

说明书

风险防控方案的生成方法、目标语言模型的训练方法、设备及存储介质

5 技术领域

本说明书涉及互联网技术领域，尤其涉及一种风险防控方案的生成方法、目标语言模型的训练方法、设备及存储介质。

背景技术

- 10 在风控领域，风险防控平台在进行风险防控时，可以针对不同类型的事件采用不同的风险防控方案。例如，配置人员可以在风险防控平台上为不同类型的事件配置不同的风险防控方案，从而在事件实时发生的过程中，风险防控平台可以运行配置人员为该事件配置的风险防控方案以防控该事件的风险。在配置过程中，由于不同配置人员的专业性水平不一，仅仅凭借配置人员的个人经验，难以为每类事件配置到准确、
- 15 全面的风险防控方案，从而影响风险防控平台对风险事件的防控性能。

背景技术部分的内容仅仅是发明人个人所知晓的信息，并不代表上述信息在本公开申请日之前已经进入公共领域，也不代表其可以成为本公开的现有技术。

发明内容

- 20 本说明书提供一种风险防控方案的生成方法、目标语言模型的训练方法、设备及存储介质，在为目标事件配置风险防控方案的过程中，推荐设备能够调用预先训练好的目标语言模型，从而获得能够相对准确、全面的概括目标事件的风险标签，从而为配置人员推荐与该风险标签对应的风险防控方案。

- 第一方面，本说明书提供一种风险防控方案的生成方法，包括：响应于检测到请
- 25 求配置目标事件对应的风险防控方案，获得所述目标事件对应的目标提示信息，所述目标提示信息至少包括以自然语言对所述目标事件存在的风险进行描述的文本信息；通过所述目标提示信息指导目标语言模型生成所述目标事件对应的 N 个目标风险标签，所述 N 为大于或等于 1 的整数，所述目标语言模型被训练为从自然语言描述的文本信息中挖掘风险标签；基于所述 N 个目标风险标签生成所述风险防控方案。

第二方面，本说明书提供一种目标语言模型的训练方法，包括：获得 M 个样本风险管理数据，并确定所述 M 个样本风险管理数据各自对应的样本风险标签，所述 M 为大于 1 的整数；基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签，生成多个训练样本，其中，所述每个训练样本包括样本提示信息
5 信息和样本标签信息，所述样本提示信息为基于所述 M 个样本风险管理数据中的至少部分样本风险管理数据得到的与风险相关的自然语言描述，所述样本标签信息包括与
所述至少部分样本风险管理数据对应的样本风险标签；以及基于所述多个训练样本训练得到所述目标语言模型。

第三方面，本说明书还提供一种电子设备，包括：至少一个存储介质，存储有至少一个指令集；以及至少一个处理器，同所述至少一个存储介质通信连接，其中，所述至少一个处理器运行时读取所述至少一个指令集，并且根据所述至少一个指令集的指示，实现如第一方面所述的风险防控方案的生成方法或如第二方面所述的目标语言模型的训练方法。
10

第四方面，本说明书提供一种计算机可读非暂时性存储介质，其中，所述计算机可读非暂时性存储介质中存储有至少一个指令集，所述至少一个指令集被至少一个处理器执行时，实现如第一方面所述的风险防控方案的生成方法或如第二方面所述的目标语言模型的训练方法。
15

由以上技术方案可知，本说明书提供的风险防控方案的生成方法、目标语言模型的训练方法、设备及存储介质，在配置人员为目标事件配置风险防控方案的过程中，
20 推荐设备能够调用预先训练好的能够基于自然语言描述的文本信息挖掘出风险标签的目标语言模型，从而通过目标提示信息指导目标语言模型生成目标事件对应的一个或多个风险标签，进而推荐设备基于所述一个或多个风险标签生成目标事件的风险防控方案，将其推荐给配置人员。也就是说，本方案在配置过程中，调用了预先训练好的目标语言模型，避免配置人员由于专业经验的限制，挖掘不到能够全面、准确的概括
25 目标事件的风险标签，从而不能基于准确的风险标签为目标事件查找到合适的风险防控方案，影响后续风险防控平台识别风险事件的准确性。

本说明书提供的风险防控方案的生成方法、目标语言模型的训练方法、设备及存储介质的其他功能将在以下说明中部分列出。根据描述，以下数字和示例介绍的内容将对那些本领域的普通技术人员显而易见。本说明书提供的风险防控方案的生成方法、
30 设备及存储介质的创造性方面可以通过实践或使用下面详细示例中所述的方法、装置

和组合得到充分解释。

附图说明

5 为了更清楚地说明本说明书实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本说明书的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 示出了根据本说明书的实施例提供一种推荐风险防控方案的应用场景示意图；

10 图 2 示出了根据本说明书的实施例提供一种电子设备的硬件结构图；

图 3 示出了根据本说明书的实施例提供一种风险防控方案的生成方法的流程图；

图 4 示出了配置界面的示意图；

图 5 示出了目标语言模型的训练方法的流程示意图；

图 6 示出了风险管理数据的层级的示意图；

15 图 7 示出了映射关系的一种示意图；

图 8 示出了映射关系的另一种示意图；

图 9 示出了数据标签对#1 的示意图；

图 10 示出了数据标签对#2 的示意图；以及

图 11 示出了基于映射关系确定风险管理数据的数据描述信息的示意图。

20 具体实施方式

以下描述提供了本说明书的特定应用场景和要求，目的是使本领域技术人员能够制造和使用本说明书中的内容。对于本领域技术人员来说，对所公开的实施例的各种局部修改是显而易见的，并且在不脱离本说明书的精神和范围的情况下，可以将这里
25 定义的一般原理应用于其他实施例和应用。因此，本说明书不限于所示的实施例，而是与权利要求一致的最宽范围。

这里使用的术语仅用于描述特定示例实施例的目的，而不是限制性的。比如，除非上下文另有明确说明，这里所使用的，单数形式“一”，“一个”和“该”也可以包括复数

形式。当在本说明书中使用时，术语“包括”、“包含”和/或“含有”意思是指所关联的整数，步骤、操作、元素和/或组件存在，但不排除一个或多个其他特征、整数、步骤、操作、元素、组件和/或组的存在或在该系统/方法中可以添加其他特征、整数、步骤、操作、元素、组件和/或组。

- 5 考虑到以下描述，本说明书的这些特征和其他特征、以及结构的相关元件的操作和功能、以及部件的组合和制造的经济性可以得到明显提高。参考附图，所有这些形成本说明书的一部分。然而，应该清楚地理解，附图仅用于说明和描述的目的，并不旨在限制本说明书的范围。还应理解，附图未按比例绘制。

- 本说明书中使用的流程图示出了根据本说明书中的一些实施例的系统实现的操作。
- 10 应该清楚地理解，流程图的操作可以不按顺序实现。相反，操作可以以反转顺序或同时实现。此外，可以向流程图添加一个或多个其他操作。可以从流程图中移除一个或多个操作。

 为了方便理解，本说明书对下文将会出现的术语进行如下解释。

- 风险防控方案：是风险防控平台在对某种类型的事件进行风险防控时采用的防控
- 15 方案，旨在减少和防止相关事件的风险的发生。风险防控平台可以理解为需要防控风险的平台。例如，金融服务平台，交易平台等。具体的，在金融服务平台，风险防控方案通常指的是在所述金融服务平台上对投资、贷款、交易等事件的风险进行评估和控制的方案。风险防控方案例如包括一个或多个防控策略，防控策略用于决定如何处理和应对事件潜在的风险。防控策略可以包括一系列规则、流程、以及规则和流程中
- 20 涉及的决策标准和方法。规则包括特定的条件、规定和限制，用于定义、约束、识别或者处理风险事件。比如，规则包括以下条件判断：交易金额是否超过阈值，用户行为是否异常等。如果针对某个事件，所述规则中的所有条件都满足的情况下，会触发风险防控平台对该事件的处理，比如触发警报，拒绝所述事件的执行完成，或者增加额外的验证等处理。

