

(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
2025 年 5 月 30 日 (30.05.2025)



(10) 国际公布号
WO 2025/107888 A1

(51) 国际专利分类号:

G06F 16/903 (2019.01)

(21) 国际申请号:

PCT/CN2024/122387

(22) 国际申请日:

2024 年 9 月 29 日 (29.09.2024)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202311570383.1 2023年11月22日 (22.11.2023) CN

(71) 申请人: 北京字跳网络技术有限公司

(**BEIJING ZITIAO NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市海淀区紫金数码园4号楼2层0207 100190 (CN)。

(72) 发明人: 黄巍(**HUANG, Wei**); 中国北京市朝阳区七

圣中街12号院融中心B1小邮局 100028 (CN)。姚璨(**YAO, Can**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局 100028 (CN)。黄树炫(**HUANG, Shuxuan**); 中国北京市朝阳区七圣中街12号院融

中心B1小邮局 100028 (CN)。马进(**MA, Jin**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局 100028 (CN)。蒋世豪(**JIANG, Shihao**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局 100028 (CN)。孟德森(**MENG, Desen**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局 100028 (CN)。

(74) 代理人: 北京世辉律师事务所(**SHIHUI PARTNERS**); 中国北京市朝阳区建国门外大街2号北京银泰中心C座42层 100022 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** APPLICATION DEBUGGING METHOD, INFORMATION QUERY METHOD, AND APPLICATION EVALUATION METHOD

(54) 发明名称: 应用调试的方法、信息查询的方法、应用评测的方法

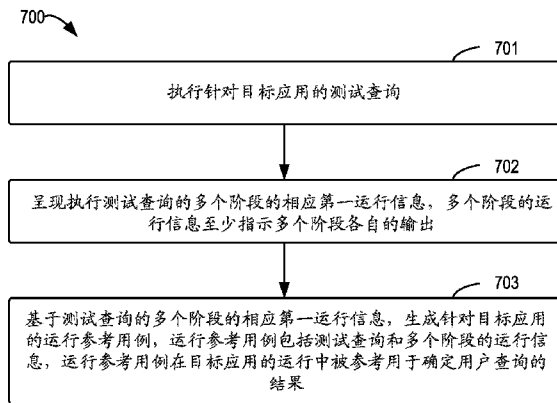


图 7A

701 Execute a test query for a target application

702 Present corresponding first operation information of a plurality of stages for executing the test query, wherein the operation information of the plurality of stages at least indicates respective outputs of the plurality of stages

703 Generate an operation reference case for the target application on the basis of the corresponding first operation information of the plurality of stages for the test query, wherein the operation reference case comprises the test query and the operation information of the plurality of stages, and the operation reference case is referenced for determining a result of a user query during the operation of the target application

(57) **Abstract:** Provided in the embodiments of the present disclosure are an application debugging method and apparatus, an information query method and apparatus, an application evaluation method and apparatus, and a device and a storage medium. The application debugging method comprises: executing a test query for a target application; presenting corresponding first operation information of a plurality of stages for executing the test query, wherein the operation information of the plurality of stages at least indicates respective outputs of the plurality of stages; and generating an operation reference case for the target application on the basis of the corresponding first operation information of the plurality of stages for the test query, wherein the operation reference case comprises the test query and the operation information of the plurality of stages, and the operation reference case is referenced for determining a result of a user query during the operation of the target application. In this way, an application-creating user can improve the quality of services provided by a target application without the need for modifying code, thereby simplifying the debugging process of the target application.

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))。

(57) 摘要: 本公开的实施例提供了应用调试的方法、信息查询的方法、应用评测的方法、装置、设备和存储介质。应用调试的方法包括: 执行针对目标应用的测试查询; 呈现执行测试查询的多个阶段的相应第一运行信息, 多个阶段的运行信息至少指示多个阶段各自的输出; 以及基于测试查询的多个阶段的相应第一运行信息, 生成针对目标应用的运行参考用例, 运行参考用例包括测试查询和多个阶段的运行信息, 运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。以此方式, 应用创建用户无需修改代码就可以提升目标应用提供服务的质量, 简化了目标应用的调试过程。

应用调试的方法、信息查询的方法、应用评测的方法

5 本申请要求 2023 年 11 月 22 日递交的、标题为“应用调试的方法、信息查询的方法、应用评测的方法”、申请号为 202311570383.1 的中国发明专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

10 本公开的示例实施例总体涉及计算机领域，特别地涉及用于应用调试的方法、信息查询的方法、应用评测的方法、装置、设备和计算机可读存储介质。

背景技术

15 随着信息技术的发展，各种终端设备可以在工作和生活等方面向人们提供各种服务。终端设备中可以部署有提供服务的应用。终端设备通过应用的用户界面呈现相应内容并实现与用户之间的交互，满足用户的各类需求。在一些情况下，用户可能在应用内发起查询请求。因此，如何提高向用户提供的查询服务的质量是所关注的问题。

发明内容

20 在本公开的第一方面，提供了一种应用调试的方法。该方法包括：执行针对目标应用的测试查询；呈现执行测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出；以及基于测试查询的多个阶段的相应第一运行信息，生成针对目标应用的运行参考用例，运行参考用例包括测试查询和多个阶段的运行信息，运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。

25 在本公开的第二方面，提供了一种信息查询的方法。该方法包括：响应于接收到针对目标应用的用户查询，基于用户查询与测试查询之间的相似度超过阈值相似度，将包括测试查询的运行参考用例确定为用户查询的目标运行参考用例，其中运行参考用例包括测试查询和测试查询的多个阶段的相应运行信息；以及基于目标运行参考用例来执行所述用户查询。

30 在本公开的第三方面，提供了一种应用评测的方法。该方法包括：生成用于所述目标应用的评测参考用例，所述评测参考用例包括评测查询、所述目标应用针对所述评测查询而生成的第一结果和对所述第一结果的评测反馈；基于所述评测查询与评测样本集中的查询之间的相似度，从所述评测样本集中选择目标评测样本，所述目标评测样本包括样本查询和由所述目标应用针对所述样本查询生成的第二结果；以及基于

所述第一结果、所述评测反馈和所述第二结果，生成针对所述目标应用的评测结果。

在本公开的第四方面，提供了一种用于应用调试的装置。该装置包括：测试查询执行模块，被配置为执行针对目标应用的测试查询；第一运行信息呈现模块，被配置为呈现执行测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出；以及运行参考用例生成模块，被配置为基于测试查询的多个阶段的相应第一运行信息，生成针对目标应用的运行参考用例，运行参考用例包括测试查询和多个阶段的运行信息，运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。

在本公开的第五方面，提供了一种用于信息查询的装置。该装置包括：运行参考用例确定模块，被配置为响应于接收到针对目标应用的用户查询，基于所述用户查询与测试查询之间的相似度超过阈值相似度，将包括所述测试查询的运行参考用例确定为所述用户查询的目标运行参考用例，其中所述运行参考用例包括所述测试查询和所述测试查询的多个阶段的相应运行信息；以及用户查询执行模块，被配置为基于所述目标运行参考用例来执行所述用户查询。

在本公开的第六方面，提供了一种用于应用评测的装置。该装置包括：评测参考用例生成模块，被配置为生成用于所述目标应用的评测参考用例，所述评测参考用例包括评测查询、所述目标应用针对所述评测查询而生成的第一结果和对所述第一结果的评测反馈；目标评测样本选择模块，被配置为基于所述评测查询与评测样本集中的查询之间的相似度，从所述评测样本集中选择目标评测样本，所述目标评测样本包括样本查询和由所述目标应用针对所述样本查询生成的第二结果；以及评测结果生成模块，被配置为基于所述第一结果、所述评测反馈和所述第二结果，生成针对所述目标应用的评测结果。

在本公开的第七方面，提供了一种电子设备。该设备包括至少一个处理单元；以及至少一个存储器，至少一个存储器被耦合到至少一个处理单元并且存储用于由至少一个处理单元执行的指令。指令在由至少一个处理单元执行时使电子设备执行第一方面和第二方面的方法。

在本公开的第八方面，提供了一种计算机可读存储介质。介质上存储有计算机程序，计算机程序被处理器执行时实现第一方面和第二方面的方法。

应当理解，该部分中所描述的内容并非旨在限定本公开的实施例的关键特征或重要特征，也不用于限制本公开的范围。本公开的其他特征将通过以下的描述而变得容易理解。

附图说明

结合附图并参考以下详细说明，本公开各实施例的上述和其他特征、优点及方面

将变得更加明显。在附图中，相同或相似的附图标记表示相同或相似的元素，其中：

图 1 示出了能够在其中实现本公开的实施例的示例环境的示意图；

图 2 示出了根据本公开的一些实施例的目标应用中各个功能的调试入口的页面的示例；

5 图 3 示出了根据本公开的一些实施例的针对目标应用中特定功能的测试查询记录的页面的示例；

图 4 示出了根据本公开的一些实施例的对目标应用进行调试的页面的示例；

图 5 示出了根据本公开的一些实施例的展示数据信息的页面的示例；

10 图 6A 示出了根据本公开的一些实施例的针对目标应用中特定功能的用例库的页面的示例；

图 6B 示出了根据本公开的一些实施例的用例详情的页面的示例；

图 7A 示出了根据本公开的一些实施例的应用调试的过程的流程图；

图 7B 示出了根据本公开的一些实施例的信息查询的过程的流程图；

图 7C 示出了根据本公开的一些实施例的应用评测的过程的流程图；

15 图 8A 示出了根据本公开的一些实施例的用于应用调试的装置的示意性结构框图；

图 8B 示出了根据本公开的一些实施例的用于信息查询的装置的示意性结构框图；

图 8C 示出了根据本公开的一些实施例的用于应用评测的装置的示意性结构框图；

以及

图 9 示出了可以实现本公开的一个或多个实施例的电子设备的框图。

20

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中示出了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反，提供这些实施例是为了更加透彻和完整地理解本公开。

25 应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

在本公开的实施例的描述中，术语“包括”及其类似用语应当理解为开放性包含，即“包括但不限于”。术语“基于”应当理解为“至少部分地基于”。术语“一个实施例”或“该实施例”应当理解为“至少一个实施例”。术语“一些实施例”应当理解为“至少一些实施例”。下文还可能包括其他明确的和隐含的定义。

30 在本文中，除非明确说明，“响应于 A”执行一个步骤并不意味着在“A”之后立即执行该步骤，而是可以包括一个或多个中间步骤。

可以理解的是，本技术方案所涉及的数据（包括但不限于数据本身、数据的获得、使用、存储或删除）应当遵循相应法律法规及相关规定的要求。

可以理解的是，在使用本公开各实施例公开的技术方案之前，均应当根据相关法律法规通过适当的方式对本公开所涉及信息的类型、使用范围、使用场景等告知相关用户并获得相关用户的授权，其中，相关用户可以包括任何类型的权利主体，例如个人、企业、团体。

5 例如，在响应于接收到用户的主动请求时，向相关用户发送提示信息，以明确地提示相关用户，其请求执行的操作将需要获得和使用到相关用户的信息，从而使得相关用户可以根据提示信息来自主地选择是否向执行本公开技术方案的操作的电子设备、应用程序、服务器或存储介质等软件或硬件提供信息。

10 作为一种可选的但非限制性的实现方式，响应于接收到相关用户的主动请求，向相关用户发送提示信息的方式，例如可以是弹窗的方式，弹窗中可以以文字的方式呈现提示信息。此外，弹窗中还可以承载供用户选择“同意”或“不同意”向电子设备提供信息的选择控件。

可以理解的是，上述通知和获得用户授权过程仅是示意性的，不对本公开的实现方式构成限定，其他满足相关法律法规的方式也可应用于本公开的实现方式中。

15 可以理解的是，本技术方案所涉及的数据（包括但不限于数据本身、数据的获取或使用）应当遵循相应法律法规及相关规定的要求。本公开实施例数字助手相关功能的启用、获取的数据、数据的处理和存储方式等，均应获得所述用户、以及与所述用户关联的其他权利主体的提前授权，并应符合相关法律法规、权利主体间协议规则的约定。

20 如本文中所使用的，术语“模型”可以从训练数据中学习到相应的输入与输出之间的关联关系，从而在训练完成后可以针对给定的输入，生成对应的输出。模型的生成可以基于机器学习技术。深度学习是一种机器学习算法，通过使用多层处理单元来处理输入和提供相应输出。神经网络模型是基于深度学习的模型的一个示例。在本文中，“模型”也可以被称为“机器学习模型”、“学习模型”、“机器学习网络”或
25 “学习网络”，这些术语在本文中可互换地使用。

