说　　明　　书

对话中的过度承诺检测方法及装置

技术领域

本说明书一个或多个实施例涉及自然语言处理技术领域，尤其涉及对话中的过度承诺检测方法及装置。

背景技术

随着科技和社会的发展，各种服务平台层出不穷。服务平台能够为人们提供越来越丰富的业务服务，以提升人们的生产和生活水平。其中，客户服务是业务售后环节必不可少的重要一环。客户服务常常会通过人工，以热线电话服务、热线邮件服务、在线语聊等方式为客户解答问题。人工客服在解答客户疑问时，可能会因为业务不熟练，或者口误、笔误等情况而出现过度承诺的问题。例如，某项退款需要收取 2%的手续费，而客服说成了1%。对客户服务内容进行质检，是提高客户服务水平的重要措施。不管使用哪种检测方式，都需要对客户的隐私数据进行有效保护。

通常，服务对话的内容量巨大。因此，希望能有改进的方案，可以对客服对话进行更加高效的过度承诺检测。

发明内容

本说明书一个或多个实施例描述了一种对话中的过度承诺检测方法及装置，可以对客服对话进行更加高效的过度承诺检测。具体的技术方案如下。

第一方面，实施例提供了一种对话中的过度承诺检测方法，包括：获取待检测的对话文本；利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；针对任意一个包含第一问题和第一答案的问答对，从知识库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；利用所述第二答案对所述第一答案进行事实性验证；当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

第二方面，实施例提供了一种计算机系统，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述程序时实现如下步骤：获取待检测的对话文本；利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；针对任意一个包含第一问题和第一答案的问答对，从知识

库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；利用所述第二答案对所述第一答案进行事实性验证；当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

第三方面，实施例提供了一种对话中的过度承诺检测装置，包括：获取模块，配置为获取待检测的对话文本；提取模块，配置为利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；匹配模块，配置为针对任意一个包含第一问题和第一答案的问答对，从知识库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；验证模块，配置为利用所述第二答案对所述第一答案进行事实性验证；确定模块，配置为当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

第四方面，实施例提供了一种计算机可读非易失性存储介质，所述存储介质存储有计算机程序，所述计算机程序被处理器执行时实现上述第一方面中任一项所述的方法。

本说明书实施例提供的方法及装置中，通过对待检测的对话文本进行问题和答案的提取，得到对话文本中的问答对；接着，从知识库中匹配得到与问答对中的问题对应的标准答案，并利用标准答案与问答对中的答案进行事实性验证，当验证两者不一致时，确定问答对中存在疑似的过度承诺问题，后续可以提交人工审核。当存在大量的客服对话时，通过上述方式可以对客户对话进行预侦查，快速锁定疑似目标，因此本说明书实施例能够对客服对话进行更加高效的过度承诺检测。

附图说明

为了更清楚地说明本发明实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单的介绍。显而易见地，下面描述中的附图仅仅是本发明的一些实施例，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 为本说明书披露的一个实施例的实施场景示意图；

图 2 为实施例提供的一种对话中的过度承诺检测方法的流程示意图；

图 3 为利用 RAG 算法确定第二答案 A1′ 的一种流程示意图；

图 4 为实施例提供的一种对话中的过度承诺检测装置的示意性框图；

图 5 为实施例提供的一种计算机系统的结构示意图。

具体实施方式

下面结合附图，对本说明书提供的方案进行描述。

图 1 为本说明书披露的一个实施例的实施场景示意图。其中包括通过热线电话服务或在线语聊服务获取到服务对话内容，对服务对话内容进行基于大模型的 QA（问题 Q、答案 A）对儿抽取。接着，采用例如 RAG（Retrieval Augmented Generation）技术，从业务规则库中得到问题 Q 的标准答案 A′，并将 QA 对儿和 A′输入大模型，基于大模型对 A 和 A′进行事实性验证。判断验证结果是否有风险，当有风险时转人工复合，当没有风险时表示核验通过。

其中，大模型是指大语言模型，即 LLM（Large Language Model），是深度学习的应用之一，尤其在自然语言处理（NLP）领域。为了实现这个目标，大语言模型需要在大量文本数据上进行训练，以学习语言的各种模式和结构。如 OpenAI ChatGPT 和 PaLM，就是大语言模型的例子。大语言模型被训练来解决常见的语言问题，如文本分类、问答、文档总结和文本生成等。在大语言模型的上下文中，“大”主要有两层含义。一方面，它指的是模型的参数数量。大语言模型的参数数量通常会非常大，达到数十亿甚至数百亿。这使得大语言模型能够学习和表示非常复杂的模式。另一方面，“大”也指的是训练数据的规模。大语言模型通常在大规模的文本数据上进行训练，这种数据可以来自互联网、书籍、新闻等各种来源。大语言模型在训练时使用了来自各种应用领域的数据，其蕴含了大量的知识，因此它们能够处理与各种应用领域相关的文本分类、问答、文档总结和文本生成等任务。这使得大语言模型具有更强的语言处理能力和更广泛的适用性。

对问答对中的答案进行的事实性验证，主要是对答案进行过度承诺检测。在服务平台向客户提供的客户服务中，服务人员因为不清楚业务规则或者口误、笔误等各种原因，可能会对客户造成过度承诺，即客服对客户问题的回答与业务规则实际上不符。服务质检是通过人工或者 AI 算法手段对商户的客户服务质量进行检查与评估的任务。服务质检能够发现和改进服务中的不足与问题，从而不断提升服务质量。例如，本来某跨境汇款从中国到美国的时间是 1 周（即 7 天），而客服人员因为不清楚规则或者口误说成了 3 天；又如，某项退款需要收取 2%的手续费，而客服说成了 1%。这种情况如果在服务质检中不及时发现并纠正，则很可能会引发客诉甚至舆情风险。但是，过度承诺这种现象不易发现，因为其承诺的内容不是固定的，这对质检算法的要求很高。

