



(12) 发明专利申请

(10) 申请公布号 CN 118551772 A

(43) 申请公布日 2024. 08. 27

(21) 申请号 202410759785.4

G06F 16/33 (2019.01)

(22) 申请日 2024.06.13

(71) 申请人 芮达

地址 201702 上海市青浦区诸光路1188弄
219号

(72) 发明人 芮达 卫岸

(74) 专利代理机构 南京禾易知识产权代理有限公司 32320

专利代理师 王丙烁

(51) Int. Cl.

G06F 40/35 (2020.01)

G06N 20/00 (2019.01)

G06F 40/295 (2020.01)

G06F 40/284 (2020.01)

G06F 16/332 (2019.01)

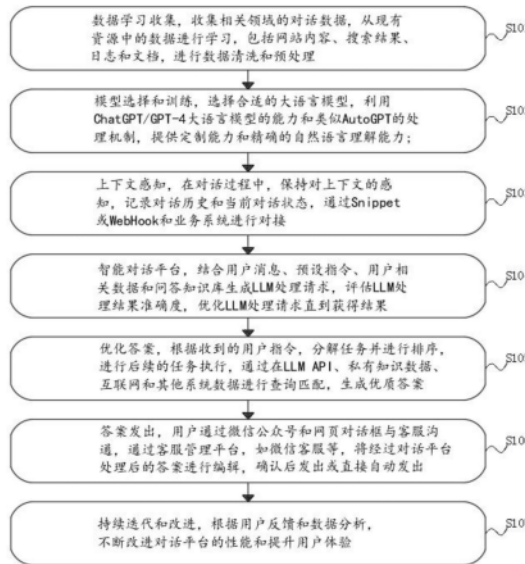
权利要求书2页 说明书5页 附图3页

(54) 发明名称

一种基于大语言模型的智能对话平台

(57) 摘要

本发明公开了一种基于大语言模型的智能对话平台,涉及人机交互技术领域,包括数据学习收集、模型选择训练、上下文感知、智能对话平台、答案发出和持续迭代和改进,充分利用现有网络发展,以大语言模型为基础,通过收集和学习相关领域的对话数据,使得生成的平台可以获得更多的领域知识和语境,提高对话的准确性和个性化程度,平台能够持续感知对话的上下文,记录对话历史和当前对话状态,以保持对话的连贯性和一致性,通过结合用户消息、预设指令、用户相关数据和问答知识库,平台能够生成LLM处理请求,并优化请求,直到获得满意的结果,以大语言模型为基础的智能对话平台具有全方面的优点,能够给用户提供准确、智能和高效的对话服务。



1. 一种基于大语言模型的智能对话平台,其特征在于,包括如下步骤:

S101、数据学习收集,收集相关领域的对话数据,从现有资源中的数据进行学习,包括网站内容、搜索结果、日志和文档,进行数据清洗和预处理;

S102、模型选择训练,选择合适的大语言模型,利用ChatGPT/GPT-4大语言模型的能力和类似AutoGPT的处理机制,提供定制能力和精确的自然语言理解能力;

S103、上下文感知,在对话过程中,保持对上下文的感知,记录对话历史和当前对话状态,通过Snippet或WebHook和业务系统进行对接;

S104、智能对话平台,结合用户消息、预设指令、用户相关数据和问答知识库生成LLM处理请求,评估LLM处理结果准确度,优化LLM处理请求直到获得结果;

S105、优化答案,根据收到的用户指令,分解任务并进行排序,进行后续的任务执行,通过在LLM API、私有知识数据、互联网和其他系统数据进行查询匹配,生成优质答案;

S106、答案发出,用户通过微信公众号和网页对话框与客服沟通,通过客服管理平台,如微信客服等,将经过对话平台处理后的答案进行编辑,确认后发出或直接自动发出;

S107、持续迭代和改进,根据用户反馈和数据分析,不断改进对话平台的性能和提升用户体验。

2. 根据权利要求1所述的一种基于大语言模型的智能对话平台,其特征在于,所述数据学习收集需要先一步确定相关领域,如医疗、金融和教育,在收集和确定领域相关的对话数据中,数据可以从现有资源中获取,包括网站内容、搜索结果、日志和文档,也可以通过爬虫程序从互联网上抓取相关数据,需要对收集到的数据进行清洗,去除重复数据、垃圾数据和错误数据,确保数据的质量和准确性,对清洗后的数据进行预处理,包括分词、去除停用词和词性标注,为后续的模型训练提供更好的输入数据。

3. 根据权利要求1所述的一种基于大语言模型的智能对话平台,其特征在于,所述模型选择训练根据需求和应用场景,选择合适的大语言模型,目前常用的大语言模型包括ChatGPT和GPT-4,具有高度的自然语言理解和生成能力,用于智能对话平台的开发,根据所选模型的要求,准备相应的训练数据,数据可以来自于S101阶段收集和预处理的对话语料库,也可以进行人工标注和编辑来满足模型训练的需求。