图 1 示出了本公开的实施例能够在其中实现的示例环境 100 的示意图。环境 100 涉及应用管理平台 110，其可以支持应用的创建和/或应用的运行。在一些实施例中，应用管理平台 110 中用于支持应用创建的部分也可以称为应用创建部分。在一些实施例中，应用管理平台 110 中用于支持应用运行的部分也可以称为应用运行部分。

30 如图 1 所示，应用创建部分可以为用户 105 提供应用的创建和发布环境。用户 105 可以被称为应用创建用户、创建者。在一些实施例中，应用创建部分可以是低代码平台，其提供应用创建的工具合集。应用创建部分可以支持对各类应用的可视化开发，从而使开发人员可以跳过手工编码的过程，加快应用的开发周期和成本。应用创建部分可以支持用户开发一个或多个类型的应用的任何适当平台，例如可以包括基于应用

平台即服务（aPaaS）的平台。这样的平台能够支持用户对应用进行高效开发，实现应用创建、应用功能调整等操作。

应用创建部分可以被部署在用户 105 的终端设备本地，和/或可以由服务端设备来支持。例如，用户 105 的终端设备可以运行有应用创建部分的客户端，该客户端可以支持用户与服务端所提供的应用创建部分的交互。在应用创建部分运行在用户的终端设备本地的情况下，用户 105 可以直接利用终端设备与本地的应用创建部分进行交互。在应用创建部分运行在服务端设备的情况下，服务端设备可以基于与终端设备之间的通信连接，实现对终端设备中运行的客户端的服务供应。应用创建部分可以基于用户 105 的操作而向用户 105 呈现相应的页面 130，以向用户 105 输出和/或从用户 105 接收与应用创建相关的信息。

在一些实施例中，应用创建部分可以关联到相应的数据库，其中存储应用创建部分所支持的应用创建过程所需的数据或信息。例如，数据库可以存储用于组成应用的各个功能模块对应的代码和描述信息等。应用创建部分还可以对数据库中的功能模块执行调用、增加、删除、更新等操作。数据库还可以存储对不同功能块可执行的操作。示例性的，在要创建应用的场景，应用创建部分可以从数据库中调用相应的功能块来搭建应用。

在本公开的实施例中，用户 105 可以在应用创建部分上根据需要创建目标应用 120，并发布目标应用 120。目标应用 120 可被发布到任一适当的应用运行部分，只要应用运行部分能够支持目标应用 120 的运行。在发布后，目标应用 120 可以用于由一个或多个用户 145 进行操作。用户 145 可以被称目标应用 120 的终端用户。在一些实施例中，目标应用 120 可以包括或者被实现为数字助手 122。

数字助手 122 可以被配置为具有智能对话。在图 1 所示的示例中，数字助手 122 可以被集成在目标应用 120 内，作为目标应用 120 的一部分来协助执行目标应用 120 内的任务处理。在其他示例中，数字助手 122 可以被配置为独立运行的应用，例如网页应用或其他类型的应用。在这样的示例中，数字助手 122 与目标应用 120 可以被视为同一应用。数字助手 122 被提供用于协助用户在不同应用、场景下的各类任务处理需求。在与数字助手 122 的交互过程中，用户输入交互消息，并且数字助手 122 响应于用户输入来提供回复消息。通常，数字助手 122 能够支持用户以自然语言方式输入提问，并基于对自然语言输入的理解和逻辑推理能力来执行任务和提供回复。

在一些实施例中，数字助手 122 可以作为用户 145 的联系人与之交互。例如，数字助手 122 可以被实现在即时消息（IM）应用中。数字助手 122 可以在与用户 145 的单聊会话中与用户 145 交互。在一些实施例中，数字助手 122 可以在包括多个用户的群聊会话中与多个用户交互。

对于每个用户 145，应用运行部分的客户端可以在客户端界面中呈现目标应用

120 或数字助手 122 的交互窗口 142，例如与数字助手 122 的会话窗口。用户 145 可以在会话窗口中输入会话消息，目标应用 120 可以基于所创建的配置信息，来确定数字助手 122 的回复消息并在交互窗口 142 中呈现给用户。在一些实施例中，取决于目标应用 120 的配置，与目标应用 120 的交互消息可以包括多模态形式的消息，诸如文本消息（例如，自然语言文本）、语音消息、图像消息、视频消息，等等。

与应用创建部分类似，应用运行部分可以被部署在每个用户 145 的终端设备本地，和/或可以由服务端设备来支持。例如，用户 145 的终端设备可以运行有应用运行部分的客户端，该客户端可以支持用户与服务端所提供的应用运行部分的交互。在应用运行部分运行在用户的终端设备本地的情况下，用户 145 可以直接利用终端设备与本地的应用运行部分进行交互。在应用运行部分运行在服务端设备的情况下，服务端设备可以基于与终端设备之间的通信连接，实现对终端设备中运行的客户端的服务供应。应用运行部分可以基于用户 145 的操作而向用户 145 呈现相应的应用页面，以向用户 145 输出和/或从用户 145 接收与应用使用相关的信息。

在一些实施例中，目标应用 120 的至少部分功能的实现，和/或目标应用 120 中的数字助手 122 的至少部分功能的实现可以基于模型来实现。在目标应用 120 的创建或运行过程中，可以调用一个或多个模型 155，例如模型 155 的能力。在目标应用 120 中，数字助手 122 可以利用模型 155 来理解用户输入，并基于模型 155 的输出来提供对用户的回复。

在创建过程中，应用创建平台 110 对目标应用 120 的测试需要利用模型 155 来确定目标应用 120 的运行结果符合预期。在运行过程中，响应于目标应用 120 的用户的不同操作请求，应用运行部分可能需要利用模型 155 来确定对用户的响应结果。

虽然被示出为独立于应用管理平台 110，但一个或多个模型 155 可以运行在应用管理平台 110 上，或者其他远端服务器。在一些实施例中，模型 155 可以是机器学习模型、深度学习模型、学习模型、神经网络等。在一些实施例中，模型可以基于语言模型（LM）。语言模型通过从大量语料中学习，能够具备问答能力。模型 155 也可以基于其他适当的模型。

管理平台 110 可以运行在适当的电子设备。这里的电子设备可以是任何类型的具有计算能力的设备，包括终端设备或服务端设备。终端设备可以是任何类型的移动终端、固定终端或便携式终端，包括移动手机、台式计算机、膝上型计算机、笔记本计算机、上网本计算机、平板计算机、媒体计算机、多媒体平板、个人通信系统（PCS）设备、个人导航设备、个人数字助理（PDA）、音频/视频播放器、数码相机/摄像机、定位设备、电视接收器、无线电广播接收器、电子书设备、游戏设备或者前述各项的任何组合，包括这些设备的配件和外设或者其任何组合。服务端设备例如可以包括计算系统/服务器，诸如大型机、边缘计算节点、云环境中的计算设备，等等。在一些实

施例中，管理平台 110 可以基于云服务来实现。

应当理解，仅出于示例性的目的描述环境 100 的结构和功能，而不暗示对于本公开的范围的任何限制。例如，虽然图 1 示出了与应用创建部分交互的单个用户和与应用运行部分交互的单个用户，但实际上可以多个用户访问应用管理平台 110 来各自创建数字助手，并且每个数字助手可以被用于与多个用户进行交互。

用户在应用内发起查询请求时，该应用可以利用机器学习模型来提供查询服务。在一开始该应用提供的查询服务的质量在可能并不高，需要经过多轮优化来提升查询服务的效果。传统的优化需要在代码层面对系统进行改进，存在成本高、耗时长、容易出现错误的问题。

为此，根据本公开的实施例提供了一种应用调试的方案。根据本公开的各种实施例，执行针对目标应用的测试查询。呈现执行测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出。基于测试查询的多个阶段的相应第一运行信息，生成针对目标应用的运行参考用例，运行参考用例包括测试查询和多个阶段的运行信息，运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。

在本公开的各种实施例中，应用创建用户只需要依据执行测试查询的多个阶段相应的运行信息，就可以生成针对目标应用的运行参考用例。在针对终端用户发起的用户查询来生成查询结果时，可以参考该运行参考用例。以此方式，应用创建用户无需修改代码就可以提升目标应用提供服务的质量，简化了目标应用的调优过程。

以下将参考附图的示例来详细描述本公开的一些示例实施例。应当理解，附图中示出的页面仅仅是示例，实际可以存在各种页面设计。页面中的各个图形元素可以具有不同的布置和不同的视觉表示，其中的一个或多个元素可以省略或被替换，并且还可以存在一个或多个其他元素。本公开的实施例在此方面不受限制。

下面继续参考附图来描述本公开的示例实施例。在下文的示例中，出于讨论的需要，从应用管理平台的角度来描述，例如图 1 所示的应用管理平台 110。应用管理平台 110 所呈现的页面，可以经由用户 105 的终端设备被呈现，并且可以经由用户 105 的终端设备来接收用户输入。

图 2 示出了根据本公开的一些实施例的目标应用 120 中各个功能的调试入口的页面 200 的示例。如图 2 所示，在用户 105 的终端设备呈现有目标应用 120 中各个功能的调试入口的页面 200。页面 200 给出了功能区域 210，其可以用于目标应用中的各种功能。目标应用中的功能例如可以包括应用数据分析功能、应用数据更新功能、知识问答功能等。但应当理解，这仅是示例性的功能，而无意任何限制。这里的目标应用可以是任何合适的的应用，例如人力资源应用、CRM (Customer Relationship Management, 客户关系管理) 应用、财务应用等。以对目标应用中的应用数据分析功能调试为例，

用户 105 触发了数据分析功能的调试控件 220，例如，用户 105 点击了调试控件 220。响应于此，可以向用户 105 呈现测试查询记录的页面。

图 3 示出了根据本公开的一些实施例的针对目标应用中特定功能的测试查询记录的页面 300 的示例。如图 3 所示，可以在用户 105 的终端设备呈现针对目标应用中特定功能的测试查询记录的页面 300。页面 300 记录了用户 105 之前创建的测试查询，用户 105 通过输入测试查询来对相应的功能进行调试。在一些实施例中，用户 105 可以重新调试先前创建的测试查询。例如，用户 105 点击了控件 310 即可重新执行测试查询“单价低于 XX 元的商品的数量是多少？”，以对目标应用中相应的功能进行调试。响应于此，可以向用户 105 呈现调试页面。在一些实施例中，用户 105 可以触发执行新的测试查询。例如，用户 105 点击了新增控件 320 即可输入新的测试查询，以对目标应用中相应的功能进行调试。响应于此，用户 105 可以进入调试页面。

图 4 示出了根据本公开的一些实施例的对目标应用进行调试的页面 400 的示例。应用管理平台 110 可以执行针对目标应用的测试查询。参考图 4，例如，用户 105 在测试查询对应的文本区域 410 输入测试查询语句，并点击控件 420，来执行针对目标应用的测试查询。用户 105 输入的测试查询语句例如可以是“单价低于 XX 元的商品的数量是多少？”、“产地是 XX 的商品的占比是多少？画一个饼状图”。应用管理平台 110 可以在分阶段调试区域 430 呈现执行测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出。

应用管理平台 110 在执行针对目标应用的测试查询之后，可以在分阶段调试区域 430 呈现执行测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出。例如，在分阶段调试区域 430 中示出了 5 个阶段的运行信息，这 5 个阶段需要执行各自的任務。在任务执行完成后，生成对应的输出 440、442、444、446 和 448。例如，第三阶段的输出 444 可以包括需要执行测试查询所需的数据表。

在一些实施例中，多个阶段的运行信息还可以指示以下至少一项：多个阶段各自的输入，多个阶段各自的运行耗时，多个阶段各自的运行日志，多个阶段各自的运行状态（即，是否运行完成）。在一些实施例中，由于机器学习模型的输入信息（例如，输入信息的长度）存在限制，因此，多个阶段的运行信息还可以指示多个阶段中的至少一个阶段中调用机器学习模型时，被提供给所述机器学习模型的输入信息。以此方式，用户 105 可以使提供给机器学习模型的输入信息满足机器学习模型的要求，从而保证机器学习模型稳定运行。

在一些实施例中，分阶段调试区域 430 中展示的多个阶段可以包括至少一个预定义阶段，至少一个预定义阶段包括生成数据库查询请求的阶段。该至少一个预定义阶段可以是应用管理平台 110 根据目标应用的特性和/或对应的功能预先设置的一些调试阶段。应用管理平台 110 执行针对目标应用的测试查询的目的可以是生成一个更好的