客户服务是指服务平台（例如某个公司）通过多种方式对客户提供（售前）咨询类或者售后（事务类）服务。热线服务中的双方，即客户和客服，通过电话、文字等对话的方式进行沟通。客户服务通常可以以热线电话服务、热线邮件服务、在线语聊服务等方式存在。在线语聊服务可以通过特定的 messager 工具以文字对话方式进行交互。热线电话服务

从客户便捷性与响应及时性的角度都较其他两种方式有明显优势，在实际场景中被普遍采用。

热线电话这种售后服务方式虽然在交互上最为便捷，但是却对服务后的质检环节提出了挑战。因为电话对话中的所有信息都以语音媒体的形式存在，质检员如果想检查一段服务对话是否有问题，起码需要亲自听完服务过程中的全部录音，且过程中容易受到客户口音和语速等因素影响而导致判断失误。这是一个效率低下且容易出错的过程。而且，从质检发现问题的角度，我们最感兴趣的是那些存在问题的对话，而正常情况下这些存在问题的对话只占总量的很小部分。如果为了捕捉这些小概率 case 而耗费大量人力去听全部服务对话的话，成本会非常高。并且，客服对话每天都在发生，数据量巨大。

过度承诺检测所处理的对话不限于客户服务对话。在其他的对话场景中也可能存在过度承诺检测的需求。这样的场景例如包括：个人之间的常规对话、师生之间的答疑对话等场景。

为了提高对对话进行检测时的效率和可行性，本说明书实施例提供了一种针对对话的过度承诺检测方法，主要包含以下阶段：对对话文本中的问答对进行抽取，以及对问答对的内容进行事实性验证等。该方法具体包括以下步骤：步骤 S210，获取待检测的对话文本；步骤 S220，利用对自然语言中信息的提取技术，从对话文本中提取问题和对应的答案，得到若干问答对；步骤 S230，针对任意一个包含第一问题和第一答案的问答对，从包含多个预设的问答对的知识库中，确定与所述第一问题匹配的第二答案；步骤 S240，利用第二答案对第一答案进行事实性验证；步骤 S250，当第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

在本实施例中，在客户服务场景中，针对过度承诺这一对客户服务质检痛点与难点问题，采用自然语言中信息的提取和本文之间的事实性验证等技术，从客服对话中检测过度承诺现象的发生，有效解决了人工质检所面临的成本高、效率低、容易漏检的问题，也克服了传统机器学习算法缺乏靶向性训练样本和开发成本高的问题。在其他应用场景中，本实施例通过自然语言中信息的提取和本文之间的事实性验证等技术，能够有效地从对话文本中检测出有问题的过度承诺，从而将这些问题反馈给对话的当事人，避免了人工检测，提高了检测效率。下文对实施例进行说明时，以客户服务场景为例进行说明。

下面结合图 2 对上述实施例的过度承诺检测方法进行详细说明。图 2 为实施例提供的一种对话中的过度承诺检测方法的流程示意图。该方法可以通过计算设备执行，计算设备可以通过任何具有计算、处理能力的装置、设备、平台、设备集群等来实现。

在步骤 S210 中，获取待检测的对话文本。

承前所述，客服对话可以来自于热线电话服务、热线邮件服务、在线语聊服务等，因此客服对话可以是文本数据，也可以是语音数据。当客服对话是语音数据时，例如在热线电话服务中，针对通过热线电话获取到的语音数据，可以采用 ASR（Acoustic Signal Recognition）等技术将该语音数据转换成文本数据。在热线邮件服务和在线语聊服务等方式中，计算设备可以直接获取到客服与客户之间的服务对话的文本数据。上述文本数据可以作为待检测的对话文本，即作为待检测过度承诺的对象数据。图 1 中的服务对话内容通常采用文本数据的形式。

在步骤 S220 中，利用对自然语言中信息的提取技术，从对话文本中提取问题和对应的答案，得到若干问答对。若干包括一个或多个。在任意一个问答对中，通过包括一个问题和一个对应的答案。若干问答对可以以文本形式存储。表 1 是两个问答对的例子。

表 1

1	问题	使用 xxx 跨境汇款 APP 汇款到美国的时间最快是几天
	答案	3 天
2	问题	使用 xx 的交易退款需求收手续费吗，比例是多少
	答案	使用 xx 的交易退款需收取 1%的手续费

表 1 中最左侧一列的数字为问答对的标号。

对话文本中的内容可能非常多，常常不是简洁的问答对的内容，会包含很多繁杂的其他内容。从繁杂的内容中提取出问题和对应的答案，是后续步骤实施的基础。需要说明的是，问答对中的问题可以是用户问题，也可以是用户诉求。例如，一个用户诉求可以是这样的形式：请告诉我某退款手续费的比例。

对话文本属于自然语言。为了从中提取到问答对，可以利用针对自然语言中信息的提取技术实现。

在一种实现方式中，可以利用预训练后的大语言模型从对话文本中提取问答对。具体在实现时，可以基于对话文本构建提示模板，为了方便描述，将该提示模板称为第一提示模板 prompt1；接着，将第一提示模板 prompt1 输入大语言模型，通过大语言模型从对话文本中提取问题和对应的答案。