- 25 风险标签：是对风险进行分类的标识信息，不同风险标签通常对应不同的风险防控方案。不同风险标签可以表示不同类型或性质的风险，或者不同风险标签可以表示不同程度的风险。例如，专业人员可以根据风险的严重程度将风险进行分类得到以下风险标签：低风险、中风险和高风险；专业人员可以根据风险的来源和类型将风险进行分类得到以下风险标签：欺诈风险、信用风险、违约风险、虚假风险、集中风险、

担保风险、逆选择风险、恶意风险、市场风险、操作风险、法律风险和流动性风险。

在金融领域中：欺诈风险指涉及欺诈行为的风险，欺诈行为例如虚假身份信息、盗用支付卡、虚构交易等。信用风险指因借款人无法按时偿还债务导致的风险，例如借款人信用记录不良、还款能力不足等。违约风险指借款人无法按时履约导致的风险。虚假风险指因提供虚假或者误导信息导致的风险，例如，提供虚假流水，资产信息等虚假信息。集中风险指因过度集中在某个或某几个借款人上而导致的风险，例如信贷组合中因过度依赖某个行业或地区的借款人导致的风险。担保风险指因担保物价值下降或者无法改变导致的风险，例如，例如，贷款的担保物价值下跌或者无法按时变现可能会增加担保风险。逆选择风险指因信息不对称导致的风险，例如，借款人提供的信息不完全或不准确，或者借款人故意隐瞒重要信息会导致逆选择风险。恶意风险指涉及恶意行为或者欺诈行为的风险，例如借款人故意制造事故、提交虚假索赔等导致的风险。市场风险指因市场因素导致的风险，例如，股市下跌、汇率波动等引起的风险。操作风险指由内部流程、系统故障、人为错误等因素导致的风险。法律风险指由法律法规、诉讼或者合同纠纷等因素导致的风险。流动性风险指由于资产无法迅速转化为现金或市场流动性不足而导致的风险，例如，持有不欢迎的低流动资产可能面临流动性风险。

提示信息：提示信息指向语言模型输入的、用于引导语言模型生成特定内容的文本信息。提示信息可以是一个或者多个语句、段落等。提示信息也可以称为提示词（prompt）。

在风控领域，配置人员在风险防控平台上为各类事件配置不同的风险防控方案。目前，许多研究致力于在实时的风险防控中，如何提高风险防控方案的运行效率以提高防控效果，对于如何为事件配置更适合所述事件的风险防控方案的关注度较低。

在实际应用中，由于不同类型事件的风险可能不同，因此风险防控平台需要为不同类型的事件配置不同的风险防控方案。例如在金融领域，对于交易类事件，配置风险防控方案 A，对于信贷类事件，配置风险防控方案 B。

例如，针对交易类事件，配置人员需要根据自身的专业经验确定交易类事件需要防控哪些风险，进而根据自身的专业经验为交易类事件配置出能够防控这些风险的风险防控方案。可见，能不能为交易类事件配置合适、准确、全面的风险防控方案完全取决于配置人员的专业经验。这样一来，配置人员为不同类型事件都配置到合适的风险防控方案的难度较大，从而影响风险防控平台对各类型事件的风险的防控性能。

鉴于此，本说明书提供一种风险防控方案的生成方法、设备及存储介质，在配置人员为目标事件配置风险防控方案的过程中，推荐设备能够调用预先训练好的目标语言模型，从而获得能够相对准确、全面的概括目标事件的风险标签，进而为配置人员推荐与该风险标签对应的风险防控方案。克服部分配置人员因专业能力的限制不能自行全面的挖掘目标事件可能存在的风险，导致无法配置出适合该目标事件的风险防控方案的问题。

下面结合图 1 对本说明书提供的生成风险防控方案的应用场景进行介绍。

图 1 示出了根据本说明书的实施例提供的一种生成风险防控方案的应用场景示意图。如图 1 所示，场景 001 可以包括：用户 100A，终端设备 200，服务器 300、管理设备 400，配置人员 100B 以及训练设备 500。

图 1 示例的应用场景包括 2 个阶段，管理阶段和运行阶段。

管理阶段：配置人员 100B 通过管理设备 400 将风险防控方案配置到服务器 300。

服务器 300 可以与终端设备 200 通信，接收来自终端设备 200 发送的数据，或者向终端设备 200 发送数据。服务器 300 可以是提供各种服务的服务器，例如对终端设备 200 安装的金融服务客户端（比如银行客户端等）提供支持的后台服务器。此时，服务器 300 可以支持用户 100A 在该金融服务客户端完成相关事件。为了使用户 100A 在金融服务客户端享受到安全的金融服务，金融服务平台需要进行风险防控，因此需要在金融服务平台为不同类型的事件配置风险防控方案。

示例性的，配置人员 100B 在配置风险防控方案的过程中，管理设备 400 可以调用目标语言模型。目标语言模型基于配置人员 100B 针对目标事件输入的目标提示信息生成目标事件对应的风险标签。管理设备 400 可以基于风险标签生成风险防控方案，并向配置人员 100B 推荐所述风险防控方案。配置人员 100B 在配置过程中，可以选择管理设备 400 推荐的所述风险防控方案，也可以自行配置风险防控方案，也可以在推荐的所述风险防控方案的基础上做一些改进。具体的，配置人员 100B 为不同类型的事件配置不同的风险防控方案。比如，配置人员 100B 基于管理设备 400 的推荐为金融服务平台的交易类事件配置风险防控方案 A，为信贷类事件配置风险防控方案 B。管理设备 400 是具有一定数据处理能力的电子设备。管理设备 400 可以执行本说明书描述的风险防控方案的生成方法的数据和指令，并可以执行或用于执行所述数据和指令。例如，管理设备 400 可以包括具有数据信息处理功能的硬件设备和驱动该硬件设备工作所需必要的程序。

可以理解，所述目标语言模型是训练设备 500 预先训练好的语言模型。训练设备 500 训练得到所述目标语言模型后，将所述目标语言模型传输到管理设备 400，从而配置人员在配置过程中可以在管理设备 400 本地调用所述目标语言模型。或者，训练设备 500 训练得到所述目标语言模型后将其传输到云端，管理设备 400 可以在云端调用
5 所述训练设备 500 上的目标语言模型。

训练设备 500 是具有一定数据处理能力的电子设备。训练设备 500 可以执行本说明书描述的目标语言模型的训练方法的数据和指令，并可以执行或用于执行所述数据和指令。例如，训练设备 500 可以包括具有数据信息处理功能的硬件设备和驱动该硬件设备工作所需必要的程序。

10 运行阶段：用户 100A 通过终端设备 200 上安装的 APP 执行事件时，所述 APP 的客户端或者服务器 300 会运行风险防控方案防控所述事件的风险。其中，上述风险防控方案可以是上述管理阶段配置得到的。

示例性的，当用户 100A 在终端设备 200 上使用金融服务客户端时，如果要执行交易类事件，金融服务客户端可以向服务器 300 获取风险防控方案 A。从而金融服务客户端运行风险防控方案 A 识别用户 100A 执行的交易事件是否有风险，以及识别所述交易事件有风险后，金融服务客户端会执行风险防控方案 A 对应的处理操作，例如身份验证操作或其他防控风险的操作。或者，当用户 100A 在金融服务客户端执行交易类事件时，服务器 300 会运行风险防控方案 A 识别用户 100A 执行的交易事件是否有风险，识别所述交易事件有风险后，可以通知金融服务客户端执行身份验证操作或其他
15 防控风险的操作。当用户 100A 在所述金融服务客户端执行信贷类事件时，金融服务客户端或者服务器 300 会运行预先配置的风险防控方案 B 识别用户 100 执行的信贷事件是否有风险，以及识别所述信贷事件有风险后，金融服务客户端会执行风险防控方案 B 对应的处理操作。用户 100 A 可以通过终端设备 200 执行相关事件。终端设备 200 除了安装有金融服务应用程序（Application，APP）之外，还可以安装其他 APP，其他
20 APP 包括但不限于网页浏览器类 APP、搜索类 APP、聊天类 APP、购物类 APP、服务类 APP、视频类 APP、理财类 APP、支付类 APP 等等。