数据库查询请求，以向用户返回更加精确的查询结果。因此，该至少一个预定义阶段还可以包括生成数据库查询请求的阶段，该阶段可以基于之前的预定义阶段的输出生成准确和高效的数据库查询请求。这里的数据库查询请求例如可以是 SQL（Structured Query Language，结构化查询语言）语句。

- 5 在一些实施例中，分阶段调试区域 430 中展示的多个阶段除了至少一个预定义阶段，还可以包括基于测试查询确定的至少一个阶段。在应用管理平台 110 运行数据库查询请求获取到对应的数据之后，还可以基于测试查询的要求确定至少一个阶段。在一个示例中，测试查询是“单价低于 XX 元的商品的数量是多少？”，该测试查询并没有附加的数据分析要求，因此应用管理平台 110 可以只确定一个阶段，例如，数据查看阶段，应用管理平台 110 可以将单价低于 XX 元的商品的数量在该阶段输出。在
10 另一示例中，测试查询是“产地是 XX 的商品的占比是多少？画一个饼状图”，该测试查询有附加的数据分析要求，因此应用管理平台 110 可以生成多个阶段，例如，生成图表阶段、数据洞察阶段等，应用管理平台 110 可以将产地是 XX 的商品的占比是
15 多少的饼图输出到生成图表阶段，应用管理平台 110 还可以将产地是 XX 的商品的占比的数据的挖掘和分析输出到数据洞察阶段。以此方式，应用管理平台 110 可以基于测试查询在预定义阶段之后确定不同的阶段，可以满足用户不同的查询需求，提升用户的查询体验。

- 在一些实施例中，应用管理平台 110 执行的针对目标应用的测试查询是针对目标应用中的第一功能的查询。由于目标应用中不同的功能具有较大的差异性，针对目标
20 应用中一个功能的调试可能对目标应用中另一个功能并不适用，因此测试查询是针对目标应用中一个特定功能的查询。

- 在一些实施例中，对于目标应用中的不同功能而言，预定义阶段可以是不同的。例如，以第一功能为例，至少一个预定义阶段可以是针对目标应用中的第一功能来配置的。应用管理平台 110 可以基于目标应用中不同功能的特性，来配置不同的至少一个
25 预定义阶段。示例性地，针对目标应用中的应用数据分析功能，应用管理平台 110 可以配置 5 个预定义阶段，分别是召回并选择术语阶段、召回并选择运行参考用例阶段、召回并选择目标数据集阶段、召回并选择目标数据集所属字段阶段、生成并执行数据库查询请求阶段。在召回并选择术语阶段可以召回并选择测试查询中使用的术语的解释。在召回并选择运行参考用例阶段可以召回并选择与测试查询相似的运行参考
30 用例。在召回并选择目标数据集阶段可以召回并选择执行测试查询所需的数据集。在召回并选择目标数据集所属字段阶段可以召回并选择执行测试查询所需的数据集中的字段。在生成并执行数据库查询请求阶段可以基于前面 4 个阶段的输出，生成并执行数据库查询请求。

上文以数据分析功能作为示例描述了阶段的划分，但这种划分仅是示例性的。此

外，针对其他的功能（例如，数据更新、知识问答等）可以根据相应功能的特性进行合适的阶段划分。以此方式，使用测试查询对目标应用中的不同功能分别进行调试，充分考虑到了不同功能之间的差异性，可以有效地提升不同功能提供的服务质量。

在呈现了执行测试查询的多个阶段的相应第一运行信息之后，应用管理平台 110 可以基于测试查询的多个阶段的相应第一运行信息，生成针对目标应用的运行参考用例。运行参考用例包括测试查询和多个阶段的运行信息，运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。例如，用户 105 可以在页面 400 中先将点击已调优控件 450，将测试查询和对应的运行信息确定已经完成了调优，再点击保存控件 460 以生成运行参考用例。在生成了运行参考用例之后，应用管理平台 110 在接收到用户 145 的用户查询时，可以参考运行参考用例来确定用户查询的结果。

在一些实施例中，应用管理平台 110 基于测试查询的多个阶段的相应第一运行信息，可以确定对多个阶段中的至少一个阶段的第一运行信息的调整。而后，应用管理平台 110 可以基于经调整的至少一个阶段的第一运行信息，重新执行测试查询。应用管理平台 110 可以呈现重新执行测试查询的多个阶段的相应第二运行信息。如果接收到用例生成确认，应用管理平台 110 可以基于相应第二运行信息，生成所运行参考用例。在至少一个阶段的第一运行信息调整之后，应用管理平台 110 可以重新执行测试查询，并在分阶段调试区域 430 中呈现重新执行测试查询的多个阶段的相应第二运行信息。参考图 4，应用管理平台 110 可以通过用户 105 对已调优控件 450 和保存控件 460 的触发接收到用例生成确认，以基于第二运行信息生成运行参考用例。

在一些实施例中，应用管理平台 110 可以接收到用户 105 针对多个阶段中的第一阶段的术语调整信息，并基于术语调整信息重新执行测试查询。该术语调整信息指示向第一阶段添加术语集或移除在所述第一阶段中选择的术语集。术语集包括对目标应用的用户群体所使用的术语的解释。用户群体所使用的术语可能是在该群体的范围内被使用的，但并不是一种通用术语，因此需要帮助机器学习模型理解这样的术语。对这种术语的解释可以包括对该术语的文字性解释，也可以包括与该术语相对应的通用术语。

示例性地，执行测试查询的第一阶段可以选择与测试查询相关的术语集，用户 105 可以在第一阶段添加术语集，也可以移除术语集。例如，用户 105 针对目标应用的测试查询为“通过 C2C 渠道销售的商品的数量是多少？”，其中 C2C 是专用术语，第一阶段选择的术语集可以包括对 C2C 的解释，用户 105 可以添加其他术语集或删除第一阶段选择的术语集。在调整完术语集后，应用管理平台 110 可以基于术语调整信息，确定在第二阶段中使用的一个或多个术语集，然后基于该一个或多个术语集重新执行测试查询。

在一些实施例中，应用管理平台 110 可以接收针对多个阶段中的第二阶段的用例

调整信息，并基于用例调整信息重新执行测试查询。用例调整信息指示向第二阶段添加参考用例或移除所述第二阶段中所选择的参考用例。示例性地，执行测试查询的第二阶段可以选择与该测试查询相似的用例，用户 105 可以在第二阶段添加参考用例，也可以移除参考用例。例如，用户 105 针对目标应用的测试查询为“单价小于 5000 元的商品的数量是多少？”，第二阶段可以基于相似度计算选择与该测试查询相似的参考用例，第二阶段选择的参考用例可以是“单价小于 8000 元的商品的数量是多少？”，用户 105 可以添加其他参考用例或移动第二阶段选择的参考用例。在调整完参考用例之后，应用管理平台 110 可以基于用例调整信息，确定在第二阶段中使用的一个或多个参考用例，然后基于该一个或多个参考用例，重新执行测试查询。

10 在一些实施例中，应用管理平台 110 可以接收针对多个阶段中的第三阶段的数据集调整信息，并基于数据集调整信息重新执行测试查询。示例性地，执行测试查询的第三阶段可以选择执行测试查询所需的数据集，用户 105 可以在第三阶段添加数据集，也可以移除数据集。例如，用户 105 针对目标应用的测试查询为“单价小于 5000 元的商品的数量是多少？”，第三阶段选择的数据集可以是商品分析表，用户 105 可以添加其他数据集，也可以删除第三阶段选择的数据集。

20 在一些实施例中，应用管理平台 110 还可以接收针对多个阶段中的第四阶段的数据集字段调整信息，并基于数据集字段调整信息重新执行测试查询。示例性地，执行测试查询的第四阶段可以选择执行测试查询所需的数据集的字段，用户 105 可以在第四阶段添加字段，也可以移除字段。例如，用户 105 针对目标应用的测试查询为“单价小于 5000 元的商品的数量是多少？”，第四阶段选择的数据集的字段是商品分析表中的商品号字段、单价字段、产地字段和生产日期字段，用户 105 可以添加其他字段，也可以删除第四阶段选择的字段。以此方式，用户可以通过图形界面的方式对测试查询的各个阶段进行调整，方便用户对于调试的理解，简化了目标应用的调试过程。

25 在一些实施例中，应用管理平台 110 在确定对至少一个阶段的第一运行信息的调整之后，可以接收阶段运行请求。阶段运行请求指示从多个阶段中的指定阶段起执行测试查询，指定阶段在至少一个阶段之前或被包括在至少一个阶段中，并且基于经调整的至少一个阶段的第一运行信息，重新执行多个阶段中的指定阶段以及指定阶段之后的阶段。示例性地，用户 105 对第二阶段进行了调整，此时指定阶段可以是第二阶段之前的第一阶段，也可以是第二阶段。在指定阶段是第一阶段的情况下，应用管理平台 110 可以基于经调整的至少第二阶段的第一运行信息，执行第一阶段和第一阶段之后的阶段（即，全部的阶段）。

例如，参考图 4，针对这五个阶段中的每个阶段都提供了用于该阶段开始重新执行的触发入口，例如触发控件 421、422、423、424、425。用户 105 可以触发控件 420 来从头开始执行全部的阶段。在指定阶段是第二阶段的情况下，应用管理平台 110 可

以基于经调整的至少第二阶段的第一运行信息,执行第二阶段和第二阶段之后的阶段。例如,继续参考图 4,用户 105 可以触发控件 422 来执行第二阶段和第二阶段之后的阶段。以此方式,用户可以灵活地运行不同的阶段,通过对比运行不同阶段的输出结果,可以验证对不同阶段进行调整的有效性。可以理解的是,尽管图 4 示出了通过控件 421、422、423、424、425 的触发来执行相应的阶段,但该触发方式仅仅是示意性的。用户 105 还可以通过其他的方式来进行相应的触发,例如,通过表征自然语言的用户输入,来触发相应阶段的执行。

在一些实施例中,在重新执行测试查询之后,应用管理平台 110 可以在分阶段调试区域 430 呈现与指定阶段以及在指定阶段之后的阶段的重新运行相关联的运行信息。例如,指定阶段是第二阶段,则可以在分阶段调试区域 430 呈现与第二阶段和第二阶段之后的阶段重新运行相关联的运行信息。

在一些实施例中,应用管理平台 110 还可以在呈现多个阶段的相应第一运行信息的同时呈现数据信息,数据信息至少指示用于执行测试查询的至少一个数据集,并且基于多个阶段的相应第一运行信息和数据信息,生成运行参考用例。参考图 4,应用管理平台 110 可以在数据信息展示区域 470 用于执行测试查询的数据信息。此外,应用管理平台 110 可以基于该数据信息或对该数据信息的调整来生成运行参考用例。

在一些实施例中,对数据信息的调整可以包括对用于执行所述测试查询的所述至少一个数据集的更新。例如,在数据信息展示区域 470 展示了用于执行测试查询的一个数据集,即商品分析表。参考图 4,用户 105 可以触发添加控件 472 以对该数据集进行更新。

在一些实施例中,对数据信息的调整可以包括对用于执行测试查询的至少一个数据集中的至少一部分的更新。在这里数据集中的一部分可以指示数据集中的字段描述。例如,在数据信息展示区域 470 展示了用于执行测试查询的一个数据集,即商品分析表。参考图 4,用户 105 可以触发修改控件 474 以对该数据集中的字段描述进行更新。

在一些实施例中,数据信息还指示用于执行测试查询所参考的术语集,术语集包括对目标应用的用户群体所使用的术语的解释。参考图 5,图 5 示出了根据本公开的一些实施例的展示数据信息的页面 500 的示例。页面 500 包括术语集展示区域 510,用户 105 可以调整术语集展示区域 510 中的内容。

在一些实施例中,数据信息还指示用于执行测试查询所参考的用户与数字助手之间的对话记录。继续参考图 5,页面 500 包括对话记录展示区域 520,用户 105 可以调整对话记录展示区域 520 中的内容。

在一些实施例中,数据信息还指示用于执行所述测试查询的参考提示词。继续参考图 5,页面 500 包括参考提示词展示区域 530,用户 105 可以调整参考提示词展示区域 530 中的内容。通过调整参考提示词,可以给机器学习模型提供更好的指引,帮助

机器学习模型更好地完成用户的任务。

在一些实施例中，数据信息还指示用于确定测试查询的候选结果的条件。继续参考图 5，页面 500 包括条件区域 540，用户 105 可以调整条件区域 540 中的相似度和单词搜索数量。例如，在知识问答场景下，数字助手会基于相似度搜索与用户相似的提问，并根据单次搜索数量向用户 145 提供预定数量的搜索结果。