4. 根据权利要求3所述的一种基于大语言模型的智能对话平台,其特征在于,所述模型选择训练包括使用准备好的数据对选择的大语言模型进行训练,训练过程中,模型会学习对输入文本的理解和生成回复的能力,需要进行调参和优化以获得更好的结果,对训练完成的模型进行评估,验证其在对话生成和理解方面的准确性和流畅性,将训练好的模型部署到智能对话平台中,供用户实际使用。

5. 根据权利要求1所述的一种基于大语言模型的智能对话平台,其特征在于,所述上下文感知包括在对话开始时,记录用户的输入作为初始上下文,随着对话的进行,将每一轮对话的用户输入和AI回复都添加到上下文中,形成对话历史,保持对话的连贯性和一致性,在生成回复之前,对上下文进行处理和分析,使用技术手段如编码器来提取上下文中的重要信息,并将其作为输入传递给模型,利用训练好的模型,根据上下文动态填充相关内容。

6. 根据权利要求5所述的一种基于大语言模型的智能对话平台,其特征在于,所述上下文动态填充需要将AI生成的回复添加到上下文中,并记录下来,在下一轮对话时,上下文包含之前的对话历史和当前对话状态,使得回复能够更好地结合上下文进行理解和生成,使

用Snippet和WebHook技术手段,通过定义接口和数据格式,将对话历史及相关上下文信息传递给业务系统,并获取系统返回的结果或信息,让AI在回复中能够使用到业务系统的数据和逻辑。

7.根据权利要求1所述的一种基于大语言模型的智能对话平台,其特征在于,所述优化答案包括当用户提出问题或需求时,首先需要对用户指令进行解析和分解,对自然语言文本进行分词、词性标注和命名实体识别,将指令转化为计算机可以理解的形式,根据用户指令,将任务进行分解,并按照优先级进行排序,包括检索信息、查询数据和执行计算,根据具体情况对任务进行调度和分配。

8.根据权利要求7所述的一种基于大语言模型的智能对话平台,其特征在于,针对所述任务,通过LLM API、私有知识库、互联网和其他系统数据源进行查询和匹配,涉及到语义理解、知识图谱查询和语言模型技术手段,获取答案,在获取到数据结果后,将其进行答案生成和排版,将查询结果转化为自然语言文本、进行语言模板填充和格式化排版操作,生成答案后,对多个答案进行排序和推荐,提供答案和推荐。

9.根据权利要求1所述的一种基于大语言模型的智能对话平台,其特征在于,所述答案发出包括在用户通过微信公众号和网页对话框提出问题时,客服人员可以立刻收到对话请求,当用户的对话请求到达对话平台时,对话平台会进行语义分析和处理,将用户的意图和需求转化为计算机可以理解的形式,再利用智能对话平台进行处理,了解用户的需求和问题,并开始进行答案编辑和确认,生成答案,答案被转发给相应的客服人员进行处理,客服人员可以对答案内容再次进行编辑和确认,补充相关链接和资料,经过编辑和确认后,客服人员将答案发送给用户,通过客服管理平台的界面直接发送,或者使用自动化的方式将答案自动发送给用户。

10.根据权利要求1所述的一种基于大语言模型的智能对话平台,其特征在于,所述持续迭代和改进包括客服人员可以记录用户的反馈和评价,进行后续对话和服务改进,帮助客服人员理解用户需求,并根据用户反馈和数据分析,安排工程师改进对话平台的性能,给客户提供服务。

一种基于大语言模型的智能对话平台

技术领域

[0001] 本发明涉及人机交互技术领域,具体为一种基于大语言模型的智能对话平台。

背景技术

[0002] 对话平台是一种基于大语言模型的智能对话平台基于人工智能技术的交互式软件平台,它可以与用户进行自然语言对话,提供各种服务和支持;

[0003] 在早期对话平台中,通常只能处理简单的固定式指令,对于复杂的语义理解和意图识别能力较弱,这导致用户只能进行有限的交互,无法完成更复杂的任务或回答更深入的问题,早期对话平台在处理对话过程中往往缺乏对上下文的深入理解,导致交流的连贯性和个性化有所欠缺,它们可能无法准确理解用户的先前提问或回答,以及对话中的隐含信息,从而无法提供更精准和个性化的回应,且通常采用固定的指令模式,用户必须按照特定的顺序提出问题或指令,这对用户的灵活性和自由度有所限制,用户可能会感到受限于预设的模板,无法自由地表达自己的需求和意图;

[0004] 这些缺点使得早期对话平台在交互和用户体验方面存在一定的限制,需要进一步提升智能化和个性化的能力来满足用户的需求,有必要对现有技术进行改进,为此,我们提出一种基于大语言模型的智能对话平台用于解决上述问题。