终端设备 200 可以向用户呈现交互界面，以便用户通过该交互界面执行一些事件。例如，响应于用户的操作，终端设备 200 向用户呈现金融服务 APP 的交互界面，用户可以在金融服务 APP 的交互界面完成相关金融事件，金融事件比如投资、贷款、基金
30 交易等事件。

终端设备 200 可以包括移动设备、平板电脑、笔记本电脑、机动车辆的内置设备或类似内容，或其任意组合。在一些实施例中，所述移动设备可包括智能家居设备、智能移动设备、虚拟现实设备、增强现实设备或类似设备，或其任意组合。在一些实施例中，所述智能家居装置可包括智能电视、台式电脑等，或任意组合。在5 一些实施例中，所述智能移动设备可包括智能手机、个人数字辅助、游戏设备、导航设备等，或其任意组合。在一些实施例中，所述虚拟现实设备或增强现实设备可能包括虚拟现实头盔、虚拟现实眼镜、虚拟现实补丁、增强现实头盔、增强现实眼镜、增强现实补丁或类似内容，或其中的任何组合。例如，所述虚拟现实设备或所述增强现实设备可能包括 AR 眼镜、头戴式显示器、VR 等。在一些实施例中，所述机动车中的内置装置10 可包括车载计算机、车载电视等。在一些实施例中，终端设备 200 可以是具有定位技术的设备，用于定位终端设备 200 的位置。

应该理解，图 1 中的终端设备 200、服务器 300、管理设备 400、训练设备 500 的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备 200、服务器 300、管理设备 400 和训练设备 500。

15 还可以理解，本说明书提供的风险防控方案的生成方法可以由图 1 中的管理设备 400 执行。本说明书提供的目标语言模型的训练过程可以由训练设备 500 执行。在一些场景下，管理设备 400 和训练设备 500 可以对应相同的硬件实体，也可以对应不同的硬件实体。

图 2 示出了根据本说明书的实施例提供的一种电子设备 600 的硬件结构图。电子设备 20 600 可以作为图 1 中的管理设备 400，也可以作为图 1 中的训练设备 500。

如图 2 所示，电子设备 600 可以包括至少一个存储介质 630 和至少一个处理器 620。在一些实施例中，电子设备 600 还可以包括通信端口 650 和内部通信总线 610。同时，电子设备 600 还可以包括 I/O 组件 660。

内部通信总线 610 可以连接不同的系统组件，包括存储介质 630、处理器 620 和通25 信端口 650。

I/O 组件 660 支持电子设备 600 和其他组件之间的输入/输出。

通信端口 650 用于电子设备 600 同外界的数据通信，比如，通信端口 650 可以用于电子设备 600 同网络 700 之间的数据通信。通信端口 650 可以是有线通信端口也可以是无线通信端口。

在一些实施例中，网络 700 可以是任何类型的有线或无线网络，也可以是其组合。比如，网络 700 可以包括电缆网络，有线网络、光纤网络、电信通信网络、内联网、互联网、局域网（LAN）、广域网（WAN）、无线局域网（WLAN）、大都市市区网（MAN）、公用电话交换网（PSTN）、蓝牙网络、ZigBee 网络、近场通信（NFC）网络或类似网络。在 5 一些实施例中，网络 700 可以包括一个或多个网络接入点。例如，网络 700 可以包括有线或无线网络接入点，如基站或互联网交换点。

存储介质 630 可以包括数据存储装置。所述数据存储装置可以是非暂时性存储介质，也可以是暂时性存储介质。比如，所述数据存储装置可以包括磁盘 632、只读存储介质（ROM）634 或随机存取存储介质（RAM）636 中的一种或多种。存储介质 630 10 还包括存储在所述数据存储装置中的至少一个指令集。所述指令集可以是计算机程序代码。当电子设备 600 为管理设备 400 时，所述计算机程序代码可以包括执行本说明书提供的风险防控方案的生成方法的程序、例程、对象、组件、数据结构、过程、模块等等。当电子设备 600 为训练设备 500 时，所述计算机程序代码可以包括执行本说明书提供的目标语言模型的训练方法的程序、例程、对象、组件、数据结构、过程、15 模块等等。

处理器 620 可以同存储介质 630 以及通信端口 650 通过内部通信总线 610 通信连接。处理器 620 用以执行上述至少一个指令集。当电子设备 600 作为管理设备 400 运行时，处理器 620 读取所述至少一个指令集，并且根据所述至少一个指令集的指示，执行本说明书提供的风险防控方案的生成方法。当电子设备 600 作为训练设备 500 运 20 行时，处理器 620 读取所述至少一个指令集，并且根据所述至少一个指令集的指示，执行本说明书提供的目标语言模型的训练方法。

处理器 620 可以是一个或多个处理器的形式，在一些实施例中，处理器 620 可以包括一个或多个硬件处理器，例如微控制器，微处理器，精简指令集计算机（RISC），专用集成电路（ASIC），特定于应用的指令集处理器（ASIP），中心处理单元（CPU），25 图形处理单元（GPU），物理处理单元（PPU），微控制器单元，数字信号处理器（DSP），现场可编程门阵列（FPGA），高级 RISC 机器（ARM），可编程逻辑器件（PLD），能够执行一个或多个功能的任何电路或处理器等，或其任何组合。

仅仅为了说明问题，在附图中电子设备 600 中仅示出一个处理器 620。然而，应当注意，本说明书中电子设备 600 还可以包括多个处理器，因此，本说明书中披露的操

作和/或方法步骤可以如本说明书所述的由一个处理器执行，也可以由多个处理器联合执行。例如，如果在本说明书中电子设备 600 的处理器 620 执行步骤 A 和步骤 B，则应该理解，步骤 A 和步骤 B 也可以由两个不同处理器 620 联合或分开执行（例如，第一处理器执行步骤 A，第二处理器执行步骤 B，或者第一和第二处理器共同执行步骤 A 和 B）。

图 3 示出了根据本说明书的实施例提供的一种风险防控方案的生成方法 P100 的流程图。如前，电子设备 600 可以执行本说明书的方法 P100。具体地，电子设备 600 中的处理器 620 可以读取存储在其本地存储介质中的指令集，然后根据指令集的规定，执行本说明书的方法 P100。方法 P100 可应用于交易领域、网络安全领域、金融领域等风险防控平台的配置过程中。为清楚、完整的描述方法 P100，后文以金融领域为例描述，即电子设备 600 是在管理阶段，供配置人员为金融服务平台的目标事件配置风险防控方案的配置设备，方法 P100 在其他领域的应用可以此类推。

如图 3 所示，风险防控方案的生成方法 P100 可以包括如下步骤。

S110：响应于检测到请求配置目标事件对应的风险防控方案，获得所述目标事件对应的目标提示信息，所述目标提示信息至少包括以自然语言对所述目标事件存在的风险进行描述的文本信息。

其中，目标事件为当前待配置风险防控方案的事件。目标事件可以是指某一个具体的事件，也可以是指某一种类型的事件。在不同的应用场景中，目标事件的指代可以有所不同。例如，在金融服务平台中，目标事件可以理解为金融服务平台中的交易类事件。后文以一个目标事件为例说明风险防控方案的整体推荐流程。本领域技术人员能够理解的，当存在多个目标事件（例如交易类事件、信贷类事件和转账类事件）时，多个目标事件的风险防控方案的推荐过程可以同时进行，或者多个目标事件的风险防控方案的推荐过程可以依次进行。

配置人员触发对目标事件的风险防控方案的配置流程后，电子设备 600 会接收到配置人员触发的配置请求，响应于所述配置请求，获得目标事件对应的目标提示信息。在本说明书中，目标提示信息为用于输入目标语言模型、并指导目标语言模型生成特定内容的文本信息。目标提示信息可以包括一个或者多个语句、段落等。