参考图 6A，图 6A 示出了根据本公开的一些实施例的针对目标应用中特定功能用例库的页面 600 的示例。页面 600 呈现了用例库，用例库中包含了已经生成的参考用例。用户 105 触发详情控件 602 可以跳转至该用例对应的用例详情的页面，参考图 6B。图 6B 示出了根据本公开的一些实施例的用例详情的页面 610 的示例。页面 610 可以包括测试查询以及多个阶段的相应第一运行信息 614。

可以理解的是，根据上文所描述的实施例，可以生成包括一个或多个运行参考用例的运行参考用例集合。

在一些实施例中，在目标应用 120 被发布之后，在目标应用 120 的运行中，可以利用运行参考用例来向用户 145 提供服务。示例性的，响应于接收到针对目标应用的用户查询，应用管理平台 110 可以从运行参考用例集合中查找与用户查询相似或匹配的测试查询。基于用户查询与某一测试查询之间的相似度超过阈值相似度，应用管理平台 110 可以确定运行参考用例为用户查询的目标运行参考用例，并基于目标运行参考用例来执行用户查询。在生成运行参考用例之后，应用管理平台 110 接收到用户 145 发起的针对目标应用的用户查询，可以通过相似度计算确定用户查询与测试查询之间的相似度，并基于该相似度确定运行参考用例为用户查询的目标运行参考用例，在确定出目标运行参考用例之后，应用管理平台 110 可以使用该目标运行参考用例来执行用户查询。以此方式，参考运行参考用例来确定用户查询的结果，可以提供给用户更加精确的查询结果，提高用户的满意度。

在一些实施例中，可以利用机器学习模型来执行用户查询。例如，应用管理平台 110 可以将所确定的目标参考用例作为针对用户查询的提示词的一部分提供给机器学习模型。而后，应用管理平台 110 可以从机器学习模型获取用户查询的执行结果。需要注意的是，这里的目标参考用例是从应用创建阶段或者应用测试阶段所生成的运行参考用例集合中所选择的。这样的目标参考用例是出于调试的目的而生成的，而不是来自用户真实使用该应用时的用例。

在一些实施例中，使用目标应用 120 的用户可以是测试人员，也即用户查询可以来自测试人员。在这种情况下，如果测试人员针对某个查询结果给出了正面反馈，可以基于该用户查询和对应的查询结果生成运行参考用例。例如，可以将得出该查询结果的运行过程的多个阶段的运行信息存储作为运行参考用例。以此方式，利用用户反馈丰富了运行参考用例的来源。

在一些实施例中，目标应用包括至少一个功能，并且测试查询和用户查询均为针对至少一个功能中的第一功能的查询。目标应用可以包括应用数据分析、应用数据更新、知识问答等功能。测试查询和用户查询需要针对同一功能，例如，测试查询和用户查询均是针对应用数据分析功能。

5 在目标应用 120 的创建或测试期间，可能需要评测目标应用 120 的性能。例如，在目标应用包括多个功能的情况下，可能需要评测这些功能中的某个功能或某些功能。

为此，在一些实施例中，应用管理平台 110 可以维护评测样本集，以用于评测目标应用 120，例如评测目标应用 120 的一个或多个功能，诸如数据分析功能、数据更新功能、知识问答功能等。评测样本集中的每个评测样本可以包括样本查询和目标应用 120 针对样本查询生成的结果（也称为样本查询结果、第二结果）。

在一些实施例中，为了获得对目标应用 120 中的某个功能的评测结果，可以利用机器学习模型获取针对评测样本中的样本查询结果的反馈。例如，机器学习模型可以从一个或多个维度对目标应用 120 生成的样本查询结果打分。

机器学习模型可以采用任何合适的方式给出反馈。作为示例，描述一种自动评测方式。可以将目标应用 120 提供服务所使用的数据源（例如，知识库、数据集等）提供给机器学习模型，使得机器学习模型从其中提取查询和对应的查询结果作为正确示例。例如，对于知识问答功能，可以提取问题和对应的答案。而后，可以在目标应用 120 中执行所提取的查询，并且由目标应用 120 生成执行结果。进而，机器学习模型可以基于答案和执行结果在各种合适的维度来对执行结果给出反馈，例如打分。评测维度的示例可以包括但不限于对用户的帮助性、语言流畅度、事实正确度、逻辑性等。

由此，通过机器学习模型针对所提取的各个问题和执行结果给出的反馈，可以确定对目标应用 120 的评测结果。例如，如果评测样本集用于知识问答功能，可以获得对知识问答功能的评测结果。这样，用户 105 可以根据评测结果来调试目标应用 120，例如调试一个或多个功能。

25 在目标应用 120 的创建或测试期间，创建人员或测试人员可能输入一些评测查询。目标应用 120 可以针对评测查询生成对应的查询结果，也称为第一结果或评测查询结果。而后，创建人员或测试人员可以对评测查询结果给出反馈，也称为评测反馈。

在一些实施例中，评测反馈可以与评测样本集结合使用，以针对目标应用 120 给出更准确的评测结果。示例性的，运行管理平台 110 可以基于创建人员或测试人员给出的评测查询，生成用于目标应用 120 的评测参考用例。评测参考用例包括来自创建人员或测试人员的评测查询、目标应用 120 针对该评测查询而生成的评测查询结果和对该评测查询结果的评测反馈。评测反馈可以是正面反馈或负面反馈，也可以指示用户 145 对查询结果的满意程度。

通常而言，针对相似的查询，期望评测结果与评测反馈是一致的或匹配的。有鉴

于此，可以从评测样本集中找出与评测查询相似的样本查询，例如相似度超过阈值相似度。包括这种相似的样本查询的评测样本也称为目标评测样本，目标评测样本还包括目标应用 120 针对该样本查询而生成的样本结果。

应用管理平台 110 可以基于评测查询结果、评测反馈和样本查询结果，生成针对目标应用 120 的评测结果。例如，可以使对目标评测样本中的样本查询结果的打分与评测反馈是匹配的。可以理解的是，评测参考用例给出了什么样的结果是创建人员或测试人员满足的或什么样的结果是创建人员或测试人员不满意的示例。

在一些实施例中，如上文所描述的，可以利用机器学习模型执行对目标应用 120 的评测。相应地，可以基于评测参考用例的至少一部分（例如，评测反馈和评测查询结果）生成提示词，并且将提示词提供给机器学习模型。机器学习模型可以生成针对目标评测样本中的样本查询结果的反馈。例如，机器学习模型可以对样本查询结果进行打分。而后可以基于该反馈，确定针对目标应用 120 的评测结果。可以理解的是，评测参考用例给出了什么样的结果是创建人员或测试人员满足的或什么样的结果是创建人员或测试人员不满意的示例。这样，机器学习模型在生成针对样本结果的反馈（例如，打分）中可以参考这样的示例。注意，在评测中所利用的机器学习模型与目标应用 120 的运行中所利用的机器学习模型可以相同，也可以不同。

在一些实施例中，如果目标应用 120 包括至少一个功能，上述用于评测的样本查询和评测查询可以是针对同一功能的查询。例如，评测样本集可以是专用于评测目标应用 120 的知识问答功能的。在这种情况下，所收集的用户查询也是针对知识问答功能的。

图 7A 示出了根据本公开的一些实施例的应用调试的过程 700 的流程图。过程 700 可以被实现在应用管理平台 110 处。下面参考图 7 描述过程 700。

在框 701，应用管理平台 110 执行针对目标应用的测试查询。

在框 702，应用管理平台 110 呈现执行所述测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出。

在框 703，应用管理平台 110 基于所述测试查询的多个阶段的相应第一运行信息，生成针对目标应用的运行参考用例，运行参考用例包括测试查询和多个阶段的运行信息，运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。

在一些实施例中，生成运行参考用例包括：确定对多个阶段中的至少一个阶段的第一运行信息的调整；基于经调整的至少一个阶段的第一运行信息，重新执行测试查询；呈现重新执行测试查询的多个阶段的相应第二运行信息；以及响应于接收到用例生成确认，基于相应第二运行信息，生成运行参考用例。

在一些实施例中，重新执行测试查询包括：接收阶段运行请求，阶段运行请求指示从多个阶段中的指定阶段起执行测试查询，指定阶段在至少一个阶段之前或被包括

在至少一个阶段中；基于经调整的至少一个阶段的第一运行信息，重新执行多个阶段中的指定阶段以及所述指定阶段之后的阶段。

在一些实施例中，呈现重新执行测试查询的所述多个阶段的相应第二运行信息包括：呈现与指定阶段以及在所述指定阶段之后的阶段的重新运行相关联的运行信息。

- 5 在一些实施例中，多个阶段的运行信息还指示以下至少一项：多个阶段各自的输入，多个阶段各自的运行耗时，多个阶段各自的运行日志，多个阶段各自的运行状态，在多个阶段中的至少一个阶段中调用机器学习模型时，被提供给机器学习模型的输入信息。

- 10 在一些实施例中，确定对多个阶段中的至少一个阶段的第一运行信息的调整包括：接收针对多个阶段中的第一阶段的术语调整信息，术语调整信息指示向第一阶段添加术语集或移除在第一阶段中选择的术语集，术语集包括对目标应用的用户群体所使用的术语的解释，并且重新执行测试查询包括：基于术语调整信息，确定在第一阶段中使用的一个或多个术语集；以及基于一个或多个术语集，重新执行测试查询。

- 15 在一些实施例中，确定对多个阶段中的至少一个阶段的第一运行信息的调整包括：接收针对多个阶段中的第二阶段的用例调整信息，用例调整信息指示向第二阶段添加参考用例或移除第二阶段中所选择的参考用例，并且重新执行测试查询包括：基于用例调整信息，确定在第二阶段中使用的一个或多个参考用例；以及基于一个或多个参考用例，重新执行测试查询。

- 20 在一些实施例中，生成针对目标应用的运行参考用例包括：在呈现多个阶段的相应第一运行信息的同时呈现数据信息，数据信息至少指示用于执行所述测试查询的至少一个数据集；以及基于多个阶段的相应第一运行信息和数据信息，生成运行参考用例。

- 25 在一些实施例中，数据信息还指示以下至少一项：用于执行测试查询所参考的术语集（其中，术语集包括对目标应用的用户群体所使用的术语的解释），用于执行测试查询所参考的用户与数字助手之间的对话记录，用于执行测试查询的参考提示词，用于确定测试查询的候选结果的条件。

在一些实施例中，对数据信息的调整包括以下至少一项：对用于执行测试查询的至少一个数据集的更新，或对用于执行测试查询的至少一个数据集中的至少一部分的更新。

- 30 在一些实施例中，过程 700 还包括：响应于接收到针对目标应用的用户查询，基于用户查询与测试查询之间的相似度超过阈值相似度，确定运行参考用例为用户查询的目标运行参考用例；以及基于目标运行参考用例来执行用户查询。

在一些实施例中，目标应用包括至少一个功能，并且其中测试查询和用户查询均为针对至少一个功能中的第一功能的查询。

在一些实施例中，多个阶段包括：至少一个预定义阶段，至少一个预定义阶段包括生成数据库查询请求的阶段；以及基于所述测试查询确定的至少一个阶段。

在一些实施例中，测试查询是针对目标应用中的第一功能的查询，并且其中至少一个预定义阶段针对第一功能来配置。

5 图 7B 示出了根据本公开的一些实施例的信息查询的过程 710 的流程图。过程 710 可以被实现在应用管理平台 110 处。下面参考图 7 描述过程 710。

在框 711，应用管理平台 110 响应于接收到针对目标应用的用户查询，基于用户查询与测试查询之间的相似度超过阈值相似度，将包括测试查询的运行参考用例确定为用户查询的目标运行参考用例，其中运行参考用例包括测试查询和测试查询的多个
10 阶段的相应运行信息。

在框 712，应用管理平台 110 基于目标运行参考用例来执行用户查询。

在一些实施例中，基于目标运行参考用例来执行用户查询包括：将目标参考用例作为针对用户查询的提示词的一部分提供给机器学习模型；以及从机器学习模型获取用户查询的执行结果。

15 在一些实施例中，如果用户查询是来自测试人员并且在应用测试阶段实施，那么过程 710 还包括：获取对用户查询的执行结果的用户反馈；以及响应于用户反馈是正向反馈，基于多个阶段的用于执行用户查询的相应执行信息，生成另外的运行参考用例。

图 7C 示出了根据本公开的一些实施例的应用评测的过程 720 的流程图。过程 720
20 可以被实现在应用管理平台 110 处。下面参考图 7 描述过程 720。

在框 721，应用管理平台 110 生成用于所述目标应用的评测参考用例，所述评测参考用例包括评测查询、所述目标应用针对所述评测查询而生成的第一结果和对所述第一结果的评测反馈。