发明内容

[0005] 本发明的目的在于提供一种基于大语言模型的智能对话平台,以解决上述背景技术中提出的技术问题。

[0006] 为实现上述发明目的,本发明采用如下技术方案:

[0007] 本发明提供一种基于大语言模型的智能对话平台,包括如下步骤:

[0008] S101、数据学习收集,收集相关领域的对话数据,从现有资源中的数据进行学习,包括网站内容、搜索结果、日志和文档,进行数据清洗和预处理;

[0009] S102、模型选择训练,选择合适的大语言模型,利用ChatGPT/GPT-4大语言模型的能力和类似AutoGPT的处理机制,提供定制能力和精确的自然语言理解能力;

[0010] S103、上下文感知,在对话过程中,保持对上下文的感知,记录对话历史和当前对话状态,通过Snippet或WebHook和业务系统进行对接;

[0011] S104、智能对话平台,结合用户消息、预设指令、用户相关数据和问答知识库生成LLM处理请求,评估LLM处理结果准确度,优化LLM处理请求直到获得结果;

[0012] S105、优化答案,根据收到的用户指令,分解任务并进行排序,进行后续的任务执行,通过在LLM API、私有知识数据、互联网和其他系统数据进行查询匹配,生成优质答案;

[0013] S106、答案发出,用户通过微信公众号和网页对话框与客服沟通,通过客服管理平台,如微信客服等,将经过对话平台处理后的答案进行编辑,确认后发出或直接自动发出;

[0014] S107、持续迭代和改进,根据用户反馈和数据分析,不断改进对话平台的性能和提升用户体验。

[0015] 优选的,所述数据学习收集需要先一步确定相关领域,如医疗、金融和教育,在收集和确定领域相关的对话数据中,数据可以从现有资源中获取,包括网站内容、搜索结果、日志和文档,也可以通过爬虫程序从互联网上抓取相关数据,需要对收集到的数据进行清洗,去除重复数据、垃圾数据和错误数据,确保数据的质量和准确性,对清洗后的数据进行预处理,包括分词、去除停用词和词性标注,为后续的模型训练提供更好的输入数据。

[0016] 优选的,所述模型选择训练根据需求和应用场景,选择合适的大语言模型,目前常用的大语言模型包括ChatGPT和GPT-4,具有高度的自然语言理解和生成能力,用于智能对话平台的开发,根据所选模型的要求,准备相应的训练数据,数据可以来自于S101阶段收集和预处理的对话语料库,也可以进行人工标注和编辑来满足模型训练的需求。

[0017] 优选的,所述模型选择训练包括使用准备好的数据对选择的大语言模型进行训练,训练过程中,模型会学习对输入文本的理解和生成回复的能力,需要进行调参和优化以获得更好的结果,对训练完成的模型进行评估,验证其在对话生成和理解方面的准确性和流畅性,将训练好的模型部署到智能对话平台中,供用户实际使用。

[0018] 优选的,所述上下文感知包括在对话开始时,记录用户的输入作为初始上下文,随着对话的进行,将每一轮对话的用户输入和AI回复都添加到上下文中,形成对话历史,保持对话的连贯性和一致性,在生成回复之前,对上下文进行处理和分析,使用技术手段如编码器来提取上下文中的重要信息,并将其作为输入传递给模型,利用训练好的模型,根据上下文动态填充相关内容。

[0019] 优选的,所述上下文动态填充需要将AI生成的回复添加到上下文中,并记录下来,在下一轮对话时,上下文包含之前的对话历史和当前对话状态,使得回复能够更好地结合上下文进行理解和生成,使用Snippet和WebHook技术手段,通过定义接口和数据格式,将对话历史及相关上下文信息传递给业务系统,并获取系统返回的结果或信息,让AI在回复中能够使用到业务系统的数据和逻辑。

[0020] 优选的,所述优化答案包括当用户提出问题或需求时,首先需要对用户指令进行解析和分解,对自然语言文本进行分词、词性标注和命名实体识别,将指令转化为计算机可以理解的形式,根据用户指令,将任务进行分解,并按照优先级进行排序,包括检索信息、查询数据和执行计算,根据具体情况对任务进行调度和分配。

[0021] 优选的,针对所述任务,通过LLM API、私有知识库、互联网和其他系统数据源进行查询和匹配,涉及到语义理解、知识图谱查询和语言模型技术手段,获取答案,在获取到数据结果后,将其进行答案生成和排版,将查询结果转化为自然语言文本、进行语言模板填充和格式化排版操作,生成答案后,对多个答案进行排序和推荐,提供答案和推荐。