其中，第一提示模板 prompt1 的内容包括上述对话文本，第一提示模板 prompt1 用于指示从对话文本中提取问题和对应的答案。大语言模型用于从对话文本中提取问题和对应的答案。大语言模型的输入包括第一提示模板 prompt1 的内容，输出包括若干问答对。

提示模板能够帮助预训练后的大语言模型“回忆”起预训练时学习到的知识，引导大语言模型进行微调，以便引导大语言模型以被希望的方式做出回答，从而更好地适应业务需求。

在本实现方式中，可以按照预设的提示模板格式，基于对话文本生成第一提示模板 prompt1。例如，第一提示模板 prompt1 可以包含第一指示部分和目标文本部分；

在构建第一提示模板 prompt1 时，可以将对话文本填充至目标文本部分，将预设的第一指示文本填充至第一指示部分。其中，第一指示文本用于指示大语言模型从目标文本部分的内容中提取问答和对应的答案。

为了使得第一提示模板 prompt1 的功能能加全面，进而使得大语言模型更有效地提取到问答对，第一提示模板 prompt1 还可以包含以下部分中的至少一种：

角色定义部分，用于定义大语言模型担任的角色；

输出格式部分，用于定义大语言模型输出问题和对应的答案时的格式；

问答对示例部分，用于向大语言模型展示问题和对应答案的例子。

对于具体的问答对提取任务，可以将预先设定的文本内容填充到上述对应的各个部分中。例如，在客户服务进行质检的场景中，可以设定以下提示模板，见表 2。

表 2

角色定义部分	你是一个服务质检人员
目标文本部分	以下是一段服务对话的内容记录： 客户：xxxxx 客服：xxxxx ……
第一指示部分	请你寻找并总结上述目标文本部分中的客户问题与对应的客服回答的 QA 对儿，QA 对儿的内容要简单精要
输出格式部分	将找出的 QA 对儿结果按照如下格式输出： 上面目标文本部分中存在___个 QA 对儿，QA 对儿内容列举如下，如没有发现 QA 对儿则不列举： Q1： A1： Q2： A2：

	……
问答对示例部分	以下是一些 QA 对的例子： 例子 1: 客户诉求（Q）：使用 xxx 跨境汇款 APP 汇款到美国的时间最快是几天 客服回答（A）： 3 天 例子 2： 客户诉求（Q）：使用 xx 的交易退款需求收手续费吗，比例是多少 客服回答（A）： 使用 xx 的交易退款需收取 1%的手续费

在表 2 中，左侧一列是提示模板中的各个部分，各部分之间的顺序不限于是表 2 中的顺序，可以根据需要调整。右侧一列给出了各个部分的具体内容，部分内容采用“x”或者省略号代替。针对不同的任务，可以相应地调整各部分的具体内容。

第一提示模板 prompt1 可以采用大模型 few-shot prompt engineering 来实现针对客服对话内容的 QA 对儿抽取的过程。Prompt Engineering，即采用提示词的方式与大语言模型进行交互，从而调用其能力来完成任务。不同目标的任务需要针对任务特点设计不同的 prompt 来完成。

步骤 S220 在实现时并不限于使用大语言模型，也可以使用其他的自然语言的信息提取技术。在一种实现方式中，可以对不同的信息抽取目标进行统一编码，自适应地生成目标结构。信息抽取模板包括实体、关系和事件等。例如，可以通过已有的 UIE（Universal Information Extraction）模型提取对话文本中的结构化信息，从而得到问答对。

在步骤 S230 中，针对任意一个包含第一问题 Q1 和第一答案 A1 的问答对，从知识库中确定与第一问题 Q1 匹配的第二答案 A1′。该知识库包含多个预设的问答对，知识库中的问答对可以是基于对应的业务规则生成的。基于业务规则生成问答对可以采用已有的技术实现，此处不再赘述。第二答案 A1′是知识库中针对第一问题 Q1 的标准答案。第一答案 A1 是客服针对第一问题 Q1 给出的答案。

在一种实现方式中，本步骤可以采用向量匹配的方式实现。首先可以针对知识库中的多个问题文本，确定对应的多个文本特征向量，将多个文本特征向量存储在向量数据库中。向量数据库中的多个文本特征向量分别是知识库中的多个问题的文本特征向量。

在从知识库中确定第二答案 A1′时，可以确定第一问题 Q1 对应的第一文本特征向量，将第一文本特征向量与向量数据库中的多个文本特征向量进行匹配，基于匹配成功的文本

特征向量对应的知识库中的答案，确定第二答案 A1′。

具体的，可以采用已有的文本编码模型对第一问题 Q1 的文本进行编码，得到第一文本特征向量，以及采用已有的文本编码模型对知识库中的多个问题文本进行编码，得到对应的文本特征向量。将第一文本特征向量与向量数据库中的多个文本特征向量进行匹配时，可以采用已有的向量相似度匹配算法进行。在匹配后可以得到多组向量之间的相似度，可以确定相似度最大的文本特征向量对应的知识库中的问答对，将该问答对中的答案作为第二答案 A1′。

在一种实现方式中，步骤 S230 可以利用检索增强生成算法 RAG（Retrieval Augmented Generation），从知识库中确定与第一问题 Q1 匹配的第二答案 A1′。