目标提示信息包括风险描述文本，风险描述文本为以自然语言对所述目标事件存在的风险进行描述的文本信息。风险描述文本是目标语言模型待分析处理的信息。目标语言模型需要从风险描述文本中挖掘出一个或者多个风险标签。除此之外，目标提

示信息还可以包括角色限定语和任务限定语。其中，角色限定语用于限定目标语言模型所充当的角色，例如“你是风控领域专家”。任务限定语用于限定所述目标语言模型需要完成的任务，例如“你需要在风险描述信息中挖掘出风险标签”。当然，目标提示信息还可以包括其他信息，用于帮助目标语言模型能够更好的完成任务。目标提示信息的内容越丰富，目标语言模型输出的内容越准确。

S120: 通过所述目标提示信息指导目标语言模型生成所述目标事件对应的 N 个目标风险标签，所述 N 为大于或等于 1 的整数，所述目标语言模型被训练为从自然语言描述的文本信息中挖掘风险标签。

目标语言模型可以是预先训练好的大数据模型，具有从自然语言描述的文本信息中挖掘、概括或抽象出风险标签的能力。目标语言模型的训练过程可以参见后文详细描述，此处不作赘述。

电子设备 600 调用目标语言模型，并向目标语言模型输入所述目标提示信息，目标语言模型会输出 N 个目标风险标签，所述 N 个目标风险标签能够从专业角度、高度概括目标事件可能涉及的所有风险。具体的，目标语言模型可以被部署在电子设备 600 中，电子设备 600 可以在本地调用目标语言模型。或者，目标语言模型可以被部署到云端，电子设备 600 可以通过 RPC (remote procedure call, 远程过程调用) 方式调用所述目标语言模型。

S130: 基于所述 N 个目标风险标签生成所述风险防控方案。

如前文描述，目标语言模型具有从自然语言描述的文本信息中挖掘、概括或抽象出风险标签的能力，因此，目标语言模型输出的 N 个目标风险标签能够表征目标事件可能涉及的风险类型。因此，在 S130 中，电子设备 600 可以基于所述 N 个目标风险标签生成目标事件对应的风险防控方案。

例如，电子设备 600 针对所述 N 个目标风险标签中的每个目标风险标签，在目标数据库中查找得到所述目标风险标签对应的风险管理数据；并基于所述 N 个目标风险标签各自对应的风险管理数据生成所述风险防控方案。

其中，风险管理数据是指服务平台中预先定义的用于进行风险管理的相关数据。风险管理数据可以包括一个或者多个数据。以金融服务平台为例，金融服务平台中预先定义了大量字段，包括但不限于：用户类型、交易类型、交易金额、交易时间、7 天内交易次数、收单币种等。在上述预先定义的字段中，与风险管理相关的字段可以称

为风险管理数据。例如，上述字段中的“7 天内交易次数”、“收单币种”等字段可以称为风险管理数据。需要说明的是，在不同的应用场景中，风险管理数据的指代有所不同，上述所列举的字段仅为一些举例说明。

继续以金融服务平台为例，金融服务平台中预先定义的每个字段均关联有风险标签。例如，字段 1 与风险标签“盗卡风险”关联，字段 2 与风险标签“盗账户风险”关联，字段 3 与风险标签“盗卡风险”关联，字段 4 与风险标签“盗账户风险”关联，等。上述字段与风险标签之间的关联关系可以存储在目标数据库中。因此，电子设备 600 针对 N 个目标风险标签中的每个目标风险标签，可以从目标数据库中查询得到与该目标风险标签关联的所有字段，并将这些字段作为该目标风险标签对应的风险管理数据。例如，风险标签“盗卡风险”对应的风险管理数据可以包括字段 1 和字段 3，风险标签“盗账户风险”对应的风险管理数据可以包括字段 2 和字段 4。进一步的，电子设备 600 在获得 N 个目标风险标签对应的风险管理数据之后，可以基于所述 N 个目标风险标签各自对应的风险管理数据，生成目标事件对应的风险防控方案。

例如，电子设备 600 在生成目标事件对应的风险防控方案的过程中，可以按照所有风险管理数据各自对应的优先级，将所述所有风险管理数据按照优先级从高到低的顺序排列，如果所述所有风险管理数据中有重复的风险管理数据，则可以进行去重操作，避免风险管理数据冗余，从而避免在执行所述目标事件对应的风险防控方案时，运行重复的风险管理数据而效率低的问题。或者，电子设备 600 还可以进行其他操作，使得生成的目标风险防控方案较为精简，且更有针对性，从而在后续运行所述风险防控方案时能够提高运行效率和防控效果。在上述 N 取值大于 1 时，即目标语言模型输出多个目标风险标签时，电子设备 600 可以分别基于每个目标风险标签对应的风险管理数据生成一个子防控方案，从而共得到 N 个子防控方案。进而，电子设备 600 将上述 N 个子防控方案进行合并得到目标事件对应的风险防控方案。该情况下，最终生成的风险防控方案相当于 N 个子防控方案的并集。每个子防控方案具有对某种特定类型风险的防控能力，而最终生成的风险防控方案可以视为 N 个子防控方案的并集，具有对 N 种类型风险的防控能力。

以 N=3 为例，电子设备 600 基于风险标签#1 对应的风险管理数据生成风险标签#1 对应的子防控方案#1，基于风险标签#2 对应的风险管理数据生成风险标签#2 对应的子防控方案#2，基于风险标签#3 对应的风险管理数据生成风险标签#3 对应的子防控方案

#3，基于子防控方案#1、子防控方案#2 和子防控方案#3 得到目标事件对应的风险防控方案。

比如，电子设备 600 在基于子防控方案#1、子防控方案#2 和子防控方案#3 得到目标事件对应的风险防控方案过程中，可以按照不同子防控方案防控的风险的严重程度将子防控方案#1、子防控方案#2 和子防控方案#3 进行合理排序生成所述目标事件对应的风险防控方案。或者，电子设备 600 还可以对子防控方案#1、子防控方案#2 和子防控方案#3 中相同的防控方案进行去重等操作。

该实施例中，目标数据库中的风险管理数据可以是基于其他风险防控平台得到的防控经验数据，配置人员可以基于目标数据库生成合适的风险防控方案。因此，可以避免配置人员采集大量事件（目标事件的同类型事件）的事件信息，并基于大量事件的事件信息深度学习得到风险防控方案。

也就是说，采用本实施例提供的方案，配置人员无需等到所述同类型事件的数量累积到一定程度，因此配置周期短。对于风险防控平台扩展的新服务场景下的新型事件，配置人员可以基于目标数据库生成与所述新型事件类似的事件的风险防控方案，将其作为新型事件的风险防控方案的参考，避免在新服务场景的推广初期产生资损。

在一些实施例中，每个目标风险标签对应的风险管理数据包括 K 个层级的数据， K 为大于 1 的整数，不同层级的数据对风险的管理范围不同。

在一些实施例中，所述 K 个层级包括下述中的至少两项：事件层级、策略层级、规则层级、特征层级和模型层级。

例如，风险标签#1 对应的风险管理数据包括上述 5 个层级的数据，不同层级的数据的管理范围不同，上述 5 个层级的排序顺序表示管理范围的从大到小。管理范围大的层级能够链接到管理范围小的层级。具体的，事件层级的数据能够链接到策略层级的数据，策略层级的数据能够链接到规则层级的数据事件层级的数据，以此类推。不同风险标签对应的风险管理数据包括的层级数量可以不同，比如风险标签#1 对应的风险管理数据包括事件层级、规则层级和特征层级的数据；风险标签#2 对应的风险管理数据包括策略层级、规则层级和模型层级的数据。

具体比如，风险标签#1 对应的风险管理数据中，事件层级的风险管理数据表示风险标签#1 对应的目标事件类型为交易类，策略层级的风险管理数据表示风险标签#1 对应的策略为支付风险防控策略。规则层级的风险管理数据包括所述支付风险防控策略