[0022] 优选的,所述答案发出包括在用户通过微信公众号和网页对话框提出问题时,客服人员可以立刻收到对话请求,当用户的对话请求到达对话平台时,对话平台会进行语义分析和处理,将用户的意图和需求转化为计算机可以理解的形式,再利用智能对话平台进行处理,了解用户的需求和问题,并开始进行答案编辑和确认,生成答案,答案被转发给相应的客服人员进行处理,客服人员可以对答案内容再次进行编辑和确认,补充相关链接和资料,经过编辑和确认后,客服人员将答案发送给用户,通过客服管理平台的界面直接发送,或者使用自动化的方式将答案自动发送给用户。

[0023] 优选的,所述持续迭代和改进包括客服人员可以记录用户的反馈和评价,进行后

续对话和服务改进,帮助客服人员理解用户需求,并根据用户反馈和数据分析,安排工程师改进对话平台的性能,给客户提供服务。

[0024] 与现有技术相比,以上一个或多个技术方案存在以下有益效果:

[0025] 本发明所提出的智能对话平台充分利用现有网络发展,以大语言模型为基础,通过收集和学习相关领域的对话数据,使得生成的平台可以获得更多的领域知识和语境,提高对话的准确性和个性化程度,其选择合适的大语言模型,ChatGPT和GPT-4,这些模型具有高度的自然语言理解和生成能力,能够满足不同应用场景的需求,平台能够持续感知对话的上下文,记录对话历史和当前对话状态,以保持对话的连贯性和一致性,通过结合用户消息、预设指令、用户相关数据和问答知识库,平台能够生成LLM处理请求,并优化请求,直到获得满意的结果,根据用户指令,平台能够分解任务并进行优先级排序,通过多个数据源进行查询和匹配,生成优质的答案,通过微信公众号和网页对话框等方式与用户进行交互,经过编辑和确认后将答案发送给用户,提供便捷的沟通方式,并且可以根据用户反馈和数据分析,平台能够不断改进性能,提升用户体验,持续优化和完善对话平台的功能;

[0026] 相较于现有的对话平台,以大语言模型为基础的智能对话平台具有全方面的优点,能够给用户提供的准确、智能和高效的对话服务。

附图说明

[0027] 构成本发明的一部分的说明书附图用来提供对本发明的进一步理解,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。

[0028] 图1是本发明的基于大语言模型的智能对话平台构件流程图。

[0029] 图2是本发明的交互平台与智能对话平台结构示意图。

[0030] 图3是本发明的智能对话平台构件模型结构示意图。

[0031] 图4是本发明的优质答案生成结构示意图。

[0032] 图5是本发明任务执行系统构件图。

[0033] 图6是本发明的系统交互示意图。

具体实施方式

[0034] 为了使本技术领域的人员更好地理解本申请方案,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分的实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都应当属于本申请保护的范围。

[0035] 请参阅图1-图6,本发明提供了一种基于大语言模型的智能对话平台,包括数据学习收集,收集相关领域的对话数据,从现有资源中的数据进行学习,包括网站内容、搜索结果、日志和文档,进行数据清洗和预处理,选择合适的大语言模型,利用ChatGPT/GPT-4大语言模型的能力和类似AutoGPT的处理机制,提供定制能力和精确的自然语言理解能力,上下文感知,在对话过程中,保持对上下文的感知,记录对话历史和当前对话状态,通过Snippet或WebHook和业务系统进行对接,智能对话平台,结合用户消息、预设指令、用户相关数据和问答知识库生成LLM处理请求,评估LLM处理结果准确度,优化LLM处理请求直到获得

结果,优化答案,根据收到的用户指令,分解任务并进行排序,进行后续的任务执行,通过在LLM API、私有知识数据、互联网和其他系统数据进行查询匹配,生成优质答案,答案发出,用户通过微信公众号和网页对话框与客服沟通,通过客服管理平台,如微信客服等,将经过对话平台处理后的答案进行编辑,确认后发出或直接自动发出,持续迭代和改进,根据用户反馈和数据分析,不断改进对话平台的性能和提升用户体验,充分利用现有网络发展,在大语言模型的基础上开发智能对话平台,给予用户最为准确和优质的服务。

[0036] 为了提高对话平台的准确度,在前期需要准确的收集数据,在数据学习收集集中需要先一步确定相关领域,如医疗、金融、教育等,有助于收集和学习与该领域相关的对话数据,在收集和确定领域相关的对话数据中,数据可以从现有资源中获取,包括网站内容、搜索结果、日志和文档,也可以通过爬虫程序从互联网上抓取相关数据,需要对收集到的数据进行清洗,去除重复数据、垃圾数据和错误数据,确保数据的质量和准确性,需要对清洗后的数据进行预处理,包括分词、去除停用词和词性标注,为后续的模型训练提供更好的输入数据,将预处理后的数据构建成对话语料库,通过收集和学习相关领域的对话数据,平台可以获得更多的领域知识和语境,提高对话的准确性和个性化程度。