在框 722，应用管理平台 110 基于所述评测查询与评测样本集中的查询之间的相
25 似度，从所述评测样本集中选择目标评测样本，所述目标评测样本包括样本查询和由所述目标应用针对所述样本查询生成的第二结果。

在框 723，应用管理平台 110 基于所述第一结果、所述评测反馈和所述第二结果，生成针对所述目标应用的评测结果。

在一些实施例中，生成针对所述目标应用的评测结果包括：将基于所述第一结果
30 和所述评测反馈而生成的提示词提供给机器学习模型；从所述机器学习模型获得针对所述目标评测样本中的所述第二结果的反馈；以及基于针对所述第二结果的反馈，确定针对所述目标应用的评测结果。

在一些实施例中，目标应用包括至少一个功能，并且其中样本查询和用户查询均为针对至少一个功能中的第二功能的查询。

图 8A 示出了根据本公开的一些实施例的用于应用调试的装置 800 的示意性结构框图。装置 800 例如可以被实现在或被包括在应用管理平台 110 中。装置 800 中的各个模块/组件可以由硬件、软件、固件或者它们的任何组合来实现。

如图所示，装置 800 包括测试查询执行模块 801，被配置为执行针对目标应用的测试查询。装置 800 还包括第一运行信息呈现模块 802，被配置为呈现执行所述测试查询的多个阶段的相应第一运行信息，多个阶段的运行信息至少指示多个阶段各自的输出。装置 800 还包括运行参考用例生成模块 803，被配置为基于所述测试查询的多个阶段的相应第一运行信息，生成针对目标应用的运行参考用例，运行参考用例包括测试查询和多个阶段的运行信息，运行参考用例在目标应用的运行中被参考用于确定用户查询的结果。

在一些实施例中，参考用例生成模块 803 包括：测试查询重新执行模块，被配置为确定对多个阶段中的至少一个阶段的第一运行信息的调整；基于经调整的至少一个阶段的第一运行信息，重新执行测试查询；呈现重新执行测试查询的多个阶段的相应第二运行信息；以及响应于接收到用例生成确认，基于相应第二运行信息，生成运行参考用例。

在一些实施例中，测试查询重新执行模块包括：阶段执行模块，被配置为接收阶段运行请求，阶段运行请求指示从多个阶段中的指定阶段起执行所述测试查询，指定阶段在至少一个阶段之前或被包括在至少一个阶段中；基于经调整的至少一个阶段的第一运行信息，重新执行多个阶段中的指定阶段以及指定阶段之后的阶段。

在一些实施例中，阶段执行模块包括：重新执行信息呈现模块，被配置为呈现与指定阶段以及在所述指定阶段之后的阶段的重新运行相关联的运行信息。

在一些实施例中，多个阶段的运行信息还指示以下至少一项：多个阶段各自的输入，多个阶段各自的运行耗时，多个阶段各自的运行日志，多个阶段各自的运行状态，在多个阶段中的至少一个阶段中调用机器学习模型时，被提供给机器学习模型的输入信息。

在一些实施例中，测试查询重新执行模块包括：第一调整信息接收模块，被配置为接收针对多个阶段中的第一阶段的术语调整信息，术语调整信息指示向第一阶段添加术语集或移除在第一阶段中选择的术语集，术语集包括对目标应用的用户群体所使用的术语的解释，并且测试查询重新执行模块进一步被配置为基于术语调整信息，确定在第一阶段中使用的一个或多个术语集；以及基于一个或多个术语集，重新执行测试查询。

在一些实施例中，测试查询重新执行模块包括：第二调整信息接收模块，被配置接收针对多个阶段中的第二阶段的用例调整信息，用例调整信息指示向第二阶段添加参考用例或移除第二阶段中所选择的参考用例，并且测试查询重新执行模块进一步被

配置为基于用例调整信息，确定在第二阶段中使用的一个或多个参考用例；以及基于一个或多个参考用例，重新执行测试查询。

5 在一些实施例中，装置 800 还包括：数据信息呈现模块，被配置为在呈现多个阶段的相应第一运行信息的同时呈现数据信息，数据信息至少指示用于执行所述测试查询的至少一个数据集；以及基于多个阶段的相应第一运行信息和数据信息，生成运行参考用例。

10 在一些实施例中，数据信息还指示以下至少一项：用于执行测试查询所参考的术语集，术语集包括对目标应用的用户群体所使用的术语的解释，用于执行测试查询所参考的用户与数字助手之间的对话记录，用于执行测试查询的参考提示词，用于确定测试查询的候选结果的条件。

在一些实施例中，数据信息呈现模块包括：数据更新模块，被配置为对用于执行测试查询的至少一个数据集的更新，或对用于执行测试查询的至少一个数据集中的至少一部分的更新。

15 在一些实施例中，装置 800 还包括：用户查询执行模块，被配置为响应于接收到针对目标应用的用户查询，基于用户查询与测试查询之间的相似度超过阈值相似度，确定运行参考用例为用户查询的目标运行参考用例；以及基于目标运行参考用例来执行用户查询。

在一些实施例中，目标应用包括至少一个功能，并且其中测试查询和用户查询均为针对至少一个功能中的第一功能的查询。

20 在一些实施例中，多个阶段包括：至少一个预定义阶段，至少一个预定义阶段包括生成数据库查询请求的阶段；以及基于所述测试查询确定的至少一个阶段。

在一些实施例中，测试查询是针对目标应用中的第一功能的查询，并且其中至少一个预定义阶段针对第一功能来配置。

25 图 8B 示出了根据本公开的一些实施例的用于信息查询的装置 810 的示意性结构框图。装置 810 例如可以被实现在或被包括在应用管理平台 110 中。装置 810 中的各个模块/组件可以由硬件、软件、固件或者它们的任何组合来实现。

30 如图所示，装置 810 包括运行参考用例确定模块 811，被配置为响应于接收到针对目标应用的用户查询，基于用户查询与测试查询之间的相似度超过阈值相似度，将包括测试查询的运行参考用例确定为用户查询的目标运行参考用例，其中运行参考用例包括测试查询和测试查询的多个阶段的相应运行信息。装置 810 还包括用户查询执行模块 812，被配置为基于目标运行参考用例来执行用户查询。

在一些实施例中，用户查询执行模块 812 包括执行结果获取模块，被配置为将目标参考用例作为针对用户查询的提示词的一部分提供给机器学习模型；以及从机器学习模型获取用户查询的执行结果。

在一些实施例中，装置 810 还包括附加运行参考用例生成模块，被配置为获取对用户查询的执行结果的用户反馈；以及响应于用户反馈是正向反馈，基于多个阶段的用于执行用户查询的相应执行信息，生成另外的运行参考用例。

图 8C 示出了根据本公开的一些实施例的用于应用评测的装置 820 的示意性结构框图。装置 820 例如可以被实现在或被包括在应用管理平台 110 中。装置 820 中的各个模块/组件可以由硬件、软件、固件或者它们的任何组合来实现。

如图所示，装置 820 包括评测参考用例生成模块 821，被配置为生成用于所述目标应用的评测参考用例，所述评测参考用例包括评测查询、所述目标应用针对评测查询而生成的第一结果和对第一结果的评测反馈。装置 820 还包括目标评测样本选择模块 822，被配置为基于评测查询与评测样本集中的查询之间的相似度，从评测样本集中选择目标评测样本，目标评测样本包括样本查询和由目标应用针对样本查询生成的第二结果。装置 820 还包括评测结果生成模块 823，被配置为基于第一结果、评测反馈和第二结果，生成针对目标应用的评测结果。

在一些实施例中，评测结果生成模块 823 包括模型反馈获得模块，被配置为将基于第一结果和评测反馈而生成的提示词提供给机器学习模型；从机器学习模型获得针对目标评测样本中的第二结果的反馈；以及基于针对第二结果的反馈，确定针对目标应用的评测结果。

在一些实施例中，目标应用包括至少一个功能，并且其中样本查询和用户查询均为针对至少一个功能中的第二功能的查询。

图 9 示出了其中可以实现本公开的一个或多个实施例的电子设备 900 的框图。应当理解，图 9 所示出的电子设备 900 仅仅是示例性的，而不应当构成对本文所描述的实施例的功能和范围的任何限制。图 9 所示出的电子设备 900 可以包括或被实现为图 1 的应用管理平台 110、图 8A 的装置 800、图 8B 的装置 810，或图 8C 的装置 820。

如图 9 所示，电子设备 900 是通用电子设备的形式。电子设备 900 的组件可以包括但不限于一个或多个处理器或处理单元 910、存储器 920、存储设备 930、一个或多个通信单元 940、一个或多个输入设备 950 以及一个或多个输出设备 960。处理单元 910 可以是实际或虚拟处理器并且能够根据存储器 920 中存储的程序来执行各种处理。在多处理器系统中，多个处理单元并行执行计算机可执行指令，以提高电子设备 900 的并行处理能力。

电子设备 900 通常包括多个计算机存储介质。这样的介质可以是电子设备 900 可访问的任何可以获取的介质，包括但不限于易失性和非易失性介质、可拆卸和不可拆卸介质。存储器 920 可以是易失性存储器（例如寄存器、高速缓存、随机访问存储器（RAM））、非易失性存储器（例如，只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、闪存）或它们的某种组合。存储设备 930 可以是可拆卸或不可拆

卸的介质，并且可以包括机器可读介质，诸如闪存驱动、磁盘或者任何其他介质，其可以能够用于存储信息和/或数据并且可以在电子设备 900 内被访问。

电子设备 900 可以进一步包括另外的可拆卸/不可拆卸、易失性/非易失性存储介质。尽管未在图 9 中示出，可以提供用于从可拆卸、非易失性磁盘（例如“软盘”）进行读取或写入的磁盘驱动和用于从可拆卸、非易失性光盘进行读取或写入的光盘驱动。在这些情况中，每个驱动可以由一个或多个数据介质接口被连接至总线（未示出）。存储器 920 可以包括计算机程序产品 925，其具有一个或多个程序模块，这些程序模块被配置为执行本公开的各种实施例的各种方法或动作。

通信单元 940 实现通过通信介质与其他电子设备进行通信。附加地，电子设备 900 的组件的功能可以以单个计算集群或多个计算机器来实现，这些计算机器能够通过通信连接进行通信。因此，电子设备 900 可以使用与一个或多个其他服务器、网络个人计算机（PC）或者另一个网络节点的逻辑连接来在联网环境中进行操作。

输入设备 950 可以是一个或多个输入设备，例如鼠标、键盘、追踪球等。输出设备 960 可以是一个或多个输出设备，例如显示器、扬声器、打印机等。电子设备 900 还可以根据需要通过通信单元 940 与一个或多个外部设备（未示出）进行通信，外部设备诸如存储设备、显示设备等，与一个或多个使得用户与电子设备 900 交互的设备进行通信，或者与使得电子设备 900 与一个或多个其他电子设备通信的任何设备（例如，网卡、调制解调器等）进行通信。这样的通信可以经由输入/输出（I/O）接口（未示出）来执行。

根据本公开的示例性实现方式，提供了一种计算机可读存储介质，其上存储有计算机可执行指令，其中计算机可执行指令被处理器执行以实现上文描述的方法。根据本公开的示例性实现方式，还提供了一种计算机程序产品，计算机程序产品被有形地存储在非瞬态计算机可读介质上并且包括计算机可执行指令，而计算机可执行指令被处理器执行以实现上文描述的方法。

这里参照根据本公开实现的方法、装置、设备和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机可读程序指令实现。

这些计算机可读程序指令可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理单元，从而生产出一种机器，使得这些指令在通过计算机或其他可编程数据处理装置的处理单元执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

可以把计算机可读程序指令加载到计算机、其他可编程数据处理装置、或其他设备上，使得在计算机、其他可编程数据处理装置或其他设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其他可编程数据处理装置、或其他设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

5 附图中的流程图和框图显示了根据本公开的多个实现的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分，模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也
10 要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

15 以上已经描述了本公开的各实现，上述说明是示例性的，并非穷尽性的，并且也不限于所公开的各实现。在不偏离所说明的各实现的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在最好地解释各实现的原理、实际应用或对市场中的技术的改进，或者使本技术领域的其他普通技术人员能理解本文公开的各个实现方式。

权利要求书

1. 一种应用调试的方法，包括：

执行针对目标应用的测试查询；

5 呈现执行所述测试查询的多个阶段的相应第一运行信息，所述多个阶段的运行信息至少指示所述多个阶段各自的输出；以及

基于所述测试查询的多个阶段的相应第一运行信息，生成针对所述目标应用的运行参考用例，所述运行参考用例包括所述测试查询和所述多个阶段的运行信息，所述运行参考用例在所述目标应用的运行中被参考用于确定用户查询的结果。