的各种判断规则。比如，交易事件涉及的用户的 7 天内交易次数是否大于或等于预设阈值 A，交易事件涉及的用户的 7 天内交易金额是否大于或等于预设阈值 B，交易事件涉及的用户的支付介质是否是该用户的常用支付介质，交易事件涉及的交易设备是否是该用户的常用设备、交易事件涉及的用户的信用指数等等。特征层级的风险管理数据包括用户的 7 天内交易次数、用户的 7 天内交易金额、用户的支付介质、交易设备和用户的信用指数等。模型层级的风险管理数据包括能够计算出用户的信用指数的模型等。

其中，特征可以理解为变量，即专业人员根据服务需要，将目标事件的事件信息中的属性值本身、或者对属性值进行加工得到的加工属性值用于规则的左值，左值称为变量。

电子设备 600 在生成目标事件对应的风险防控方案之后，可以向配置人员推荐该风险防控方案，从而给配置人员的配置过程提供帮助，提高风险防控方案的配置效率。

在一些实施例例中，在配置界面中显示所述风险防控方案，并在所述配置界面中显示操作控件，所述操作控件被配置为：在所述操作控件被触发时，将所述风险防控方案配置到所述目标事件。

例如，图 4 所示，目标事件为交易类事件，配置人员可以在配置界面的窗口 10 选中交易类事件。选中后跳转到窗口 20，配置人员在窗口 20 的输入位置输入目标提示信息，输入目标提示信息后可以点击回车或者确定等操作控件，表示输入完成。输入完成后跳转到窗口 30，窗口 30 中会显示电子设备 600 推荐的风险防控方案的具体内容，比如策略 A、规则 A 等。

窗口 30 包括操作控件“应用”和“自定义”。如果配置人员选中操作控件“应用”，则表示将窗口 30 显示的风险防控方案直接配置到交易类事件。如果配置人员选中操作控件“自定义”，则表示配置人员不选择窗口 30 显示的风险防控方案，自己重新为交易类事件定义风险防控方案。

上述方案通过采用可视化交互方式向配置人员展示风险防控方案，并在交互页面中提供将风险防控方案直接应用到目标事件的能力，能够提高配置人员在配置过程中的便利性，提高配置效率，并降低对配置人员的风控经验的要求。

可以理解，在执行方法 P100 之前，需要训练得到目标语言模型，下面介绍目标语言模型的训练过程。

电子设备 600 可以通过方法 P200 训练得到所述目标语言模型，如图 5 所示，方法 P200 包括如下步骤：

S210：获得 M 个样本风险管理数据，并确定所述 M 个样本风险管理数据各自对应的样本风险标签，每个样本风险管理数据对应的样本风险标签的数量可以是一个或者
5 多个， M 为大于 1 的整数。

其中，样本风险管理数据和 P100 中提及的风险管理数据的数据类型和格式类似，样本是为了强调这部分风险管理数据是用于训练目标语言模型的。样本风险标签与 P100 中提及的目标风险标签也是类似，只是说，目标风险标签是基于目标事件得到，
10 样本风险标签是基于其他事件或者其他途径得到的，用于训练目标语言模型的风险标签。

电子设备 600 确定所述 M 个样本风险管理数据各自对应的样本风险标签可以通过多种方式，下面以其中 2 种方式示例说明。

方式 1

在一些实施例中，电子设备 600 获得初始数据库，所述初始数据库包括所述 M 个
15 样本风险管理数据，所述初始数据库中的数据按照管理范围由大到小的顺序被划分为 K_1 个层级， K_1 大于或等于 K ， K_1 个层级的层级内容可以参考 P100 中关于 K 个层级的相关描述，不再赘述。

可以理解，所述 M 个样本风险管理数据包括的每个样本风险管理数据包括的层级数量可以不同。所述初始数据库为数据处理前的所述目标数据库。

例如，图 6 所示， M 个样本风险管理数据中的样本风险管理数据#1 包括的层级为
20 5 层，每个层级之间的数据链路为第 1 层级（事件层级）从下往上到第 5 层级（模型层级）之间的链接链路。其中，第 1 层级的风险管理数据可以链接到第 2 层级的风险管理数据，第 2 层级的风险管理数据可以链接到第 3 层级的风险管理数据，以此类推。

在一些实施例中，电子设备 600 获得映射关系，所述映射关系包括 K_1 个层级中的
25 每个层级、以及每个层级各自对应的样本风险标签，每个层级对应的样本风险标签的数量可以是一个或多个。

例如，图 7 所示，箭头左边表示的是初始数据库中的数据层级结构，其中，有些层级对应层级风险标签，有些层级没有对应的层级风险标签。电子设备 600 可以基于初始数据库中的数据链路，清洗风险标签和层级之间的映射关系，在清洗的过程中，

可以将管理范围大的层级对应的层级风险标签叠加到管理范围小的层级，从而达到风险标签逐层细化的效果。

在一些实施例中，针对第 1 层级，将第 1 层级对应的层级风险标签，作为所述第 1 层级对应的样本风险标签。针对第 i 层级，i 依次取 2 至 K：若第 i 层级没有对应的层级风险标签，则将第 i-1 层级对应的样本风险标签，作为所述第 i 层级对应的样本风险标签；或者若第 i 层级有对应的层级风险标签，则将第 i-1 层级对应的样本风险标签和第 i 层级对应的层级风险标签，作为所述第 i 层级对应的样本风险标签。

例如，图 7 所示，电子设备 600 从箭头左边的数据结构清洗得到箭头右边的映射关系#1 的过程中，第 1 层级是事件层级，对应的层级风险标签为风险标签 1，则将风险标签 1 作为映射关系#1 中第 1 层级对应的样本风险标签。第 2 层级是策略层级，对应的层级风险标签为风险标签 2，将第 1 层级对应的样本风险标签（风险标签 1）和第 2 层级对应的层级风险标签（风险标签 2）作为映射关系#1 中第 2 层级对应的样本风险标签。第 3 层级是规则层级，对应的层级风险标签为风险标签 3，将映射关系#1 中第 2 层级对应的样本风险标签（风险标签 1、风险标签 2）以及第 3 层级对应的层级风险标签（风险标签 3）作为映射关系#1 中第 3 层级对应的样本风险标签。第 4 层级是特征层级，没有对应的层级风险标签，因此将映射关系#1 中第 3 层级对应的样本风险标签（风险标签 1、风险标签 2 和风险标签 3）作为映射关系#1 中第 4 层级对应的样本风险标签。以此类推，得到映射关系#1 中第 5 层级（模型层级）对应的样本风险标签。

可以理解，电子设备 600 获得多个所述映射关系#1 后，可以将多个所述映射关系#1 存储到初始数据库中，从而初始数据库更新为所述目标数据库。可见，从初始数据库到目标数据库的过程中，没有改变初始数据库中的层级结构，在复用初始数据库的基础上，基于清洗后得到的多个映射关系#1 更好的梳理了初始数据库中各个风险管理数据管理的风险类型。每个映射关系#1 可以对应一个事件，或者说每个映射关系#1 可以对应一条数据链路。此外，映射关系#1 中还可以包括每个层级对应的样本风险管理数据的数据标识，数据标识可以用于在目标数据库中查找所述数据标识对应的样本风险管理数据的存储位置。

电子设备 600 获得的所述映射关系包括一个或多个映射关系#1，其中，每个映射关系#1 包括的层级数量不同，所述映射关系包括的 K_1 个层级可以理解为电子设备 600 获得的所述一个或多个映射关系#1 中层级数量最大的那个映射关系#1 中包括的所有层

级。

电子设备 600 获得 M 个样本风险管理数据以及所述映射关系后，针对所述 M 个样本风险管理数据中的每个样本风险管理数据，确定所述样本风险管理数据所属的目标层级，从所述映射关系中查询所述目标层级对应的样本风险标签，并将所述目标层级对应的样本风险标签作为所述样本风险管理数据对应的样本风险标签。