[0037] 为了使得对话平台可以充分理解使用者所发出的自然语言,在模型选择训练中需要根据需求和应用场景,选择合适的大语言模型,目前常用的大语言模型包括ChatGPT和GPT-4,具有高度的自然语言理解和生成能力,可以用于智能对话平台的开发,根据所选模型的要求,准备相应的训练数据,数据可以来自于S101阶段收集和预处理的对话语料库,也可以进行人工标注和编辑以满足模型训练的需求,使用准备好的数据对选择的大语言模型进行训练,训练过程中,模型会学习对输入文本的理解和生成回复的能力,在模型训练过程中,需要进行调参和优化以获得更好的结果,对训练完成的模型进行评估,验证其在对话生成和理解方面的准确性和流畅性,将训练好的模型部署到智能对话平台中,供用户实际使用,大数据模型具有高度的自然语言理解和生成能力,能够满足不同应用场景的需要,充分理解使用者的需求。

[0038] 为了保持使用者在使用对话平台时的沟通连续性,其中上下文感知包括在对话开始时,记录用户的输入作为初始上下文,随着对话的进行,将每一轮对话的用户输入和AI回复都添加到上下文中,形成对话历史,保持对话的连贯性和一致性,在生成回复之前,对上下文进行处理和分析,使用技术手段如编码器来提取上下文中的重要信息,并将其作为输入传递给模型,利用训练好的模型,根据上下文动态填充相关内容,将AI生成的回复添加到上下文中,并记录下来,在下一轮对话时,上下文会包含之前的对话历史和当前对话状态,使得回复能够更好地结合上下文进行理解和生成,使用Snippet或WebHook技术手段,通过定义合适的接口和数据格式,将对话历史及相关上下文信息传递给业务系统,并获取系统返回的结果或信息,让AI在回复中能够使用到业务系统的数据和逻辑,平台能够持续感知对话的上下文,记录对话历史和当前对话状态,以保持对话的连贯性和一致性,提供更加个性化和精准的回复,使用者与对话平台沟通也更为准确,能够直接查看到自身的历史沟通记录。

[0039] 为了保证对话平台所寻找的答案准确性,其中优化答案包括当用户提出问题或需求时,首先需要对用户指令进行解析和分解,对自然语言文本进行分词、词性标注、命名实体识别等处理,将指令转化为计算机可以理解的形式,根据用户指令,将任务进行分解,并

按照优先级进行排序,这些任务包括检索信息、查询数据和执行计算,根据具体情况对任务进行调度和分配,针对每一个任务,通过LLM API、私有知识库、互联网和其他系统数据源进行查询和匹配,涉及到语义理解、知识图谱查询、语言模型生成等技术手段,以获取最优质的答案,在获取到数据结果后,将其进行答案生成和排版,将查询结果转化为自然语言文本、进行语言模板填充和格式化排版操作,使得答案可读性更高、易于理解,在生成答案后,对多个答案进行排序和推荐,根据用户指令,平台能够分解任务并进行优先级排序,通过多个数据源进行查询和匹配,生成最为优质的答案。

[0040] 为了进一步的提高用户的使用舒适性,当用户通过微信公众号或网页对话框提出问题或需求,客服人员可以立刻收到对话请求,用户的对话请求到达对话平台时,对话平台会进行语义分析和处理,将用户的意图和需求转化为计算机可以理解的形式,再利用智能对话平台进行处理,根据对话请求的内容,了解用户的需求和问题,并开始进行答案编辑和确认,进而生成优质答案,优质答案被转发给相应的客服人员进行处理,客服人员可以对答案内容再次进行编辑和确认,包括对答案进行调整、补充相关链接或资料,以提供更全面和准确的答复,经过编辑和确认后,客服人员将答案发送给用户,可以通过客服管理平台的界面直接发送,或者使用自动化的方式将答案自动发送给用户,持续迭代和改进包括客服人员会可以记录用户的反馈和评价,以便后续对话和服务的改进,帮助客服人员更好地理解用户需求,根据用户反馈和数据分析,工程师可以不断改进对话平台的性能,用来给客户提供更优质的服务,整个智能对话平台可以通过多种方式与用户进行交互,来提供便捷的沟通方式,且能够改进自身性能,以适应不断发展改变的社会,来进一步的提升用户体验,在保证自身对话平台可以一直提供高质量的复位,可以增强用户黏性和满意度,不断增加对话平台用户,可以进一步促进自身智能对话平台的发展。

[0041] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,根据本发明的技术方案及其发明构思加以等同替换或改变,都应涵盖在本发明的保护范围之内。