10 2. 根据权利要求 1 所述的方法，其中生成所述运行参考用例包括：

确定对所述多个阶段中的至少一个阶段的第一运行信息的调整；

基于经调整的所述至少一个阶段的第一运行信息，重新执行所述测试查询；

呈现重新执行所述测试查询的所述多个阶段的相应第二运行信息；以及

15 响应于接收到用例生成确认，基于所述相应第二运行信息，生成所述运行参考用例。

3. 根据权利要求 2 所述的方法，其中重新执行所述测试查询包括：

接收阶段运行请求，所述阶段运行请求指示从所述多个阶段中的指定阶段起执行所述测试查询，所述指定阶段在所述至少一个阶段之前或被包括在所述至少一个阶段中；以及

20 基于经调整的所述至少一个阶段的第一运行信息，重新执行所述多个阶段中的所述指定阶段以及所述指定阶段之后的阶段。

4. 根据权利要求 3 所述的方法，其中呈现重新执行所述测试查询的所述多个阶段的相应第二运行信息包括：

25 呈现与所述指定阶段以及在所述指定阶段之后的阶段的重新运行相关联的运行信息。

5. 根据权利要求 1 所述的方法，其中所述多个阶段的运行信息还指示以下至少一项：

所述多个阶段各自的输入，

所述多个阶段各自的运行耗时，

30 所述多个阶段各自的运行日志，

所述多个阶段各自的运行状态，

在所述多个阶段中的至少一个阶段中调用机器学习模型时，被提供给所述机器学习模型的输入信息。

6. 根据权利要求 2 所述的方法，其中确定对所述多个阶段中的至少一个阶段的

第一运行信息的调整包括：

接收针对所述多个阶段中的第一阶段的术语调整信息，所述术语调整信息指示向所述第一阶段添加术语集或移除在所述第一阶段中选择的术语集，术语集包括对所述目标应用的用户群体所使用的术语的解释，并且

5 其中重新执行所述测试查询包括：

基于所述术语调整信息，确定在所述第一阶段中使用的一个或多个术语集；以及基于所述一个或多个术语集，重新执行所述测试查询。

7. 根据权利要求 2 所述的方法，其中确定对所述多个阶段中的至少一个阶段的第一运行信息的调整包括：

10 接收针对所述多个阶段中的第二阶段的用例调整信息，所述用例调整信息指示向所述第二阶段添加参考用例或移除所述第二阶段中所选择的参考用例，并且

其中重新执行所述测试查询包括：

基于所述用例调整信息，确定在所述第二阶段中使用的一个或多个参考用例；以及

15 基于所述一个或多个参考用例，重新执行所述测试查询。

8. 根据权利要求 1 所述的方法，其中生成针对所述目标应用的运行参考用例包括：

在呈现所述多个阶段的相应第一运行信息的同时呈现数据信息，所述数据信息至少指示用于执行所述测试查询的至少一个数据集；以及

20 基于所述多个阶段的相应第一运行信息和所述数据信息，生成所述运行参考用例。

9. 根据权利要求 8 所述的方法，其中所述数据信息还指示以下至少一项：

用于执行所述测试查询所参考的术语集，所述术语集包括对所述目标应用的用户群体所使用的术语的解释，

用于执行所述测试查询所参考的用户与数字助手之间的对话记录，

25 用于执行所述测试查询的参考提示词，

用于确定所述测试查询的候选结果的条件。

10. 根据权利要求 8 所述的方法，其中对所述数据信息的调整包括以下至少一项：

对用于执行所述测试查询的所述至少一个数据集的更新，或

对用于执行所述测试查询的所述至少一个数据集中的至少一部分的更新。

30 11. 根据权利要求 1 所述的方法，还包括：

响应于接收到针对所述目标应用的用户查询，基于所述用户查询与所述测试查询之间的相似度超过阈值相似度，确定所述运行参考用例为所述用户查询的目标运行参考用例；以及

基于所述目标运行参考用例来执行所述用户查询。

12. 根据权利要求 11 所述的方法, 其中所述目标应用包括至少一个功能, 并且其中所述测试查询和所述用户查询均为针对所述至少一个功能中的第一功能的查询。

13. 根据权利要求 1 所述的方法, 其中所述多个阶段包括:

至少一个预定义阶段, 所述至少一个预定义阶段包括生成数据库查询请求的阶段;

5 以及

基于所述测试查询确定的至少一个阶段。

14. 根据权利要求 13 所述的方法, 其中所述测试查询是针对所述目标应用中的第一功能的查询, 并且其中所述至少一个预定义阶段针对所述第一功能来配置。

15. 一种信息查询的方法, 包括:

10 响应于接收到针对目标应用的用户查询, 基于所述用户查询与测试查询之间的相似度超过阈值相似度, 将包括所述测试查询的运行参考用例确定为所述用户查询的目标运行参考用例, 其中所述运行参考用例包括所述测试查询和所述测试查询的多个阶段的相应运行信息; 以及

基于所述目标运行参考用例来执行所述用户查询。

15 16. 根据权利要求 15 所述的方法, 其中基于所述目标运行参考用例来执行所述用户查询包括:

将所述目标参考用例作为针对所述用户查询的提示词的一部分提供给机器学习模型; 以及

从所述机器学习模型获取所述用户查询的执行结果。

20 17. 一种应用评测的方法, 包括:

生成用于所述目标应用的评测参考用例, 所述评测参考用例包括评测查询、所述目标应用针对所述评测查询而生成的第一结果和对所述第一结果的评测反馈;

基于所述评测查询与评测样本集中的查询之间的相似度, 从所述评测样本集中选择目标评测样本, 所述目标评测样本包括样本查询和由所述目标应用针对所述样本查询生成的第二结果; 以及

25 基于所述第一结果、所述评测反馈和所述第二结果, 生成针对所述目标应用的评测结果。

18. 根据权利要求 17 所述的方法, 其中生成针对所述目标应用的评测结果包括:

将基于所述第一结果和所述评测反馈而生成的提示词提供给机器学习模型;

30 从所述机器学习模型获得针对所述目标评测样本中的所述第二结果的反馈; 以及基于针对所述第二结果的反馈, 确定针对所述目标应用的评测结果。

19. 根据权利要求 17 所述的方法, 其中所述目标应用包括至少一个功能, 并且其中所述样本查询和所述用户查询均为针对所述至少一个功能中的第二功能的查询。

20. 一种用于应用调试的装置, 包括:

测试查询执行模块，被配置为执行针对目标应用的测试查询；

第一运行信息呈现模块，被配置为呈现执行所述测试查询的多个阶段的相应第一运行信息，所述多个阶段的运行信息至少指示所述多个阶段各自的输出；以及

5 运行参考用例生成模块，被配置为基于所述测试查询的多个阶段的相应第一运行信息，生成针对所述目标应用的运行参考用例，所述运行参考用例包括所述测试查询和所述多个阶段的运行信息，所述运行参考用例在所述目标应用的运行中被参考用于确定用户查询的结果。

21. 一种用于信息查询的装置，包括：

10 运行参考用例确定模块，被配置为响应于接收到针对目标应用的用户查询，基于所述用户查询与测试查询之间的相似度超过阈值相似度，将包括所述测试查询的运行参考用例确定为所述用户查询的目标运行参考用例，其中所述运行参考用例包括所述测试查询和所述测试查询的多个阶段的相应运行信息；以及

用户查询执行模块，被配置为基于所述目标运行参考用例来执行所述用户查询。

22. 一种用于应用评测的装置，包括：

15 评测参考用例生成模块，被配置为生成用于所述目标应用的评测参考用例，所述评测参考用例包括评测查询、所述目标应用针对所述评测查询而生成的第一结果和对所述第一结果的评测反馈；

20 目标评测样本选择模块，被配置为基于所述评测查询与评测样本集中的查询之间的相似度，从所述评测样本集中选择目标评测样本，所述目标评测样本包括样本查询和由所述目标应用针对所述样本查询生成的第二结果；以及

评测结果生成模块，被配置为基于所述第一结果、所述评测反馈和所述第二结果，生成针对所述目标应用的评测结果。

23. 一种电子设备，包括：

至少一个处理单元；以及

25 至少一个存储器，所述至少一个存储器被耦合到所述至少一个处理单元并且存储用于由所述至少一个处理单元执行的指令，所述指令在由所述至少一个处理单元执行时使所述电子设备执行根据权利要求 1 至 19 中任一项所述的方法。

24. 一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序可由处理器执行以实现根据权利要求 1 至 19 中任一项所述的方法。