图 3 为利用 RAG 算法确定第二答案 A1′的一种流程示意图。其中包括知识库文本处理阶段、检索阶段和答案生成阶段。大模型可以通过包括通用语料和领域语料在内的训练数据进行微调。在知识库文本处理阶段，可以采用文本编码模型对知识库中的多个问题文本进行编码，得到对应的文本特征向量，并将得到的文本特征向量存储在向量数据库中。对向量数据库构建倒排序索引和/向量索引。

在检索阶段，将用户问题（即第一问题 Q1）输入主入口，对其进行结构化信息的提取，其中包括意图理解和语义化处理，得到第一文本特征向量。接着，进行基于关键词检索、向量检索和结构化检索的检索过程。该过程即是将用户问题的第一文本特征向量与向量数据库中的向量进行匹配的过程。在匹配时，可以按照文本相关性、语义相关性、类目相关性和静态分，对检索到的内容进行排序，得到包含 top N 个待选问题的检索结果。基于该检索结果和用户问题构建提示模板，prompt 的具体内容包含该检索结果、用户问题以及指示大模型找到最匹配的待选问题的指示。将该提示模板输入大语言模型，通过大语言模型从待选问题中确定最准确的待选问题作为与用户问题匹配的知识库问题。知识库中该最准确的待选问题对应的答案即作为第二答案 A1′，将答案返回给主入口。

RAG 是一种利用大模型补充外部知识，从而提升大模型知识容量与知识真实性的技术。RAG 将检索技术和大模型结合起来，一方面可以补充大模型所没有掌握到的领域知识，另一方面还可以解决大模型内部知识的时效性问题，同时在将外部知识与大模型内部知识的结合上可以用 Prompt Engineering 来控制，从而减轻大模型的幻觉问题。

在步骤 S240 中，利用第二答案 A1′对第一答案 A1 进行事实性验证，也就是判断第一答案 A1 与第二答案 A1′之间是否存在事实上的一致性。当确定第一答案 A1 与第二答案 A1′存在事实上的一致性时，表明第一答案 A1 通过事实性验证；当第一答案 A1 与第

二答案 A1′不存在事实上的一致性时，表明第一答案 A1 未通过事实性验证。

在步骤 S250 中，当第一答案 A1 未通过事实性验证时，确定对应的问答对存在过度承诺问题。本步骤是通过算法预侦查出的对话文本中的过度承诺问题，后续可以作为疑似问题转人工复查。

在步骤 S240 中，可以采用多种方式确定第一答案 A1 与第二答案 A1′之间是否存在事实上的一致性。例如，可以确定第一答案 A1 与第二答案 A1′分别对应的文本特征向量，基于向量之间的相似度匹配确定两者之间是否存在事实上的一致性。

为了更加准确地确定两个答案之间是否存在事实上的一致性，可以利用强大的大语言模型进行事实性验证。具体的，可以基于第一答案 A1、第二答案 A1′构建第二提示模板 prompt2；接着，将第二提示模板 prompt2 输入大语言模型，通过大语言模型核验两个答案之间的一致性。

其中，第二提示模板 prompt2 的内容包括上述第一答案 A1 和第二答案 A1′，第二提示模板 prompt2 用于指示核验第一答案 A1 与第二答案 A1′之间的一致性。这里，大语言模型用于核验两个答案之间的一致性。

在生成第二提示模板 prompt2 时，可以按照预设的提示模板格式生成。例如，第二提示模板 prompt2 可以包括：第二指示部分、待核验文本部分和标准答案部分。在构建第二提示模板 prompt2 时，可以将第一答案 A1 填充至待核验文本部分，将第二答案 A1 填充至标准答案部分，将预设的第二指示文本填充至第二指示部分。

其中，第二指示文本用于指示大语言模型核验待核验文本部分和标准答案部分之间的一致性。

为了使得第二提示模板 prompt2 的功能能加全面，进而使得大语言模型更有效地核验两个答案，第二提示模板 prompt2 还可以包含以下部分中的至少一种：

角色定义部分，用于定义大语言模型担任的角色；

输出格式部分，用于定义大语言模型输出核验结果时的格式。

第二指示部分还可以包含核验时的提示，例如提示在核验时需要大语言模型对数字部分特别在意等。

对于具体的答案核验任务，可以将预先设定的文本内容填充到上述对应的各个部分中。例如，在客户服务进行质检的场景中，可以设定以下提示模板，见表 3。

表 3

角色定义部分	你是一个服务质检人员

待核验文本部分	下面列举的内容是一段服务对话中客户的问题或者诉求（记为 Q），针对该问题 Q 客服人员作出的回答（记为 A），以及系统的知识库中对于该问题的答案（记为 A′）： Q：xxxxx A：xxxxx A′：xxxxx
第二指示部分	请你利用你的专业知识，判断对于这个用户问题 Q，客服的答案 A 与系统的答案 A′是否为一个意思。你可以忽略 A 与 A′在一些不影响大意的用词风格与语气词上的差异，但请特别小心 A 与 A′在数字上的差别
输出格式部分	A 与 A′的意思是否一致：{是/否}

在表 3 中，左侧一列是提示模板中的各个部分，各部分之间的顺序不限于是表 3 中的顺序，可以根据需要调整。右侧一列给出了各个部分的具体内容，部分内容采用“x”或者省略号代替。针对不同的任务，可以相应地调整各部分的具体内容。

本实施例中使用的大语言模型可以采用同一个大语言模型实现。上述步骤 S210 和 S220 是对对话文本中的问答对进行抽取的阶段，步骤 S230~S250 是对问答对的内容进行事实性验证的阶段。