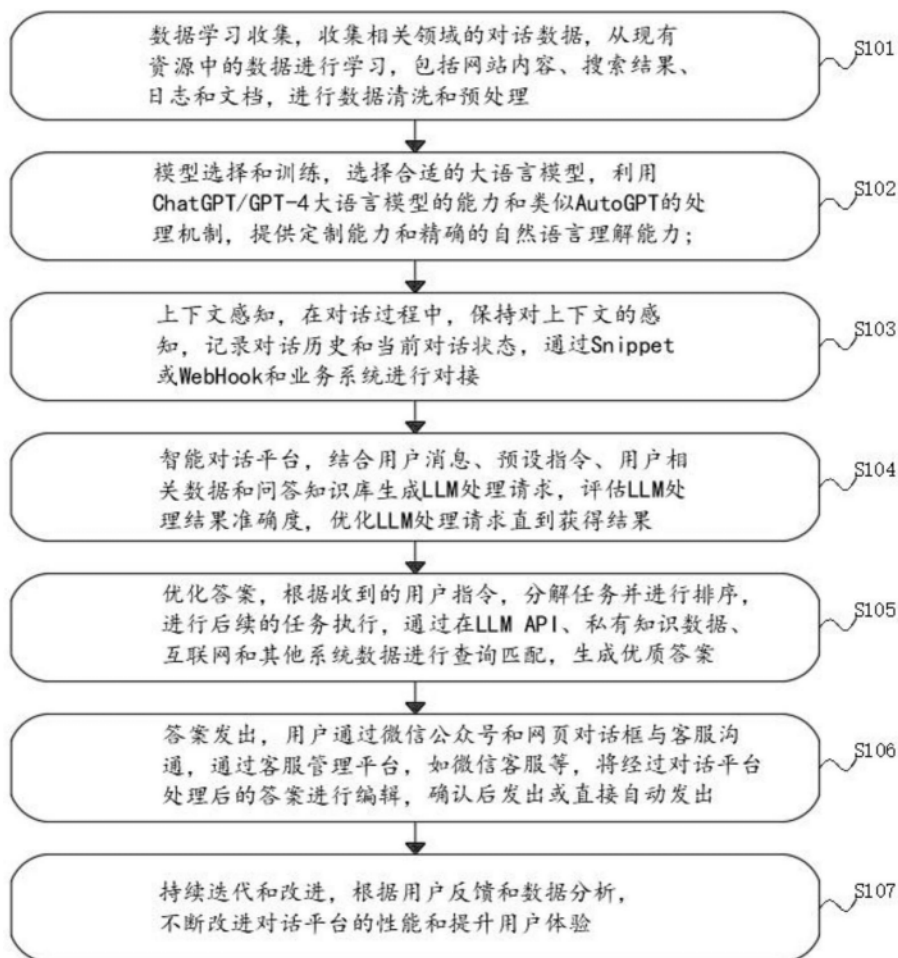


图1

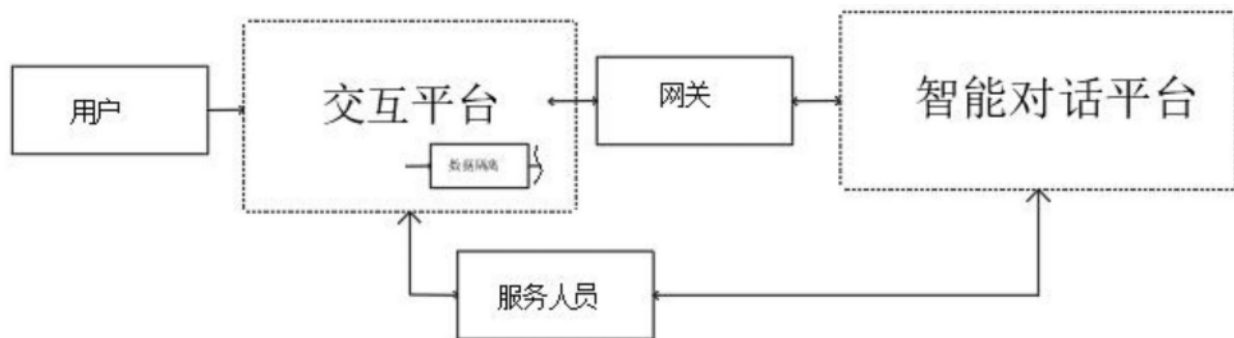


图2