以所述 M 个样本风险管理数据中的样本风险管理数据#1 为例，电子设备 600 可以从所述一个或多个映射关系#1 中确定与样本风险管理数据#1 对应的映射关系#1，所述样本风险管理数据#1 对应的映射关系#1 例如图 7 的右边所示。电子设备 600 可以在图 7 示例的映射关系#1 中确定样本风险管理数据#1 所属的目标层级，若样本风险管理数据#1 为策略层级，则电子设备 600 可以确定欺诈风险和支付风险为样本风险管理数据#1 对应的样本风险标签。所述 M 个样本风险管理数据中的其他样本风险管理数据对应的样本风险标签的确定方式以此类推，不再赘述。

至此，电子设备 600 确定了所述 M 个样本风险管理数据各自对应的样本风险标签。

方式 2

在一些实施例中，电子设备 600 获得如图 8 所示的映射关系#2，所述映射关系#2 包括目标数据库中所有样本风险管理数据（例如模型 1、模型 2、特征 1、特征 2、规则 1、规则 2、策略 1、策略 2、事件 1、事件 2 等）和风险标签的对应关系。除此之外，映射关系#2 中还可以包括每个样本风险管理数据对应的数据标识，数据标识可以用于在目标数据库中查找所述数据标识对应的样本风险管理数据的存储位置。本说明书提及的数据标识可以理解为数据索引，用于数据检索。

电子设备 600 确定所述 M 个样本风险管理数据后，以样本风险管理数据#1 为例，若样本风险管理数据#1 为策略 2，则电子设备 600 可以从图 8 所示的映射关系#2 中确定样本风险管理数据#1 对应的风险标签为恶意风险和虚假风险。以此类推，电子设备 600 确定所述 M 个样本风险管理数据中的其他样本风险管理数据对应的样本风险标签。

S220: 基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签，生成多个训练样本，其中，所述每个训练样本包括样本提示信息和样本标签信息，所述样本提示信息为基于所述 M 个样本风险管理数据中的至少部分样本风险管理数据得到的与风险相关的自然语言描述，所述样本标签信息包括与所述至少部分样本风险管理数据对应的样本风险标签。

可以理解，所述样本提示信息至少包括基于所述 M 个样本风险管理数据中的至少部分样本风险管理数据得到的样本风险描述文本，样本风险描述文本可以理解为以自然语言对所述部分样本风险管理数据存在、管理或者相关的风险进行描述的文本信息。电子设备 600 基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签生成 $M1$ 个训练样本， $M1$ 大于或等于 M 。样本提示信息的具体内容结构和形式与目标提示信息类似，样本风险标签和目标风险标签的类型和形式类似，不再赘述。

在一些实施例中，服务器 500 基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签，生成多个数据标签对，所述多个数据标签对中的部分数据标签对中的每个数据标签对中包括：一个样本风险标签、以及所述样本风险标签对应的样本风险管理数据，或者，所述多个数据标签对中的另一部分数据标签对中的每个数据标签对中包括多个样本风险标签、以及所述多个样本风险标签各自对应的样本风险管理数据。

例如，电子设备 600 基于图 8 示例的映射关系#2 确定支付风险对应的风险管理数据为模型 2、特征 1、策略 1 和事件 1，则基于支付风险、模型 2、特征 1、策略 1 和事件 1 生成如图 9 所示的数据标签对#1。

再例如，电子设备 600 基于图 8 所示的映射关系#2 确定支付、恶意、欺诈、盗卡、盗账户和虚假等风险标签对应的风险管理数据为模型 1、特征 1、规则 1、策略 1、事件 1、模型 2、特征 2、规则 2、策略 2 和事件 2，则基于支付、恶意、欺诈、盗卡、盗账户和虚假等风险标签和所述风险标签对应的所有风险管理数据生成如图 10 所示的数据标签对#2。

以此类推，电子设备 600 可以生成多个数据标签对#1 和多个数据标签对#2，并基于所述多个数据标签对#1 和多个数据标签对#2 生成所述 $M1$ 个训练样本。

在一些实施例中，针对所述多个数据标签对中的每个数据标签对，电子设备 600 获得所述数据标签对中的各个样本风险管理数据对应的数据描述信息，所述数据描述信息包括样本风险管理数据的数据名称和/或数据描述。电子设备 600 基于所述数据描述信息生成一个训练样本的所述样本提示信息，所述样本提示信息包括样本风险描述文本。除此之外，所述样本提示信息还可以在样本风险描述文本的基础上增加前缀以限制目标语言模型的知识领域，前缀例如角色限定语和/或任务限定语，角色限定语、任务限定语的具体内容可参考前文的相关描述，不再赘述。

例如，电子设备 600 基于图 11 右边所示的映射关系获得数据标签对#1 中风险标签对应的所有样本风险管理数据，并在目标数据库中查找所述所有样本风险管理数据对应的数据名称和数据描述，基于所述数据名称和数据描述生成数据标签对#1 对应的训练样本#1 的样本提示信息#1。以此类推，电子设备 600 获得数据标签对#2 中每个样本风险管理数据对应的数据名称和数据描述，基于所述数据名称和数据描述生成数据标签对#2 对应的训练样本#2 的样本提示信息#2。

电子设备 600 基于所述数据标签对#1 中的样本风险标签#1，生成所述训练样本#1 的样本标签信息#1，基于样本提示信息#1 和所述样本标签信息#1 生成训练样本#1。

训练样本#1 例如：

```
10      {  
        “prompt”：“你是风控领域的策略专家，可以将以下风险描述文本概括为风险标签，只给出结论，不用解释\营销事件中，欺诈行为可能出现在各个环节，例如虚假宣传、虚假承诺、提供虚假优惠等手段，以诱导用户购买产品或服务”  
        “summary”：“盗卡”  
15      }
```

其中，训练样本#1 的数据结构例如为 json 结构，prompt 为样本提示信息#1，prompt 包含风险描述文本信息、角色限定语和任务限定语。Summary 为样本标签信息#1，包含 1 个样本风险标签：“盗卡”。电子设备 600 基于所述数据标签对#2 中的样本风险标签#2，生成所述训练样本#2 的样本标签信息#2，基于样本提示信息#2 和所述样本标签信息#2 生成训练样本#2。

训练样本#2 例如：

```
      {  
        “prompt”：“你是风控领域的策略专家，可以将以下风险描述文本概括为风险标签，只给出结论，不用解释\营销事件中，欺诈行为可能出现在各个环节，例如虚假宣传、虚假承诺、提供虚假优惠等手段，以诱导用户购买产品或服务，该场景下也会有个人信息泄露的风险，包括银行卡号、密码、个人姓名和住址等，这些个人信可能被盗用者用于其他非法活动，比如身份盗窃、欺诈或其他犯罪行为”  
25      “summary”：“盗卡#盗账户#异常行为#欺诈”  
      }
```

其中，训练样本#2 的数据结构例如为 json 结构，prompt 为样本提示信息#2，prompt 包含风险描述文本信息、角色限定语和任务限定语。Summary 为样本标签信息#2，包含 4 个样本风险标签：“盗卡”、“盗账户”、“异常行为”和“欺诈”。

可以理解，电子设备 600 可以通过上述方式生成类似训练样本#1 和训练样本#2 的所述 M1 个训练样本。在具体生成过程中，可以先生成多个训练样本中每个训练样本中的样本标签信息和样本提示信息中的风险描述文本信息，再为所述每个训练样本增加角色限定语和任务限定语等固定内容。电子设备 600 可以预先设置训练样本的固定结构，电子设备 600 在生成训练样本的过程中可以在所述固定结构的对应位置填充对应内容，以生成训练样本。

在上述实施例中，电子设备 600 可以通过增加训练样本的数量和多样性来提高目标语言模型的泛化能力，以增强目标语言模型对复杂风险的理解。即电子设备 600 可以构造一些由多类服务应用场景下的风险描述文本信息和多个样本标签信息组成的训练样本，从而增强目标语言模型的训练样本。