上述两个阶段通过合理设置对应的提示模板，借助大语言模型强大的自然文本的分析和处理能力，很好地实现了问答对的提取以及答案的事实性验证，攻克了对话内容 QA 对儿抽取以及答案事实性验证两个技术难点。上述实施例无需对传统机器学习算法模型进行开发与调优，开发与维护成本低，且对不同的服务对话质检场景具有普遍适用性。上述实施例也可以不必对大语言模型进行微调，直接使用大语言模型的通用文本内容理解能力即可完成任务的执行。

上述实施例可以有多种落地应用：1）作为服务对话后服务质量质检的预过滤环节，来提升后续人工质检的靶向性和效率，并降低总体质检成本。2）用于对客服对话过程中的对话内容进行实时安全检测，如果发现有过度承诺情况的出现，则及时上升到专员处理。

本说明书中，第一问题、第一答案、第一提示模板、第一指示部分等词语中的“第一”，以及文中相应的“第二”（如果存在）等，仅仅是为了区分和描述方便，并不具有任何限定意义。

上述内容对本说明书的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按照示出的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，或者可能是有利的。

图 4 为实施例提供的一种对话中的过度承诺检测装置的示意性框图。该装置实施例与图 2 所示方法实施例相对应。该装置部署在计算设备中。计算设备可以通过任何具有计算、处理能力的装置、设备、平台、设备集群等来实现。该装置 400 包括：获取模块 410，配置为获取待检测的对话文本；提取模块 420，配置为利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；匹配模块 430，配置为针对任意一个包含第一问题和第一答案的问答对，从知识库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；验证模块 440，配置为利用所述第二答案对所述第一答案进行事实性验证；确定模块 450，配置为当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

在一种实施方式中，所述提取模块 420 包括：第一构建子模块和提取子模块（图中未示出）。其中，第一构建子模块，配置为基于所述对话文本构建第一提示模板；所述第一提示模板用于指示从所述对话文本中提取问题和对应的答案。提取子模块，配置为将所述第一提示模板输入大语言模型，通过所述大语言模型从所述对话文本中提取问题和对应的答案。

在一种实施方式中，所述第一提示模板包含第一指示部分和目标文本部分；所述第一构建子模块，具体配置为将所述对话文本填充至所述目标文本部分；将预设的第一指示文本填充至所述第一指示部分；其中，所述第一指示文本用于指示大语言模型从所述目标文本部分的内容中提取问答和对应的答案。

在一种实施方式中，所述第一提示模板还包含以下部分中的至少一种：角色定义部分，用于定义所述大语言模型担任的角色；输出格式部分，用于定义所述大语言模型输出问题和对应的答案时的格式；问答对示例部分，用于向大语言模型展示问题和对应答案的例子。

在一种实施方式中，所述匹配模块 430 包括：嵌入子模块、匹配子模块和确定子模块（图中未示出）。其中，嵌入子模块，配置为确定所述第一问题对应的第一文本特征向量；匹配子模块，配置为将所述第一文本特征向量与向量数据库中的多个文本特征向量进行匹配。所述向量数据库中的多个文本特征向量分别是知识库中的多个问题的文本特征向量。确定子模块，配置为基于匹配成功的文本特征向量对应的知识库中的答案，确定所述第二

答案。

在一种实施方式中，所述匹配模块 430 具体配置为利用检索增强生成算法，从所述知识库中确定与所述第一问题匹配的第二答案。

在一种实施方式中，所述验证模块 440 包括：第二构建子模块和核验子模块（图中未示出）。其中，第二构建子模块，配置为基于所述第一答案、所述第二答案构建第二提示模板；所述第二提示模板用于指示核验所述第一答案与所述第二答案之间的一致性。核验子模块，配置为将所述第二提示模板输入大语言模型，通过所述大语言模型核验所述第一答案与所述第二答案之间的一致性。

在一种实施方式中，所述第二提示模板包括：第二指示部分、待核验文本部分和标准答案部分；所述第二构建子模块，具体配置为将所述第一答案填充至所述待核验文本部分，将第二答案填充至所述标准答案部分，将预设的第二指示文本填充至所述第二指示部分。第二指示文本用于指示所述大语言模型核验所述待核验文本部分和所述标准答案部分之间的一致性。

上述装置实施例与方法实施例相对应，具体说明可以参见方法实施例部分的描述，此处不再赘述。装置实施例是基于对应的方法实施例得到，与对应的方法实施例具有同样的技术效果，具体说明可参见对应的方法实施例。

The specification also provides a computer-readable non-transitory storage medium, wherein the storage medium stores a computer program, which, when executed by a processor, can be used to perform one or more steps of one or more methods described or illustrated herein or provides functionality described or illustrated herein. Herein, the computer-readable non-transitory storage medium or media may include one or more semiconductor-based or other integrated circuits (ICs) (such, as for example, field-programmable gate arrays (FPGAs) or application-specific ICs (ASICs)), hard disk drives (HDDs), hybrid hard drives (HHDs), optical discs, optical disc drives (ODDs), magneto-optical discs, magneto-optical drives, floppy diskettes, floppy disk drives (FDDs), magnetic tapes, solid-state drives (SSDs), RAM-drives, SECURE DIGITAL cards or drives, any other suitable computer-readable non-transitory storage media, or any suitable combination of two or more of these, where appropriate. A computer-readable non-transitory storage medium may be volatile, non-volatile, or a combination of volatile and non-volatile, where appropriate.