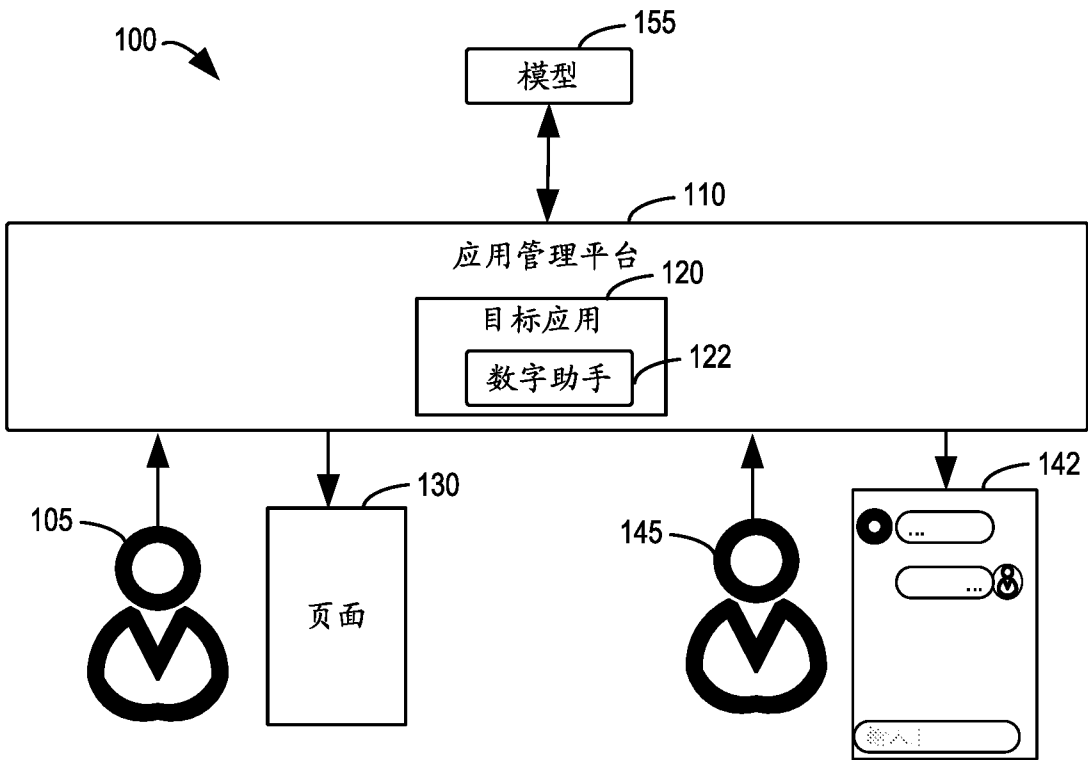


图 1

200

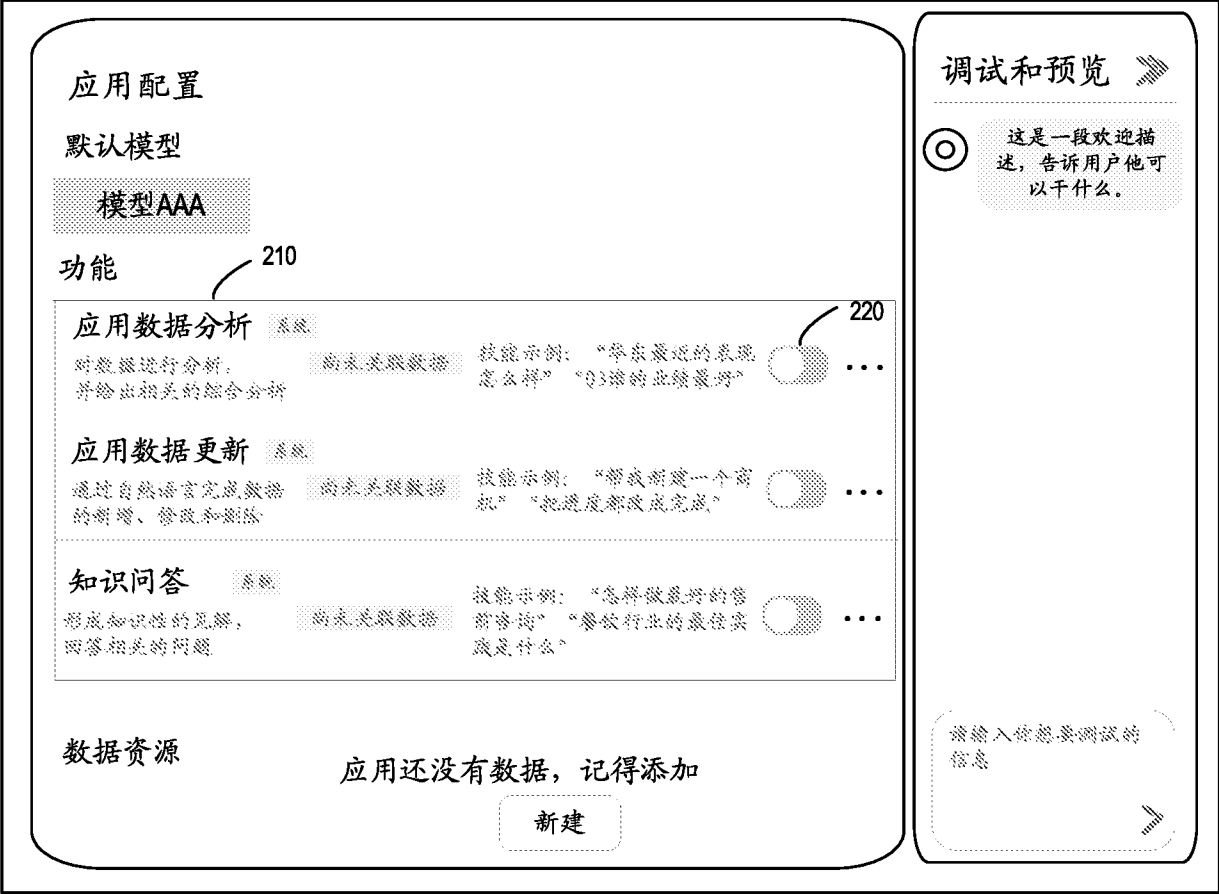


图 2

300



图 3

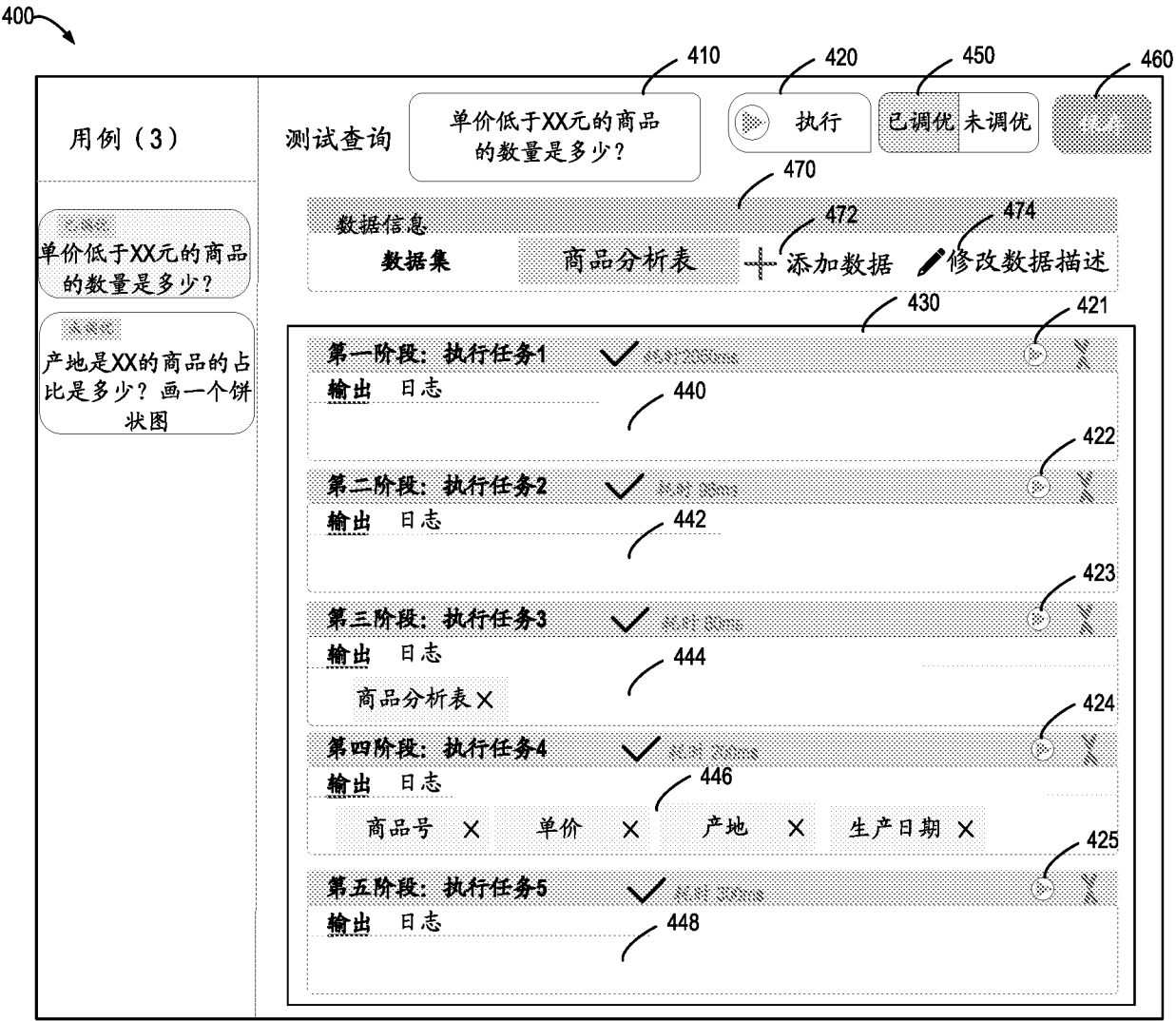


图 4

500

参数配置

术语集

对话记录

参考提示词

-对话要求:

1. 回答前, 请先判断背景知识是否足够回答问题, 如果无法回答, 请直接答复: “抱歉, 我不清楚这个问题的答案”。

2. 使用背景知识回答问题。

你的回答应该实事求是, 结构清晰, 以确保用户理解。将你的答案组织成清晰的步骤或要点, 系统地解决问题的各个方面。

条件

相似度

0

0.6

1

单次搜索数量

1

7

20

调试

图 5

600

XXX功能调试台

测试查询

用例库

记录ID	用户输入	SQL生成记录	最近更新时间	操作
10001230	单价低于XX元的商品的数量是多少?	Select xxx from commodity	2023-09-22 19:00:00	详情 删除
10012311	产地是XX的商品的数量是多少?	Select xxx from commodity	2023-09-22 18:55:00	详情 删除