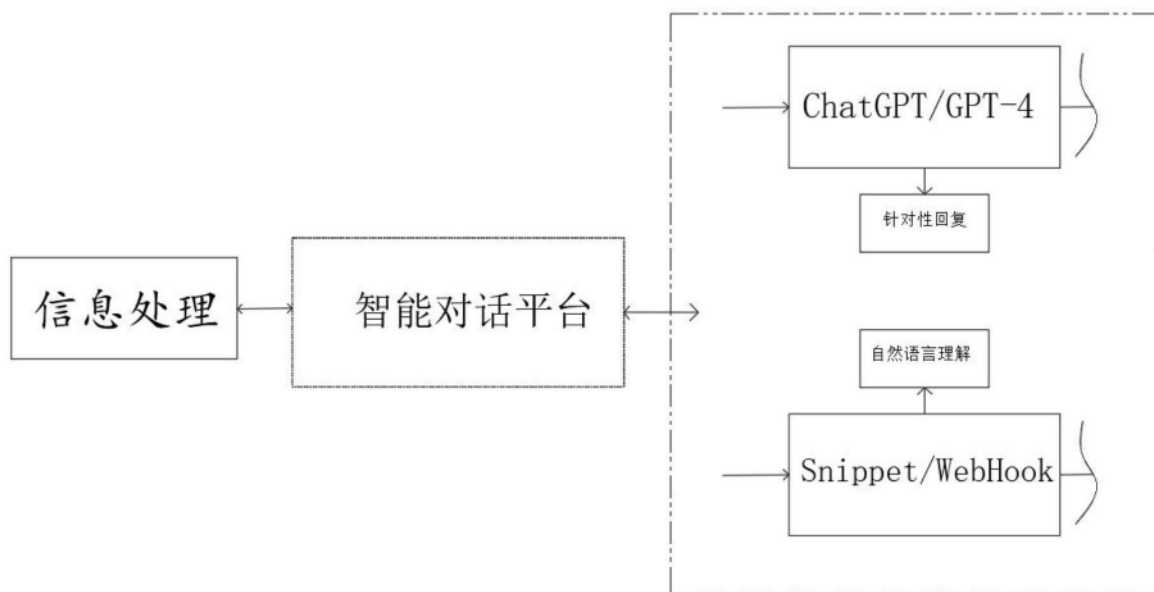


图3

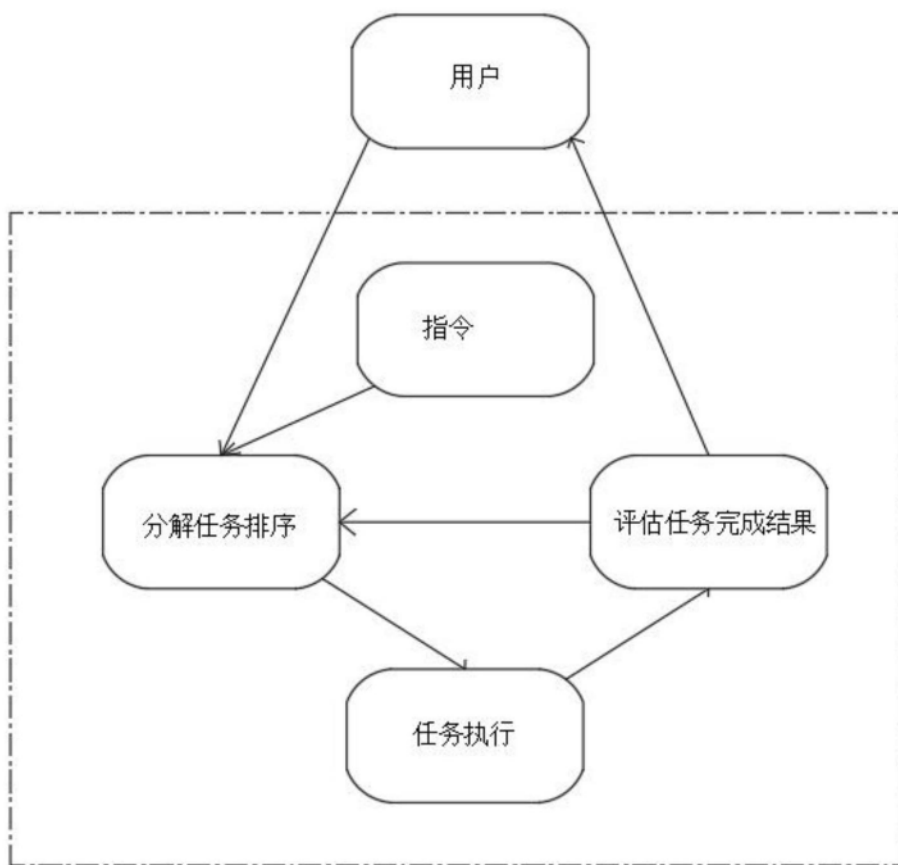


图4