The specification also provides a computer system. FIG. 5 illustrates an example computer system 1000. In particular embodiments, one or more computer systems 1000 perform one or more steps of one or more methods described or illustrated herein. In particular embodiments, one or

more computer systems 1000 provide functionality described or illustrated herein. In particular embodiments, software running on one or more computer systems 1000 performs one or more steps of one or more methods described or illustrated herein or provides functionality described or illustrated herein. Particular embodiments include one or more portions of one or more computer systems 1000. Herein, reference to a computer system may encompass a computing device, and vice versa, where appropriate. Moreover, reference to a computer system may encompass one or more computer systems, where appropriate.

This disclosure contemplates any suitable number of computer systems 1000. This disclosure contemplates computer system 1000 taking any suitable physical form. As example and not by way of limitation, computer system 1000 may be an embedded computer system, a system-on-chip (SOC), a single-board computer system (SBC) (such as, for example, a computer-on-module (COM) or system-on-module (SOM)), a desktop computer system, a laptop or notebook computer system, an interactive kiosk, a mainframe, a mesh of computer systems, a mobile telephone, a personal digital assistant (PDA), a server, a tablet computer system, or a combination of two or more of these. Where appropriate, computer system 1000 may include one or more computer systems 1000; be unitary or distributed; span multiple locations; span multiple machines; span multiple data centers; or reside in a cloud, which may include one or more cloud components in one or more networks. Where appropriate, one or more computer systems 1000 may perform without substantial spatial or temporal limitation one or more steps of one or more methods described or illustrated herein. As an example and not by way of limitation, one or more computer systems 1000 may perform in real time or in batch mode one or more steps of one or more methods described or illustrated herein. One or more computer systems 1000 may perform at different times or at different locations one or more steps of one or more methods described or illustrated herein, where appropriate.

In particular embodiments, computer system 1000 includes a processor 1002, memory 1004, storage 1006, an input/output (I/O) interface 1008, a communication interface 1010, and a bus 1012. Although this disclosure describes and illustrates a particular computer system having a particular number of particular components in a particular arrangement, this disclosure contemplates any suitable computer system having any suitable number of any suitable components in any suitable arrangement.

In particular embodiments, processor 1002 includes hardware for executing instructions, such as those making up a computer program. As an example and not by way of limitation, to execute instructions, processor 1002 may retrieve (or fetch) the instructions from an internal register, an internal cache, memory 1004, or storage 1006; decode and execute them; and then write one or

more results to an internal register, an internal cache, memory 1004, or storage 1006. In particular embodiments, processor 1002 may include one or more internal caches for data, instructions, or addresses. This disclosure contemplates processor 1002 including any suitable number of any suitable internal caches, where appropriate. As an example and not by way of limitation, processor 1002 may include one or more instruction caches, one or more data caches, and one or more translation lookaside buffers (TLB s). Instructions in the instruction caches may be copies of instructions in memory 1004 or storage 1006, and the instruction caches may speed up retrieval of those instructions by processor 1002. Data in the data caches may be copies of data of 80 in memory 1004 or storage 1006 for instructions executing at processor 1002 to operate on; the results of previous instructions executed at processor 1002 for access by subsequent instructions executing at processor 1002 or for writing to memory 1004 or storage 1006; or other suitable data. The data caches may speed up read or write operations by processor 1002. The TLBs may speed up virtual-address translation for processor 1002. In particular embodiments, processor 1002 may include one or more internal registers for data, instructions, or addresses. This disclosure contemplates processor 1002 including any suitable number of any suitable internal registers, where appropriate. Where appropriate, processor 1002 may include one or more arithmetic logic units (ALUs); be a multi-core processor; or include one or more processors 1002. Although this disclosure describes and illustrates a particular processor, this disclosure contemplates any suitable processor.

In particular embodiments, memory 1004 includes main memory for storing instructions for processor 1002 to execute or data for processor 1002 to operate on. As an example and not by way of limitation, computer system 1000 may load instructions from storage 1006 or another source (such as, for example, another computer system 1000) to memory 1004. Processor 1002 may then load the instructions from memory 1004 to an internal register or internal cache. To execute the instructions, processor 1002 may retrieve the instructions from the internal register or internal cache and decode them. During or after execution of the instructions, processor 1002 may write one or more results (which may be intermediate or final results) to the internal register or internal cache. Processor 1002 may then write one or more of those results to memory 1004. In particular embodiments, processor 1002 executes only instructions in one or more internal registers or internal caches or in memory 1004 (as opposed to storage 1006 or elsewhere) and operates only on data in one or more internal registers or internal caches or in memory 1004 (as opposed to storage 1006 or elsewhere). One or more memory buses (which may each include an address bus and a data bus) may couple processor 1002 to memory 1004. Bus 1012 may include one or more memory buses, as described below. In particular embodiments, one or more memory management

units (MMUs) reside between processor 1002 and memory 1004 and facilitate accesses to memory 1004 requested by processor 1002. In particular embodiments, memory 1004 includes random access memory (RAM). This RAM may be volatile memory, where appropriate. Where appropriate, this RAM may be dynamic RAM (DRAM) or static RAM (SRAM). Moreover, where appropriate, this RAM may be single-ported or multi-ported RAM. This disclosure contemplates any suitable RAM. Memory 1004 may include one or more memories 1004, where appropriate. Although this disclosure describes and illustrates particular memory, this disclosure contemplates any suitable memory.