S230：基于所述多个训练样本训练得到目标语言模型。

在一些实施例中，服务器 500 获得预训练语言模型，预训练语言模型例如 ChatGLM-6B。

电子设备 600 对所述预训练语言模型进行调整，得到待训练语言模型。

例如，电子设备 600 对所述预训练语言模型进行 P-Tuning v2 微调，即对所述预训练语言模型进行参数剪枝，得到一个更加小巧、服务效率更高的轻量级模型。

具体地，P-Tuning v2 使用自适应的剪枝策略，对所述预训练语言模型中的参数进行裁剪，去除其中不必要的冗余参数。然后，对于被剪枝的参数，P-Tuning v2 使用特定的压缩方法，能够有效压缩参数，减少所述预训练语言模型微调的总参数量。

可以理解，在通过 P-Tuning v2 对所述预训练语言模型微调得到的所述待训练语言模型，更加轻便、高效，同时在微调过程中尽可能地让所述待训练语言模型拥有所述预训练语言模型的性能。这样既可以加快所述待训练语言模型的训练和推理速度，也可以在待训练语言模型被训练更新为所述目标语言模型后，在使用目标语言模型的过程中减少所述目标语言模型对电子设备 600 的内存和计算资源的消耗，使得所述目标语言模型能够适用于各种实际的配置场景中。

电子设备 600 得到待训练语言模型后，通过所述多个训练样本对所述待训练语言模型进行训练，得到所述目标语言模型。

例如，在训练前，训练人员可以根据电子设备 600 得到的所述多个训练样本的实际情况设置训练样本的最大输入长度（max_source_length）、最大输出长度（max_target_length）、样本提示信息的长度（pre_seq_len）和学习率（learning rate，LR）等参数，学习率还可以称为步长。

在具体的训练过程中，训练人员可以将所述多个训练样本分为训练集和验证集，并将所述训练集中的所有训练样本分为 n 个 epoch（轮数）， n 大于或等于 1。一个 epoch 包括 n_1 个 batch（批次），一个 batch 包括的训练样本的数量等于一个 epoch 包括的全部训练样本的数量 n_2 与 n_1 的比值。依次将每个 epoch 包括的 n_2 个训练样本分 batch 输入所述待训练语言模型进行训练。在一个 epoch 训练过程中，每个 batch 训练所述待训练语言模型后，计算损失函数的梯度，基于计算得到的梯度更新所述待训练语言模型的参数。所述损失函数为表示待训练语言模型输出的风险标签和训练样本中的样本标签信息之间差值的函数。

在一个 epoch 训练完成后，将验证集包括的部分训练样本输入在 batch 训练过程中更新了参数的所述待训练语言模型中，根据输出结果评估更新后的所述待训练语言模型的性能指标，根据所述性能指标的评估结果调整学习率、所述待训练语言模型的结构等。若性能指标的评估结果不理想，则进入下个 epoch 训练。

按照上述流程，循环多个 epoch 训练，直至基于验证集得到的性能指标达到预期或者所述待训练语言模型收敛时，停止继续训练，基于最后一个 epoch 训练更新得到的所述待训练语言模型即所述目标语言模型。

可以理解，所述目标语言模型在管理设备 400 实际的使用过程中，训练设备 500 可以根据目标语言模型的使用反馈再次训练更新所述目标语言模型。例如，针对实际任务，训练设备 500 生成一个测试集，测试集包括与上文提及的训练样本类似的多个测试样本，并将多个测试样本输入所述目标语言模型，计算所述目标语言模型的准确率、精确率、召回率等指标，进而评估目标语言模型在实际任务中的性能，从而决定是否要继续训练更新所述目标语言模型。如果决定要继续训练更新所述目标语言模型，则训练设备 500 按照上文描述的方式生成多个训练样本，按照上文描述的训练方法再次训练更新所述目标语言模型，将更新后的所述目标语言模型发布到云端，以便管理

设备 400 调用。

综上所述，本说明书提供的风险防控方案的生成方法、目标语言模型的训练方法、设备及存储介质，在配置人员为目标事件配置风险防控方案的过程中，推荐设备能够调用预先训练好的能够基于自然语言描述的文本信息挖掘出风险标签的目标语言模型，

5 从而通过目标提示信息指导目标语言模型生成目标事件对应的一个或多个风险标签，进而推荐设备基于所述一个或多个风险标签生成目标事件的风险防控方案，将其推荐给配置人员。也就是说，本方案在配置过程中，调用了预先训练好的目标语言模型，避免配置人员由于专业经验的限制，挖掘不到能够全面、准确的概括目标事件的风险标签，从而不能基于准确的风险标签为目标事件查找到合适的风险防控方案，影响后

10 续风险防控平台识别风险事件的准确性。并且，配置人员在配置过程中，有目标语言模型的协助，可以降低配置人员配置的上手成本，即在配置人员的专业能力一般的情况下，也能基于推荐为目标事件配置到较为合适的风险防控方案。例如，对于目标平台新扩展的服务类型涉及的新型事件，配置人员也可以在目标语言模型的协助下，挖掘到新型事件的风险标签，从而找到较为合适新型事件的风险防控方案。

15 需要说明的是，上述描述方法 P100、方法 P200 的过程中，以金融领域的应用场景举例，因此具体描述中采用的是该交易领域中会涉及的术语，例如提示信息的举例、风险标签的举例，风险管理数据的举例等。如果将方法 P100、方法 P200 应用到前文提及的其他领域，所涉及的术语可以适应性的调整，此处不一一枚举。

本说明书另一方面提供一种非暂时性存储介质，存储有至少一组用来推荐风险防

20 控方案、或者训练目标语言模型的可执行指令。当所述可执行指令被处理器执行时，所述可执行指令指导所述处理器实施本说明书所述的方法 P100 或 P200 的步骤。在一些可能的实施方式中，本说明书的各个方面还可以实现为一种程序产品的形式，其包括程序代码。当所述程序产品在电子设备 600 上运行时，所述程序代码用于使电子设备 600 执行本说明书描述的方法 P100 或 P200 的步骤。用于实现上述方法的程序产品

25 可以采用便携式紧凑盘只读存储器（CD-ROM）包括程序代码，并可以在电子设备 600 上运行。然而，本说明书的程序产品不限于此，在本说明书中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统使用或者与其结合使用。所述程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、

30 或半导体的系统、装置或器件，或者任意以上的组合。可读存储介质的更具体的例子

包括：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。

所述计算机可读存储介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读存储介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。可读存储介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF 等等，或者上述的任意合适的组合。

可以以一种或多种程序设计语言的任意组合来编写用于执行本说明书操作的程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在电子设备 600 上执行、部分地在电子设备 600 上执行、作为一个独立的软件包执行、部分在电子设备 600 上、部分在远程计算设备上执行、或者完全在远程计算设备上执行。

上述对本说明书特定实施例进行了描述。其他实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者是可能有利的。

综上所述，在阅读本详细公开内容之后，本领域技术人员可以明白，前述详细公开内容可以仅以示例的方式呈现，并且可以不是限制性的。尽管这里没有明确说明，本领域技术人员可以理解本说明书需求囊括对实施例的各种合理改变，改进和修改。这些改变，改进和修改旨在由本说明书提出，并且在本说明书的示例性实施例的精神和范围内。

权 利 要 求 书

1. 一种风险防控方案的生成方法，包括：

响应于检测到请求配置目标事件对应的风险防控方案，获得所述目标事件对应的
5 目标提示信息，所述目标提示信息至少包括以自然语言对所述目标事件存在的风险进行描述的文本信息；

通过所述目标提示信息指导目标语言模型生成所述目标事件对应的 N 个目标风险标签，所述 N 为大于或等于 1 的整数，所述目标语言模型被训练为从自然语言描述的文本信息中挖掘风险标签；以及