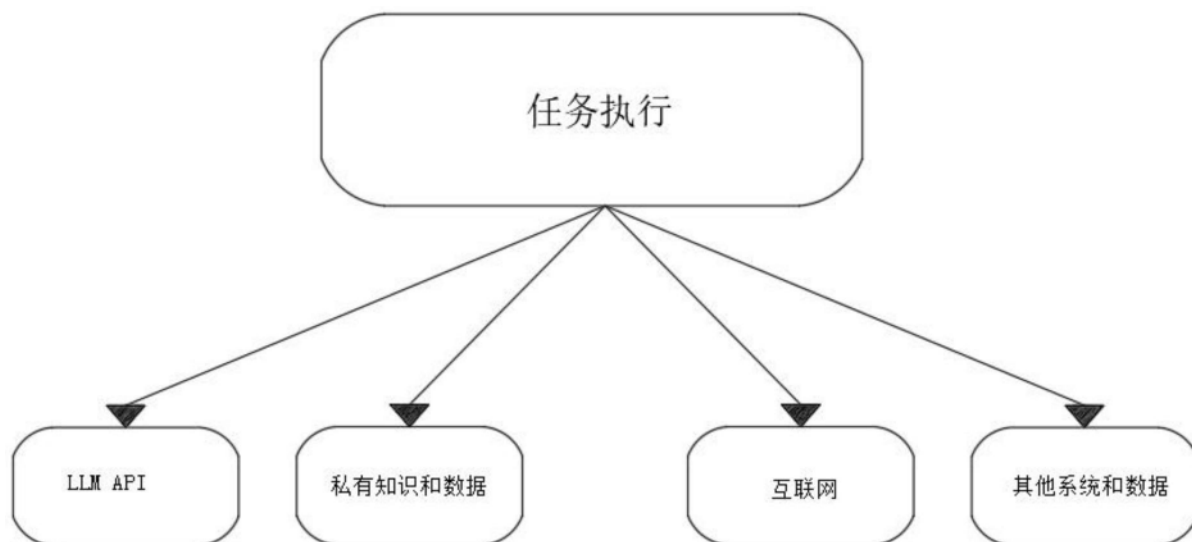


图5

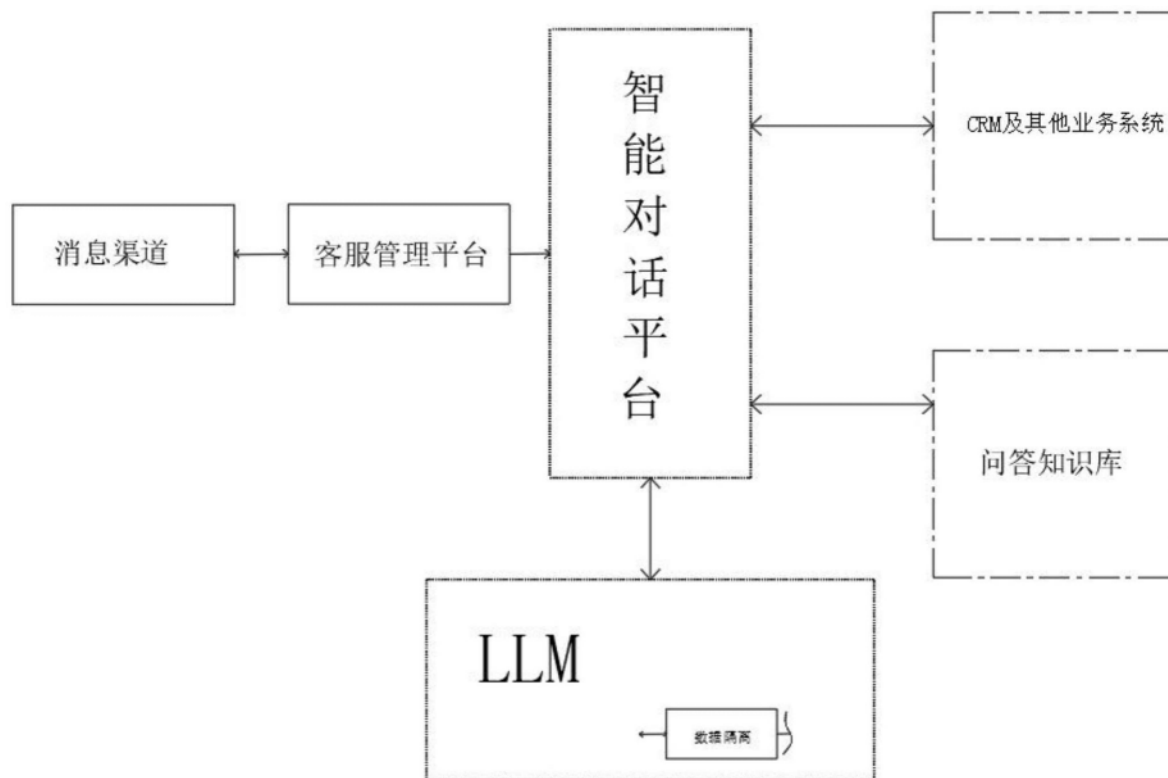


图6