In particular embodiments, storage 1006 includes mass storage for data or instructions. As an example and not by way of limitation, storage 1006 may include a hard disk drive (HDD), a floppy disk drive, flash memory, an optical disc, a magneto-optical disc, magnetic tape, or a Universal Serial Bus (USB) drive or a combination of two or more of these. Storage 1006 may include removable or non-removable (or fixed) media, where appropriate. Storage 1006 may be internal or external to computer system 1000, where appropriate. In particular embodiments, storage 1006 is non-volatile, solid-state memory. In particular embodiments, storage 1006 includes read-only memory (ROM). Where appropriate, this ROM may be mask-programmed ROM, programmable ROM (PROM), erasable PROM (EPROM), electrically erasable PROM (EEPROM), electrically alterable ROM (EAROM), or flash memory or a combination of two or more of these. This disclosure contemplates mass storage 1006 taking any suitable physical form. Storage 1006 may include one or more storage control units facilitating communication between processor 1002 and storage 1006, where appropriate. Where appropriate, storage 1006 may include one or more storages 1006. Although this disclosure describes and illustrates particular storage, this disclosure contemplates any suitable storage.

In particular embodiments, I/O interface 1008 includes hardware, software, or both, providing one or more interfaces for communication between computer system 1000 and one or more I/O devices. Computer system 1000 may include one or more of these I/O devices, where appropriate. One or more of these I/O devices may enable communication between a person and computer system 1000. As an example and not by way of limitation, an I/O device may include a keyboard, keypad, microphone, monitor, mouse, printer, scanner, speaker, still camera, stylus, tablet, touch screen, trackball, video camera, another suitable I/O device or a combination of two or more of these. An I/O device may include one or more sensors. This disclosure contemplates any suitable I/O devices and any suitable I/O interfaces 1008 for them. Where appropriate, I/O interface 1008 may include one or more device or software drivers enabling processor 1002 to drive one or more of these I/O devices. I/O interface 1008 may include one or more I/O interfaces 1008, where

appropriate. Although this disclosure describes and illustrates a particular I/O interface, this disclosure contemplates any suitable I/O interface.

In particular embodiments, communication interface 1010 includes hardware, software, or both providing one or more interfaces for communication (such as, for example, packet-based communication) between computer system 1000 and one or more other computer systems 1000 or one or more networks. As an example and not by way of limitation, communication interface 1010 may include a network interface controller (NIC) or network adapter for communicating with an Ethernet or other wire-based network or a wireless NIC (WNIC) or wireless adapter for communicating with a wireless network, such as a WI-FI network. This disclosure contemplates any suitable network and any suitable communication interface 1010 for it. As an example and not by way of limitation, computer system 1000 may communicate with an ad hoc network, a personal area network (PAN), a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN), or one or more portions of the Internet or a combination of two or more of these. One or more portions of one or more of these networks may be wired or wireless. As an example, computer system 1000 may communicate with a wireless PAN (WPAN) (such as, for example, a BLUETOOTH WPAN), a WI-FI network, a WI-MAX network, a cellular telephone network (such as, for example, a Global System for Mobile Communications (GSM) network), or other suitable wireless network or a combination of two or more of these. Computer system 1000 may include any suitable communication interface 1010 for any of these networks, where appropriate. Communication interface 1010 may include one or more communication interfaces 1010, where appropriate. Although this disclosure describes and illustrates a particular communication interface, this disclosure contemplates any suitable communication interface.

In particular embodiments, bus 1012 includes hardware, software, or both coupling components of computer system 1000 to each other. As an example and not by way of limitation, bus 1012 may include an Accelerated Graphics Port (AGP) or other graphics bus, an Enhanced Industry Standard Architecture (EISA) bus, a front-side bus (FSB), a HYPERTRANSPORT (HT) interconnect, an Industry Standard Architecture (ISA) bus, an INFINIBAND interconnect, a low-pin-count (LPC) bus, a memory bus, a Micro Channel Architecture (MCA) bus, a Peripheral Component Interconnect (PCI) bus, a PCI-Express (PCIe) bus, a serial advanced technology attachment (SATA) bus, a Video Electronics Standards Association local (VLB) bus, or another suitable bus or a combination of two or more of these. Bus 1012 may include one or more buses 1012, where appropriate. Although this disclosure describes and illustrates a particular bus, this disclosure contemplates any suitable bus or interconnect.

还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的

包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于存储介质和计算设备实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。

本领域技术人员应该可以意识到，在上述一个或多个示例中，本发明实施例所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。

以上所述的具体实施方式，对本发明实施例的目的、技术方案和有益效果进行了进一步的详细说明。所应理解的是，以上所述仅为本发明实施例的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的技术方案的基础之上所做的任何修改、等同替换、改进等，均应包括在本发明的保护范围之内。