10 基于所述 N 个目标风险标签生成所述风险防控方案。

2. 根据权利要求 1 所述的方法，其中，所述基于所述 N 个目标风险标签生成所述风险防控方案，包括：

针对所述 N 个目标风险标签中的每个目标风险标签，在目标数据库中查找得到所
15 述目标风险标签对应的风险管理数据；以及

基于所述 N 个目标风险标签各自对应的风险管理数据生成所述风险防控方案。

3. 根据权利要求 2 所述的方法，其中，所述风险管理数据包括 K 个层级的数据，所述 K 为大于 1 的整数，不同层级的数据对风险的管理范围不同。

20

4. 根据权利要求 2 所述的方法，其中，所述基于所述 N 个目标风险标签各自对应的风险管理数据生成所述风险防控方案，包括：

针对所述 N 个目标风险标签中的每个目标风险标签，基于所述目标风险标签对应的风险管理数据生成所述目标风险标签对应的子防控方案；以及

25 基于所述 N 个目标风险标签各自对应的子防控方案生成所述风险防控方案。

5. 根据权利要求 1 所述的方法，其中，所述目标提示信息还包括：

角色限定语，用于限定所述目标语言模型所充当的角色；以及

任务限定语，用于限定所述目标语言模型需要完成的任务。

30

6. 根据权利要求 1 所述的方法，其中，所述方法还包括：

在配置界面中显示所述风险防控方案，并在所述配置界面中显示操作控件，所述操作控件被配置为：在检测到所述操作控件被触发时，将所述风险防控方案配置到所述目标事件。

5

7. 一种目标语言模型的训练方法，包括：

获得 M 个样本风险管理数据，并确定所述 M 个样本风险管理数据各自对应的样本风险标签，所述 M 为大于 1 的整数；

10 基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签，生成多个训练样本，其中，所述每个训练样本包括样本提示信息和样本标签信息，所述样本提示信息为基于所述 M 个样本风险管理数据中的至少部分样本风险管理数据得到的与风险相关的自然语言描述，所述样本标签信息包括与所述至少部分样本风险管理数据对应的样本风险标签；以及

基于所述多个训练样本训练得到所述目标语言模型。

15

8. 根据权利要求 7 所述的方法，其中，所述 M 个样本风险管理数据按照管理范围由大到小的顺序被划分为 K_1 个层级，所述 K_1 为大于 1 的整数，不同层级的数据对风险的管理范围不同；以及

所述确定所述 M 个样本风险管理数据各自对应的样本风险标签，包括：

20 获得映射关系，所述映射关系包括所述 K_1 个层级中的每个层级、以及所述每个层级各自对应的样本风险标签，以及

针对所述 M 个样本风险管理数据中的每个样本风险管理数据，确定所述样本风险管理数据所属的目标层级，从所述映射关系中查询所述目标层级对应的样本风险标签，并将所述目标层级对应的样本风险标签作为所述样本风险管理数据对应的样本风险标签。

25

9. 根据权利要求 8 所述的方法，其中，所述 K_1 个层级中的至少部分层级对应有层级风险标签，所述至少部分层级包含第 1 层级；所述获得映射关系，包括：

针对第 1 层级，将所述第 1 层级对应的层级风险标签，作为所述第 1 层级对应的样本风险标签，以及

30

针对第 i 层级，所述 i 依次取 2 至 K_1 ：

若所述第 i 层级没有对应的层级风险标签，则将所述第 $i-1$ 层级对应的样本风险标签，作为所述第 i 层级对应的样本风险标签，或者

5 若所述第 i 层级有对应的层级风险标签，则将所述第 i 层级对应的层级风险标签和第 $i-1$ 层级对应的样本风险标签，作为所述第 i 层级对应的样本风险标签。

10. 根据权利要求 7 所述的方法，其中，所述基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签，生成多个训练样本，包括：

10 基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自对应的样本风险标签，生成多个数据标签对，每个数据标签对中包括：

一个样本风险标签、以及所述样本风险标签对应的样本风险管理数据，或者

15 多个样本风险标签、以及所述多个样本风险标签各自对应的样本风险管理数据；以及

基于所述多个数据标签对生成所述多个训练样本。

11. 根据权利要求 10 所述的方法，其中，所述基于所述多个数据标签对生成所述多个训练样本，包括：

20 针对所述多个数据标签对中的每个数据标签对，

获得所述数据标签对中的各个样本风险管理数据对应的数据描述信息；

基于所述数据描述信息生成一个训练样本的所述样本提示信息；

基于所述数据标签对中的样本风险标签，生成所述训练样本的所述样本标签信息；以及

25 基于所述样本提示信息和所述样本标签信息生成所述训练样本。

12. 根据权利要求 7 所述的方法，其中，所述基于所述多个训练样本训练得到所述目标语言模型，包括：

获得预训练语言模型；

30 对所述预训练语言模型进行调整，得到待训练语言模型；以及

利用所述多个训练样本对所述待训练语言模型进行训练，得到所述目标语言模型。

13.一种电子设备，包括：

至少一个存储介质，存储有至少一个指令集，用于进行风险防控方案的生成；以

5 及

至少一个处理器，同所述至少一个存储介质通信连接，其中，所述至少一个处理器运行时读取所述至少一个指令集，并且根据所述至少一个指令集的指示：

响应于检测到请求配置目标事件对应的风险防控方案，获得所述目标事件对应的目标提示信息，所述目标提示信息至少包括以自然语言对所述目标事件存在的风险进行描述的文本信息，

10

通过所述目标提示信息指导目标语言模型生成所述目标事件对应的 N 个目标风险标签，所述 N 为大于或等于 1 的整数，所述目标语言模型被训练为从自然语言描述的文本信息中挖掘风险标签，以及

基于所述 N 个目标风险标签生成所述风险防控方案。

15

14. 根据权利要求 13 所述的电子设备，其中，为了生成所述风险防控方案，所述至少一个处理器：

针对所述 N 个目标风险标签中的每个目标风险标签，

在目标数据库中查找得到所述目标风险标签对应的风险管理数据；以及

20

基于所述 N 个目标风险标签各自对应的风险管理数据生成所述风险防控方案。

15. 根据权利要求 14 所述的电子设备，其中，所述风险管理数据包括 K 个层级的数据，所述 K 为大于 1 的整数，不同层级的数据对风险的管理范围不同。

25

16. 根据权利要求 14 所述的电子设备，其中，为了生成所述风险防控方案，所述至少一个处理器：

针对所述 N 个目标风险标签中的每个目标风险标签，

基于所述目标风险标签对应的风险管理数据生成所述目标风险标签对应的子防控方案；以及

30

基于所述 N 个目标风险标签各自对应的子防控方案生成所述风险防控方案。

17. 根据权利要求 13 所述的电子设备，其中，所述目标提示信息还包括：
角色限定语，用于限定所述目标语言模型所充当的角色；以及
任务限定语，用于限定所述目标语言模型需要完成的任务。

5

18. 根据权利要求 13 所述的电子设备，其中，所述至少一个处理器：

在配置界面中显示所述风险防控方案，并在所述配置界面中显示操作控件，所述
操作控件被配置为：在检测到所述操作控件被触发时，将所述风险防控方案配置到所
述目标事件。

10

19. 一种电子设备，包括：

至少一个存储介质，存储有至少一个指令集，用于进行目标语言模型的训练；以
及

至少一个处理器，同所述至少一个存储介质通信连接，其中，所述至少一个处理
器运行时读取所述至少一个指令集，并且根据所述至少一个指令集的指示：

15

获得 M 个样本风险管理数据，并确定所述 M 个样本风险管理数据各自对
应的样本风险标签，所述 M 为大于 1 的整数；

基于所述 M 个样本风险管理数据、以及所述 M 个样本风险管理数据各自
对应的样本风险标签，生成多个训练样本，其中，所述每个训练样本包括样本
提示信息和样本标签信息，所述样本提示信息为基于所述 M 个样本风险管理
数据中的至少部分样本风险管理数据得到的与风险相关的自然语言描述，所述
样本标签信息包括与所述至少部分样本风险管理数据对应的样本风险标签；以
及

20

基于所述多个训练样本训练得到所述目标语言模型。

25

20. 一种计算机可读非暂时性存储介质，其中，所述计算机可读非暂时性存储介
质中存储有至少一个指令集，所述至少一个指令集被至少一个处理器执行时，实现如权
利要求 1-12 中任一项所述的方法。

30