标注方法需要大量计算资源进行任务描述提取以及基于时间戳进行数据关联，因此考虑将标注子系统200实现的功能交由独立的计算设备执行。

应理解，图10中示出的计算设备700C的功能也可以由多个计算设备700完成。同样，计算设备700D的功能也可以由多个计算设备700完成。

本申请实施例还提供了一种计算机可读存储介质。所述计算机可读存储介质可以是计算设备能够存储的任何可用介质或者是包含一个或多个可用介质的数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘）等。该计算机可读存储介质包括指令，所述指令指示计算设备执行上述应用于数据标注系统10用于执行数据标注方法。

本申请实施例还提供了一种包含指令的计算机程序产品。所述计算机程序产品可以是包含指令的，能够运行在计算设备上或被储存在任何可用介质中的软件或程序产品。当所述计算机程序产品在至少一个计算设备上运行时，使得至少一个计算设备执行上述数据标注方法。

最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的保护范围。

权 利 要 求 书

1.一种数据标注方法，其特征在于，所述方法包括：

获取用户输入的自然语言指令和所述用户通过交互设备触发的交互操作数据，以及获取所述应用运行时的交互界面数据，所述交互操作数据包括至少一个交互操作的操作描述，所述交互界面数据包括交互界面视频或多个交互界面图；

根据所述自然语言指令提取至少一个任务的任务描述；

根据所述自然语言指令的时间戳、所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，获得标注数据，所述目标交互界面从所述交互界面视频中提取或者是多个交互界面图中的至少一个，所述标注数据用于训练或测试人工智能 AI 代理。

2. 根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

根据所述交互界面视频或所述多个交互界面图，提取所述至少一个交互操作导致的界面变化数据；

对所述界面变化数据进行语义识别，获得所述至少一个交互操作关联的任务的效果描述；

所述将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，包括：

将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面、所述目标任务的效果描述关联。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，所述用户包括多个用户，所述多个用户通过不同通道接入所述应用；

所述根据所述自然语言指令的时间戳、所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，包括：

根据所述自然语言指令的时间戳和通道标识、所述交互操作数据的时间戳和通道标识以及所述交互界面数据的时间戳和通道标识，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联。

4. 根据权利要求 3 所述的方法，其特征在于，所述多个用户通过不同终端接入所述应用，所述通道标识包括终端标识。

5. 根据权利要求 3 所述的方法，其特征在于，所述多个用户通过相同终端接入所述应用，所述多个用户的交互设备连接至相同终端，所述通道标识包括所述交互设备的设备标识。

6. 根据权利要求 1 至 5 任一项所述的方法，其特征在于，所述方法还包括：

获取所述应用的驱动逻辑；

所述将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，包括：

根据所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个交互操作中目标交互操作与所述交互界面数据中目标交互界面关联；

根据所述应用的驱动逻辑将所述至少一个任务中目标任务的任务描述与所述目标交互操作或所述目标交互界面关联。

7. 根据权利要求 1 至 6 任一项所述的方法，其特征在于，所述交互操作包括键盘操作、鼠标操作、屏幕触摸操作、体感操作或控制器操作中的至少一种。

8. 根据权利要求 1 至 7 任一项所述的方法，其特征在于，所述应用包括游戏应用、设计应用或绘图应用中的一种或多种。

9. 一种数据标注系统，其特征在于，所述系统包括：

采集子系统，用于获取用户输入的自然语言指令和所述用户通过交互设备触发的交互操作数据，以及获取所述应用运行时的交互界面数据，所述交互操作数据包括至少一个交互操作的操作描述，所述交互界面数据包括交互界面视频或多个交互界面图；

标注子系统，用于根据所述自然语言指令提取至少一个任务的任务描述；

所述标注子系统，还用于根据所述自然语言指令的时间戳、所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联，获得标注数据，所述目标交互界面从所述交互界面视频中提取或者是多个交互界面图中的至少一个，所述标注数据用于训练或测试人工智能 AI 代理。

10. 根据权利要求 9 所述的系统，其特征在于，所述标注子系统还用于：

根据所述交互界面视频或所述多个交互界面图，提取所述至少一个交互操作导致的界面变化数据；

对所述界面变化数据进行语义识别，获得所述至少一个交互操作关联的任务的效果描述；

所述标注子系统具体用于：

将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面、所述目标任务的效果描述关联。

11. 根据权利要求 9 或 10 所述的系统，其特征在于，所述用户包括多个用户，所述多个用户通过不同通道接入所述应用；

所述标注子系统具体用于：

根据所述自然语言指令的时间戳和通道标识、所述交互操作数据的时间戳和通道标识以及所述交互界面数据的时间戳和通道标识，将所述至少一个任务中目标任务的任务描述与所述至少一个交互操作中目标交互操作、所述交互界面数据中目标交互界面关联。

12. 根据权利要求 11 所述的系统，其特征在于，所述多个用户通过不同终端接入所述应用，所述通道标识包括终端标识。

13. 根据权利要求 11 所述的系统，其特征在于，所述多个用户通过相同终端接入所述应用，所述多个用户的交互设备连接至相同终端，所述通道标识包括所述交互设备的设备标识。

14. 根据权利要求 9 至 13 任一项所述的系统，其特征在于，所述标注子系统还用于：

获取所述应用的驱动逻辑；

所述标注子系统具体用于：

根据所述交互操作数据的时间戳和所述交互界面数据的时间戳，将所述至少一个交互操作中目标交互操作与所述交互界面数据中目标交互界面关联；

根据所述应用的驱动逻辑将所述至少一个任务中目标任务的任务描述与所述目标交互操作或所述目标交互界面关联。

15. 根据权利要求 9 至 14 任一项所述的系统，其特征在于，所述交互操作包括键盘操作、鼠标操作、屏幕触摸操作、体感操作或控制器操作中的至少一种。

16. 根据权利要求 9 至 15 任一项所述的系统，其特征在于，所述应用包括游戏应用、设计应用或绘图应用中的一种或多种。

17. 一种计算设备集群，其特征在于，所述计算设备集群包括至少一台计算设备，所述至少一台计算设备包括至少一个处理器和至少一个存储器，所述至少一个存储器中存储有计算机可读指令；所述至少一个处理器执行所述计算机可读指令，以使得所述计算设备集群执行如权利要求 1 至 8 中任一项所述的数据标注方法。

18. 一种计算机可读存储介质，其特征在于，包括计算机可读指令；所述计算机可读指令用于实现权利要求 1 至 8 任一项所述的数据标注方法。

19. 一种计算机程序产品，其特征在于，包括计算机可读指令；所述计算机可读指令用于实现权利要求 1 至 8 任一项所述的数据标注方法。