图 6A

610

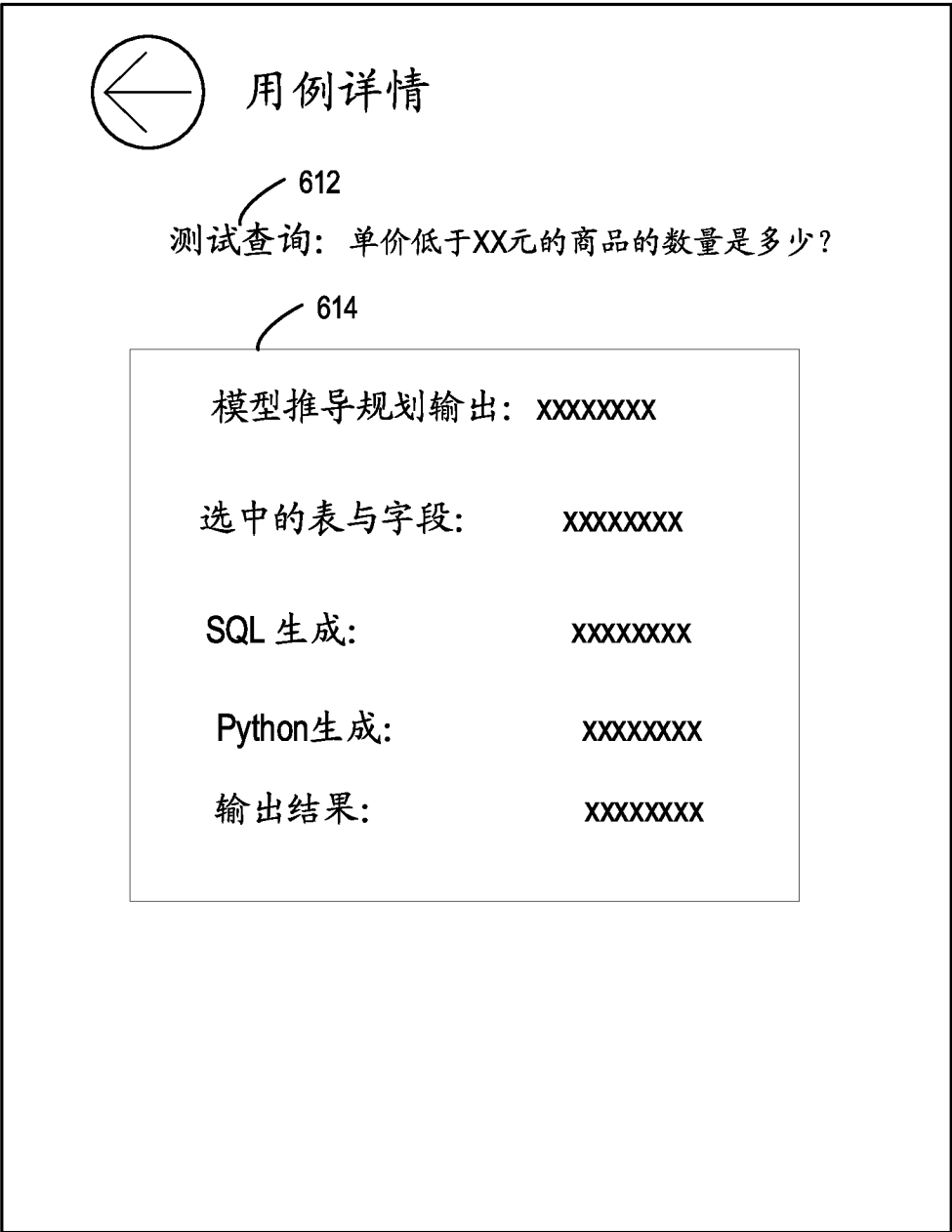


图 6B

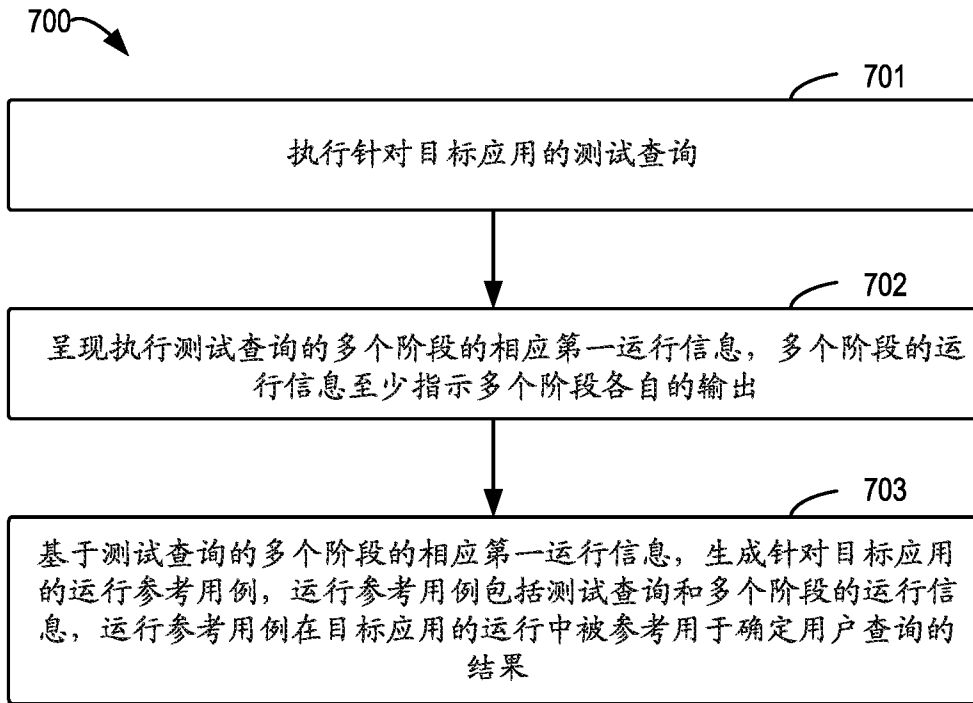


图 7A

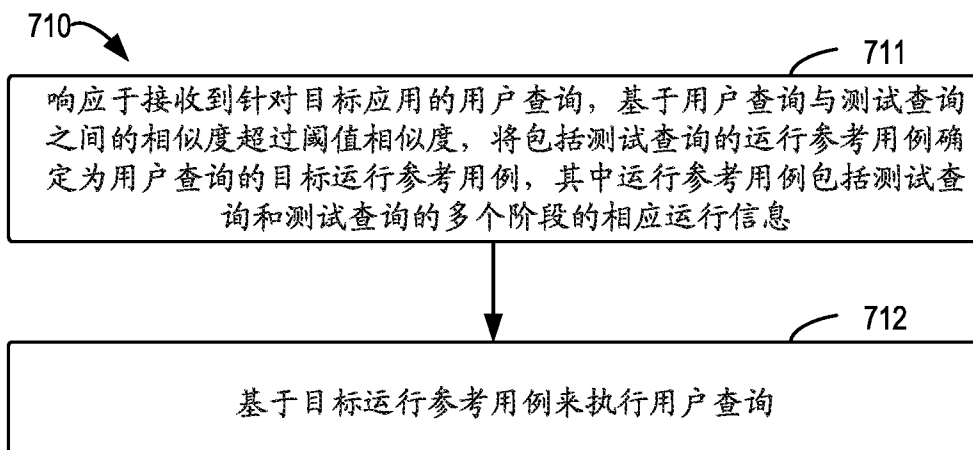


图 7B

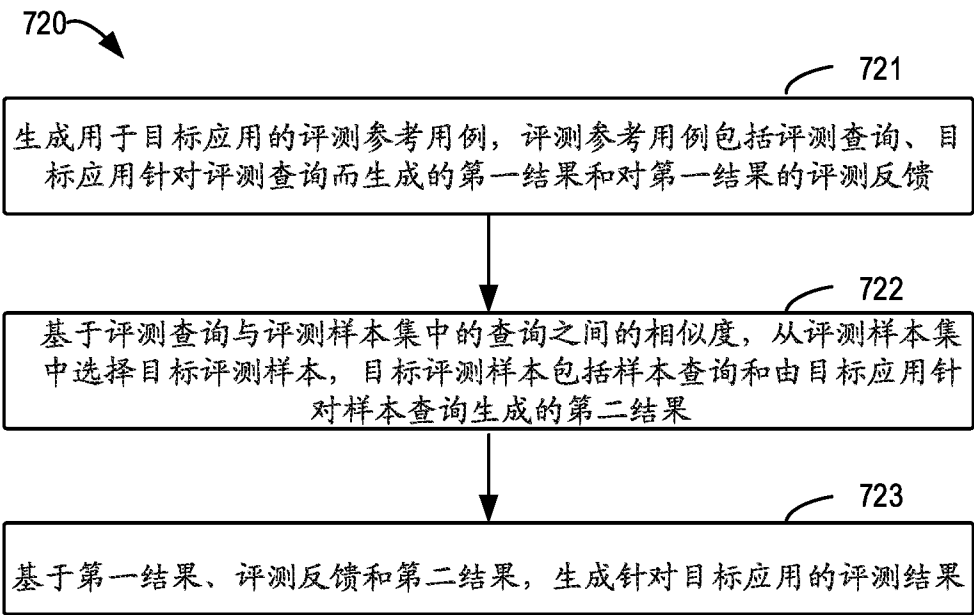


图 7C

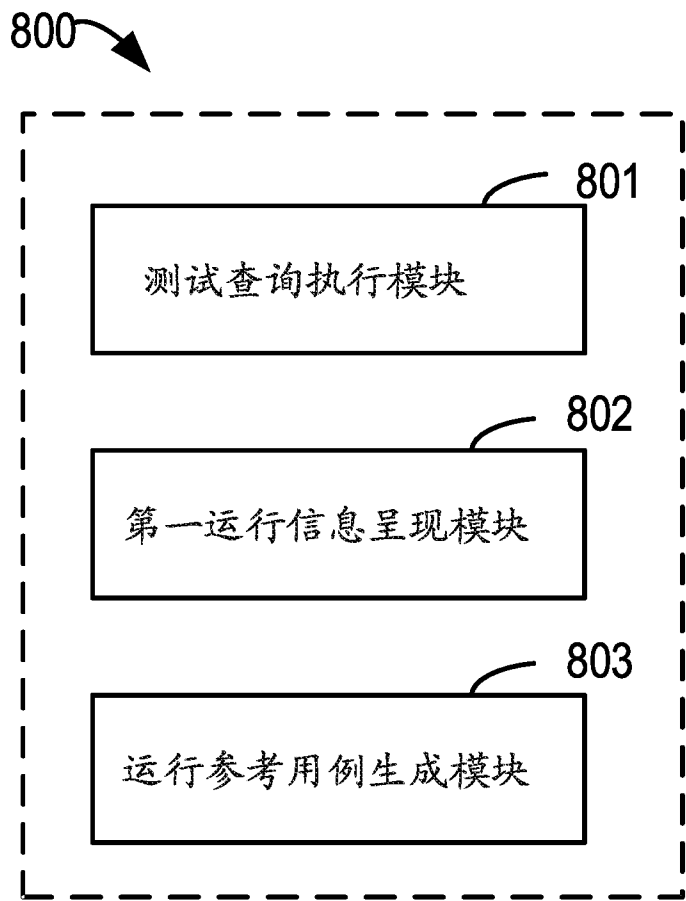


图 8A

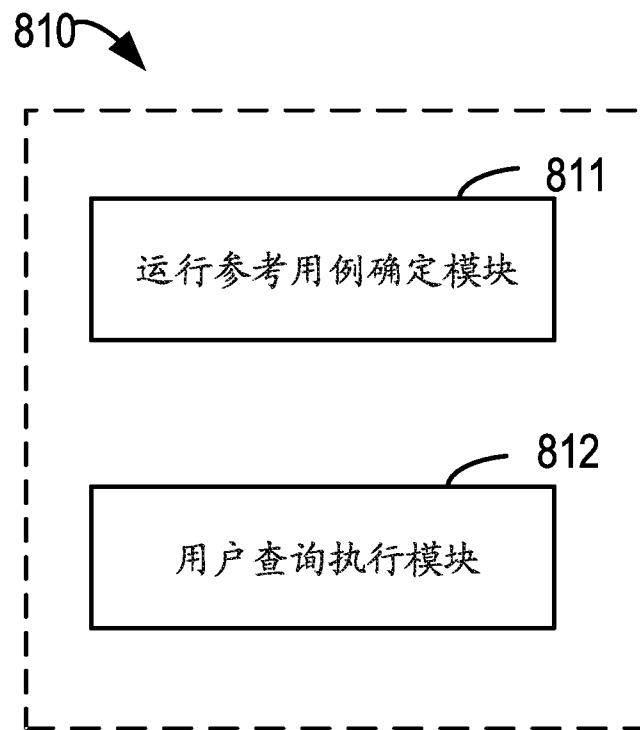


图 8B

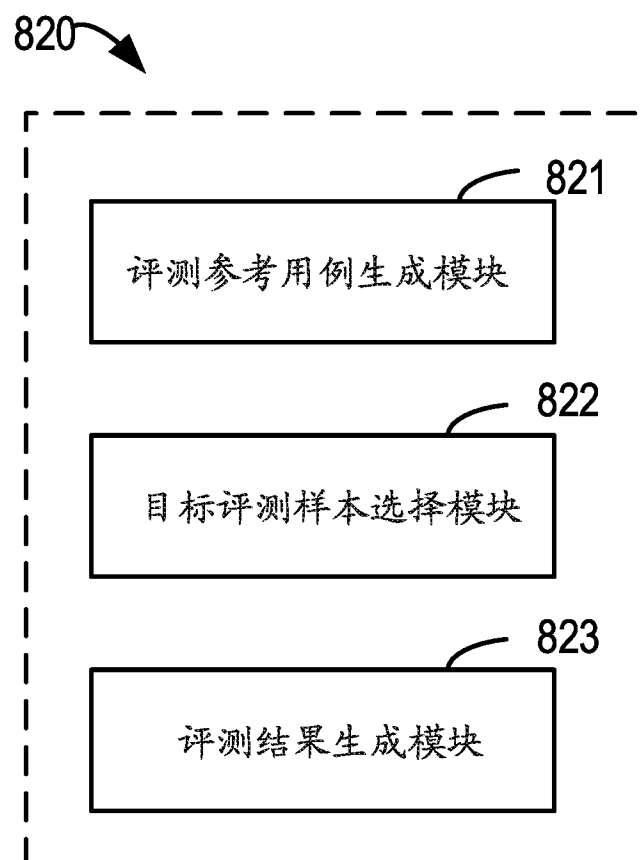


图 8C

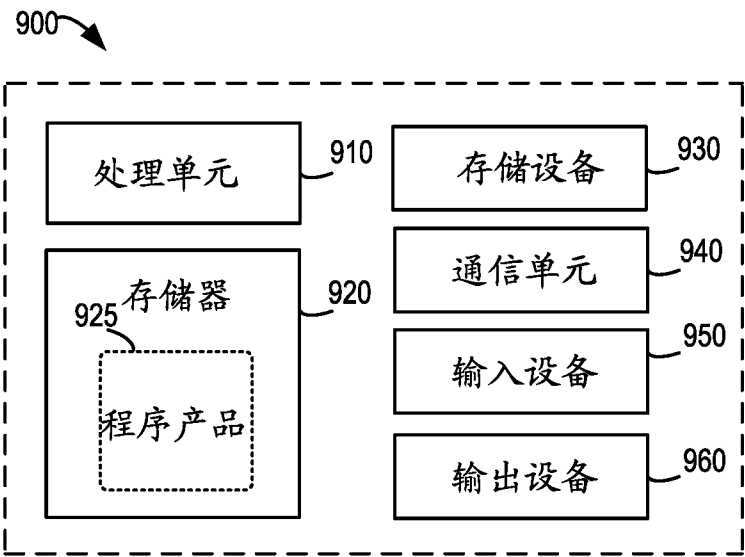


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/122387

A. CLASSIFICATION OF SUBJECT MATTER G06F 16/903(2019.01)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, G06N 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, ENTXT, DWPI, CNKI: 应用, 程序, 软件, 代码, 调试, 查询, 参考, 输出, 阶段, 步骤, APP, program, software, code, debug, query, reference, output, phase, step		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105740312 A (DASSAULT SYSTEMES SE) 06 July 2016 (2016-07-06) description, paragraphs [0004] and [0035]	15-16, 21, 23-24
A	CN 103164334 A (INTERNATIONAL BUSINESS MACHINES CORP.) 19 June 2013 (2013-06-19) entire document	1-24
A	US 2023251960 A1 (OPTUM, INC.) 10 August 2023 (2023-08-10) entire document	1-24
☐ 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 03 November 2024		Date of mailing of the international search report 20 November 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/CN2024/122387

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105740312	A	06 July 2016	US	2016188696	A1	30 June 2016
				JP	2016126770	A	11 July 2016
				EP	3038018	A1	29 June 2016
CN	103164334	A	19 June 2013	US	2013159975	A1	20 June 2013
US	2023251960	A1	10 August 2023	None			

A. 主题的分类

G06F 16/903(2019.01)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, G06N

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))

CNTEXT,ENTXT,DWPI,CNKI: 应用,程序,软件,代码,调试,查询,参考,输出,阶段,步骤, APP, program, software, code, de-bug, query, reference, output, phase, step

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 105740312 A (达索系统公司) 2016年7月6日 (2016 - 07 - 06) 说明书第[0004],[0035]段	15-16, 21, 23-24
A	CN 103164334 A (国际商业机器公司) 2013年6月19日 (2013 - 06 - 19) 全文	1-24
A	US 2023251960 A1 (OPTUM, INC.) 2023年8月10日 (2023 - 08 - 10) 全文	1-24

☐ 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11月3日

国际检索报告邮寄日期

2024年11月20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

庄湧

电话号码 (+86) 010-53961296

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/CN2024/122387

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	105740312	A	2016年7月6日	US	2016188696	A1	2016年6月30日
				JP	2016126770	A	2016年7月11日
				EP	3038018	A1	2016年6月29日
CN	103164334	A	2013年6月19日	US	2013159975	A1	2013年6月20日
US	2023251960	A1	2023年8月10日		无		