权　　利　　要　　求　　书

1. 一种对话中的过度承诺检测方法，包括：

获取待检测的对话文本；

利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；

针对任意一个包含第一问题和第一答案的问答对，从知识库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；

利用所述第二答案对所述第一答案进行事实性验证；

当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

2. 根据权利要求 1 所述的方法,所述从所述对话文本中提取问题和对应的答案的步骤，包括：

基于所述对话文本构建第一提示模板；所述第一提示模板用于指示从所述对话文本中提取问题和对应的答案；

将所述第一提示模板输入大语言模型，通过所述大语言模型从所述对话文本中提取问题和对应的答案。

3. 根据权利要求 2 所述的方法,所述第一提示模板包含第一指示部分和目标文本部分；

所述基于所述对话文本构建第一提示模板的步骤，包括：

将所述对话文本填充至所述目标文本部分；

将预设的第一指示文本填充至所述第一指示部分；其中，所述第一指示文本用于指示大语言模型从所述目标文本部分的内容中提取问答和对应的答案。

4. 根据权利要求 3 所述的方法，所述第一提示模板还包含以下部分中的至少一种：

角色定义部分，用于定义所述大语言模型担任的角色；

输出格式部分，用于定义所述大语言模型输出问题和对应的答案时的格式；

问答对示例部分，用于向大语言模型展示问题和对应答案的例子。

5. 根据权利要求 1 所述的方法，所述从知识库中确定与所述第一问题匹配的第二答案的步骤，包括：

确定所述第一问题对应的第一文本特征向量；

将所述第一文本特征向量与向量数据库中的多个文本特征向量进行匹配；所述向量数据库中的多个文本特征向量分别是知识库中的多个问题的文本特征向量；

基于匹配成功的文本特征向量对应的知识库中的答案，确定所述第二答案。

6. 根据权利要求 1 所述的方法，所述从知识库中确定与所述第一问题匹配的第二答案的步骤，包括：

利用检索增强生成算法，从所述知识库中确定与所述第一问题匹配的第二答案。

7. 根据权利要求 1 所述的方法，所述利用所述第二答案对所述第一答案进行事实性验证的步骤，包括：

基于所述第一答案、所述第二答案构建第二提示模板；所述第二提示模板用于指示核验所述第一答案与所述第二答案之间的一致性；

将所述第二提示模板输入大语言模型，通过所述大语言模型核验所述第一答案与所述第二答案之间的一致性。

8. 根据权利要求 7 所述的方法，所述第二提示模板包括：第二指示部分、待核验文本部分和标准答案部分；所述基于所述第一答案、所述第二答案构建第二提示模板的步骤，包括：

将所述第一答案填充至所述待核验文本部分；

将所述第二答案填充至所述标准答案部分；

将预设的第二指示文本填充至所述第二指示部分；所述第二指示文本用于指示所述大语言模型核验所述待核验文本部分和所述标准答案部分之间的一致性。

9. 一种计算机系统，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述程序时实现如下步骤：

获取待检测的对话文本；

利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；

针对任意一个包含第一问题和第一答案的问答对，从知识库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；

利用所述第二答案对所述第一答案进行事实性验证；

当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

10. 根据权利要求 9 所述的计算机系统，所述从所述对话文本中提取问题和对应的答案的步骤，包括：

基于所述对话文本构建第一提示模板；所述第一提示模板用于指示从所述对话文本中提取问题和对应的答案；

将所述第一提示模板输入大语言模型，通过所述大语言模型从所述对话文本中提取问

题和对应的答案。

11. 根据权利要求 10 所述的计算机系统，所述第一提示模板包含第一指示部分和目标文本部分；

所述基于所述对话文本构建第一提示模板的步骤，包括：

将所述对话文本填充至所述目标文本部分；

将预设的第一指示文本填充至所述第一指示部分；其中，所述第一指示文本用于指示大语言模型从所述目标文本部分的内容中提取问答和对应的答案。

12. 根据权利要求 11 所述的计算机系统，所述第一提示模板还包含以下部分中的至少一种：

角色定义部分，用于定义所述大语言模型担任的角色；

输出格式部分，用于定义所述大语言模型输出问题和对应的答案时的格式；

问答对示例部分，用于向大语言模型展示问题和对应答案的例子。

13. 根据权利要求 9 所述的计算机系统，所述从知识库中确定与所述第一问题匹配的第二答案的步骤，包括：

确定所述第一问题对应的第一文本特征向量；

将所述第一文本特征向量与向量数据库中的多个文本特征向量进行匹配；所述向量数据库中的多个文本特征向量分别是知识库中的多个问题的文本特征向量；

基于匹配成功的文本特征向量对应的知识库中的答案，确定所述第二答案。

14. 根据权利要求 9 所述的计算机系统，所述从知识库中确定与所述第一问题匹配的第二答案的步骤，包括：

利用检索增强生成算法，从所述知识库中确定与所述第一问题匹配的第二答案。

15. 根据权利要求 9 所述的计算机系统，所述利用所述第二答案对所述第一答案进行事实性验证的步骤，包括：

基于所述第一答案、所述第二答案构建第二提示模板；所述第二提示模板用于指示核验所述第一答案与所述第二答案之间的一致性；

将所述第二提示模板输入大语言模型，通过所述大语言模型核验所述第一答案与所述第二答案之间的一致性。

16. 根据权利要求 15 所述的计算机系统，所述第二提示模板包括：第二指示部分、待核验文本部分和标准答案部分；所述基于所述第一答案、所述第二答案构建第二提示模板的步骤，包括：

将所述第一答案填充至所述待核验文本部分；

将所述第二答案填充至所述标准答案部分；

将预设的第二指示文本填充至所述第二指示部分；所述第二指示文本用于指示所述大语言模型核验所述待核验文本部分和所述标准答案部分之间的一致性。

17. 一种对话中的过度承诺检测装置，包括：

获取模块，配置为获取待检测的对话文本；

提取模块，配置为利用对自然语言中信息的提取技术，从所述对话文本中提取问题和对应的答案，得到若干问答对；

匹配模块，配置为针对任意一个包含第一问题和第一答案的问答对，从知识库中确定与所述第一问题匹配的第二答案；所述知识库包含多个预设的问答对；

验证模块，配置为利用所述第二答案对所述第一答案进行事实性验证；

确定模块，配置为当所述第一答案未通过事实性验证时，确定对应的问答对存在过度承诺问题。

18. 一种计算机可读非易失性存储介质，所述存储介质存储有计算机程序，所述计算机程序被处理器执行时实现上述权利要求 1-8 任一项所述的方法。